2-2020

Predicting Pilot Misperception of Runway Excursion Risk Through Machine Learning Algorithms of Recorded Flight Data

Edwin Vincent Odisho II

PREDICTING PILOT MISPERCEPTION OF RUNWAY EXCURSION RISK THROUGH MACHINE LEARNING ALGORITHMS OF RECORDED FLIGHT DATA

By

Edwin Vincent Odisho II

A Dissertation Submitted to the College of Aviation
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Aviation

Embry-Riddle Aeronautical University
Daytona Beach, Florida
February 2020

PREDICTING PILOT MISPERCEPTION OF RUNWAY EXCURSION RISK THROUGH MACHINE LEARNING ALGORITHMS OF RECORDED FLIGHT DATA

By

Edwin Vincent Odisho II

This Dissertation was prepared under the direction of the candidate's Dissertation Committee Chair, Dr. Dothang Truong, and has been approved by the members of the dissertation committee. It was submitted to the College of Aviation and was accepted in partial fulfillment of the requirements for the
Degree of
Doctor of Philosophy in Aviation

Dothang Truong, Ph.D.
Committee Chair

Robert E. Joslin, Ph.D.
Committee Member

Steven Hampton, Ed.D.
Associate Dean of Research and Graduate Studies, College of Aviation

David A. Esser, Ph.D.
Committee Member

Alan J. Stolzer, Ph.D.
Dean, College of Aviation

Robert W. Maxson, Ph.D.
Committee Member (External)

02/27/2020
Date

ABSTRACT

Researcher: Edwin Vincent Odisho II

Title: **PREDICTING PILOT MISPERCEPTION OF RUNWAY EXCURSION RISK THROUGH MACHINE LEARNING ALGORITHMS OF RECORDED FLIGHT DATA**

Institution: Embry-Riddle Aeronautical University

Degree: Doctor of Philosophy in Aviation

Year: 2020

The research used predictive models to determine pilot misperception of runway excursion risk associated with unstable approaches. The Federal Aviation Administration defined runway excursion as a veer-off or overrun of the runway surface. The Federal Aviation Administration also defined a stable approach as an aircraft meeting the following criteria: (a) on target approach airspeed, (b) correct attitude, (c) landing configuration, (d) nominal descent angle/rate, and (e) on a straight flight path to the runway touchdown zone. Continuing an unstable approach to landing was defined as *Unstable Approach Risk Misperception* in this research. A review of the literature revealed that an unstable approach followed by the failure to execute a rejected landing was a common contributing factor in runway excursions.

Flight Data Recorder data were archived and made available by the National Aeronautics and Space Administration for public use. These data were collected over a four-year period from the flight data recorders of a fleet of 35 regional jets operating in the National Airspace System. The archived data were processed and explored for evidence of unstable approaches and to determine whether or not a rejected landing was executed. Once identified, those data revealing evidence of unstable approaches were processed for the purposes of building predictive models.

SAS™ Enterprise Miner® was used to explore the data, as well as to build and assess predictive models. The advanced machine learning algorithms utilized included: (a) support vector machine, (b) random forest, (c) gradient boosting, (d) decision tree, (e) logistic regression, and (f) neural network. The models were evaluated and compared to determine the best prediction model. Based on the model comparison, the decision tree model was determined to have the highest predictive value.

The Flight Data Recorder data were then analyzed to determine predictive accuracy of the target variable and to determine important predictors of the target variable, *Unstable Approach Risk Misperception*. Results of the study indicated that the predictive accuracy of the best performing model, decision tree, was 99%. Findings indicated that six variables stood out in the prediction of *Unstable Approach Risk Misperception*: (1) glideslope deviation, (2) selected approach speed deviation (3) localizer deviation, (4) flaps not extended, (5) drift angle, and (6) approach speed deviation. These variables were listed in order of importance based on results of the decision tree predictive model analysis.

The results of the study are of interest to aviation researchers as well as airline pilot training managers. It is suggested that the ability to predict the probability of pilot misperception of runway excursion risk could influence the development of new pilot simulator training scenarios and strategies. The research aids avionics providers in the development of predictive runway excursion alerting display technologies.

DEDICATION

I would like to dedicate this work to my parents and my wife. My parents, Edwin and DeAnn Odisho, instilled upon me the value of education and the pursuit of knowledge. The encouragement you provided in my formative years, to ask why and seek answers, laid the foundation for a lifelong journey in aviation. Without your support and love, this journey would not have been possible. Although I lost you in the middle of the program, your presence in my life persists and will be with me until we meet again.

I would like to share this dedication with my wife, Bettina Odisho. Bettina was with me every step of the way on this journey. Without you I would not have had the intestinal fortitude to begin this endeavor, much less the shared responsibility of balancing family, full time flying, and pursuing a Ph.D. During the challenging times, your love and encouragement gave me the courage and motivation to succeed.

ACKNOWLEDGEMENTS

To successfully complete an endeavor of this magnitude is not an individual accomplishment. I received immeasurable support, encouragement, guidance, advice, and help. I would first like to acknowledge my family. The day to day effort required to complete this program would not have been sustainable without complete and comprehensive teamwork at home. The demands of the program necessitated sacrifice and selflessness on behalf of my wife, Bettina and daughters, Natalie and Sofia. I am forever grateful to you all.

I have been fortunate to receive coaching, guidance, instruction, and mentorship from teachers, coaches, flight instructors, professors, and others in my lifelong journey of learning. No one has been more important or valuable to me than my mentor and committee chairman, Dr. Dothang Truong. His unique ability to constructively criticize, encourage improvement, and maintain high standards was crucial to my development on this journey of scholarship. I have never met a finer gentleman in my nearly 40 years in aviation. I feel fortunate and blessed to have had the opportunity to work with Dr. Truong.

Dr. Dothang Truong was assisted by a very special group of scholars on my dissertation committee. External committee member, Dr. Robert Maxson, is Director, National Centers for Environmental Prediction, and is an expert in data mining and predictive modeling. Dr. David Esser is an acknowledged expert in flight data analysis programs such as FOQA. Dr. Robert "Buck" Joslin, a fellow Marine Corps aviator and test pilot, is an expert in governmental regulatory issues associated with flight deck technology integration. Semper Fidelis!

A special acknowledgement is greatly warranted for several people whose importance and value to the program could not be stressed enough. Mrs. Susie Sprowl, Administrative Assistant, School of Graduate Studies, College of Aviation, and Ms. Katie Esguerra, Marketing and Admissions Coordinator, School of Graduate Studies, College of Aviation, greatly assisted me in my transition back to student status after many years out of academia. I would also like to acknowledge Mrs. Jan Neal, Lead Instructional Designer, School of Graduate Studies, College of Aviation, for her assistance in the production aspect of my dissertation document.

I would be remised if did not acknowledge each of my professors in the program: Dr. Bruce Conway, (Management of Systems Engineering); Dr. Haydee Cuevas (Human Factors in Aviation, Quantitative Research Methods in Aviation, User-Centered Design in Aviation); Dr. Mark Friend, Academic Advisor (Aviation Safety Management Systems); Dr. Steven Hampton, Associate Dean of Research and Graduate Studies, College of Aviation, (Current Practices/Future Trends in Aviation); Dr. Kadie Mullins, (Instructional Design in Aviation); Dr. Alan Stolzer, Dean College of Aviation (Topics in Safety Management Systems); and Dr. Dothang Truong, Dissertation Committee Chairman (Advanced Quantitative Data Analysis, Research Methods, Quantitative and Qualitative Data Analysis, Operations Research and Decision Making). Although I was not fortunate enough to have taken a class from Dr. Scott Winter, I would like to acknowledge the important role he played in my development in the program. Thank you all for your encouragement and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The Federal Aviation Administration (FAA), the National Transportation Safety Board (NTSB), the Flight Safety Foundation (FSF), and the International Air Transport Association (IATA) have identified the continuation of an unstable approach to a landing as a hazard that has contributed to runway excursion (RE) accidents and incidents. The FAA (2003) defined a RE as a landing attempt that results in an overrun or veer off the runway surface. The IATA Accident database indicated that 61% of all aviation accidents from 2012-2016 occurred during the approach and landing phases of flight. IATA also claimed that 16% of those accidents contained unstable approach contributory factors (IATA, 2017). Consequently, the NTSB has issued numerous safety recommendations to enhance runway safety, which have been consistently included in recent NTSB Most Wanted List of Transportation Safety Improvements (NTSB, 2019a). A review of recent NTSB accident investigation reports produced evidence that aircraft operators have not fully developed effective risk mitigation strategies concerning REs (FAA, 2014, 2015; NTSB, 2000, 2001, 2014a, 2014b, 2016, 2019b).

Boeing Commercial Airplanes (BCA) (2017) has compiled data on commercial aircraft accidents worldwide since 1959. Boeing reported that the highest percentage of fatal accidents over the last 10 years occurred during the approach and landing phases of flight as shown in Figure 1. Boeing also emphasized the contrast of relatively low flight time in the approach and landing phases of flight and the higher percentage of fatal accidents relative to other phases of flight (BCA, 2017).

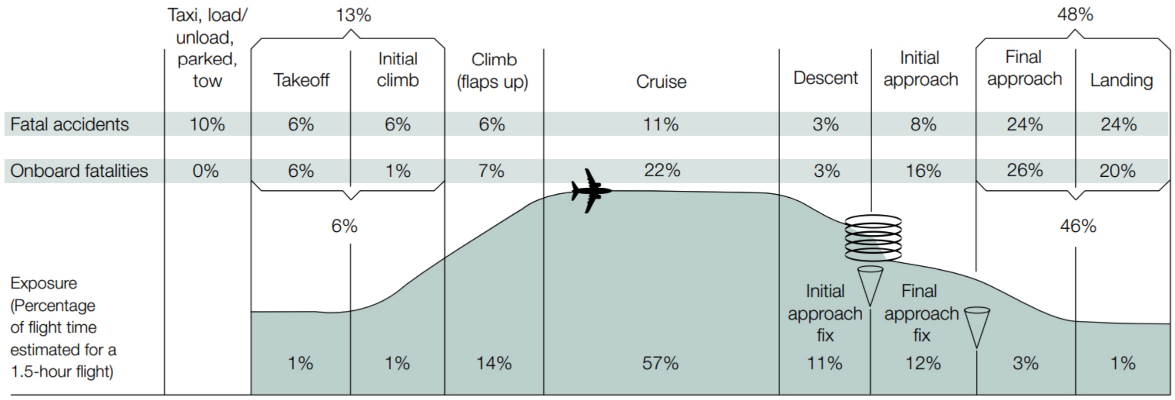

Figure 1. Fatal accidents and onboard fatalities by phase of flight from 2007–2016. Percentages may not sum to 100% due to numerical rounding. Reprinted from "Statistical Summary of Commercial Jet Airplane Accidents Worldwide Operations: 1959-2016," Aviation Safety, 2017. Copyright 2017 by Boeing Commercial Airplanes, p. 20. Adapted with permission. Source: www.skybrary.aero

To facilitate safety risk mitigation strategies for commercial airline operators, the FAA commissioned a working group, the Commercial Aviation Safety Team (CAST), in 2002. One of the key recommendations from the CAST was the drafting of Advisory Circular (AC) 120-71A, *Standard Operating Procedures for Flight Deck Crewmembers* (subsequently replaced with AC 120-71B). This Advisory Circular introduced stabilized approach criteria, based on aircraft glide path, energy state, and configuration for landing (FAA, 2003). The FAA subsequently removed the stabilized approach criteria from AC 120-71A when it was updated to AC-120-71B. Although the specific criteria for the stabilized approach concept was not listed subsequent documentation, the FAA provided guidance on stable approaches in AC-91-79A, *Mitigating the Risks of a Runway Overrun Upon Landing*. In this AC, the FAA presented a case study on an unstable approach scenario as well as listing unstabilized approaches as the primary contributory factor in runway excursions (FAA, 2009, p. 3). Although FAA stable approach criteria have not been updated or modified in subsequent advisory circulars, both AC-120-71A and AC-

91-79A have been referred to in FAA documentation regarding unstable approaches. For example, the FAA advised readers to "refer to AC-120-71" in its description of stabilized approaches in AC-120-108, *Continuous Descent Final Approach,* (FAA, 2011, p. 2) and also made a similar suggestion in FAA Safety Briefing 18-09, *FAA Stabilized Approach and Go-Around Concept,* but referred readers instead to AC-91-79A (FAA, 2018, p. 2). References to specific FAA stable approach criteria were made to those listed in AC-120-71A (FAA, 2003) in this research.

The FAA also discussed unstable approaches in recommendations made to pilots concerning energy state management techniques in Advisory Circular 120-111, *Upset Recovery and Prevention Training*. In this AC, the FAA asserted that proper energy state management was a critical component in flight path management associated with stable approaches. (FAA, 2017a). Figure 2 shows an overview of a stabilized approach.

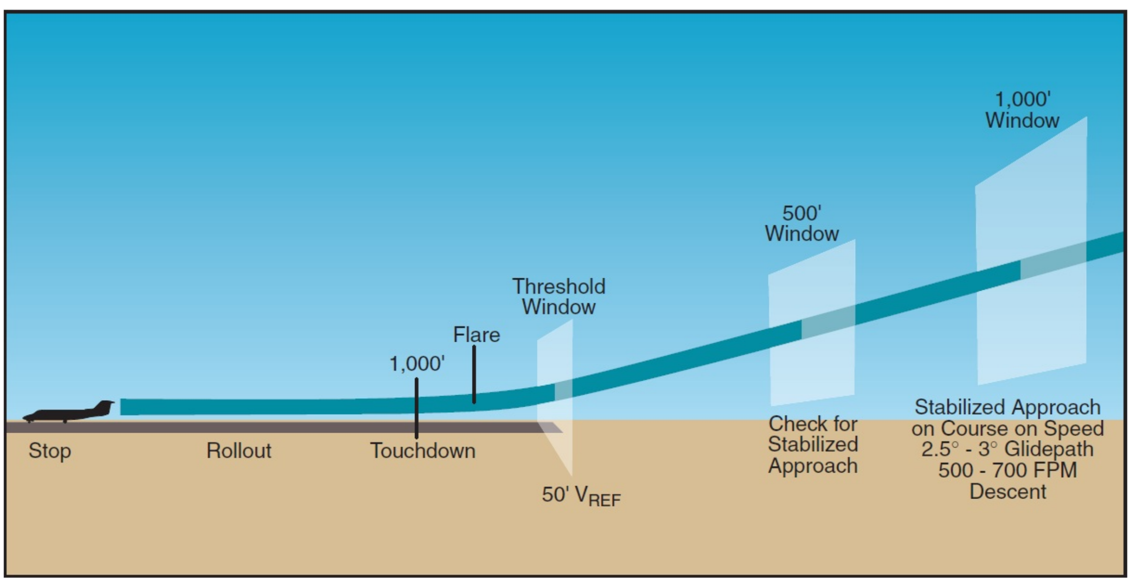

Figure 2. Stabilized approach. Reprinted from "Air Traffic Bulletin Procedures (ATB 2019-1)," by Federal Aviation Administration Air Traffic Procedures, April 2019, p. 2. Retrieved from https://www.faa.gov/air_traffic/publications/media/atb_april_2019.pdf

Additionally, in Advisory Circular 91-79A the FAA informed operators of the importance of safety risk mitigation strategies regarding runway excursions and highlighted concerns associated with unstable approaches. The FAA asserted that exceedances in stable approach criteria could be contributory factors in runway excursions (FAA, 2013). For example, approach airspeed exceedance could cause a long landing, contributing to a runway overrun.

In 2008, the FAA formed the Runway Safety Council in collaboration with industry, to address hazards associated with runway safety. One of the stated goals of this cooperative effort was to decrease the number and severity of REs (FAA, 2008). The FAA asserted that REs play a crucial role in the overall risk-based scope of runway safety, with over 30 percent of REs resulting in accidents (FAA, 2008). The FAA and industry stakeholders subsequently developed action plans to reduce REs focused on identifying important factors that contribute to REs, using a data-driven approach (FAA, 2015).

An NTSB (2013) report presented details on the importance of aeronautical decision making (ADM) in the unstable approach/rejected landing process in an accident involving an Embraer 505 (also known as the Embraer Phenom 300 light jet) regional jet. The primary contributory cause of the accident was the failure of the pilots to execute a rejected landing when faced with evidence of an unstable approach (NTSB, 2013). The NTSB noted that the correct approach reference speed was 110 knots indicated airspeed (KIAS); however, data from the flight data recorder (FDR) indicated an actual approach speed of 158 KIAS. The approach speed exceedance contributed to a RE as the aircraft experienced a runway overrun, resulting in destruction of the aircraft

Runway excursion caused by unstable approach to a landing was also stated as the primary contributory factor in an average of 10 accidents and incidents per year from 2005 to 2016, resulting in the NTSB issuing Safety Recommendations A-08-16 through A-08-20 (FAA, 2014a, 2014b; NTSB, 2016). These Safety Recommendations detailed significant deficiencies in industry-based initiatives mitigating the risk of runway excursions through pilot training alone, substantiated with evidence presented in NASA Aviation Safety Reporting System (ASRS) pilot reports (NTSB, 2008). The suggested lapse in ADM, which occurs when a pilot elects to continue an unstable approach to landing, thus risking a runway excursion, was defined as *Unstable Approach Risk Misperception (UARM)* in the research.

Several examples of accidents with UARM-like contributory factors have been reported by the NTSB:

- Asiana Air Flight 214 crashed at the San Francisco International Airport, due to a RE (veer-off) caused by exceedance of glidepath stabilized approach criteria (well below glidepath). The accident resulted in destruction of the Boeing 777 as well as fatal injuries to three passengers (NTSB, 2014).
- UPS Flight 1354 crashed while attempting a night instrument approach to a landing at the Birmingham-Shuttlesworth International Airport, Birmingham, Alabama in August 2013. The NTSB reported the primary cause of the accident was the flight crew continuing an unstable approach to landing, resulting in destruction of the Airbus A300 as well as fatally injuring the sole occupants, the two pilots. The report noted that the A300

exceeded stable approach criteria based on excessive glidepath deviations (below glidepath), resulting in the aircraft impacting the ground short of the runway (NTSB, 2014).

- Federal Express (FedEx) Flight 14 crashed during landing at the Newark International Airport, Newark, New Jersey in July 1997, resulting in destruction of the McDonnell Douglas MD-11. The NTSB stated that the probable cause of the accident was the Captain's lapse in ADM, continuing a landing with evidence of exceedance of stabilized approach criteria (excessive descent rate). The unstable approach resulted in a hard, bounced landing, leading to a loss of directional control on the runway, and ultimately a RE (NTSB, 2000).

Based on these and other related aviation accidents and incidents associated with REs, the NTSB (2019b) issued Safety Alert 077 advising pilots that failure to reject a landing associated with an unstable approach could result in not only a RE, but also loss of control and/or collision with terrain. In this Safety Alert, the NTSB advised pilots who face evidence of an unstable approach at 500 ft in visual conditions, to execute a rejected landing. The NTSB also advised pilots to beware of operational pressures and continuation bias to continue a landing attempt when unstable, and reiterated the importance of performing a rejected landing when faced with evidence of an unstable approach (NTSB, 2019b).

These examples of lapses in ADM, which resulted in the occurrence of UARM and REs, provide evidence of an aviation hazard. The FAA explained the importance of ADM as it related to pilot risk management in AC 60-22, *Aeronautical Decision Making.*

In this document, the FAA detailed how pilot risk mismanagement could lead to aviation incidents and accidents (FAA, 1991). Orasansu et al. (2001) provided details on the criticality of risk perception in ADM with a discussion of several perspectives on risk perception factors such as: (a) organizational pressures, (b) pilot experience levels, (c) job responsibilities, and (d) mental modeling. The researchers described pilot risk tolerance and perception as they relate to ADM as depicted in Figure 3.

Figure 3. Risk management decision-making process. From "Pilot's Handbook of Aeronautical Knowledge (FAA-H-8083-25)," by Federal Aviation Administration, 2016 (https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak/media/04_phak_ch2.pdf). In the public domain.

Additionally, the FAA (2016) asserted that the goal of risk management "is to proactively identify safety-related hazards and mitigate the associated risks" (p. 2-3). The FAA continued to describe how the development of good risk assessment skills were necessary for pilots to demonstrate successful ADM.

Hunter (2005) provided much of the foundational research on pilot risk perception of hazards. Hunter defined pilot risk perception as the cognitive ability to appraise and discern risk involved in the formulation of an environmental mental model. The researcher detailed the misconception of risk perception when this appraisal of a situation is in error. When pilots either underestimate the risk inherent in the situation or overestimate their own capabilities, pilot risk perception error is probable. You and Han (2013) build on the work of Hunter (2005) with the assertion that the effects of airline pilot risk perception on threat and error management (TEM) are significant. Effective risk perception aspects of ADM enable pilots to successfully identify hazards while addressing the cognitive demands inherent to flight operations.

Orasanu et al. (2001) described how inappropriate risk perception could contribute to lapses in ADM. The researchers described how *continuation errors*, also referred to as *continuation bias* by Dismukes (2010), could occur when pilot ADM and SA do not evolve and adapt to a dynamic environment. For example, if evidence of an unstable approach becomes apparent to the flight crew and they elect to continue to landing, as originally planned, Dismukes (2010) and Orasanu et al. (2001) assert that *continuation bias* may have been experienced.

Pilots have been trained to compare actual aircraft performance variables, such as (a) indicated airspeed, (b) descent rate, (c) angle of bank, and (d) engine thrust, with stable approach criteria recommended by the FAA and further customized by each operator. Based on this assessment, pilots are expected to execute a rejected landing if stable approach criteria have not been met (Moriarty & Jarvis, 2014). Although the literature indicated that several variations of the term *rejected landing* have been used interchangeably (i.e., missed approach or go-around), for purposes of standardization, the term *rejected landing* was used in this research when referring to the procedure pilots perform to abandon a landing attempt.

Although pilots have been trained to reject landings based on evidence of an unstable approach, a study by Giles (2013) on compliance with Standard Operating Procedures (SOP), an aspect of Aeronautical Decision Making, stated the following:

> For the most part pilots will comply with SOP, but when they (1) don't agree with SOP, (2) don't understand SOP or the risks associated with not complying with SOP, or (3) don't feel adequately trained to know what SOP is, it is difficult to motivate them to comply. (p.2)

Hence, pilots may not comply with stabilized approach criteria if they do not perceive the risk of a runway excursion associated with not executing the required rejected landing. The FAA suggested that one possible consideration to this risk misperception was that 97% of unstabilized approaches have resulted in a safe landing, although 10% of these *safe landings* exceeded some parameter (e.g. landing long). Regardless, the FAA suggested that non-compliance with an SOP was indicative of ineffective ADM (FAA, 2016).

With the continued advancements in FDR and Cockpit Voice Recorder (CVR) technology on commercial airliners, large volumes of data have been recorded and archived at a very rapid pace (Walker, 2017). W. Vogt, G. Vogt, Gardner, and Haeffele (2014) define *big data* as data so large in volume that it would be impossible for one person to code and analyze in less than one year without utilizing a computer. FDR technology was originally used to assist accident investigations with mathematical analytical techniques concurrently evolving as FDR technology improved. As a result of these developments, new applications in safety risk management (SRM) emerged. Examples of these SRM processes include the Flight Operations Quality Assurance (FOQA) and FDM programs (Treder & Crane, 2004).

The advent of large data gathering methods has also provided the impetus for the continued development of advanced data analytical tools. Other industries have utilized advanced techniques in computing capability to develop complex mathematical algorithms with the capability to handle large data (Tufféry, 2011). Recently, aviation researchers have begun to utilize these advanced data analytical tools in the exploration of large flight data (Li, Das, Hansman, Palacios, & Srivastava, 2015). Machine learning (ML) techniques have emerged as a preferred technique to rapidly analyze large volumes of flight data (Koteeswaran, Malarvizhi, Kannan, Sasikala, & Geetha, 2017).

The purpose of the research was to utilize machine learning techniques to explore large flight data in order to predict UARM. The exploration focused on the approach and landing phases of flight, specifically on unstable approaches. Variable selection processes were based on the stable approach criteria as defined in AC 120-71A (FAA, 2003) and AC-91-79A (FAA, 2014). Variables were defined using the recorded flight data parameters including: (a) target approach speed deviation, (b) flap position, (c) landing gear position, (d) engine speed, and (e) glide path deviation. Additional criteria were based on: (a) vertical and lateral position of the aircraft with reference to the landing runway, (b) energy state, and (c) landing configuration. The information gathered in the data analysis was used to predict the probability of the pilot misperceiving the runway excursion risk of continuing an unstable approach to landing. Pilot misperception was represented by non-compliance (either intentional or not) with standard operating procedures regarding FAA guidance on required actions when faced with evidence of an unstabilized approach. Machine learning techniques were used to populate and compare various predictive models, and to determine the most accurate model, which was then used to make predictions of the probability of the manifestation of the target dependent variable, UARM. The occurrence of UARM contradicts best safety practices as recommended by the NTSB and the guidance by the FAA, which is considered the minimum requirement in the operations specifications (OPSPECS) of any air carrier (FAA, 2003, 2014; NTSB, 2016, 2019b).

Background

The background on stable approaches began in 1997 with NTSB Safety Recommendation A-97-85 that requested the FAA require all 14 CFR Part 121 and 135 operators to provide guidance for pilots regarding critical safety-of-flight decision-making, particularly regarding stabilized approaches. A Part 121 air carrier (i.e. airliners) is an alias for scheduled passenger/freight operations and a Part 135 carrier comprises only commuter and on-demand operations. In response to the NTSB recommendations, the FAA issued Flight Standards Handbook Bulletin for Air Transportation (HBAT) 98-22, *stabilized approaches*. A key component of this document was the requirement for all 14 CFR Part 121 and 135 operators to establish defined criteria for stabilized approaches and also to train pilots to perform rejected landings if stabilized approach conditions were not met (NTSB, 2001). Although unstable approaches were also a known hazard with general aviation (GA) aircraft, these operators were considered out of scope because data have only been obtained for a Part 121 carrier.

Despite these initiatives, American Airlines flight 1420 crashed during a landing attempt in June 1999 at the Little Rock National Airport in Little Rock, Arkansas. The McDonnell Douglas MD-82 aircraft overran the runway resulting in destruction of the aircraft. The Captain and 10 passengers were fatally injured. In addition to attempting to land in spite of evidence indicating exceedance of aircraft operating manual (AOM) crosswind limitations, the aircraft was not in the correct landing configuration (i.e. spoilers were not armed), as required for a stabilized approach. The spoilers are normally armed to automatically deploy upon touchdown, reducing lift/increasing drag and assisting in aircraft deceleration. The spoilers were particularly important on this flight as the runway was wet and the increased drag could have assisted in the prevention of hydroplaning (i.e. tires losing contact with the runway surface, on a thin layer of water). Because the spoilers were not armed, upon touchdown they did not automatically deploy, or extend, resulting in excessive rollout speeds and hydroplaning, which contributed to the aircraft being unable to stop prior to overrunning the runway (NTSB, 2001).

The NTSB (2001) noted that when the AA 1420 accident occurred in 1999, the only written guidance available to the crew concerning the stabilized approach concept was a vaguely worded description of a *landing technique* in the carriers' SOPs. The NTSB (2001) stated that:

> The only stabilized approach guidance provided in aircrew training at AA stipulated that the minimum recommended stabilized approach altitudes for IFR and visual flight rules (VFR) conditions were 1,000 and 500 feet, respectively,

> and that landing flaps were to be selected by 1,000 feet above ground level. Before descending below the specified minimum stabilized approach altitude, the airplane was to be in the final landing configuration (gear down and final flaps), on approach speed, on the proper flightpath, at the proper sink rate, and at stabilized thrust; these conditions were expected to be maintained throughout the rest of the approach. However, the guidance did not define what was meant by "on" approach speed, "on" the proper flightpath, and "at" the proper sink rate. In addition, the guidance did not describe the necessary flight crew actions if the stabilized approach criteria were not met. Information presented in the "Techniques" section was not considered by American to be required procedures but rather suggested ways of accomplishing a task. (p.160)

The FAA responded to the recommendations by the NTSB with the development of the *stabilized approach concept*. The fundamental premise of the stabilized approach concept was that a general description of the aircraft state in the final approach and landing phases of flight should be based on three main aspects: (a) aircraft position on glide path and lateral extended runway centerline, (b) energy state, and (c) landing configuration (FAA, 2003). More restrictive criteria were left to the discretion of the operator and with the approval of each operator's FAA Principal Operations Inspector (POI). The POI is tasked with ensuring air carrier compliance with their FAA approved Operator SOPs.

Campbell, Schroeder, Shah, and Zaal (2018) provided additional information regarding the collaboration between the FAA, NASA, and the NTSB on unstable approaches and pilot rejected landing ADM. The researchers detailed NTSB assertions that AC 91-79A did not provide specific guidance on rejected landing requirements as recommended, which resulted in the NTSB closing the recommendation in 2012 with an unsatisfactory response. Campbell et al. (2018) contended that previous studies had not accurately investigated the root causes of the lack of compliance regarding rejected landings following an unstable approach. The researchers insisted that stable approach criteria were too complex and restrictive to the operational environment (Campbell et al., 2018).

In 2000, the FAA developed the first advisory circular on standard operating procedures (SOPs), now universally recognized as a basic component in an organization's safety management system (SMS) (FAA, 2003). An organization's SOPs are the foundation to effective crew performance and help pilots maintain an accurate mental model of an aviation task. The FAA has provided air carriers with guidance that a rejected landing is a successful outcome when given evidence of an unstable approach (FAA, 2014). Analysis of FDR data in air carriers shows that the frequency of unstable approaches was 4% in 2009. Additionally, line operations safety audit (LOSA) jump seat observers on the flight decks of 4532 commercial flights between 2002 and 2006 reported, based on visual observation of flight instrument indications, that 5% of approaches were unstable and of those only 4% of unstable approaches resulted in a rejected landing (Moriarty & Jarvis, 2014).

The Flight Safety Foundation (2009) concluded that the number of rejected landings greatly underestimates the number of unstable approaches. This evidence was based on data gathered not only by the FSF but also by the collaborative industry based Commercial Aviation Safety Team, formed by the FAA in 2008 to address runway safety

(FAA, 2008). Conclusions made by both the FAA and the FSF suggest that current risk mitigation strategies have fallen short of stated objectives by the NTSB and FAA in their collaboration on the Runway Safety Council (RSC) (FSF, 2009). One of the main objectives of the RSC was to reduce the risk of REs (FAA, 2014). Although aeronautical decision making, human error, and situation awareness have been well represented in the literature, little work has been presented regarding the use of machine learning to predict probability of pilot misperception of the runway excursion hazard, when faced with evidence of an unstable approach.

Statement of the Problem

Runway excursions are an aviation safety risk associated with hazards inherent in unstable approaches. The NTSB described problems of continuing unstable approach to landing in Safety Alert 077 (2019b). In this document, the NTSB (2019b) listed problems associated with unstable approaches as:

- Failure to establish and maintain a stabilized approach, or continuing an unstabilized approach, could lead to landing too fast or too far down the runway, potentially resulting in a runway excursion, loss of control, or collision with terrain.
- Regardless of the type of aircraft, the level of pilot experience, or whether the flight was being conducted under instrument flight rules or visual flight rules, an unstabilized approach was a key contributor to runway excursions, loss of control, and terrain collisions control. (p. 1)

The FAA and NTSB have consistently identified unstable approaches as one of the most frequent causal factors in aircraft runway excursions, with flight data indicating an average of 10 accidents per year from 2005 to 2016 (FAA, 2014, 2016; NTSB, 2010, 2016). The FSF (2009) presented details concerning the hazards associated with unstable approaches and the risk of runway excursion that resulted when pilots elected to continue to landing. The FAA, NASA, and the NTSB have confirmed the existence of the hazard and have made efforts to address the issue with guidance to air carriers regarding the following: (a) aircrew training, (b) SOP enhancement, and (c) safety mitigation strategies (e.g. pilot simulator training scenarios). However, an analysis of flight data gathered via LOSA, FOQA, and NASA ASRS voluntary pilot reports revealed that this hazard continued to exist (FAA, 2003, 2014; FSF, 2009; NTSB, 2014a, 2014b, 2016, 2019b). The NTSB has communicated its concern that even though air carriers now have stabilized approach guidance, as described in FAA Flight Standards HBAT 98-22, runway excursions have continued to occur in part due to lapses in pilot perception of the risk when faced with evidence of an unstable approach (FSF, 2009; NTSB, 2001, 2016, 2019b).

Purpose Statement

The purpose of this research was to utilize machine learning techniques to explore large flight data in order to predict UARM. The study had two main objectives: (a) use machine learning algorithms to develop a prediction model for UARM, and (b) determine variables that contribute to the prediction of UARM. Predictive models were constructed based on advanced machine learning algorithms using 186 recorded flight data variables. Specific machine learning techniques applied to the flight data included: (a) decision tree, (b) logistic regression, (c) neural network, (d) support vector machine, (e) random forest, and (f) gradient boost machine algorithms. The flight data were recorded by FDRs on a fleet of 35 regional jets over a period of four years (2001-2004). NASA had de-identified these data and made them available to the public. These data points were analyzed to identify unstable approaches and to construct prediction models. Once the models were built and validated, the model with the highest predictive score was used to predict the probability of UARM, which could be used to identify RE hazard. Additionally, SAS™ EM® software was used to rank flight data variables in order of importance to the occurrence of UARM. SAS™ EM® defined variable worth as the rank order (from 0 to 1) of input variables determined by the Chi-square statistic and described the strength of the relationship between categorical input variables and the target variable. SAS™ EM® used *binning* to derive categorical input variables from continuous input variables (Sarma, 2013).

Significance of the Study

The research helps to enhance the effectiveness of commercial airline pilot simulator training as a hazard mitigation strategy by utilizing scenarios involving unstable approaches. Given the ability to predict UARM and the identification of flight variables most important in the prediction of UARM, airline training managers can evaluate and improve pilot ADM specifically to mitigate runway excursions.

Theoretical implications. The development of prediction models based on the application of ML algorithms to recorded flight data was a seminal study that focused on using data mining of data to build models to predict a desired or undesired event. Additionally, the results of predictive algorithms could be used to detect lapses in decision making in other high risk fields such as medicine (e.g., surgery). For example, medical professionals perform many of the similar tasks requiring decisions to be made based on safety of the patient. This decision-making ability relies on the management of risk and the perception of risk versus an estimate of one's ability to complete the task; the predictive algorithm could provide the capability to the medical professional to mitigate and reduce such risk The ability to predetermine exceedance could also contribute to the evolution of pilot alerting technologies, such as Honeywell's SmartLanding™ software algorithms that increase pilot SA of the aircraft state in the approach and landing phases of flight.

Practical implications. Key beneficiaries of the research are airline pilot simulator training programs and airline Safety Management System managers. The ability of airline pilot training managers to not only predict UARM but also identify hazardous trends in aircraft state variables involved in ADM could have a positive impact on airline safety risk mitigation strategies inherent in pilot simulator training programs, such as developing realistic runway excursion scenarios. Results of the study could be used to further refine not only FAA (2014) stabilized approach criteria but also in the oversight of air carrier pilot training programs.

Safety Management Systems managers could use the results of the study to improve SRM effectiveness, as required under 14CFR Part 5. Because SMS programs have traditionally relied on hazard identification using accident and incident reports rather than proactive measures, predictive capabilities could be beneficial. The ability to predict UARM could provide SMS managers with a predictive tool that would enhance safety risk mitigation effectiveness.

Research Questions

The study was exploratory and data-driven in nature, based on the following research questions (RQ):

- RQ 1: How can the application of data-mining and machine learning techniques to recorded flight data be used to predict the probability of *Unstable Approach Risk Misperception* by the pilot?
- RQ 2: What flight data variables are the most important predictors of pilot misperception of a runway excursion hazard as evidenced by continuing an unstable approach to a landing?

Delimitations

Exceedance criteria described in FAA AC 120-71A were considered the threshold for determining an unstable approach. Reference approach speed criteria excluded Category A approach speeds (i.e. ≤ 90 knots), as that category generally applies to helicopters (no stall speed) and light GA airplanes certified under 14 CFR §23.49 (FAA, 2012).

Limitations and Assumptions

Limitations. The data were limited to the 186 flight variables provided by the NASA public access website for four years (2001-2004) of flight operations by 35 regional jet aircraft. No data were available regarding passenger configuration, which is used by the FAA to describe regional jet commercial aircraft (less than 100 passengers) (FAA, 2005). Because CVR data were not available for the study, CRM influence on pilot misperception could not be analyzed. Additionally, pilot/automation interface was also not available. Because only FDR aircraft state data were available, the study could not consider any other variables that may have contributed to pilot UARM, such as weather (e.g. turbulence, wind shear, cross-winds), emergency or abnormal conditions (e.g. low fuel, engine or flight control anomalies), runway conditions (e.g. contamination with snow, water, lights) or visual illusions. In addition to weather considerations, day/night flight conditions were not provided and as such, were not considered in the analysis. No data were available to indicate if any of the approaches resulted in an actual runway excursion.

Assumptions. The 186 flight data variables were sufficient to develop predictive models. The data were redacted for any identifying information such as specific air carrier, aircraft type, airports, and name/type of instrument approach, hence assumptions pertaining to certain approach parameters such as approach speed, glideslope and landing configuration were made. Because FAA guidelines have allowed for more restrictive criteria to be developed by an air carrier, it was assumed that the air carrier had an SOP that followed the FAA stable approach guidance at least as restrictive as those criteria defined in AC 120-71A. Although there was no regulatory definition of regional jets, the FAA used a passenger configuration of less than 100 passengers to describe RJs in AC 150/525-4b, *Runway Length Requirements for Airport Design* (FAA, 2005). It was assumed that the flight data were sampled from RJs configured for less than 100 passengers.

Approaches were assumed to be conducted on a three-degree glideslope, and any flap setting greater than zero was assumed to be a proper landing configuration. Pilot indications of stabilized approach criteria were assumed to be provided with standard transport aircraft flight instruments. For example, descent rate, airspeed, and glidepath indications were assumed to be provided to the pilots on industry standard pilot display technology, such as electronic flight instrument systems (EFIS), primary flight display (PFD) and navigational display (ND) avionics. Target reference approach speed range was assumed to be from 105 to 140 knots indicated airspeed and was based on FAA approach category airspeed determination characteristics detailed in Title 14 CFR, Chapter I, Subchapter F, Part 97, Subpart A., § 97.3 (FAA, 2012). Additionally, target approach speed was assumed to be calculated based on a zero-wind condition. Pilots flying the aircraft represented in the study were fully qualified professional pilots.

Summary

The FAA and NTSB have identified unstable approaches as one of the primary contributory factors to runway excursion hazards (FAA, 2014; NTSB, 2000, 2001, 2014, 2014, 2016, 2019b). In the effort to enhance runway safety, the FAA has stipulated that operators adhere to criteria defining stable approaches (FAA, 2003, 2014). Data indicated that although unstable approaches still occurred, pilots may not have always followed the FAA guidance by performing a rejected landing (FAA, 2014; FSF, 2009).

Non-compliance, whether intentional or not, of FAA approved air carrier OPSPECS and SOPs concerning stabilized approaches, suggested a lapse in pilot ADM, and often had been included as a primary contributory factor in accidents and incidents involving runway excursions (NTSB, 2016, 2019b). With the advent and deployment of advanced digital data recording devices, required under 14CFR §91.609 for all air carriers with an operating certificate, opportunities exist to sample and analyze recorded flight data. Concurrently, recent developments in complex mathematical machine learning algorithms have improved research capability regarding the analysis of these flight data (Oehling & Barry, 2019). Data mining techniques, both exploratory and predictive, have provided aviation researchers the tools necessary to both analyze these large flight data and also to predict abnormal flight occurrences.

The results of the study present an example of aviation research using machine learning to predict *Unstable Approach Risk Misperception*. Subsequent chapters present a review of relevant peer-reviewed research, including gaps in the literature. Six machine learning algorithms were used for the analysis to identify which most accurately modeled the prediction of the probability of pilot misperception of runway excursion risk, as well as to identify the stabilized approach criteria flight variables associated with frequent non-compliance of rejected landing guidelines. Finally, recommendations for further research are made, based on how large flight data monitoring can be used to improve and enhance aviation safety through training, procedures, and aircraft flight instrument design.

Definitions of Terms

AVSKD	Aviation Safety Knowledge Discovery Process. The process of analyzing aviation data, beginning with the collection of raw FOQA or FDM data via aircraft flight data recorders, though several phases: data preparation, detection, feature selection, and knowledge discovery. For the purposes of the research study, the AVSKD process concludes with the assessment of predictive models, as described in Chapter III (Mathews et al., 2013).

Data Mining	Data mining is the set of methods and techniques for exploring and analyzing large data sets, in order to find certain unknown or hidden rules, associations or tendencies. It is the art of extracting information (knowledge) from the data. For the purposes of the study, predictive data mining techniques are used to extrapolate new information based on the present information (Tufféry, 2011, p. 4).
Decision Tree	A decision tree represents a hierarchical segmentation of the data and is composed of a set of rules that can be applied to partition the data into disjoint groups (Sarma, 2013, p. 196).
Energy State Management	The interrelationship between kinetic energy (airspeed), potential energy (altitude), and chemical energy (power). Refers to pilot energy state management technique options available for pilots to change or maintain a safe and stable energy state, including external factors and corrective techniques (FAA, 2017a, p. 3).
Gradient Boost Machine	A form of ML technique, using ensemble learning, used in the construction of predictive models,

	generally classification and regression. Typically, weaker decision trees are used in the ensemble.
Neural Networks	A neural network is a complex nonlinear function of inputs, divided into different layers and different units within each layer. A large number of nonlinear functions can be generated and fitted to the data by means of different architectural specifications (Sarma, 2013, p. 362). This architecture can be based on that of the brain, organized in neurons and synapses, and takes the form of interconnected units (or formal neurons), with each continuous input variable corresponding to a unit at a first level, called the input layer, and each category of a qualitative variable also corresponding to a unit of the input layer (Tufféry, 2011, p. 217).
Random Forest	A form of ML learning, using ensemble learning for purposes of classification and regression. Builds models consisting of multiple decision trees for training and produces a mode of classes or mean prediction of the individual decision tree models.
Runway Excursion	This term is limited to veer off or overrun from the runway surface that occurs while an aircraft is

	landing, based mainly on an unstable approach (FAA, 2014).
SAS® Enterprise Miner™	SAS® Enterprise Miner™ is a software package consisting of different levels of data, such as textual or numeric, and was used for the construction and analysis of predictive models. SAS EM utilizes machine-learning algorithms that streamline the data mining process and create highly accurate predictive and descriptive models that are based on analysis of vast amounts of data (Sarma, 2013).
Stabilized Approach Concept	"A stabilized approach is characterized by a constant-angle, constant-rate of descent approach profile ending near the touchdown point, where the landing maneuver begins" (FAA, 2003, Appendix 2, para 2). The energy state, landing configuration, aircraft location approach criteria are applied at 500 ft height above touchdown.
Standard Operating Procedures	Aircrew procedures developed by an airline for normal, abnormal, and emergency procedure compliance to ensure safe, efficient, and on-time flight performance (Giles, 2013).
Support Vector Machine	A machine learning model using algorithms that analyze data for classification and regression

	analysis. SVMs can perform both linear and non-linear classification using the *kernel trick.* SVMs can be used in both supervised and non-supervised approaches, in addition to clustering techniques in data analysis (Lauer & Bloch, 2008).
Unstable Approach Risk Misperception	Pilot lapses in aeronautical decision making occurring when evidence of an unstable approach exists, and the pilot elects to continue the approach to a landing, risking a runway excursion (FAA, 2014).

List of Acronyms

AC	Advisory Circular
ADM	Aeronautical Decision Making
ADMS	Aircraft Diagnostic and Maintenance System
ADS-B	Automatic Dependent Surveillance-Broadcast
AGL	Above Ground Level
AOM	Aircraft Operating Manual
ASPM	Aviation System Performance Metrics
ASRS	Aviation Safety Reporting System
AVSKD	Aviation Safety Knowledge Discovery Process
BADA	Base of Aircraft Data
BCA	Boeing Commercial Airplanes
BLS	Bureau of Labor Statistics

CART	Classification and Regression Tree
CAST	Commercial Aviation Safety Team
CEDAR	Comprehensive Electronic Data Analysis and Reporting
CFR	Code of Federal Regulations
CI	Cost Index
CRM	Crew Resource Management
CVR	Cockpit Voice Recorder
DH	Decision Height
DWH	Data Warehouse
EFIS	Electronic Flight Instrument System
FAA	Federal Aviation Administration
FDM	Flight Data Monitoring
FDR	Flight Data Recorder
FOQA	Flight Operations Quality Assurance
FSF	Flight Safety Foundation
GA	General Aviation
HAS	Hazardous Attitude Scale
HBAT	Handbook Bulletin for Air Transportation
HIP	Human Information Processing
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
IFR	Instrument Flight Rules

ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
LOSA	Line Operations Safety Audit
MKAD	Multiple kernel anomaly detection
ML	Machine Learning
MLM	Multilevel Modeling
MTS	Multi-variate Time Series Search
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
ND	Navigational Display
NOAA	National Oceanic Atmospheric Administration
NTSB	National Transportation Safety Board
OPSPECS	Operations Specifications
PFD	Primary Flight Display
POI	Principal Operations Inspector
RE	Runway Excursion
SA	Situation Awareness
SMS	Safety Management Systems
SOP	Standard Operating Procedures
SVM	Support Vector Machine
TAWS	Terrain Awareness and Warning System
TEM	Threat Error Management
TOD	Top of Descent

UARM	Unstable Approach Risk Misperception
UPS	United Parcel Service
VFR	Visual Flight Rules
VIPR	Vehicle Integrated Prognostics Reasoner
VMC	Visual Meteorological Conditions
WOW	Weight on Wheels

CHAPTER II

REVIEW OF THE RELEVANT LITERATURE

The following chapter presents a review of extant literature pertaining to the topic of the research and consists of three sections: (a) FAA guidance for stabilized approaches, (b) a review of data and text mining methodologies in aviation research, and (c) a description of the machine learning algorithms and data processing techniques that were applied in this research.

Federal Aviation Administration Guidance for Stabilized Approaches

The FAA has asserted that stabilized approaches are one of the most important factors in safe landings. One of the products that resulted from a working group study by the Commercial Aviation Safety Team was the creation of FAA Advisory Circular (AC) 120-71A: *Standard Operating Procedures for Flight Deck Crewmembers* (2003). In this AC, the FAA describes a stabilized approach as one in which all landing checklists and approach procedures have been completed, the aircraft is in landing configuration, on constant rate of descent, with the engines providing stable thrust, and in a position to make a normal landing on the runway in use. Appendix 2 of the AC provides other specific details of a stabilized approach:

- Flight should be stabilized by 1000’ Height Above Touchdown (HAT) in Instrument Meteorological Conditions (IMC) and by 500’ HAT in Visual Meteorological Conditions (VMC).
- The airplane is on the correct track.
- The airplane is in the proper landing configuration (i.e. landing gear, flaps and slats, and speed brakes).

- After glide path intercept, the pilot flying requires no more than normal bracketing corrections to maintain the correct track and desired profile (3° descent angle, nominal) to landing within the touchdown zone.
- The airplane speed is within the acceptable range specified in the approved operating manual used by the pilot (e.g. Vref).
- The rate of descent is no greater than 1000 feet per minute (fpm).
- If an expected rate of descent greater than 1000 fpm is planned, a special approach briefing should be performed.
- Power setting is appropriate for the landing configuration selected and is within the permissible power range for approach specified in the approved operating manual used by the pilot (p. A2.1)

The FAA (2003) allows for nominal bracketing adjustments related to engine thrust, descent rate and angle of bank. Recommended ranges allow for more restrictive limitations, but are provided as follows:

- Angle of bank less than 30°
- Descent rate ± 300 fpm from target
- Operator specified thrust management, per flight manual
- Momentary exceedances are acceptable, but continuous exceedance is not considered acceptable (p. A2.2).

In the aftermath of the accident involving American Airlines Flight 1420 on June 1, 1999, the NTSB (2001) recommended that the FAA further define stabilized approach criteria. In response to the NTSB recommendations, the FAA provided a brief summarization of FAA stabilized approach oversight efforts. The FAA asserted that approach gates could be customized by a carrier as milestone points in which flight crew are to assess performance criteria during an approach, in order to maintain situation awareness concerning stabilized approach indications. These approach gates are predetermined intervals at which the flight crew compare aircraft glidepath, lateral track, and airspeed data against stabilized approach criteria. It is at these intervals where the flight crew must use the information determined from the stabilized approach criteria to make the decision to either continue the approach to landing or to execute a rejected landing (NTSB, 2001).

The FAA assigns a Principal Operations Inspector (POI) to provide regulatory oversight and guidance to each air carrier. Among other functions, the POI applies federal oversight to the carrier on the stabilized approach concept stated in FAA Order 8900.1 (2007) as follows:

- Airspeed within 5 knots of approach speed at the 100-foot decision height (DH),
- The flight deck remains within the lateral confines of the runway at the 100-foot DH,
- After passing the outer marker (OM), the glidepath deviation does not exceed one half of full deflection, and
- After passing the middle marker (MM), no unusual changes in aircraft occur. (p. 4-221)

Turbojet aircraft operators must incorporate procedures that are based on stabilized approach criteria set forth in FAA Order 8900.1, as well as the recommended guidelines provided in FAA AC 120-71A (subsequently AC 91-79A). Additionally, operator standard operating procedures (SOPs) may incorporate more restrictive stabilized approach criteria than that provided by FAA guidance from these two documents. Each carrier must provide flight crew training and SOP materials which contain a description of acceptable deviations from glidepath and lateral track when covering approach and landing procedures. Once the operator's training programs are approved by the FAA, the carrier is not free to revise these procedures without approval from their POI.

The purpose of the approach gate criterion is to provide the flight crew with target values to fly, as displayed on flight deck instruments (to assess the feasibility and safety of continuing the approach to landing or to execute a rejected landing). In its report on AA Flight 1420, the NTSB notes that the pilots should have executed a rejected landing during the final approach, when stabilized approach criteria were not met. The failure of the flight crew to configure the landing flap configuration before reaching 1,000 feet AGL, and their failure to maintain a normal rate of descent, combined with deteriorating weather conditions, decreased the safety margin enough that the pilots should have executed a rejected landing (NTSB, 2001).

Unstable approaches and runway excursions. The NTSB (2013) presents additional details on the importance of ADM in the rejected landing process in its report on the accident involving an Embraer 505 regional jet. The NTSB listed the primary contributory cause of the accident as failure of the pilots to execute a rejected landing when faced with evidence of an unstable approach. The correct approach reference speed was 110 knots indicated airspeed (KIAS), and the FDR data indicated that an actual approach speed of 158 KIAS was flown. The approach speed exceedance contributed to a RE as the aircraft experienced a runway overrun, resulting in destruction of the aircraft. Runway excursion caused by unstable approach landing was stated as the primary contributory factor in this accident, as well as approximately 10 other accidents and incidents per year from 2005 to 2016 resulting in the NTSB issuing Safety Recommendations A-08-16 through A-08-20 (FAA, 2014a, 2014b; NTSB, 2016). These Safety Recommendations detailed significant deficiencies in industry-based initiatives mitigating the risk of runway excursions with pilot training substantiated with evidence presented in NASA ASRS pilot reports (NTSB, 2008).

Several recent examples of accidents with UARM-like contributory factors are described by the NTSB:

- Asiana Air Flight 214 crashed at the San Francisco International Airport, due to a RE caused by exceedance of glidepath stabilized approach criteria. The accident resulted in destruction of the Boeing 777 as well as fatal injuries to three passengers (NTSB, 2014).
- UPS Flight 1354 crashed while attempting an approach to landing at the Birmingham-Shuttlesworth International Airport, Birmingham, Alabama in August 2013. The NTSB lists the primary cause of the accident as the flight crew continuing an unstable approach to landing, resulting in destruction of the Airbus A300 as well as fatally injuring the sole occupants, the two pilots. The NTSB reports that the A300 exceeded stable approach criteria based on incorrect landing configuration and excessive glidepath deviations, resulting in the aircraft impacting the ground short of the runway (NTSB, 2014).

- Federal Express (FedEx) Flight 14 crashed during landing at the Newark International Airport, Newark, New Jersey in July 1997, resulting in destruction of the McDonnell Douglas MD-11The NTSB states that the probable cause of the accident was the Captain's lapse in ADM, continuing a landing with evidence of exceedance of stabilized approach criteria (i.e. excessive descent rate). The unstable approach resulted in a hard landing, bounce, loss of control, and ultimately a RE (NTSB, 2000).

These examples of lapses in aeronautical decision making, which result in the occurrence of UARM and REs, provide evidence of an aviation hazard. Large amounts of flight operations data have been collected with the advent of Flight Data Monitoring technologies. The evolution of advanced and complex data processing algorithms has provided aviation researchers with the opportunity to explore what patterns or relationships might exist in these large flight data.

Pilot risk perception and risk tolerance. A key point in the study is the prediction of pilot risk misperception. The FAA (1991) relates pilot risk management to task accomplishment in AC 60-22, *Aeronautical Decision Making,* with the self-assessment technique of asking oneself "Is the success of the task worth the risk?" (p. 22). Orasanu et al. (2001) describe a relative lack of aviation research on pilot risk perception and risk tolerance. The researchers continue to describe the importance of better understanding of pilot risk perception and risk tolerance in the ADM process. Martinussen and Hunter (2010) assert that pilot risk assessment and management are crucial aspects of pilot ADM. They then define pilot risk perception as "recognition of the risk inherent in a situation" (p. 198). Hunter (2005) professes that pilots can be prone to display poor risk judgement and substantially underestimate risk. The researcher's conclusions are based on evidence of pilots *pressing on* when faced with evidence of deteriorating performance conditions, while underestimating the impact of external factors to the aircraft and overestimating their self-capacity to accomplish certain tasks. The conclusions of Hunter (2005) are in agreement with those reached by both Orasansu et al. (2001) and Dismukes (2010) regarding the propensity of pilots to exhibit lapses in ADM regarding *continuation bias* (i.e., *pressing on* or *continuation errors*). Hunter (2005) states that risk perception can be mediated by both pilot self-assessment, as well as more accurate mental modeling of the environment. Martinussen and Hunter (2010) conclude that risk perception is primarily a cognitive activity and involves the accurate perception and projection of aircraft state and external factors, and the resulting mental model, to maintain a high level of situation awareness.

Hunter (2005) provides further evidence of pilot risk perception measurement using a Hazardous Attitude Scale (HAS). Airline pilots were presented with 10 different aviation scenarios and provided alternative solutions to assess ADM. Hunter concludes that poor risk perception was a more significant variable than poor risk tolerance. For example, pilots who experienced significantly more hazardous events generally rated hazardous scenarios less risky than those pilots who had experienced fewer hazardous events. Hunter subsequently deduces that poor correlation between pilot hazardous experiences and estimation of risk supports this supposition.

Hunter (2005) further expands the definition of pilot risk perception as the cognitive ability of the pilot to appraise and discern risk, while involved in the process of formulating an environmental mental model. The researcher goes on to describe the misconception of risk perception when this appraisal of a situation is in error. When the pilot either underestimates the risk inherent in the situation or overestimates his/her own capabilities, pilot risk perception error is probable.

You and Han (2013) discuss the effects of risk perception and flight experience on airline pilot attitudes relative to safety operational behaviors. They also state that pilot risk perception is a crucial pilot attribute regarding hazards. Their research affirms their supposition that pilot risk perception enables pilots to mitigate risk, while addressing the cognitive demands inherent to flight operations.

Benbassat and Abramson (2002) provide research on pilot landing risk perception with their study on landing accidents and pilot risk perception. The researchers analyzed over 6,000 NTSB accident reports whose contributory factors included runway excursions. The researchers also surveyed student and GA pilots on their perceptions of risk in the landing flare maneuver. Their findings corroborate those of Hunter (2005) regarding pilot overestimation of risk inherent in landings. The researchers recommend further research involving pilot perception of risk in the approach and landing phases of flight.

Ju, Ji, Lan, and You (2017) describe how narcissistic personality issues factor into risk perception and how overoptimistic expectations affect pilot perception of risk in Chinese pilots. Recommendations based on their findings include the necessity for pilots to accurately compare risk estimates with actual risk. They further describe the challenges inherent in the measurement and assessment of aviation risk. However, they assert that the measurement of pilot risk perception is less difficult and can be determined based on optimism bias to reflect risk estimation. Thus, while their study focused on optimism bias, they recommend further research on other cognitive biases which could significantly affect accuracy of risk perception. IATA (2017) corroborates results of the study with the assertion that pilots may sometimes continue an unstable approach to landing due to factors such as peer pressure, organizational pressure to meet schedule, and perceptions of company policy rejected landing decision making. IATA also suggests that pilot perception risk associated with the rejected landing maneuver is higher than continuing an unstable approach to landing.

Campbell, Schroeder, Shah, and Zaal (2018) conducted experimental research at the NASA Ames Research Center on perceptions of pilot risk concerning unstable approach recovery to landing and risk perception of various unstable approach conditions. Research was conducted using 36 professional airline pilot subjects and three full motion Level-D flight simulators: B747, B737, and A330. Level D flight simulators were the highest rated of four ratings (A through D) and provided full motion feedback to the pilots by means of a motion platform. Level D simulators also provided accurate flight control feedback to pilots and simulated other aircraft systems including avionics and advanced electronic flight instruments (EFIS). The FAA provided certification approval guidance to air carriers in AC 61-136A, *FAA Approval of Aviation Training Devices and Their Use for Training and Experience* (FAA, 2014b). The researchers utilized an experimental design that removed the rejected landing decision-making process in order to assess landing performance under various unstable approach conditions. The stated objective of the research was to determine if more effective rejected landing criteria were possible. One ancillary purpose of this research was to investigate pilot landing performance under various approach states to determine which limitations were most critical in pilot landings. Results indicated that energy state management parameters associated with excessive approach speeds and descent rate were the most important predictors in pilot landing performance. Although not listed in experimental controls, energy state management associated with engine power settings was observed to be an important predictor of landing performance and was recommended by the researchers to be more closely scrutinized in future research (Campbell, Schroeder, Shah & Zaal, 2018).

Aviation Research Using Data and Text Mining Methods

Data mining methods. With the rapid accumulation of large data becoming more available to researchers, opportunities for the exploration of these data have manifested themselves. FOQA and FDR data provide an appropriate source for the application of data mining methods. Much of the early work of examining these large flight data was accomplished with the intent of validating, comparing, and identifying the most accurate models of the system being evaluated. As the body of knowledge grew, so did the techniques and applications of data mining. Early examples of research applying data mining methods to flight data show that much of the work was done on validating complex algorithms, with limited results on the identification of patterns or significant relationships hidden in the large data. The work of Mugtussidis (2000) provides one such example of the limitations inherent in early efforts to analyze flight data. Mugtussidis describes the challenges associated with feature selection and proposed a classification process using estimated probability density functions. Limitations in the research represent obstacles typical of flight data analysis techniques before the advent and deployment of advanced ML algorithms. The author recommends that as computational capabilities improve, that more optimal search techniques be utilized. Mugtussidis continues to recommend that flights are segmented using cluster analysis allowing unusual events to be more accurately discovered. As the research progressed, more work began to evolve into the use of data mining to provide results which could be applied to the identification of hazards. Finally, with the development of data mining techniques, the feasibility of building complex predictive models evolved. Li, Das, Hansman, Palacios, and Srivastava (2015) present research concerning cluster-based anomaly detection to identify abnormal flight events. The results were then examined by subject matter experts, who then classified the events regarding level of hazard. Their research was enabled using large data gathered from over 26,000 commercial airline flights. Clustering techniques were then applied to the data to identify anomalies in the takeoff and landing phases of flight. The research was designed using two experimental methods: one to sample 91 flight parameters in the effort to identify abnormal flight events, and a second to evaluate three different data clustering algorithms. Limitations to the study include a vague description of what constitutes abnormal flight events, as well as those variables of interest in the clustering analysis, and the lack of a clearly defined target variable. It is also unclear what coding was used to build the models used in the evaluation of the algorithms. Additionally, the SMEs used in the evaluation of the abnormal flight events do not apply any standard criteria in their analysis.

Wang, Wu, and Sun (2014) provide a landmark study on the use of Tianjin Airlines B737 QAR based flight data research on RE impact factors. The researchers determine that pilot flare technique and weather factors are significantly correlated to RE hazards. Although the research is based on a limited sample size, the findings indicate that reactive modeling can provide important results regarding feature selection. Nanduri and Sherry (2016) build on the research of Wang et al. but limit their study to the use simulated FOQA-like flight data to investigate landing excursion hazards into San Francisco International Airport. The researchers use X-Plane simulated flight data to construct recurrent NN models using 21 flight variables. The researchers were able to demonstrate the performance of NN models in anomaly detection of FOQA-like data but used a very small sample size and applied vague exceedance criteria. They recommend that future research increase the number of flight variables and build models with different feature combinations.

Aslaner, Unal, and Iyigun (2016) present research based on the application of data mining methods to FDR data in order to identify safety issues in commercial flight operations. Specifically, cluster analysis techniques were applied to a sample of landing phase of flight operations airline data, although the source of the sample and population were not identified. Dynamic time warping (DTW) was introduced as a cluster analysis technique based on unsupervised learning and was used to examine FDM data and unstable approaches were filtered. A key factor of this research is that DTW was shown to adequately classify landings at different airports and runways and that different events could be grouped with respect to the similarities. Another meaningful aspect of the study is that the DTW method, which was used to approximate the distances between the landing performance variables, was not adversely affected by small variations in the same type of data. Limitations of the study are the relatively small sample size, poorly defined population, and lack of generalizability. Also, the data source was not specifically identified, and the phase of flight was poorly defined by reliable criteria.

Friso, Richard, Visser, Vincent, and Bruno (2018) build on Li's research with their paper on the use of ML methods to predict abnormal runway occupancy times, based on radar data patterns. Sampling and data sources were gathered using final approach radar data and A-SMGCS runway data consisting of 78,321 flights at Paris Charles de Gaulle airport and were compared with 500,000 flights at Vienna airport. Machine learning was used in feature selection and regression was used to observe the important precursors, which were identified from the top 10 features. The study focused on two different time window predictions. The first one made predictions based on abnormal arrival runway occupancy time (AROT) and associated precursors. The second one made predictions on arrival sequences for terminal air traffic flow based on a one to two-hour window. The researchers argue that the usefulness of these predictions could lead to an improvement in runway safety and throughput. The significance of the results of the research was the demonstration of the use of combined ML techniques to forecast arrival runway occupancy time (AROT) per flight. A limitation of the study was the restriction of the scope to arrival runway occupancy data.

Shi, Guan, Zurada, and Manikas (2017) detail how data mining (DM) methods can be utilized in aviation safety management system programs. Specifically, data-mining methods were applied to identify risk factors in commercial aviation by sampling pilot narratives from the NASA ASRS. The researchers initially used topical mining methods to convert the pilot narratives to model input. ML algorithms were then used to incrementally build and assess classification models for risk factor identification. Three different classification algorithms were evaluated. Results indicated the effectiveness for inclusion in an organization's SMS. A limitation of the study was the focus was primarily on the justification of the evaluation methods for various algorithms, rather than on the identification of aviation safety hazard applications.

Finally, Maxson (2018) and Truong, Friend, and Chen (2018) apply data mining methods in their research on flight delay prediction. These studies are particularly significant regarding the building, evaluation, and comparison of various models with the intent of the prediction of flight delays. Specifically, Truong et al. (2018) used FAA Aviation System Performance Matrix (ASPM) data, which were sampled in order to apply decision tree and Bayesian inference modelling, in order to predict the probability of a flight-delay incidents. Maxson (2018) focused his analysis on arrival delays based on input variables related to weather phenomena. Although the sampling of Truong et al. was limited to on-time data at Newark and Miami International Airport and Maxson considered only 10 airports in the NAS, the demonstration of data mining methods in the prediction of a target variable was found to be significant.

Text mining methods. Matthews et al. (2013) developed research on both topic and data mining. The researchers applied these methods to both pilot narrative safety reports as well as flight data. Significant aspects of the study include the description of the Aviation Safety Knowledge Discovery Process (AVSKD). The context of their study demonstrates the use of a structured knowledge discovery process, which harmonizes DM methods in order to identify precursors to aviation safety incidents. Their research used data samples based on raw FOQA data, from aircraft operating in the NAS collected from takeoff to landing. The typical flight generated 5000-6000 samples and 350 flight parameters. DM methods include scalable multiple-kernel learning algorithm for anomaly detection. Significant findings of the research include the application and validation of data mining methods to large commercial aviation datasets to detect precursors to aviation safety events. A limitation of the study was the restriction to data from one aircraft fleet type, thus limiting the generalizability of the results.

Christopher, Vivekanandam, Anderson, Markkandeyan, and Sivakumar (2016); Koteeswaran, Malarvizhi, Kannan, Sasikala, and Geetha (2017); and Kuhn (2018) present the application of text mining techniques to accident data reports. Arockia et al. (2016) developed a classification model for aviation risk mitigation techniques using decision tree methods. While Koteeswaran et al. (2017) also applied several data mining methods to accident report narratives, distinct aspects of the research include the investigation of correlation-based feature selection (CFS) using an oscillating search technique (OST). This technique was used to select attributes that could be important factors in contributory accident causal identification. Both studies utilized feature selection, which is normally done by searching attribute subsets and evaluating each one, and decision tree and Bayesian networks for classification of aircraft accident factors. The studies contrast in the findings of model accuracy effectiveness. Christopher et al. (2016) concluded that decision tree models performed the best regarding classification accuracy and lower error rates in misclassification, while Koteeswaran et al. (2017) determined that cluster computing, using the feature selection algorithm, Improved Oscillating Correlation based Feature Selection (IOCFS), was superior to decision tree models, in terms of classification accuracy. Limitations to these studies are the reactive nature of hazard identification and the lack of predictive capabilities inherent in their modelling techniques.

Finally, Kuhn (2018) builds on previous research with the use of structural topic modeling NASA ASRS data. A key feature of the research is that although the study continues with previous work on the classification of accident causal factors, results also revealed the ability to identify and predict relationships in aircraft accident contributory factors. Even though the study was limited to ASRS narratives involving fuel system and landing gear anomalies, results also revealed unstable approaches to San Francisco International Airport. The unintended identification of unstable approach instances provides an opportune juncture in the body of knowledge for the investigation of flight data monitoring (FDM) or FOQA data for the purpose of addressing unstable approach occurrence and how to predict these occurrences.

Data Mining and Machine Learning Techniques

While data mining is a relatively broad term, a key descriptor includes computer-driven data analysis techniques and applies artificial intelligence in the exploration of large data in order to discover patterns or relationships that might be used in predictive modelling. Large, messy data sources may be “cleaned” and structured in a more feasible format for the purposes of applying data mining techniques. Models can then be constructed in order to represent theoretical relationships of interest. The models can subsequently be used as a comparison tool to try and justify evidence for further analysis (Dubey, Kamath, & Kanakia, 2016).

Gera and Goel (2015) assert that data mining is part of a more general process based on the discovery of knowledge pertaining to large data. They further describe the idea that several sources of data can be exploited simultaneously and introduce the concept of dynamic and static sources of data. Dynamic data sets, like those generated by FDR data of commercial aircraft operations in the NAS, are particularly important to aviation research. The use of dynamic data could be appropriate in applications with continuously changing environmental conditions inherent in the NAS.

Al Ghoson (2011) compares the strengths and weaknesses of several commercially available tools in the application of DM techniques: SAS® Enterprise Miner™, SPSS® Clementine™, and the IBM DB2® Intelligent Miner™. The researchers assert that many business application packages favor decision tree and clustering analyses in the decision-making process. Their research criteria included the following in their evaluation: a) performance, b) functionality, c) usability, and d) auxiliary task support. Based on these criteria, the researcher concludes that SAS® Enterprise Miner™ encompasses nearly all aspects of data mining, to include: text mining, simulation, predictive modeling, optimization, and experimental designs.

Bharadwaj et al. (2013) detail a multifaceted process of discovering and describing unusual events as Anomaly Detection. They use the term synonymously with unusual occurrences, outliers, and surprises. The researchers further assert that the process encompasses several attributes, including the type of anomaly, the nature of the data, and the handling of uncertainty inherent in the system. Contextual anomalies describe abnormal occurrences as defined by guidelines or expectations. For example, an unstable approach is considered an anomaly in the context of this research, as defined by any exceedance of limitations presented in FAA AC 120-71A and later AC 91-79A. Hence, anomaly detection can be described as occurrences that do not fall into normal regions of expectations or standards. In terms of complex systems with multiple regions of normal behaviors, such as flight operations (takeoff, cruise, descent, and approach and landing phases of flight), anomaly detection describes operations that do not fall into these regions. The researchers further assert that abnormal behaviors may appear as clusters that are discernible from normal clusters. Thus, these clusters can become the framework which describes clustering algorithms models.

Das, Mathews, and Lawrence (2011) and Gorinevsky, Mathews and Martin (2012) discuss machine learning methods as a basis for addressing data-driven anomaly detection problems associated with large data. Gorinevsky et al. (2012) describe the evolution of ML from traditional statistical process control (SPC) to supervised ML methods. Supervised ML methods utilize training data labeling in order to build predictive models to detect abnormalities. Examples of classification and regression supervised anomaly detection methods include: (a) decision tree, (b) neural network, (c) logistic regression, (d) SVM, (e) random forest, and (f) gradient boosting algorithms.

Decision tree. DTs are a flowchart type of decision support tool and have become popular in machine learning. It provides a visual representation in decision analysis also used in modeling event outcomes and is commonly used in ML processes to display algorithms consisting of conditional control statements (Tufféry, 2011). Decision tree models consist of internal nodes, which represent analysis on model attributes. Branches represent the results of the process, with each leaf node representing classification labels. Tufféry (2011) describes advantages of decision tree models as easy to understand and analyze, and can be conveniently combined with other analysis techniques. However, disadvantages include being relatively unstable and often inaccurate when compared with other techniques. Additionally, decision tree models tend to over fit and tend to bias favoring attributes with categorical variables.

Tufféry (2011) describes decision tree as an iterative process that divides a given population into segments, based on variables that encompass distinctive qualities within the population. The process begins with the formation of the root, or parent, node. Subsequent nodes are called child nodes, which are then further segregated into intermediate nodes, if applicable. The end cycle consists of terminal nodes, or leaves, which, when integrated with previous segments, indicate a branch of the tree. Training data are then used to calculate probabilities for each node. These probabilities are based on a node rule, which is established using the selection variable value set. These value sets are referred to as targets and a decision refers to the selection of a variable at each node. With business decision making applications, decision trees are typically used for the purposes of minimizing loss, maximizing profit, or to reduce classification error (Maxson, 2018; Sarma, 2013). Maxson (2018) provides details on the use of decision tree predictive modeling in his study focusing on the prediction of arrival rates at 10 airports in the NAS.

Misclassification, lift, and ASE are used to determine tree value. Ultimately, the goal is to minimize cost and maximize profit regarding decisions. ASE is only appropriate in cases with a continuous target variable. Training, validation, and test datasets are partitioned based on the size of the sample. In cases of relatively small sample size, 40/30/30 or 50/25/25 percentage splits are used to train, validate, and test data subsets. The validation dataset is also commonly referred to as the pruning dataset. Larger training datasets usually result in consistent parameter estimates. The training dataset: assists selection based on specific guidelines at a node, conducts probability estimates at each node, and determines the decision variable value at the node. The tree is pruned using the validation dataset and the optimal tree generates the highest profit. Tree worth is determined by comparing splitting values. Performance assessment is based on the training data set and is used in model comparison (Maxson, 2018; Sarma, 2013).

Neural network. Often referred to as artificial neural networks (NN), they are models that process information between multiple layers. Tufféry (2011) describes the broad application of neural networks in the application of clustering, classification, and predictive model designs. The neural network model represents a series of source nodes that circuitously transport data to output layers of neurons. Additionally, intermediate layers may be found between the source nodes and output neurons that process data prior to the outer layers. Sarma (2013) asserts that a NN model transforms variables and performs model estimation. NN modeling is based on an iterative process, where the data source can be transformed and processed.

SAS Enterprise Miner has a menu of options which allow for the selection of a combination of hidden layer or target layer function, with each generating a new NN model (Sarma, 2013). The AutoNeural node can select activation functions for a variety of multilayer networks. The DM Neural node selects the highest performance rated input variables to fit a non-linear solution using R^2 assessment of the linear regression on the important input factors (Maxson, 2018).

Regression. With the examination of a binary target variable, logistic regression is recommended by Tufféry (2011) based on several factors. First, logistic regression can handle dependent variables with two values, without making as restrictive assumptions necessary. Logistic regression has been shown to be highly reliable, with the reliability relatively straightforward to monitor using several available statistical indicators. Logistic regression is also highly generalizable, widely interpretable, and robust. Tufféry (2011) describes several advantages of logistic regression: appropriate for discrete, qualitative, or continuous independent variables; ordinal or nominal dependent variables; requires less restrictive assumptions (compared with linear regression) of multinormality or homoscedasticity of the independent variables; and can provide very accurate models (Tufféry, 2011). Logistic regression also allows for interactions among independent variables. This is important to the research in that unstable approach criteria involve several independent variables. One of the most significant advantages of logistic regression is that it directly models a probability, which is a key point in the research. Although logistic regression has some disadvantages, such as the requirement that the explanatory variables must be linear independent, and sensitivity to missing values of continuous variables, these disadvantages are expected to have little impact to the research. Tufféry (2011) continues to suggest that regression analysis is very useful when the collinear relationships could exist among the important predictor variables or the observations are exceeded by the selection of variables.

Logistic regression analysis, or logit regression, consists of a logistic function to model a binary dependent target variable. The binary dependent target variable in the study, UARM, has only two possible outcomes, the presence of UARM, or the lack of presence of UARM, valued as “0” or “1”. In the logistic regression model, the logarithm of the odds for the presence of UARM, or “1”, is a linear combination of the independent, or predictor variables. The predictor variables can be either continuous or categorical. The logistic function converts the log-odds to probability, with the unit of measurement being a *logit,* or logistical unit. Logistic regression was binomial in this research, meaning the target variable has only two possible outcomes, hence the target variable is a binary categorical variable. Binary logistic regression was used to predict the odds of UARM occurring based on the values of the independent predictor variables. Because logistic regression is used to predict a categorical target variable, rather than a continuous target variable, as in linear regression, the assumptions of linear regression may be discarded, particularly normal distribution of residuals, are violated (Tufféry, 2011). Tufféry (2011) continues to assert that logistic regression is analogous to linear regression, but differs in the relationship between the independent and dependent variables. Significant differences include the prediction of values are probabilities, “0” or “1”, rather than the values of the outcomes themselves. Subsequent descriptions of model fitting and regression coefficient estimating were provided in detail, as well as an examination of the contribution of individual predictors of UARM. The odds ratio was used to examine predictor effects on the exponential function of the regression coefficient (Tufféry, 2011). Several of the most notable tests of significance regarding important predictor variables are the likelihood ratio test and the Wald statistic (Tufféry, 2011).

Truong, Friend, and Chen (2018) support logistic regression methods with a contrasting description of multiple logistic regression. Maxson (2018) provides evidence supporting techniques used by the team, even though his research uses linear regression modeling with a continuous target variable. The researchers assert that multiple logistic regression is commonly used in predictive modeling and is an appropriate method to use to predict the value of a dependent variable based on multiple independent (predictor) variables. However, linear regression necessitates that no missing values be present, as well as the very restrictive assumptions on normality, linearity, homoscedasticity, and non-multicollinearity. Although Tufféry (2011) continues to assert that while logistic regression is a form of multiple regression, an important distinction contrasting with the work of Maxson (2018) is that it has an outcome variable that is a categorical variable and predictor variables which are continuous or ordinal. Binary variables are present if only two categorical outcomes can result; with more than two possibilities, the regression is considered multi-nominal, or polychromous. The linearity assumptions are implicitly violated when a categorical dependent variable is chosen. However, this issue can be overcome by transforming the data logarithmically. In other words, the binary dependent categorical variable is transformed into a continuous curve. While different approaches can be taken, generally it is assumed that the dependent variable is based on probabilistic outcomes that vary from zero to one. In a binary logistic regression, if the outcome probability is close to zero, the determination is that outcome "Y" did not occur, whereas if the outcome probability is close to one, the outcome "Y" did occur.

Sarma (2013) reports the SAS® Enterprise Miner™ software defaults to a logistic regression if the target variable is binary, ordinal, or nominal. If the target variable is binary, the regression defaults to logit link; if there are more than two categorical outcomes for an ordinal target variable a cumulative logits link is used, if there are more than two categorical outcomes for a nominal target variable a generalized logits link is employed. These include: (a) Akaike Information Criterion (AIC), (b) Schwarz Bayesian Criterion (SBC), (c) Validation Error, (d) Validation Misclassification, (e) Cross

Validation Error, (f) Cross Validation Misclassification, (g) Validation Profit/Loss, (h) Profit/Loss, (i) Cross Validation Profit/Loss, or (j) no classification criterion.

Support vector machine (SVM). This is a machine learning algorithm technique form of supervised learning. SVMs can be used in data analysis for classification and regression. A SVM model is built using a training algorithm involving an example training set of data. The data is partitioned into two categories, and the SVM algorithm can build a model that assigns data into one of the two categories. SVM models can be used as both non-probabilistic binary linear classifiers, as well as with probabilistic classifications. A SVM model provides a visual map of data points in space, depicted so that the data points representing the separate categories indicate a clear gap, as wide as possible. The data is then mapped into the model so that it can be predicted to fall into one of the two categories. SVMs can be used for both linear and non-linear classification. A kernel trick is used in non-linear classification SVMs, in order to depict the data into multi-dimensional feature (Chidambaram & Srinivasagan, 2018).

Oehling and Barry (2019) present the use of ML techniques to detect unknown occurrences in flight data, generated by approximately three hundred aircraft, from six different Airbus A320 fleets and sub-fleets, for over 1000 flights per day, from March 2013 to March 2016. The researchers introduce methods enhancing the safety knowledge discovery process. They continue to describe ML in terms of algorithms which learn from the data. The researchers assert that effective uses have been demonstrated with software which builds models, based on input data, rather than a predefined model which was encoded in the software during algorithm development. The study divides ML subcategories into both supervised and unsupervised learning. The contrasting categories are distinct in that supervised learning is based on previous knowledge of the solution. Unsupervised learning does not possess pre-knowledge but is commonly used to structure large data in clusters, or outliers. Other examples of unsupervised learning include: (a) outlier detection, (b) extreme value analysis, (c) probabilistic and statistical model-based approaches, (d) proximity-based approaches, (e) angle-based approaches, and (f) artificial neural network modeling. Specific goals of their ML methods addressed the following requirements:

- Detect unknown occurrences: The model should not rely on pre-determined criteria but use the entire data to find safety-related events. Contrasting exceedance monitoring systems, the goal is to detect previously unknown false negatives.
- Handle large data: The model should be able to process millions of flights and provide safety departments results within two to three days.
- Handle diverse data: The model should not be restricted to a limit on aircraft type or airports.
- Produce useful results: The model should have significant practical and theoretical implications. (Oehling and Barry, 2019, p. 90)

Lauer and Bloch (2008) assert that a key element in the incorporation SVMs as a state-of-the-art performance application regards two types of prior knowledge: class-variance and knowledge of the data. Class-variance applies to transformations, permutations and in domains of input space, contrasting with knowledge of unlabeled data, imbalances in the training set, or the quality of the data. The researchers continue to describe a recent method, which was developed for support vector regression, and considers prior knowledge on arbitrary regions of the input space. Significant contributions to the literature include the importance of prior knowledge in ML techniques implies that to gain improvements in model performance, some prior knowledge about the research problem is necessary. Specifically, Lauer and Bloch (2007) present three methods for incorporating prior knowledge:

- Sample methods: incorporate prior knowledge by either generating new data or modifying existing data accountability;
- Kernel methods: incorporate prior knowledge in the kernel function either by creating a new kernel or selecting the most appropriate one;
- Optimization methods: incorporate prior knowledge in the problem formation either with problem constraints or by defining a new formulation based intrinsically on the prior knowledge. (p. 1584)

Biswas, Mack, Mylaraswamy, and Bharadwaj (2013) describe machine learning approaches as a basis for addressing data-driven anomaly detection problems. The researchers assert that supervised and unsupervised machine learning methods can be effective to detect anomalies in nominal situations. Decision tree classifiers, neural network, regression, SVM, random forest, and gradient boosting are presented as examples of both types of anomaly detection methods. The researchers present a contrasting anomaly detection algorithm, which uses a semi-supervisory learning approach to explore fleet-wide aircraft flight data segments or phases in order to discover deviations from a nominal model of the data. The researchers continue to assert that human experts assist in the semi-supervised modeling process because of their effectiveness in preventing classifications errors based on differentiating criteria between nominal and abnormal.

Das et al. (2010, 2011) identify one-class SVM as a popular semi-supervised anomaly detection technique. The researchers describe the one-class SVM as an extension of SVM applications, which optimizes the classifier for a single class label. This optimization performs data segregation based on boundary criteria, based on the training data, and can suffer from limited information for the SME. Another difficulty described by the researchers concerns noisy training data creating a poor decision boundary in the classification process, thus introducing classification error in the modeling process. One mitigation strategy is for the SME to clearly define segregation criteria (Biswas et al., 2013).

A One-Class SVM classifier can be constructed from a Multiple Kernel Anomaly Detection (MKAD), semi-supervised method for anomaly detection. The algorithm processes the flight data into symbolic feature sequences, so that measures can be applied to comparing the similarity between samples. Das et al. (2010, 2011) describe this transformation as a significant challenge in the knowledge discovery process. The MKAD approach has been proven effective to the application of anomaly detection in FOQA data to a fleet of aircraft. The assumption that SVM is constructed from nominal data and used to discriminate and segregate non-nominal data. The researchers continue to show that SVM modeling has been effective in anomaly detection of flight data, with examples including high energy approaches, pilot responses to external environmental conditions, and high energy, low altitude flight conditions.

Mendes (2012) and Smart (2012) provide examples of supervised learning ML techniques using SVM to investigate 629 automatic landings and 1518 flights into one airport, respectively. Their research indicates that anomaly detection using ML techniques could be more efficient and effective when compared to conventional FDM methods of exceedance algorithm analysis. Ju and Tian (2012) discuss in more detail how the introduction of knowledge-based SVM via nearest point can incorporate prior knowledge in support vector machines. They assert that SVMs can be highly effective in the data mining process by constructing hyperplanes with a large separation margin, and hence, lower generalization error of the classification. The researchers demonstrate how effective measures can be used with prior knowledge by transforming boundary points in order to compute the shortest distances between the original training data and the knowledge data sets.

Hu, Zhou, Xie, and Chang (2016) establish a model to predict the occurrence of hard landing flight safety events. The researchers based their study of the use of QAR-collected flight data and determined that nine aircraft variables were relevant in landing phase. The featured variables used in the study were (a) radio altimeter height AGL, (b) aircraft pitch angle, (c) aircraft pitch angle rate of change, (d) groundspeed, (e) longitudinal distance to touchdown, (f) elevator flight control surface displacement, and (g) vertical acceleration. The team selected the flight variable *radio altimeter height AGL* to partition the flight data, and vertical speed was selected as the output variable. Thus, vertical acceleration was determined to be the target variable, or the predictive target, and the other variables were considered input data for the SVM model. Factor analysis was used to select relevant input variables to build the predictive model. A SVM predictive model was then constructed to predict the occurrence of hard landings. Important results of the study include the confirmation that increased efficiency and accuracy of the predictive model, based on SVM techniques, can be provided by solving feature selection and parameter optimization issues. The researchers assert that feature selection could be enhanced by recursive feature elimination. Additionally, results of the study indicate that parameter optimization was solved in practice by using the application of a grid-search algorithm. This grid-search algorithm was shown to select the model parameters, which subsequently set the range for higher accuracy. The researchers conclude that SVM prediction accuracy improvements could be demonstrated through optimized parameters, thus improving the prediction rate of hard landings.

Ju and Tian (2012) introduce knowledge-based SVM via nearest points (NPKBSVM), to address the context of prior knowledge in SVMs. Their research indicates that SVMs can play a significant role in data mining methods. They continue to assert that the construction of hyperplanes can be used for classification and regression, among other tasks. Their research describes that, regarding classification problems, increased separation can be accomplished by the hyperplane, which lowers the generalization error of the classifier by increasing the margins. Concurrently, they build on the work of Lauer and Bloch (2008) with the introduction of the kernel function, which addresses non-linearity. They also indicate that several methods of SVMs will continue to emerge because of demands, which will allow SVMs to solve problems efficiently and effectively.

Chidambaram and Srinivasagan (2018), Ju and Tian (2012), and Lauer and Bloch (2008) agree that prior knowledge and nonlinear sets can be integrated into SVMs as linear constraints. The researchers also claim that this can lead to optimization challenges that necessitate the use of sophisticated convex optimization tools. Incorporating the use of prior knowledge into the computation process of shortest distance as measured from difference in the training data points and the knowledge sets produces an amended set of training data. The researchers conclude that current SVM tools can be used to achieve significantly improved optimization levels using prior knowledge advantages.

Lauer and Bloch (2008) summarize how the integration of prior knowledge into SVMs can be crucial in order to enhance the performance of SVMs in many applications. The researchers provide a review of studies that utilizes the two general types of prior knowledge into classification tasks: *class-invariance* and *knowledge on the data*. The former describes transformation invariances in domains of input space, and the latter contains knowledge on the quality of the data or accuracy issues in the training data set. Lauer and Bloch (2008) review the uses of prior knowledge with a brief description:

> Prior knowledge refers to all information about the problem available in addition to the training data. Several classifiers incorporate the smoothness assumption that a test pattern similar to one of the training samples tends to be assigned to the same class. Also, choosing the soft-margin version of SVMs can be seen as a use of prior knowledge on the non-divisibility of the data or the presence of outliers and noise in the training set. However, in both cases, these assumptions are intrinsically made by the SV learning and are thus excluded from the definition of prior knowledge. In machine learning, the importance of prior knowledge can be seen from the no free lunch theorem, which states that all the algorithms perform

> the same when averaged over the different problems. Thus, it implies that to gain in performance one must use a specialized algorithm that includes some prior knowledge about the problem at hand. (p. 1,581)

Key contributions of Lauer and Bloch (2008) to the literature include their conclusions that the inclusion of prior knowledge via SME input expert knowledge can improve future SVM model performance. The researchers go on to assert that this in turn could also improve classification performance, thus enhanced practical implications. They recommend continued research in order to explore other forms of prior knowledge, together with optimized algorithms for their implementations. The researchers also describe how the combination of different types of knowledge might be explored for practical applications.

Finally, Mendes (2012) builds on the research regarding SVM methods with his study on anomaly detection using ML classifiers exploring flight data. The researcher used SVM techniques to investigate automatic landings in A320 aircraft over a period of two months, exploring 359 flight events. Specifically, the researcher investigated how both classification models can predict both normal and abnormal flight characteristics of an aircraft autoland system as determined through the analysis of flight data. The research uses exceedance-based criteria set by the aircraft manufacturer then explores the use of algorithms to detect atypical or outlier values. The study details how this approach allows the detection of those situations in an unsupervised learning environment, thus increasing efficiency. Important results of the study include the finding that principal component analysis (PCA) improved the correlation between dimensions. A linear relationship between the features was enhanced through the reduction of variance, thus leading to an increased anomaly detection rate. The researcher also recommended that future research be conducted into the cases that PCA detected, in addition to the detected SVM cases. The researcher concluded that PCA and SVM, when used in combination, provided an optimized solution regarding the final case labels. Finally, the study indicates that a collaboration of these methods allows robust detection improvement.

Random forest. Sometimes referred to as random decision forests, RFs are another example of machine learning techniques using ensemble learning methods for regression and classification. They operate by building numerous decision trees in the data training phase of the AVSKD data processing model. Advantages of RFs include the ability to mitigate the tendency of decision trees of overfitting their training data set. Breiman (2001) describes the construction of numerous, unrelated decision trees using the Classification and Regression Tree (CART) model. CART was developed by Breiman with trees for both classification and regression purposes. The basic premise of CART is that the product of the process is the determination of where to split the trees, using ensemble methods to accomplish the construction of multiple trees. Breiman introduces RFs as a classifier focusing on bootstrap aggregating, which constructs the trees by iteratively sampling the training data set, and eventually forming a consensus predictive value Breiman (1999). Breiman asserts that one of the advantages of this bootstrapping technique is better model performance, achieved by decreasing model variance, while not allowing an increase in the bias. Breiman continues to describe another advantage of bootstrapping as a way of addressing correlation issues with similar trees, achieved by using sampling strategies from different training sets (Breiman, 2001). RFs are distinguishable from bagging in that RFs create a random subset during feature selection, thus identifying strong features regarding prediction of the target variable indicating strength of correlation among predictor and target variables.

Mathews et al. (2013) detail several challenges in the AVSKD data processing model, particularly those challenges dealing with missing values. They provide research indicating that missing values are often a significant factor in the examination of FDM data. Hapfelmeier and Ulm (2014) provide research on mitigation strategies when considering missing data in RF model development. The researchers detail variable selection strategies when considering missing values. They suggest that RFs could be utilized in the variable selection process to enhance data analysis and predictive accuracy. The researchers continue to suggest that several solution strategies could be used to address missing values: multiple imputation, case analysis, and a significance metric. Results of their study indicate that RFs achieve the best results with a self-contained metric in selecting relevant variables regarding the predictive model construction. The researchers conclude that when data contains missing predictor values, measures like multiple imputation or self-contained metrics could be used. The researchers recommend that when using RF modeling with missing values, self-contained metrics can be used successfully to select variables which are of relevance for prediction.

Paul and Dupont (2015) build on the work on Breiman (1999, 2001) with their study on statistically significant feature selection problems. The team presents the results of experiments using RFs to demonstrate how these modelling techniques can address problems associated with feature selection of predictor variables. The researchers assert that RFs can be used to analyze embedded variables selected in the feature selection process. They propose a statistical process in order to measure variable importance regarding significance in the interaction with other variables in RFs. Results of the study demonstrate the correct identification of relevant variables, using Breiman's index of importance (Breiman, 2001). Significant results of the study include the ability to calculate p-values during the iterative RF model building process. In this way, the researchers detail how to discover significant predictor values while minimizing false identification. The researchers conclude that the RF predictive modelling processes consistently perform better than other modelling techniques, without sacrificing performance.

Genuer, Poggi, Tuleau-Malot, and Villa-Vialaneix (2017) and Singh, Gupta, Sevakula, and Verma (2016) conducted research comparing RFs with other ML algorithms, as applied to very large data. Genuer et al (2017) investigated and compared the performance of several ensemble ML techniques, such as linear regression models, clustering methods and bootstrapping schemes. The researchers build upon theories presented by Breiman (2001), who introduced RFs as decision trees integrated with aggregation and bootstrap ideas, which are effective regarding the consideration of regression problems and two-class as well as multi-class classification analysis. Genuer et al. (2017) also compare and contrast several variants of RF modeling algorithms on datasets consisting of 120 million observations. Variations in the modeling include parallel implementation, several different forms of bootstrapping, and various subsampling techniques. Additionally, Genuer et al. (2017) included a sample experiment detailing the application to real-world data related to flight delays. These data were used mainly for descriptive purposes rather than aviation research. Singh et al. (2016) contrast this approach, while focusing on Gaussian mixture model (GMM), logistic regression, and RF classifiers used in mining large data. The researchers analyzed these algorithms while comparing run time, test accuracy, and number of mappers on large data sets. The study focused on three big data sets comprised of over two million sets of data. Results of the study detail RFs perform best in terms of accuracy. Although Genuer et al. (2017) also assert that RFs perform best when trees are diversified, the researchers recommend that improvement in RF performance should be attained when improvements regarding diversity are defined in initial RF iterations.

Lee, Park, and Jung (2014) describe aviation research using RFs and similarity measures to build predictive models that determine aircraft system fault detection. The researchers conducted research on unmanned aircraft vehicle (UAV) fault detection systems using a fault detection algorithm they developed. Fault decision was conducted with the calculation of prioritized similarity measure. The impact of predictor variables was determined and weighted using RFs. The researchers identified 89 data variables in feature selection, 39 being variables associated with normal operations, and 51 variables detailing non-normal flight conditions. Prior knowledge and familiarity with the data were used as factors in the determination of impact variables. The fault detection models were built using both lateral and longitudinal flight control surface related variables. Results achieved with the performance of RFs indicate the impact of each feature or parameter. Several flight tests were conducted in order to validate the fault detection models. Significant results of the study include the demonstration of the feasibility of fault detection algorithm using RFs. Additionally, high fault detection rates were observed, as well as the validation of similarity measure decision results. The researchers note that the flight control system of the UAV could be enhanced without the requirement of additional sensors, thus saving aircraft weight and complexity.

Gregorutti, Michel, and Saint-Pierre (2015) propose a new approach to the use of RFs, with emphasis on multiple Functional Data Analysis (FDA) and the grouped variable importance measure, used in their study to predict aircraft landing distance. The researchers assert that several groupings of basis coefficients could be utilized considering specific functional disposition. Unique to the research model is the approach to the selection RF algorithm, when considering the grouped variable importance. The computational requirements of the RF algorithm were decreased by regrouping the coefficients, which contrasts to the iterative elimination of grouping coefficients usually demonstrated in RF modeling by eliminating one coefficient in each iteration. The team also describe how simulation studies demonstrate scale of the grouped importance regarding measurement for FDA. The research team then applied the resulting RF algorithm in the feature selection of important factors in long landings to explore and predict aircraft landing distances. The team consulted with several SMEs, who determined that 23 variables were relevant to landing distance. 1868 flights were used, operating at one airport by one airline, to collect data. RFs as well as a permutation-based importance measures were used for variable groupings. The researchers also compared and assessed models using bagging, neural network, and SVM ML techniques. Key contributions of the study include the unique application the team used to select operational variables using the measure of grouped variable importance and a predictive modeling approach. Additionally, the RFs were adapted in a backward context regarding iterative elimination. This variable selection technique was also adapted to the unique measure of grouped importance in order to predict the target variable, landing distance. The researchers conclude that future studies could incorporate their group variable importance projection technique into flight data analysis to assist airline flight training managers to develop SOPs as well as enhance pilot training.

Blair, Lee, and Davies (2017) use RFs to build models used by aircraft to detect real time inflight aircraft damage. The researchers developed techniques for detection using real-time classification of flight trajectories distinguishing normal from abnormal aircraft. The team demonstrated an efficient computational approach using RFs, with the key factor being the ability to provide accuracy necessary for fault detection and classification. The methods were tested using a full motion 757 flight simulator at the NASA ARC research center. The team used RFs in their approach to detect fault in six flight scenarios in the simulator. Principal Component Analysis (PCA) was used in the classification study as a dimension reduction technique. The chosen methodology integrated several statistical techniques: principal component analysis, random forest, and cross-validation, which the team asserts produced a reliable and fast classifier. The team then used sliding window approach, which the team describes as a sequential approach to the problem, while concurrently learning temporal and probability thresholds, by training a validation data set. The team claims their methods resulted in a 98.7% predictive accuracy rate, and also provided fault detection at half-second time intervals. This key result from the study indicates that faster fault detection times could be achieved with increased computational capabilities, which would ultimately assist the pilot in real-time SA enhancement.

Tong, Yin, Wang, and Zheng (2018) analyze and process QAR using RFs to predict the landing speed of commercial aircraft. The researchers collected data from the QAR devices onboard Airbus 300 aircraft from a single operator based on 2000 flight segments. The team asserts that missing data was a significant challenge in the data processing task. In order to address this challenge, the researchers resampled the data a one frame per second. The team then used RFs to sort the 69 candidate features in order of importance and selected the 20 most relevant features regarding landing speed. The RF process resulted in 250 trees determining the top 20 features. Top features included engine speed (thrust), angle of attack, and descent rate. The research team asserts that their most significant result was the ability to predict landing speed, which they describe as a causal factor in landing accidents. Additionally, the team describes how RFs were able to extract the most important features. The team recommends that AI be incorporated into future studies in order to process the sensor data.

Lv, Yu, and Zhu (2018) provide research into hazards associated with RE based on excessive landing distance. The researchers used SVM, RF, and logistic modeling of QAR data from 6,000 Chinese Airlines A320 flights. The researchers present results that indicate excessively long landings are significantly correlated with pilot flare technique and runway distance. Limitations in the study include lack of standards applied in the feature selection phase of the knowledge discovery process as well as the lack of SME involvement in the selection of impact variables.

Finally, Kumar and Ghosh (2019) detail the use of CART and RFs in their research into predicting unsteady aerodynamics model using quasi-stall flight test data. The researchers assert that RFs were more effective regarding scalability that NN models. The team developed RFs model specifically to model lift, drag, and pitching moment aerodynamic models. The team chose to use non-parametric models using the flight data. They describe that non-parametric models do not include physical parameters or parametric equations like parametric models do. The team present CART and RF data-driven modeling approaches by understanding the gathered flight data and using this prior knowledge in feature selection. The researchers agree that advantages of RFs include avoiding overfitting problems often encountered by decision tree, diversity regarding both classification and regression, and important predictor identification based on training data. Results compared favorably with maximum likelihood estimation (MLE) predictions. The team declare that the key conclusion in the study was that RFs are superior to CART, with each method more desirable alternative to the parametric approach inherent in MLE. The team states that RFs could be scaled and applied to many other applications of nonlinear modeling.

Gradient boost machine. One of several recently developed machine learning algorithm-based techniques, gradient boost machines (GBMs) are used mainly for classification and regression. The product of this technique is prediction models formed by integrating weaker prediction models, usually decision tree. GBMs typically build the model in an iterative process. Additionally, GBMs are able to generalize the integration of the weaker individual models, which can produce an optimized differential loss function (Breiman, 1997). Breiman (1997) and Freidman (2001) developed early gradient boosting optimization algorithms, mainly based on cost analysis. Friedman (2001) continued research using optimization algorithms, conducting significant research on algorithms expressly for work on regression using gradient boosting. Although his research initially observed boosting algorithms that optimized cost functions over function space, this functional perspective of gradient boosting preceded development of boosting algorithms in many areas of ML, including aviation research exceeding classification and regression. GBMs integrate multiple weak models into one strong model, with the purpose of predicting a target variable probability using supervised learning. A training data set is typically used to teach a model to predict the target variable in an iterative process that can minimize the mean squared error, for example, in a regression problem. In effect, the iterative process continues to improve as the error is minimized in each iteration and produces a residual, which is a negative gradient. In effect, GBMs incorporate the negative residual into the next iteration. Similar to other supervised learning problems, GBMs produce an output variable and a vector of input variables. The vector, or joint probability distribution, uses the training data set to develop a function in order to minimize loss. The work of Breiman and Friedman has provided aviation researchers with foundational tools to investigate and explore problems associated with predictive model construction using GBMs. A brief description of extant literature pertaining to GBM techniques, as applied to aviation, is provided. To date, the general area of research has been accomplished concerning cost savings and optimization, with some work done on flight path prediction.

Alligier and Gianazza (2018) provide aviation research using GBMs to build predictive models focusing on ground-based aircraft trajectory. Their research models the forces imposed on an aircraft to predict future points of the trajectory of the flight path. They assert that previous knowledge was necessary regarding aircraft mass, thrust, and speed. The researchers applied ML methods in order build models to predict mass and speed as constructs, thus improving the ability to predict their target variable, aircraft trajectory. The data used in the research was provided by the Eurocontrol Base of Aircraft Data (BADA), which was utilized to provide default values for the model parameters. The data were acquired on ADS-B from The OpenSky Network in 2017, using the climbing segments of 11 types of aircraft, resulting in millions of flights worldwide, operating from 1520 airports. The ML techniques were used to predict the most significant factors necessary to calculate the predicted trajectory. The researchers used GBMs to build predictive models for each specific aircraft type. As mentioned, previous knowledge was then required to construct a training data set, with known operational factors. The training set was constructed by separating, or partitioning, the original large set of data. The researchers claim that they achieved two of their stated goals: a non-optimistically biased result, and the demonstration of the use of explanatory, real-time variables in a performance assessment. Significant contributions to the body of knowledge were achieved regarding bias: the climbing segments were not limited by altitude restrictions, and very large data were included in the study, using millions of climbing segments of the 11 most popular aircraft types in commercial aviation. Thus, the researchers demonstrated that ML techniques can be effectively scaled to very large data. The researchers detail that further significant practical implications apply to air traffic control, which could benefit from the results of the study, regarding the ability to predict climb trajectory and vertical separation. Additionally, the ability to predict climb trajectory real-time could enhance flight path track error and improve top of climb prediction.

Research by Thompson (2017) using FAA Comprehensive Electronic Data Analysis and Reporting (CEDAR) data provides an example of the use of ADS-B data in modeling unstable approaches. Thompson (2017) asserts that ADS-B data could be augmented with weather observation data for development of real-time stable approach models. The researcher asserts that these real-time models could be used by air traffic control to calculate the probability of occurrence of an unstable approach and rejected landing. Thompson (2017) also indicates that the capability for ATC to predict an unstable approach and rejected landing could improve safety by decreasing interference with air traffic flow in normal operations. The researchers concur that the ability to predict the occurrence of rejected landings improves air traffic safety. Alligier and Gianazza (2018) and Thompson (2017) agree that future work should include analysis of specific airport operations, at which error in prediction was higher, in order to help improve unstable approach model accuracy.

Achenbach and Spinler (2018) introduce ensemble modeling, combining elements of linear regression and GBM algorithms to generate predictive models. The researchers provide airline arrival time and cost index optimization predictions on European airspace short haul flights. Cost index (CI) is an airline metric used to predict fuel cost and flight time of a flight segment. The researchers describe the cost savings airlines can attain through accurate flight planning based on CI predictive accuracy. The researchers assert that increased efforts by aviation researchers to minimize costs associated with flight delays have faced challenges due to inability to accurately combine arrival time predictions with CI models. Contrasting previous studies, the researchers chose to use a ML ensemble to predict arrival time, but rather than predicting arrival time once the aircraft was airborne, they chose to use gate departure to predict arrival time. Achenbach and Spinler (2018) continue to build on the work of Breiman (1997) and Friedman (2001) using GBMs based on European flight data from 2015 and 2016, from over 200 European airports. The researchers identified important predictor variables regarding weather data, airport congestion, flight levels, and other basic flight planning data. The predictive modeling then focused on arrival time with various CI values. Significant contributions of their study include the first attempt to combine dynamic CI values with arrival time predictions. This combination produced greatly improved ability to airline managers to minimize total cost of a flight segment. Additionally, the study demonstrated the increase in accuracy by using ensemble GBMs. These key contributions provide evidence that ensemble GBMs can improve aircraft arrival time predictions while considering both linear and non-linear relationships of important impact variables, along with the interdependence of these variables. Limitations to the study are that it focused on one airline. The researchers recommend future research include several airlines, in order to apply their model to airlines of different sizes and city pairs. Additionally, CI was calculated based on cost per unit of time, when in reality, often cost of time cannot be determined per minute for all variables. Lack of enroute weather conditions, including wind effects, was a delimitation of the study, and could significantly affect results. Regarding future research, they recommend that extensions of their work include improved cost estimates for not only delays but also cost per unit time be further investigated. The researchers assert that ML techniques should be incorporated with predictive modeling in future studies with the ability to combine cost blocks in order to improve CI optimization.

Kang and Hansen (2018) build on the work of Achenbach and Spinler (2018) with their research on the application of GBMs to improve commercial airline fuel burn prediction. The researchers built predictive models using prediction intervals (PIs) to mitigate the uncertainty of model predictions. The researchers determined that annual cost savings to airlines could approximate $60 million annually, just for one domestic US airline. Additional benefits include the reduction by 428 million kg of CO_2 annually per airline. Data for the study were collected from three sources: flight level performance data from the FAA Aviation System Performance Metrics (APSM) database, the terminal area forecast (TAF) weather information from the National Oceanic and Atmospheric Administration (NOAA) database and flight and fuel statistics from one U.S. airline. The researchers included all available predictive variables to build the models. The target variable for their study was actual fuel burn for a flight segment. The FAA ASPM database was used to gather historical data for the largest 77 US airports, then descriptive statistics were generated to create flight time data between different city pairs. The researchers then describe the various advantages that advanced computing power and new ML techniques have allowed. They describe GBMs with the ability to iteratively improve predictive accuracy by starting with a weak learner base, then continuously adding them together. The researchers contrast GBM ML techniques with other ensemble learning methods such as Bagging, Random Forest, and Stacking. Similar to Achenbach and Spinler (2018), the researchers determined that a significant limitation to their study was the lack of real-time enroute weather data. They assert that current data made available by the airlines is based on the flight planning system (FPS), which is subject to prediction errors, which could significantly affect fuel burn values. The researchers recommend, in future studies, the investigation of predictive model construction, focusing on airline dispatcher decision making. The researchers detail the affects that human decision-making processes could have on the fuel planning process, and raise the question of HIP and ADM regarding predictive model applications.

Finally, Gallego, Gómez, Sáez, Orenga, and Valdés (2018) produced research with the objective of investigating the effects of operational input variables on the vertical flight path trajectory prediction. Additionally, the research team desired to determine what important input variables should be included in the feature selection segment of the knowledge discovery process. The study was based on the use of a data warehouse (DWH) program to construct a comprehensive information management layer. The aircraft trajectory flow models were based on these data, applied to the Barcelona International Airport. ML techniques were used to determine the flow patterns within the DWH. Contrasting other similar studies, the researchers chose to use a set of multilevel linear models (MLMs) which were adopted to investigate vertical aircraft flight path trajectories on descent into the airport. Key discriminators included operational vertical procedures, airline specific procedures, and unique flow patterns to the airport. The MLMs performed linear regression tasks, which included training the independent variables for different groups. The target variable was the rate of descent from the top of descent (TOD) to Flight Level 250. The researchers indicate that results show a correlation between rate of descent and the location of the TOD. The researchers continue to conclude that based on the MLM results, key input variables were determined to be the position of TOD and flow factor, regarding flight path trajectory prediction. Additionally, specific airline operational procedures (SOPs) were not found to be significant factors in the prediction of flight path trajectory. The researchers recommend that future research investigate different airspace sectors for flow similarities. The researchers also agree with Kang and Hansen (2018) and Achenbach and Spinler (2018) that future research should incorporate weather-related data for improving accuracy of the predictive models. Finally, the researchers recommend that GBMs, as well as other data-driven approaches, should be explored in order to more accurately predict aircraft flight path trajectory.

Table 1 presents a summary of comparative aspects of predictive modeling techniques.

Table 1

Comparison of Prediction Methods

Prediction Method	Number of Predictors	Interrelationships among Predictors	Detection of Hidden Patterns in Large Data
multiple regression	limited	continuous	no
logistic regression	limited	binary or nominal	no
simultaneous regression	limited	continuous	no
Decision tree	many	continuous binary nominal	yes
conventional data mining	many	continuous binary nominal	yes
data mining using Bayesian inference	many	continuous binary nominal	yes
support vector machine	many	continuous binary nominal	yes
neural network	many	continuous binary nominal	yes
gradient boost model	many	continuous binary nominal	yes
random forest	many	continuous binary nominal	yes

Adapted from "Applications of Business Analytics in Predicting Flight On-Time Performance in a Complex and Dynamic System," by D. Truong, M. A. Friend, & H. Chen, 2018. *Transportation Journal, 57*(1), 24-52. Copyright 2018 by the *Transportation Journal*.

Gaps in the Literature

A review of the extant aviation research literature on topics pertinent to the study was conducted. The three areas of focus were: (a) federal guidelines and oversight of hazards associated with unstable approaches and runway excursions, (b) aviation research conducted on pilot risk perception and risk tolerance, and (c) aviation research using predictive modelling based on advanced ML techniques applied to large FDR data. While the review describes many examples of aviation research in each of these topics, the recommendations among the aviation researchers conducting the most recent relevant work (approximately last five years) describes gaps and opportunities for future research and is presented here.

Unstable approach and runway excursion hazards. The FAA, NTSB, and FSF have provided oversight, guidance, and/or recommendations to operators regarding the hazards associated with mitigating the risk of runway excursions. The FAA has listed unstable approaches as one of the most common causal factors (FAA, 2014). The FSF and NTSB corroborate this assertion that stable approaches (and safe landings) begin early in the approach planning phase of flight (FSF, 2009; NTSB, 2016, 2019b). These organizations have called for improved pilot training initiatives, enhanced CRM training, as well as research into risk mitigation strategies for operators to avoid the hazards associated with unstable approaches (FAA, 2017b; NTSB, 2016, 2019b). Recent aviation accidents have demonstrated that unstable approaches continue to be causal factors. The NTSB has recommended that the aviation industry respond to the hazard of unstable approaches with improvements in pilot training, as well as the development of CRM techniques to enhance pilot risk assessment and perception in flight operations (NTSB, 2013, 2019b).

Pilot risk perception and risk tolerance. You and Han (2013) describe how the safe operational behavior of pilots can be affected by HF characteristics such as ADM, HIP, SA, and interpersonal communications and teamwork attitudes. The researchers build on the previous work of Hunter (2005) regarding the correlation between pilot risk perception and hazardous events. They used the Risk Perception Scale developed by Hunter to conduct surveys of pilot attitudes associated with perception of risk and the relative levels of safety inherent in airline operations. Results of the study indicate that present research on pilot risk perception and risk tolerance vary with pilot perception of locus of control, or the belief that one has a direct effect of the outcome of a situation. The researchers continue to assert that once a pilot achieves a certain threshold of flight experience, measured in total flight time, then perception of internal locus of control diminishes. Important recommendations for future research regarding pilot risk perception are that future research be conducted to identify key factors that contribute to inaccurate perceptions of risk, as well as the exploration of effects of organizational safety culture on safe operational behaviors.

Ju, Ji, Lan, and You (2017) conducted research addressing recommendations by You and Han (2013) by investigating what factors affect pilot perception of risk. Specifically, the study explored the relationship between narcissistic personality and optimism in aviation risk perception. A key component of the research was to determine whether self-promotion mediated the personality traits of narcissism and over-optimism in pilot risk perception. Results of the study indicated that narcissism had a significant effect on risk perception among pilots, in that overestimation of promotion focus predicts underestimation of risk. The researchers limited their study to that of optimism bias and recommend that future studies address other cognitive biases. The research team also recommends that active airline pilots be included in the study, rather than the inclusion of only flight students. Agreeing with You and Han (2013), the researcher concurs that organizational safety culture effects on pilot risk perception be explored.

Predictive modeling using recorded flight data. The literature provides several examples of research based on anomaly detection to identify abnormal flight events. Bharadwaj et al. (2013) detail a multifaceted process of discovering and describing unusual events as Anomaly Detection. They use the term synonymously with unusual occurrences, outliers, and surprises. The researchers further assert that the process encompasses several attributes, including the type of anomaly, the nature of the data, and the handling of uncertainty inherent in the system. Contextual anomalies describe abnormal occurrences as defined by guidelines or expectations. For example, an unstable approach is considered an anomaly in the context of this research, as defined by any exceedance of limitations presented in FAA AC 120-71A and later refined in AC 91-79A. Hence, anomaly detection can be described as occurrences that do not fall into normal regions of expectations or standards. In terms of complex systems with multiple regions of normal behaviors, such as flight operations (takeoff, cruise, descent, and approach and landing phases of flight), anomaly detection describes operations which do not fall into these regions. The researchers further assert that abnormal behaviors may appear as clusters that are discernible from normal clusters. Thus, these clusters can become the framework which describes clustering algorithms models.

Building upon this work, Li et al. (2015) and Aslaner, Unal, and Iyigun (2016) applied clustering techniques to flight data to identify anomalies in the takeoff and landing phases of flight. The research was designed using two experimental methods: one to sample 91 flight parameters in the effort to identify abnormal flight events and a second to evaluate three different data clustering algorithms. Limitations to the study include a vague description of what constitutes abnormal flight events, variables of interest in the clustering analysis, and the lack of a clearly defined target variable. It is also unclear what coding was used to build the models used in the evaluation of the algorithms. Additionally, the SMEs used in the evaluation of the abnormal flight events do not apply any standard criteria in their analysis. The researchers agree that future research should investigate the application of advanced ML techniques to large FDM data to try and identify previously unknown anomalies.

Gera and Goel (2015) suggest that data mining is part of a more general process based on the discovery of knowledge pertaining to large data. They further describe the idea that several sources of data can be exploited simultaneously and introduce the concept of dynamic and static sources of data. Dynamic data sets, like those generated by FDR data of commercial aircraft operations in the NAS, are particularly important to aviation research. The research team recommends that future research use large FDM data to explore the continuously changing environmental conditions inherent in the NAS.

Tong et al. (2018) analyze and process QAR using RFs to predict the landing speed of commercial aircraft. The researchers collected data from the QAR devices onboard Airbus A300 aircraft from a single operator based on 2000 flight segments. The research team asserts that their most significant result was the ability to predict landing speed, which they describe as a causal factor in landing accidents. Additionally, the team describes how RFs were able to extract the most important features. The team recommends that AI be incorporated into future studies in order to process the sensor data.

Gallego et al. (2018) produced research with the objective of investigating the effects of operational input variables on the vertical flight path trajectory prediction. Additionally, the research team desired to determine what important input variables should be included in the feature selection segment of the knowledge discovery process. ML techniques were used to determine the flow patterns within the DWH. Contrasting other similar studies, the researchers chose to use a set of Multilevel Linear Models (MLMs), which were adopted to investigate vertical aircraft flight path trajectories on descent into the airport. Key discriminators included: operational vertical procedures, airline specific procedures, and unique flow patterns to the airport. The researchers recommend that future research investigate different airspace sectors for flow similarities. The researchers also agree with Kang and Hansen (2018) and Achenbach and Spinler (2018) that future research should incorporate weather-related data for improving accuracy of the predictive models. Finally, the researchers recommend that GBMs, as well as other data-driven approaches, should be explored in order to more accurately predict aircraft flight path trajectory.

Finally, Oehling and Barry (2019) present the use of ML techniques to detect unknown occurrences in flight data, generated by approximately three hundred aircraft, from six different Airbus A320 fleets and sub-fleets, for over 1000 flights per day, from March 2013 to March 2016. The researchers introduce methods enhancing the safety knowledge discovery process. They continue to describe ML in terms of algorithms which learn from the data. The researchers assert that effective uses have been demonstrated with software which builds models, based on input data, rather than a predefined model which was encoded in the software during algorithm development. The study divides ML subcategories into both supervised and unsupervised learning. Supporting the recommendations of Aslaner et al. (2016), Bharadwaj et al. (2013), and Li et al. (2015), the researchers recommend that future studies using advanced ML methods be applied to large FDM data to detect unknown anomalies.

Walker (2017) details the development of the modern QAR. The ability to gather large FDM data with advances in QAR technology has encouraged new developments in advanced data driven techniques such as advanced ML methods. The large data being gathered daily in the NAS by these QAR devices has presented an opportunity for the exploration and investigation of these data. Aviation research using data mining of QAR data has progressed from strictly exceedance based anomaly detection, to both semi-supervised and unsupervised ML methods. The researcher makes an important assertion regarding the concept of *leading* vs. *lagging* indications of measurable precursors to accidents or incidents. Walker makes the distinction between *reactive* discovery based on lagging indicators and *proactive* discovery based on leading indicators. This distinction is significant regarding the transition from reactive to predictive modeling of flight data. Walker continues to describe the evolution of the use of flight data collection devices from post-accident analysis to actively accessing, analyzing, and preventing anomalous flight events.

Das et al. (2010) conducted research using known exceedance criteria and applied the criteria to QAR data using MKAD and CAD to detect anomalous events on commercial aircraft. The data driven study compared several different ML algorithms to detect previously unknown anomalous events. Similar to the study conducted by Li et al. (2016), a structured process was not used in the cleaning, examination, or processing of the FDM data. In both of these cases, SMEs were used to design exceedance criteria, which were then used to identify variables of interest, prior to the application of cluster analysis based ML techniques. These works focused on the identification of anomalous events, as defined by SMEs, rather than a standardized aviation knowledge discovery process based on industry standard or federal guidelines.

Bharadwaj et al. (2013) led a NASA Aviation Safety project to develop data driven, supervised and unsupervised techniques to enhance the diagnostic capabilities of in-flight systems' reasoners. The researchers build on the work of Li et al. (2016) and Das et al. (2010) using supervised clustering techniques to detect anomalous events. The team advances the data process model with a data driven analysis using feature analysis based on nominal data frame comparison. Similar to previous works, the research is limited by the necessity of SME developed exceedance criteria.

Theoretical Foundation

Aviation safety knowledge discovery process. Matthews et al. (2013) introduce the aviation safety knowledge discovery (AVSKD) process, which was used as the model for the research. The AVSKD process describes the entire process for analyzing aviation data from raw FOQA data to reporting of predictive models. The AVSKD framework illustrated in Figure 4 was adapted as the framework for the methodology for the research.

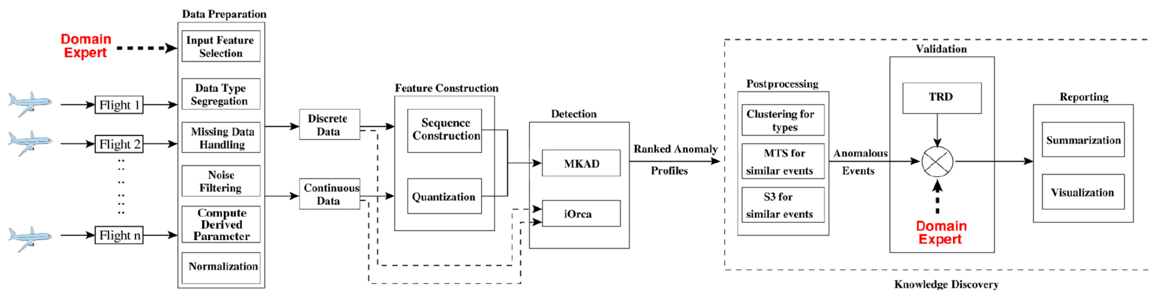

Figure 4. Aviation safety knowledge discovery (AVSKD) process. From "Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms," by B. Matthews, S. Das, K. Bhaduri, K. Das, R. Martin, and N. Oza, 2013, *Journal of Aerospace Information Systems, 10*(10), p. 469. Copyright 2013 by the Journal of Aerospace Information Systems.

The rectangles in Figure 3 on the far left denote the raw flight operational quality assurance (FOQA) data inputs from the aircraft. The data preparation module is where the domain expert typically identifies the types of data and selects, segregates, and normalizes the data variables. Because FAA criteria were used in the feature selection process, these functions as well as SME input, were not required. Additionally, in the feature construction module, the separated discrete and continuous data parameters undergo sequence construction and quantization. The goal of the detection module is to identify anomalous events of interest (outliers) at both the fleet and flight levels using the open-source multiple-kernel anomaly detection (MKAD) and index-Orca (iOrca) knowledge discovery algorithms, respectively. Then ranked data profiles are created from the processed datasets. During the knowledge discovery process in the post-processing module, the frequency and severity of the distance-based events are determined using the multivariate time series (MTS) tool and a sequence similarity search (S3). The anomalies are typically validated using domain field experts, text reports database (TRD), and flight crew interviews; however, FAA exceedance criteria precluded the necessity of a domain field expert, and TRD and flight crew narratives were not available. The last step of the knowledge discovery process is the generation of a report summarizing and visualizing the results.

Data and preprocessing. Raw FDM, or QAR data, from each aircraft, are collected for input into the AVSKD process. Each flight data record is constructed as a matrix with rows corresponding to time sampling, and columns to specific parameters. The sample rate can vary based on complexity of the QAR, with typical rates of 1 Hz and a flight record generating approximately 5000-6000 samples and up to approximately 300 parameters. Both discrete and continuous parameters are gathered. Data preparation is then conducted, with the raw data put through a data preparation module, which performs feature selection, data type segregation, missing data processing, noise filtering, and normalization.

Anomaly detection algorithms. Considering AVSKD is a flexible process, different types of ML algorithms can be integrated into the process. Thus, the algorithms selected and applied to the AVSKD process can vary based on the type of research being conducted. The six ML algorithms previously described were used to build predictive models based on cases determined to contain anomalous unstable approach events. Hui and Fanxing (2012) describe the difficulties encountered in data processing and introduce a symbolic aggregate approximation (SAX) algorithm to enhance the accuracy and performance of anomaly detection in QAR data. Even though the SAX algorithm experienced less than desirable accuracy, the research is indicative of the crucial role of algorithm in QAR flight data anomaly detection capability.

Knowledge discovery. Once anomalies have been discovered, the database is examined for frequency of events and for the purposes of predictive model building. Mathews et al. (2013) assert that validation of the events through domain expertise is typical and considered industry standard. Because FAA guidance was followed as the standard for feature selection, it was evident that the data variables used in the assessment of FAA exceedance criteria were valid. The final step in the AVSKD process is the reporting system. The reporting system consisted of figures representing important impact factors in the prediction of the target variable, UARM. The displays are percentage contributions for discrete variables, graphical plots for continuous variables, and percentage contributions of each continuous variable.

Sample, explore, modify, model, assess (SEMMA). The SAS Institute (Patel & Thompson, 2013) recommends using the SEMMA modeling process with SAS® Enterprise Miner™. Specifically, the SEMMA acronym stands for *Sample*, *Explore, Modify, Model,* and *Assess*. The SEMMA process is iterative in nature, and the *Sample* or *Explore* stages were repeated after assessment of the model in order to make changes and then repeat the *Model, Modify*, and *Assess* processes (Maxson, 2018). This process was adapted for the study as illustrated in Figure 5.

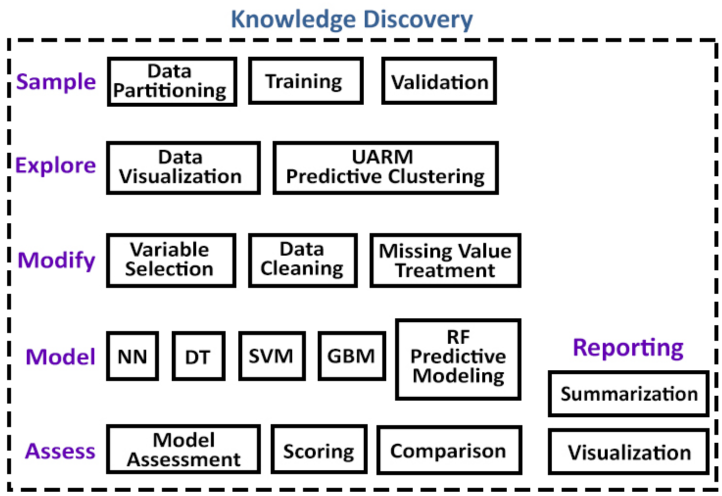

Figure 5. Sample, Explore, Modify, Model, Assess (SEMMA) knowledge discovery. UARM = Unstable Approach Risk Misperception; NN = neural network; DT = decision tree; SVM = support vector machine; GBM = gradient boost machine; RF = random forest. Adapted from "Prediction of Airport Arrival Rates Using Data Mining Methods" (Doctoral Dissertation) by R. W. Maxson, 2018, p. 70. Copyright 2018 by R. W. Maxson.

Sample. To begin the process, the data were inputted into the data mining software as input variables and a target variable is selected, e.g. UARM. The data were then partitioned into model training and validation subsets. The SAS© Enterprise Miner™ is compatible with many data input formats.

Explore. Once the data were uploaded into SAS Enterprise Miner™, the data were examined for missing variables, outliers, and/or skewed or peaked distributions. Several tools were available to use for these purposes; "StatExplore" was used for clustering, correlation, graphical investigation, and variable selection. The operational input and target variable were selected in the *Sample* module, and an iterative process included modifications after the data were inspected in the *Explore* step.

Modify. The purpose of this step was to prepare the data for model construction. This task could have been accomplished in several ways: inputting missing values, filtering the selected data, adding data to the set of data, merging the data with other sources, or partitioning the input data into smaller subsets. In this step, the data was further processed based on model construction requirements. Using the *Model* menu, "AutoNeural" was used to determine the most appropriate action for the NN, for example. Similarly, regarding regression model construction, data could have been transformed or imputed using dummy variables for categorical variables, and missing values could have either been removed or imputed (Maxson, 2018; Sarma, 2013).

Model. In this step, the prepared data were used to construct the models: decision trees, neural networks, regression, SVM, RFs, and GBMs. The software allowed flexibility in model construction depending on the data being used and the problem being investigated. Also, simultaneous model evaluations and comparisons were possible (Maxson, 2018; Sarma, 2013).

Assess. Finally, the models were compared and assessed using the model comparison function in the "Assess" menu grouping. Model assessment capability includes performance scores which are used to rank the models. ASE or misclassification rate could have been used to score the models and Receiver Operating Characteristic

(ROC) and lift curves in the assessment of model performance (Maxson, 2018; Tufféry, 2011).

Summary

Although hazards have been identified with the continued occurrences of runway excursions, a reliable and valid representation of rejected landing decision making based on unstable approach criteria has not been fully investigated (Koteeswaran et al., 2017). Limitations inherent in qualitative methods are evident in the literature, with examples of text mining narratives from LOSA field observations, NASA voluntary ASRS reports, and surveys. The extant literature does not indicate that there is research that addresses or exploits the large amount of aircraft data available (Matthews et al., 2013). Recent studies have described that aviation researchers have begun to realize the application of advanced data mining techniques as an appropriate and powerful tool to handle these voluminous data being generated daily by FDM and FOQA programs (Li et al., 2015; Puranik & Mavris, 2018; Shi et al., 2017). However, these recent studies have focused primarily on the validation and evaluation of advanced mathematical algorithms and leave analysis of safety mitigation information for either further research or with the qualitative assessment of a subject matter expert (Arockia et al., 2016; Li et al., 2015; Puranik, & Mavris, 2018).

Finally, several salient aspects of the literature become evident based on the review focusing on large data and prediction of in-flight anomalies. The first factor that becomes apparent is the large amount of data that are being recorded by advanced digital flight recorders on every commercial airline flight in the NAS. Airlines are encouraged by the FAA to voluntarily participate in the FOQA program. FOQA was designed to improve safety in commercial aviation by allowing airlines and pilots to share de-identified aggregate information with the FAA so that the FAA can monitor national trends in aircraft operations and focus its resources to address risk issues (e.g., flight operations, air traffic control (ATC), airports) (FAA, 2004). Although voluntary (in the United States), the FOQA program has resulted in very large amounts of flight data that have not been accessed on a scale appropriate for these data. Even though pilot safety reports, accident reports, and safety debrief narratives constitute a large amount of data, the literature indicates that these data have only been explored with the use of text mining and qualitative methods. Second, while a review of the literature indicates that various statistical analytical methods have revealed clear patterns in the prediction of pilot performance, these data have not been fully exploited in order to fully investigate significant relationships of the predictors. Third, the extant literature indicates that once the relationship connecting pilot performance to flight anomaly variable inputs was explained, that future pilot performance in the approach and landing phases of flight could be estimated. Finally, although much of the existing literature presents results on the evaluation of complex algorithms applied to large data, subject matter experts have been required to analyze the results and apply them to aviation problems. A review of the existing literature indicates a gap in the research in the application of predictive modeling techniques, particularly the application of these techniques to the prediction of probability of occurrence of factors contributing to pilot misperception of risk. Table 2 presents a summary of key aspects of related literature. This gap in the body of knowledge created an opportunity to provide a contribution with the research.

Table 2

Summary of Primary Aviation Research Using Recorded Flight Data

Study	Machine Learning Model	Data Source	Modeling Approach	Target Variable	Limitations	Findings
Li, et al., 2015	Cluster analysis	FDM	Reactive	Abnormal events	Small sample	Anomaly detection
Mathews, et al., 2013	Cluster analysis	FOQA	Reactive	Abnormal events	Excludes weather data	Anomaly detection
Truong, et al.., 2018	DT & Bayesian network	FAA ASPM	Predictive	On-time performance	Only two airports	Prediction of target
Treder, 2004	DT & NN	FOQA	Reactive	Exceedance criteria	Proof of concept	Anomaly detection
Christopher, et al.., 2016	DT	Mishaps	Reactive	Classifcation	Limited accident data	Classification model
Tong, 2018	RF	QAR	Predictive	Landing speed	Small sample	High predictive accuracy
Oehling, 2019	NN	FOQA	Predictive	Unknown safety events	One aircraft type	Unsupervised knowledge benefits

Note. DT = Decision Tree. NN = Neural Network. RF = Random Forest. FDM = Flight Data Monitoring. FOQA = Flight Operational Quality Assurance. ASPM = Aviation System Performance Matrix. QAR = Quick Access Recorder.

CHAPTER III

METHODOLOGY

The purpose of this research was to utilize data mining techniques to explore a large-volume database of flight data recorder (FDR) data from commercial flight operations to predict *Unstable Approach Risk Misperception* (UARM). The focus of this exploration was on the data generated beginning at the assessment window (500 ft AGL) to a point of either a landing or a rejected landing. The complete list of variables recorded by the FDRs is listed in Appendix C. Variables were defined using the recorded flight data parameters, with FAA AC-120-71A providing guidance for variable selection for the UARM algorithm. For example, (a) target approach speed deviation, (b) flap position, (c) landing gear position, (d) engine speed, (e) altitude above ground level (AGL), and (f) glide path deviation are variables stated in the FAA stable approach criteria categories. Adherence to stable approach criteria was determined based on the data, including: (a) the vertical and lateral position of the aircraft with reference to the landing runway, (b) energy state, and (c) landing configuration. The information gathered in the data analysis was then used to predict the probability of the pilot misperceiving the runway excursion risk of continuing an unstable approach to landing. Pilot misperception was represented by the decision to continue to a landing even when evidence exists of exceedance in any one or more of the flight data variables from the stabilized approach criteria, and for purposes of this research, was referred to as UARM. Data mining techniques were used to populate and compare various predictive models and to determine the most accurate model, which was then used to make predictions of the target variable.

The research was exploratory and data-driven in nature, based on the following research questions:

- How can the application of data-mining and machine learning techniques to recorded flight data be used to predict the probability of *Unstable Approach Risk Misperception* by the pilot?
- What flight data variables are the most important predictors of pilot misperception of a runway excursion hazard as evidenced by continuing an unstable approach to a landing?

The AVSKD data processing model was used to address data sampling, partitioning, and validation. This chapter describes in detail the step by step process that was used in the processing of the data and the predictive model construction. The SEMMA process was used to process the flight data sets integral to the SAS® EM™ predictive modeling software.

Research Method Selection and Design

The research methodology chosen for the study was selected to build and test prediction models of UARM using large flight data. The research utilized predictive data mining methods, processes, and applications. Tufféry (2011) describes data mining as a method for exploring and analyzing large data with the purpose of discovering unknown or hidden patterns or relationships. In the research, data mining techniques were used to explore the approach and landing phases of flight for evidence of unstable approach criteria exceedance, as well as the presence, or not, of UARM. The flight data was explored to determine how unstable approaches and UARM could be measured. In order to answer the research questions, and to predict the probability that UARM would occur during an unstable approach, the following were used to build the models: (a) logistic regression, (b) decision tree, (c) neural network, (d) support vector machine, (e) gradient boost machine, and (f) random forest. Flight-related variables representing approach speed, glideslope deviation (i.e. vertical), localizer deviation (i.e. horizontal), aircraft landing configuration, engine thrust, vertical rate of descent, and altitude above ground level were anticipated to be the main focus of the models (see Table 3). Additionally, the occurrence of a rejected landing, or continued approach to landing, when confronted with evidence of an unstable approach, was also included. Flight variables for weight on wheels (WOW, > 0) and radar altimeter (RALT, > 0) were used to determine whether or not a rejected landing was executed.

Data mining was chosen to be the appropriate method for the research based on the goal of exploring very large flight data. The AVSKD data processing model was used as a framework for the exploration of large data collected by the flight data recorders on aircraft operating in the NAS. The data sets used for this research were assembled by NASA and are publicly available. Mathews et al (2013) have validated the AVSKD process.

A description of the data source, samples, data mining software, and analytical techniques proposed for the research are presented. As stated in the research questions, the intent of the research was to discover how unstable approaches and UARM could be measured when examining large data. Predictive data mining algorithms were used to estimate the veracity of this approach by constructing, testing, and comparing these models.

Population/Sample

The NASA gathered the data from one regional airline, from one type of multi-engine transport category aircraft.

- Population: U.S. domestic regional airlines.
- Sampling frame: FDR data from 35 regional jets, over a period of four years, from 2001-2004.
- Sample size: Data was comprised of 186 flight variables and approximately 152,000 rows of data (representing approach and landing or rejected landing events). A pilot study was accomplished to validate the algorithm developed to determine UARM. The data used in the pilot study was sampled solely from one aircraft tail number data set. The complete data set was then partitioned into training and validation samples. 2004 data was used for scoring the final model.

As previously described, sampling of the data included only the portion of the flight from 500 feet above ground level, during the approach phase, to a point when either a landing, or a rejected landing, was evident. Although 186 variables were available in the FDM data, those variables affecting one of three areas (energy state, aircraft configuration, and aircraft relative position to the landing runway) were initially examined in the feature selection process. These flight data variables are presented in tabulated form and listed in Appendix C, for convenience of reference.

Population and sampling frame. For the study, the population was limited to one regional jet operator in the National Airspace System. NASA has made available data collected from a single operator of regional jets, with only one model of aircraft (unknown type/model/series) represented in the research. Although data were available from these 35 aircraft over a period of four years, the sampled data was restricted to the approach and landing phase of flight, beginning at 500 AGL until either a landing or a rejected landing was performed. Figure 6 shows the points at which data was collected.

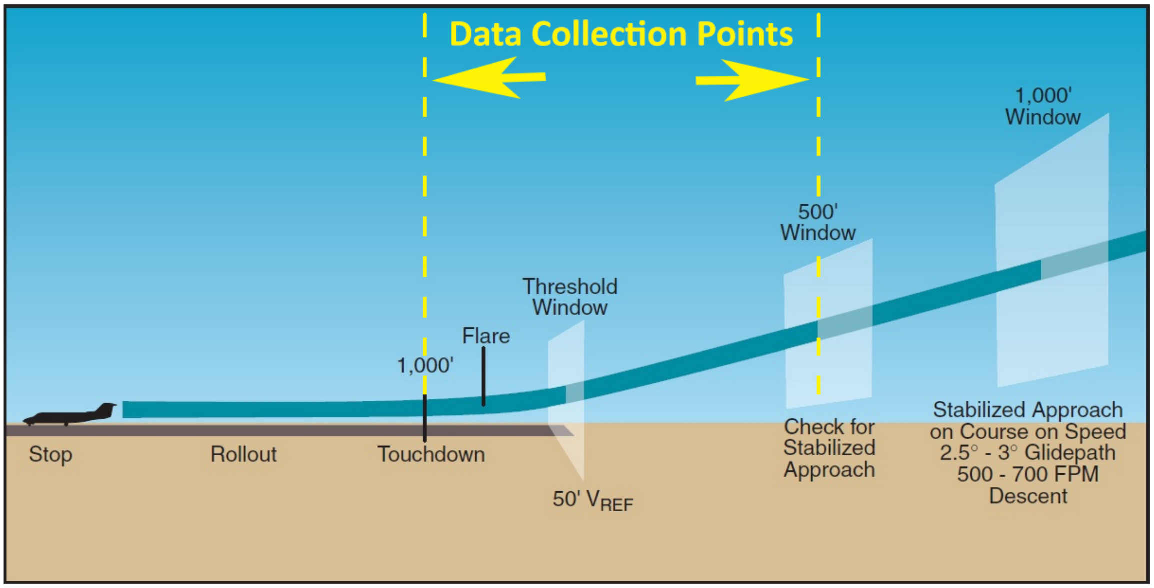

Figure 6. Data collection points. Adapted from "Air Traffic Bulletin Procedures (ATB 2019-1)," by Federal Aviation Administration Air Traffic Procedures, April 2019, p. 2. Retrieved from https://www.faa.gov/air_traffic/publications/media/atb_april_2019.pdf

Sample size. The data were sampled at several different rates, depending on the variable. Sample rates varied from 0.25 to 4 Hz. The data sets were partitioned into two separate files. One file consisted of three years (2001-2003) of data (194.3 Megabytes) and a second sample that consisted of one year (2004) data (18.7 Megabytes). The data were sampled from each flight to identify evidence of unstable approaches, with a total of 152,000 rows of data, representing approach and either landing or rejected landing events.

Sampling strategy. The strategy used to provide a sampling frame was based on the flight variables described in FAA AC 91-79A. The purpose of the sampling strategy was to extract only those data sets that provided evidence of unstable approaches, as defined by the FAA AC. The collected data sample was then assessed for evidence of not only an unstable approach, but for evidence of UARM.

Data Collection Process

The data collection process followed a framework adapted from Mathews et al. (2013) and Maxson (2018). The data was collected and archived by NASA and made available for public use. The AVSKD process began with the processing and cleaning of large flight data and then progressed to feature selection. Feature selection of flight variables representing FAA unstable approach criteria variables was based on energy state, aircraft landing configuration, and aircraft position relative to the runway as defined by FAA AC 91-79A.

Procedures.

- The data were made available by NASA and are accessible to the public on the NASA DASHlink website. Recorded data from 35 transport category aircraft of the same type and model were collected over a four-year period (2001-2004).
- The data are available to the public without permission or any restrictions. All flight data was de-identified regarding individual aircrew, airline, aircraft type, operating area, types of approaches, and any other possible identifying features such as airports or unique terrain or weather patterns.

- The Mathematical Laboratory (MATLAB®) files (mat) for each of the 35 aircraft tail numbers were downloaded using the DASHlink shell executable. Some errors in file nomenclature were noted and corrected in the downloading process. For example, tail number 679 did not exist as labelled. Additionally, one file originally included in data from tail number 666 was mislabeled, and it actually belonged in the file for tail number 674. Data processing started by downloading and verifying that data for each tail number were complete and properly labeled. Data files were uncompressed, with 232 gigabytes (GB) of data contained in 180,159 (.mat) files and was found to be organized with file names based on each aircraft tail number. The data was found to be uniform, with the same sample rate for all of the variables contained in each of the (.mat) files. NASA was able to provide information on missing values, and none were discovered in the exploration of the data.

  The next step in the data process was to partition the data and extract only that portion of the flight data pertinent to the study. As previously described, only that data from 500 ft AGL to either a go-around or landing was examined. An algorithm was developed to extract the unstable approaches and UARM from the data. Once the algorithm was developed, flight data representing unstable approaches were partitioned into training and validations data sets for the purpose of creating input files for predictive modeling in the SAS® EM™ software package.

Data Coding and Algorithm Development

Flight-related variables representing: (a) approach speed, (b) glideslope deviation (i.e. vertical), (c) localizer deviation (i.e. horizontal), (d) aircraft landing configuration, (e) engine thrust, (f) vertical rate of descent, and (g) altitude above ground level were anticipated to be the main focus of the models.

The archived data have been collected for each flight in matrix format with each row corresponding to snapshot in time, and each column corresponding to each flight variable. The acquisition of these data represented an opportunity for aviation research, with the application of data mining techniques as a strategy to discover empirical relationships between the variables captured in large data. The data was explored to determine instances when FAA stable approach criteria were not met, and then whether a landing or rejected landing was executed. As presented in Figure 7, only the approach to a landing, or rejected landing, was examined in the exploration of the data.

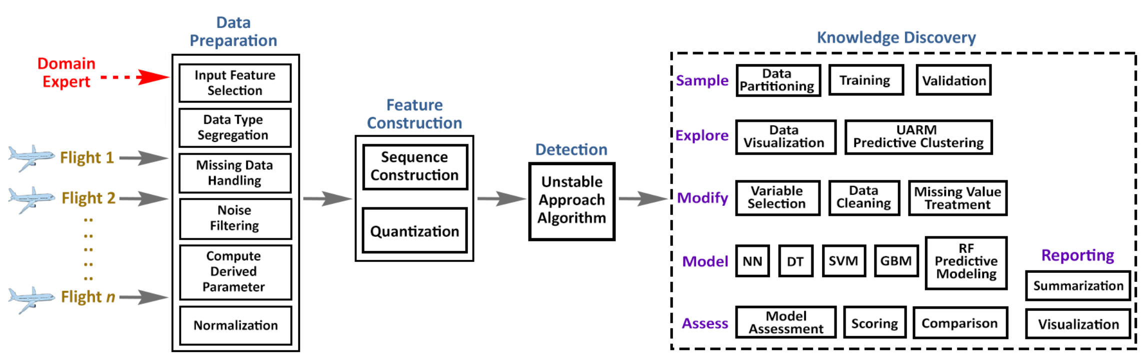

Figure 7. Research procedure framework. UARM = Unstable Approach Risk Misperception; NN = neural network; DT = decision tree; SVM = support vector machine; GBM = gradient boost machine; RF = random forest. Adapted from "Discovering Anomalous Aviation Safety Events Using Scalable Data Mining Algorithms," by B. Matthews, S. Das, K. Bhaduri, K. Das, R. Martin, and N. Oza, 2013, *Journal of Aerospace Information Systems, 10*(10), p. 469. Copyright 2013 by the Journal of Aerospace Information System, and from "Prediction of Airport Arrival Rates Using Data Mining Methods" (Doctoral Dissertation) by R. W. Maxson, 2018, p. 70. Copyright 2018 by R. W. Maxson.

The data were cleaned and processed by a data processing specialist, who is a computational mathematics and aerospace engineering student at Embry-Riddle Aeronautical University, following the AVSKD shown in Figure 5. The data preparation module performed several functions: (a) feature selection, (b) segregation, (c) missing data processing, (d) noise filtering, and (e) normalization. The FAA unstable approach criteria were used to validate the choice of variables (feature construction) that are most relevant for the analysis, as presented in Figure 5. Considering FDR data often have missing data, out-of-bounds variables, noisy recordings and amplitude spikes, the next step in the data process applied several data quality filters to clean the data. Finally, the continuous data variables were either normalized using z score or 0 or 1 normalization, or categorical data variables (gear or flap position) were converted into a dichotomous representation (Mathews et al., 2013).

A description of the data preprocessing follows and includes details regarding the conversion of data into format and structure for subsequent analysis. Primary tasks in this process included coding of flight data variables and development of an algorithm that was used to extract unstable approach evidence as well as UARM. Kumar, Tan, and Steinbach (2018) and Zhao, Bin, and Wang (2017) describe the process of extracting useful information from large data. The researchers provide details of how data mining techniques can be used to discover interesting knowledge automatically from large data sets (Kumar, Tan, & Steinbach, 2018; Zhao, Bin, &Wang, 2017). Kumar et al. (2018) continue to describe how the *Knowledge Discovery in Databases* (KDD) refers to the process of transforming raw data into *knowledge* through a series of procedural steps, to include data preprocessing, data mining, and post processing. Although Zhao et al. (2017) describe shortcomings in subjective threshold selection in flight parameter exceedance based designs, Kumar et al. (2018) assert that a sequential-based process including feature selection and construction depend primarily on successful data segmentation and extraction. Kumar et al. (2018) also insist that data preprocessing encompasses more than half of the KDD tasks.

The first step in the preprocessing of data required the raw FDR data from each aircraft to be collected. Lommadi (2019) classifies data structures based on the organization of the data in a computer's memory. FDR data containing various datatypes such as linked, ordered, and/or unordered lists are considered to be heterogeneous. The first step in the coding process was to structure the data in matrix form, with rows of time samples and columns of observed flight parameters. The heterogeneous data were collected from takeoff to landing and consisted of 186 parameters approximately 152,000 rows representing the approach and landing or rejected landing events. NASA performed initial preparation of the raw FDR data in the data preparation module of the AVSKD process. Functions performed include: (a) feature selection, (b) data type partitioning, (c) missing data identification and processing, (d) noise filtering, and (e) normalization. Feature selection was performed with flight variables chosen based on relevance to unstable approach criteria as previously described. NASA then performed additional preprocessing functions by applying quality filtering and cleaning. Derived parameters were then used to discover unreported aircraft states: energy state, landing configuration, and location relative runway, as defined along with exceedance criteria, and presented in Table 3.

Sampling strategy included taking an average of the flight variable values over five seconds to decrease likelihood of inaccurate identification of unstable approaches. Additionally, target approach reference speed criteria were based on approach speed categories defined by the FAA under Title 14, Code of Federal Regulations, Chapter I, Subchapter F, Part 97, Subpart A., § 97.3 (FAA, 2012).

Table 3

Summary of Unstable Approach Criteria Using Recorded Flight Data

Unstable Approach Construct	Flight Data Variables	Exceedance Criteria
Energy State	IVV	IVV > 1000 FPM
	GS	110 Knots > GS >150 Knots
	ALTR	ALTR > 1000 FPM
	CAS	110 > CAS > 150 Knots
Landing Configuration	LGDN	LGDN not down
	PLA1	Any 3 PLA = 0
	PLA2	
	PLA3	
	PLA4	
Relative Runway Location	DA	DA > 3 *SD*
	LOC	LOC > 3 *SD*
	GLS	GLS > 3 *SD*

Note. IVV = Inertial Vertical Velocity. FPM = Feet Per Minute. GS = Ground Speed. ALTR = Altitude Rate. CAS = Calibrated Airspeed. LGDN = Landing Gear Down. PLA = Power Lever Angle. DA = Drift Angle. LOC = Localizer. GLS = Glideslope. *SD* = Standard Deviation.

One of the primary tasks was to identify unstable approaches. In order to identify unstable approaches, a snapshot was taken at 500 ft AGL to perform an assessment of the sampled data against FAA stable approach criteria. This task was performed by taking a moving average time value of the radio altimeter variable (RALT), over a 5 second window, as suggested by Ravindran and Meht (2018), who provide details on the process of partitioning large data structures. The researchers suggest that the goal of partitioning is to divide key values in a distributed system. Chen, Zhang, Zhao, and Xia (2017) introduce and validate this method in their research modeling the approach phase of flight for Airbus 321 aircraft. The researchers describe how they eliminated influences of dimension on variable relationships by unifying the sample rate. The research team performed this task of unifying the sample rate by averaging the mean of the variable per second (Chen et al., 2017). Takahashi and Delisle (2018) present an alternative coding procedure for Airbus 320 unstable approach modeling using a time-step integrating point-mass simulation. This coding technique requires tabular aerodynamic files, propulsion data file, and assumes flight under nominal trimmed conditions. These requirements precluded this option, so the moving time average values introduced by Chen et al. (2017) was deemed more appropriate. The script identifier was then used to segregate the data less than 500 AGL. Next, the coding process was used to identify points where the aircraft has just passed through 500 feet (on ascent or descent). Data coding was used to manipulate the data so that only descending (not ascending, i.e. takeoff) flight path variables were used, thus representing the approach and landing phases of flight, as described by Kommadi (2019). The researcher describes how search algorithms can be developed to retrieve information stored in large sources of data. Kommadi (2019) also presents details of the search algorithm, which can be enhanced to discover multiple values related to search criteria. The coding process was used to further determine if a landing or rejected landing was performed. In order to accomplish this task, an event window was analyzed to determine if weight on wheels (LMG or RMG) was greater than zero, indicating a landing, and if it equals zero, then a rejected landing is indicated. A complete description of the coding process is listed in Appendix B.

Both flow chart and pseudo code was used to represent the algorithm in the research. Kommadi (2019) describes how keywords, documentation, and action tasks can assist in the visualization of the algorithm. The researcher continues to compare and contrast flow chart representations of algorithms. Algorithms depict the process of problem solving, decision making, and logic applications. Flow charts use symbols in series to provide a visual representation of the problem to be considered and the AVSKD data process was applied to FDR data. The process followed analysis based on consideration of approach/unstable approach evidence, and the decision to either continue to landing or to execute a rejected landing. The flow chart representing this logic process is presented in Figure 8.

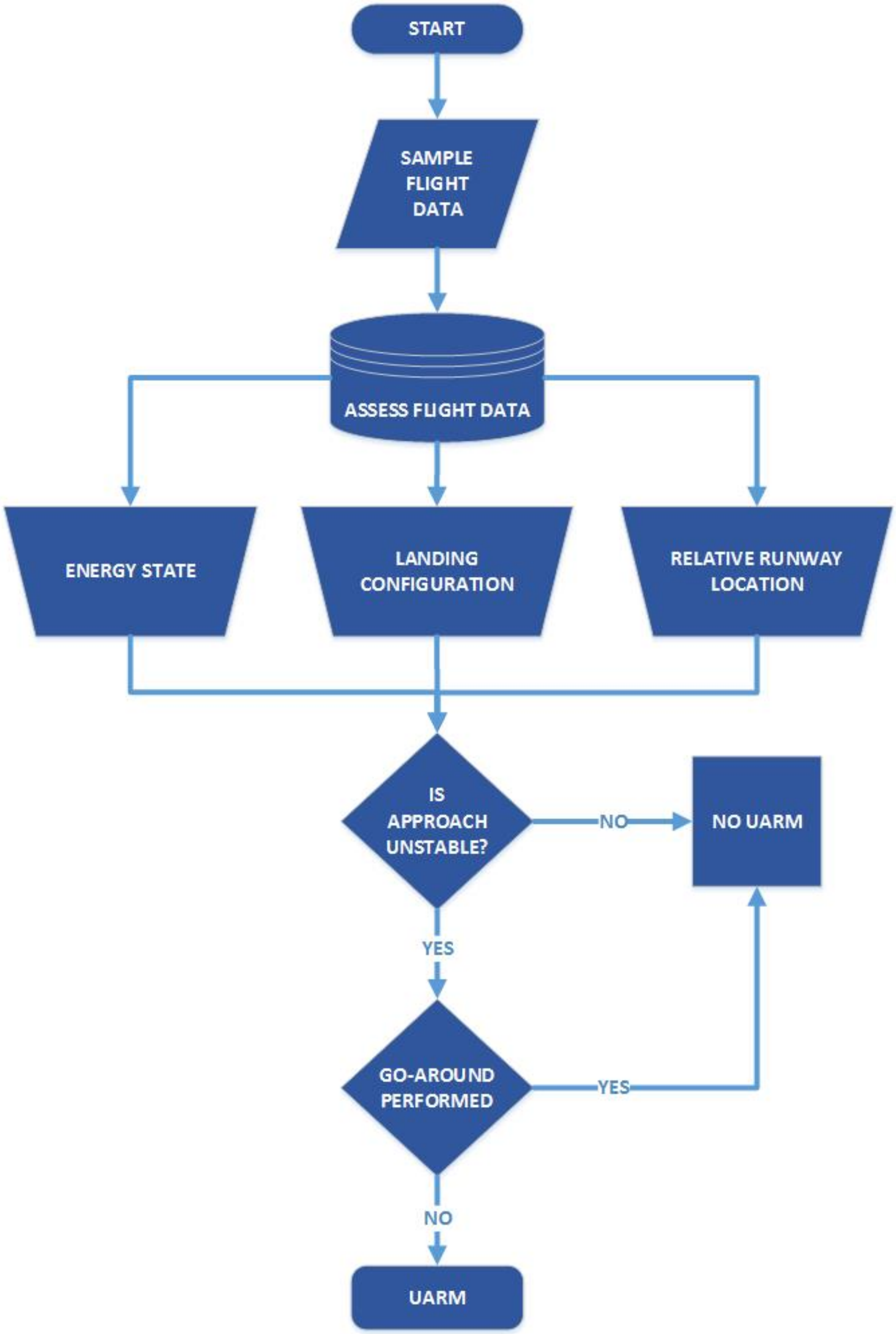

Figure 8. Unstable Approach Risk Misperception Algorithm Flow Diagram.

Once the data were cleaned and formatted, knowledge discovery algorithms was applied to determine which flights contain evidence of unstable approaches. Mathews et al. (2013) assert that the AVSKD is a very flexible process. The researchers continue to describe the development and application of several anomaly detection algorithms, which have been successfully demonstrated in various aspects of anomaly detection in recorded flight data.

Once evidence of unstable approaches was discovered using a simple algorithm based anomaly detection technique, the frequency and result (presence or not of UARM) was then determined. Mathews et al. (2013) detail the use of a tool called the Multivariate Time-Series Search (MTS) which can be used to search for similar anomaly evidences in the entire database:

> The MTS tool takes as input the data files (which can be very large), a query in the form of a multivariate time series defined over a small subset of variables, and a threshold ϵ specifying the radius of the search with respect to the query. It returns all the location pairs in the form of (file number, time instance) where a match has been found. In the preliminary phase, an index is built on the dataset. This consists of selecting a random point from the dataset (called the reference point) and then rearranging all the other points in the dataset according to distance to this fixed point. (p. 5)

Data Mining Process

As described in the literature review, Sarma (2013) presents an outline of the data analysis approach. The data process at this point in the AVSKD process then entered the feature selection mode. The algorithm constructed for the “detection” phase of the process was also applied to allow for the process to continue to the “detection” phase, where the models were trained and validated, as depicted in Figure 4. These procedures, as well as those detailed by the SAS Institute, which recommends using a SEMMA modeling process, comprised the methods used for building the predictive models. SEMMA includes *Sample, Explore, Modify, Model*, and *Assess*. The SAS® Enterprise Miner™ software package was utilized for this purpose based on the data mining methodology described. The SEMMA process is iterative in nature and the repetition of variable selection was conducted based on gained familiarity and relationships among variables as they were discovered. Thus, the *Model, Modify*, and *Assess* processes was repeated as modeling strategies evolved.

Using the NASA FDM database, this exploratory research extracted aircraft performance and position data from flights that were determined to contain evidence of an unstable approach. These data were then loaded into the SAS Enterprise Miner© for further analysis. The SEMMA process was incorporated into the *Knowledge Discovery* phase of the AVSKD model, as depicted in Figure 4. In the *Sample* step, data was partitioned into training and validation samples with the ratio of 50/50. In the *Explore* step, data was explored to examine missing values, histograms, and outliers. The *Modify* step considered data imputation and transformation as needed. In the *Model* step, the prepared data was used to construct the models. Finally, the models were compared using the model comparison function found under the *Assess* menu grouping. Model comparison capability presents multiple performance scores to rank the models; misclassification rate, Receiver Operating Characteristic (ROC), and lift curve analysis was then used as the basis for model validation and comparison and are described in the *Reliability and Validity* assessment section below.

Independent and dependent variables were selected, and model analyses was performed as outlined by Sarma (2013). In the *Model* step, the data was introduced to the models in two steps. The first step was to train and validate the models using the 2001-2003 FDR data. The final data set to be introduced was the segregated FDR addresses on only those flights that were identified to have indications of unstable approach criteria exceedance. In all steps, the selection of input and target variables was guided by the measurable criteria described in FAA AC-120-71A. Once selected, the same target and input variables were used for each unstable approach event throughout the research. The results of each modeling approach were then compared and scored.

In the *Assess* step, the performance of each model (decision tree, logistic regression, neural network, support vector machine, gradient boost machine, and random forest) were assessed for each flight containing evidence of an unstable approach. Misclassification rate was used to compare these models. Although the model with lowest misclassification rate was anticipated to be used for determining the best model, Truong et al. (2018) assert that ROC and Lift curves can also be used to determine the best model if MR rates are not substantially different. Predictive power of the selected model was also evaluated using Lift chart and ROC charts. The objective of this process was to construct a predictive model that could predict the probability of UARM at the highest level of accuracy. A sensitivity and specificity analysis was performed in order to determine which percentage of true and false positive occurrences of UARM as well as true and false negative occurrences of UARM.

Once the models were evaluated and compared and one was identified as most accurately representing the probability of UARM, this *best* model was then used to further analysze the relationships between the independent variables to the target variable. It was expected that this investigation could reveal the impact of unknown factors to pilot decision making lapses in pilot risk misperception.

Apparatus and materials. The FDR data were accumulated, archived, and made available by NASA for public aviation research via their DASHlink website. Subsequent materials consisted of several mathematical based algorithms contained in MATLAB® software. The only materials utilized in the study pertain directly to the treatment and analysis of the data as described.

Sources of the data. This research uses flight data extracted from 35 aircraft publicly available from the NASA DASHlink website (NASA, 2012; see https://c3.nasa.gov/DASHlink/projects/85/resources/?type= ds). NASA has accumulated de-identified aggregate flight recorded data, giving researchers the ability to proactively identify and analyze trends and target resources to reduce operational risks in the National Airspace System (NAS) (NASA, 2012). Permission of the site manager is not required to access or utilize the data.

Ethical Consideration

This data source provided an opportunity for aircraft operators to examine actual flight data from aircraft operating in the NAS. Access to the data was provided on the NASA DASHlink website and contains FDR data from 35 different regional jets, of a single type and airline, operating from 2001 to 2004. The data contained flight parameters detailing aircraft dynamics, system performance, and other engineering parameters, but they did not provide any information that could have been used to identify a particular airline or aircraft manufacturer. These data were not included in any airline FOQA program. The appropriate parties have allowed NASA to provide the data to the general public for the purpose of research, in order to promote aviation safety (NASA, 2012). Because the research utilized de-identified archival data, internal review board review and approval were not required.

Aviation Safety Knowledge Discovery Process

The FDR and all data contained therein were de-identified by NASA prior to release for research, hence the archived data measurement device is considered to be a valid and reliable representation of FAA certified FDR equipment. These measurement and collection devices provide real time flight data and are measured and collected by FDR and CVR devices, as well as the more recently developed Quick Access Recorder (QAR). Walker (2017) asserts that the QAR has developed into a superior data collection device and is now considered industry standard. This development of the QAR has enabled data to be conveniently extracted from commercial transport aircraft with recent systems integrated into wireless systems (Walker, 2017).

Categories. The FAA provides guidance on criteria to be used for unstable approach assessment. The categories of flight variables represent both continuous and categorical variables. For example, continuous variables represent approach speed and descent rate. Categorical variables represent flap and landing gear position information. The categories selected for the study were those most appropriate for the exploration and investigation of unstable approaches and evidence of UARM. As described in the review of the literature, FAA and SOP guidelines dictate that unstable approaches and the rejected landing decision-making process should be determined by the evidence, or not, of an unstable approach. The assessment required by the pilots is based on flight data variables which have been described in FAA AC 91-79A. The unstable approach criteria were based on those flight data variables which constitute three constructs: aircraft energy state, aircraft position relative to the landing runway, and landing configuration (e.g., landing gear, flaps, speedbrakes/spoilers).

Variables and scales. The variables of interest were selected based on the measurement of the three constructs used to define and describe an unstable approach, as well as the identification of UARM. A description of literature supporting the selection and scale of the variables as well as the dependent variable, is provided in a review of the literature in Chapter II. Table 3 and Table C6 list the variables and scale used in the study.

A target variable, UARM*,* representing the occurrence of pilots continuing an unstable approach to landing was identified and treated as a binary variable. For example, when the aircraft is on approach at 500 feet AGL, an assessment window for the pilot to examine specific flight parameters and assess aircraft performance based on the FAA stable approach criteria opens. As part of the landing checklist, the pilot monitoring (PM) observes and calls out deviations (if any) on descent rate, approach speed, glide path position, localizer, and landing configuration (e.g. landing gear down, flaps in landing position, and speed brakes stowed). In the event that uncorrected deviations exist and exceed FAA criteria, a rejected landing maneuver is recommended by the FAA and standard operating procedures (IATA, 2017).

As defined earlier, UARM occurs if a pilot elects to continue an unstable approach to landing. For the purposes of the quantitative analysis of the data, UARM was coded, based on using an arbitrary *K*-way categorical variable to be expressed as a separate possibility, based on each of the *K* possible outcomes. The target variable, UARM, is a categorical variable, with *K* equal to two. Categorical variables with only two possible outcomes are known as *binary, or dichotomous, variables.* UARM was either present (1) or not present (0). Analyses was conducted such that only $g - 1$ (g is the number of groups, two in the study) was not coded. The purpose of coding UARM in this way was to prevent redundancy while still representing the data set. In this case, with only two groups, the group that was not be coded is the group of least interests, when UARM does not occur. The rest of the flight data variables representing the FAA stable approach criteria, were utilized as input variables and were anticipated to be continuous or categorical. For example, approach speed was expected be continuous, based on numerical values while landing gear position was expected to be categorical (i.e. either up or down).

Data Analysis Approach and Process

The NASA FDM data sets were constructed separately by each aircraft tail number as large comma separated variable files and was opened using Microsoft Excel© 2016. Within each aircraft file, the data was sequentially ordered by year, date, and hour. None of the models used require normal distribution of data. As previously described, the AVSKD process provided the framework for data preparation, feature selection, and anomaly detection. The following sections will detail the SEMMA process that was applied to the data. The Stat Explore node in SAS® EM was then used to inspect for (a) outliers, (b) missing values, (c) skewness, and (d) kurtosis. As an overview, and for each aircraft, the data was treated as follows:

- Download FDM for 2001 to 2003
- Separately download FDR data for 2004,
- Train and validate the models using the three 2001 to 2003 data sets,
- After building and comparing the models, test the models using 2004 data and
- Score the models using the "Score" function in SAS® EM™.

Model steps. The following sections detail the step by step process of the SEMMA procedures that was used to input the flight data into SAS® EM™ for the purposes of building, comparing and assessing predictive models. Specific models constructed include: (a) decision tree, (b) logistic regression, (c) neural network, (d) support vector machine, (e) gradient boost machine, and (f) random forest.

Decision tree. Advantages of decision trees include being based on logical rules and the ability to tolerate missing values, non-linear relationships, and are easy to interpret. Disadvantages include being prone to instabilities with a tendency to "over-fit" the solution and can be difficult with simple linear relationships (Tufféry, 2011). A decision tree represents hierarchical structures of variables, including the parent node and child nodes. Visual representations represent the presence, or effects of relationships between the binary target variable, UARM, and independent variables. The root node is split into two or more branches, based on the condition that best separates the individuals of each class. If the variable is categorical, the branches represent separate classes of the root node. Conversely, if the variable is continuous, then the branches represent specific ranges of the node. At each split, a separation condition was used to determine how to split the parent into child nodes. This iterative process, referred to as recursive partitioning, continued until an assigned termination condition was satisfied. Common splitting criteria include: Chi-square, Gini, and entropy (Tufféry, 2011). Default settings were selected based on the coding of the independent variables. These default settings set for decision trees modelled in SAS® Enterprise Miner™, with a maximum branch size were determined by the coding process as well. For example, initially, the maximum depth could be six, with a minimum categorical size of 5.

The decision tree modelling process involved the application of a series of relatively straightforward rules, in which observations are assigned to a segment relating the value of a single input variable. The process was iterative in nature and repeated until a hierarchy results. The resulting hierarchy, or tree, contained the segments, which were referred to nodes. The tree root node consisted of the complete data set. Leaves were created from all of the branches with the final node being referred to as the leaves. A decision, or predicted value, was made for all leaf values. The decision tree (DT) node was used to classify observations and for prediction of the occurrence of the target variable, UARM. Advantages of decision tree predictive modelling are that they are easily interpreted and handle missing data well.

Splitting criterion rules were generated using training data in the Decision Trees machine learning model. Once these splitting rules were generated, these rules were then used to determine predictions using other data sets. Splitting criterion rules were then validated using statistical significance tests *F*-test or Chi-square tests. Gini values, reduction in variance, entropy, and P-values could have also been used in the rule set for stopping the DT model.

As shown in Table 2, the Decision Tree node consisted of the flight variables and the target variable, UARM. The Decision Tree node was appropriate for the binary target variable in the study. Missing values or negative frequency values were excluded and were not truncated.

Tables C1 and C2 provide a detailed description of node functions as applied to Decision Tree and Logistic Regression modeling in SAS® EM. These examples were representative of node functions in each of the ML techniques that were used in the construction, training, validation, and assessment phases of the AVSKD research model.

The Variable Importance Table was used to indicate the relative value of the tree node output variable importance. The sum of the squared errors was then used in the Decision Tree node for UARM, based on the training data and was indicated with the Gini Index. Table C2 lists what the Results of the Decision Tree predictive models.

The Fit Statistics table displays the statistics for the training, validation, and test data sets. The Classification chart used a bar chart to present the classification results for UARM. The Score Ranking Overlay chart was a presentation of training and validation overlay plots based on statistics used to create the model. The horizontal axis displays the observation depth. The curves for the Best Measures represent the model that predicts the target correctly.

The score rankings matrix plot is an overlay of selected statistics for the models using training and validation data. Lift plots were constructed for all reporting variables using a Data Partition node. Best Measures were then used to represent the model with the highest predictive probabilities for the observations.

The Score Distribution chart was used to present model scores based on the prediction of UARM. The target variable predictions were binned, based on the number of buckets, and indicate the best scores for the highest percentage of correct predictions and conversely, lower scores for lower percentages of predictions.

Logistic regression. In order to predict and explain a binary variable, as opposed to a metric dependent measure, logistic regression modeling was utilized. Similar to multiple regression, this form of logistic regression variate can also be used to represent a single multivariate relationship, regarding the use of regression-like coefficients to represent the significance, or impact, of each predictor variable. When attempting to identify and predict which group a target variable belongs, logistic regression is appropriate (Tufféry, 2011). Because the purpose of logistic regression is the prediction of a two-group (g = 2) dependent variable, it was the proper choice for this research and was less affected by basic assumptions, such as normality of the variables, than linear regression.

The Regression Node (RN) was used to fit logistic regression predictive models to the NASA FDM data set in the AVSKD data processing flow. The purpose of the logistic regression modelling was to predict the probability that UARM occurred based on one or more effects of the independent variables described in Table 3.

Prior to building regression models, the data mining tasks were performed as described in Figure 3.

Neural network. Also commonly referred to as artificial neural networks (ANNs), neural networks are data reduction models based on nonlinear regression and were developed to replicate features used by the neurological functions of the human brain. The basic structure of ANN models is a combination of relatively simple computing elements (neurons or units) into an integrated system. ANNs can be used to predict a target variable in aviation research and can be particularly useful when large data is available to train the model, and a mathematical based relationship has been developed relating input to output variables.

Data mining tasks that were accomplished to train the ANN models include:

- Sample the Input Data — The Sampling node extracts sample data used to train the ANN that can be generalized to the data
- Create Partitioned Data Sets — The Data Partition node is used to partition the data sample into training, validation, and test data sets. These sets were then used to learn network weights, select architecture, and for model assessment. The test data set was also used to estimate generalization error.
- Use Only the Important Variables – Prior knowledge was used to select only important inputs regarding the target variable. This step was only be accomplished after using the Explore and multi-plot nodes to eliminate bias.
- Transform Data and Filter Outliers - Transform Variables node was used in the case that several transformations are necessary in a single variable. For example, log, exponential, square root, inverse, and square could have been necessary. Transformation and optimal binning were also used for target variable transformation.
- Impute Missing Values - The Neural Network node allows for exclusion of missing values or if all target criteria are missing. This node was used to replace missing values with mean or median values, as necessary. Dummy variables could also have been used (either 0 or 1) as inputs to the ANN model.

The Neural Network node was used to train the ANN models. Training an ANN model required an objective function which was the total error combined with a penalty function divided by frequency. Maximum likelihood was used for the statistical estimation method and is the negative log likelihood that is minimized, as opposed to maximized (Sarma, 2013).

Advantages of neural networks include the ability to tolerate non-linear data; however, they tend to over-fit and are negatively influenced by poor variable selection. Neural networks present causal relationships among factors. Neural network modeling was used to incorporate factors included in the FAA unstable approach criteria, the outcome of which included not only the prediction accuracy and impact factors, but also a causal network of these factors. A causal network using neural inference modeling is a graph of nodes and arcs that form a visual representation of factor relationships. Each node represents a random variable, and an arc represents a direct relationship among variables. The nodes can either be continuous or discreet and represent variables of interest. Each node represents an association of probability distribution, and the relationships among the nodes are described by conditional probability distributions (Truong, Friend, & Chen, 2018).

Support vector machine. The High Performance (HP) SVM Node was used to build predictive models and was advantageous because it was able to handle binary, ordinal, nominal and interval input variables and one binary target variable. Data with missing values was ignored (Tuffery, 2011). Tuffery (2011) describes the selection of the *optimal hyperplane* used to mitigate the infinite possible classification solutions in the case of linear separated observations. The selection of the optimal hyperplane influences both the correct fit and robustness of the model. The optimal hyperplane also maximizes the width of separation between observations, with points on the boundaries of separation known as *support points or vectors.* In the case that observations could be clearly separated a classification error could have been added to the separation term.

Tuffery (2011) presents advantages of SVM models including: ability to use a kernel function to model non-linear phenomena, high degree of predictive precision, and robust modeling capability due to the use of an *optimal hyperplane*. Tuffery (2011) also lists disadvantages of SVMs as: a lack of transparency, sensitivity to kernel selections, and lengthy computational time, possibility of overfitting.

The HP SVM node menu included the following options: Node ID, import, browse, explore and properties. Additional menu options were similar for other models, with key differences in the interior point options which allowed for a choice of kernel or polynomial functions to be selected. Kernel options allow for separation of observations using linear functions while the degree of polynomial functions could have been selected instead. For a detailed description of steps in the SVM model see Table C5.

Gradient boost machine. The Gradient Boosting Node was used to partition the data based on an algorithm that detected values for the target variable, UARM. The algorithm assigned weights to partitioned data, with weights varying with target variable similarity, using recursive partitioning. The predictive model was then constructed based on a combination of the partitioned data. The model was then evaluated based on the target variable goodness of fit statistics, rather than the individual partitioned data sets. This combination of partitioned data sets was expected to create a better model than was possible with the combination of partitioned data sets.

GBM modeling utilized resampled data sets in an iterative process to produce output that was based on a weighted average of the re-sampling. GBM is similar to DT boosting which combines a group of trees to form one predictive model. A residual is expected to result from each DT, which could have then been fitted to the previous tree and could have been defined regarding a derivative of a loss function. Because the target variable was binary, logistic loss, or negative binomial log-likelihood represented the loss function.

Results menu options are displayed in the same manner detailed in the DT modeling process detailed in Table C2. The GBM node allowed two different process techniques regarding variable importance assessment: split-based and observation-based. The split-based technique utilized a reduced value in the summation of squares based on all nodes that were used. The observation-based technique was based on an increase in fit statistic value. This technique addressed the decrease in value usually associated with correlated values.

Random forest. The HP Forest node was used to construct predictive models combining trees and was referred to as a forest. Training was achieved in random forest (RF) models using a random sampling of all possible input variables in the splitting node. Additionally, sampling training data did not use replacement in any observations.

One advantage of RF modeling was that RF models were adept at handling binominal, categorical target variables. In this case, posterior probabilities within the RF model were achieved through the average of posterior probabilities of individual DT models to predict target variable probability. The HP Forest Node was also used to predict the probability of target variable observation by voting on the category that the individual DT models most often predicted.

Another advantage of RF modeling was that by averaging DT models with different training samples, predictions on one particular training sample are minimized. This *out-of-bag* sample was more reliable than individual DT models. Additionally, overfitting, a common problem with DT models was minimized in RF modeling and decreased overfitting the sample.

Options in the HP Forest node menu included: Maximum Number of Trees, Number of Variables to Consider in Split Search, and Proportion of Obs in Each Sample. "Number of Variables to Consider in Split Search" was an option within the HP Forest node which could have been scaled to select the number of input variables to be considered in the RF model (See Table C7 for a descriptive list of the HP Forest node).

Bagging is a term associated with the HP Forest node. This term refers to the term "bootstrap bagging" when replacement sampling was used. Observations in the training sample was then referred to as "bagged" samples when associated with a specific DT.

The HP Forest node was assigned observations to a single leaf in each DT model in the forest. The individual DT model was then used to make a prediction on the target variable. The HP Forest node then averaged all of the individual predictions. For the target variable, UARM, the posterior probability was expected to the proportion of the category in the bagged training observations (see Table C7 for a descriptive list of the "Results" options).

Model comparison. The Model Comparison function in SAS Enterprise Miner© provided a reliable and valid reference for initial results and could have been used to interpret potentially inconsistent criteria across the multiple models being analyzed. In the research, under assessment reports, the number of bins was initially set to 20, a ROC chart was selected, and the selection statistic employed was set to cumulative lift. The model output results for each selected event were then reported based on misclassification rate for the validation data, rather than ASE. Although ASE was a preferred model diagnostic because it provided common estimates of performance for the predictive modeling, because the target variable, UARM, is binary, misclassification rate was more appropriate. In the cases where the target variable (UARM) was binned (e.g., a multinomial logistic regression) statistics such as misclassification rate or a confusion matrix for class variable outputs were used to determine the best model (Sarma, 2013).

For the study, the basic data mining outputs were:

- Misclassification rate for validation data, ROC and Lift chart
- Relative variable worth for each variable

Scoring. The models were tested by using a scoring technique. SAS® scripts were then created using the Score assessment function and a different data set was loaded to test the model predictive capability. The Score Node was used to manage the scoring process of the models. Scoring was accomplished with the generation of a scoring code that was produced from training runs of each model. The scoring code was then used to assign a score to the data set which in turn resulted in the creation of predicted values regarding the probability prediction of the target variable, UARM. The results of scoring were then reported using scoring tables. Scoring tables listed both input and output variables. The score node train properties function in SAS allowed for flexibility in the selection input variables. The Explore window enabled the selection among several menu options; however, because of the exploratory nature of the research, initial training runs were executed by accessing the entire list of 186 flight data observation values for input by the score node. Output variables were created by the score code. Once training runs were completed, input data was introduced from a later range of dates (e.g., 2001-2003 aircraft data) and were evaluated using the models developed from these data sets.

Descriptive statistics Representative descriptive statistics were determined, they were segregated into class and interval variables. The FDR descriptive statistics were listed in appendices and tabulated for ease of observation. The descriptive statistics were further explored and presented along with the data mining results. Frequency statistics describing total number of approaches, number of unstable approaches, and rejected landings are also provided. Additionally, descriptive statistics pertaining to the flight variables comprising the three unstable approach constructs were provided, including mean, *SD,* and histogram presentations.

Reliability assessment method. With the data mining approach used in this research, reliability is strongly determined by the quality of the quantitative input data. Because these data have been collected using previously FAA certified recording devices, the data were considered to be reliable. Because the data source had been considered as credible, reliability testing consisted of the evaluation of the results in the different machine learning techniques to be utilized. Results were compared between training and validation samples using misclassification rates, ROC charts, Cumulative Lift charts and leaf statistics for all models used in the study. Similar results were observed in each of these techniques. Thus, reliability was demonstrated when the FDR data was analyzed in different models with similar results (Tuffery, 2011).

Validity assessment method. Mathews et al. (2013) assert that once anomalous events are detected using the knowledge discovery algorithm, validation is generally performed in three ways: (a) domain expert input, (b) ASRS pilot narrative reports, and/or (c) pilot interviews. As the AVSKD process was validated in previous research, domain expertise is represented by using FAA guidelines as the basis of feature selection. One advantage of using FAA unstable approach criteria was that any potential bias was eliminated regarding feature selection.

In order to perform validation assessment using SAS Enterprise Miner, the data sample was partitioned into *training, validation,* and *test.* The *training* data was used to develop the models. The *validation* data set was used to validate the models and then to select the best one. The *test* data set was used for an independent assessment of the selected model (Sarma, 2013). Additionally, the *score* assessment function was used to further enhance the validity of the results. Using the *score* assessment function, a SAS® *scripts* technique was created and a different data set was loaded to test the model predictive capability.

In the development of predictive models, validating the models was important for the results to be generalizable and reliable. Further details involving the process of feature selection, data partitioning, and the iterative nature of training and validating the models are described in the SEMMA process application section of the study. The *model comparison* node was used to compare the models. Misclassification rate, Receiver Operating Characteristic (ROC) curves, and Cumulative Lift (CL) charts are among assessment techniques that were used to evaluate the models. ROC charts were developed in the 1940's to evaluate and compare predictive models. ROC charts provide a visual depiction distinguishing between true positive and true negative predictions of the target variable. Cumulative Lift charts provide a visual depiction of the percent of captured responses within each percentile bin to the average percent of responses for each model. CL charts provide a visual representation that depicts the advantages of using the model over the random prediction of the target variable without the use of predictive modeling. Misclassification rate is a measure of accuracy of the model, or how well the model correctly predicts UARM. Truong, Friend, and Chen (2018) describe how misclassification rate can be used to ensure validation accuracy of prediction. Because the misclassification rate = 1 – (validation accuracy), lower misclassification rates indicate higher validation accuracy. The researchers continue to describe how a ROC curve indicates validity using the *validation* data ROC curve. For example, the ROC curve indicates the contrast between sensitivity and specificity. *Sensitivity* depicts the probability of UARM, while *specificity* reflects the probability of UARM not occurring. The ROC curve was compared to the baseline straight line, and this comparison was used to provide an assessment of the best predictive model. In the ROC chart, random selection is represented by a diagonal line (line of no discrimination) which divides ROC space into good classification performance (above) and poor classification performance (below). The area under the curve (AUC) equals probability randomly chosen positive results higher than a randomly chosen negative instance. Larger AUC indicates stronger predictive power of the model. CL charts show a comparison between random prediction and model predictive performance. A horizontal line intersecting the y-axis at one depicts random prediction with no model.

Summary

NASA has provided public access to FDR data, which was gathered from 35 regional jets operating in the NAS from 2001-2004. Data mining techniques were utilized to address the nature of the research. Specifically, the AVSKD process provided the framework for the treatment of the data as well as the discovery of unstable approaches and evidence of the occurrence or the absence of UARM. The SEMMA process was then be applied to the data, with the construction and evaluation of predictive models. These NASA FDR data were modeled using: (a) decision trees, (b) neural networks, (c) logistic regression, (d) SVM, (e) RF, and (f) GBM predictive models. Model performance was compared and tested, with the goal of predicting the probability of UARM in the event an unstable approach occurs. Once the models were analyzed, a determination of how DM techniques could be utilized to predict UARM, and which are the most that contribute to the probability of UARM, was determined.

CHAPTER IV

RESULTS

FDR data were achieved and made available by NASA for public use. These FDR data were collected from a fleet of 35 regional jets operating in the NAS from 2001 to 2004. The research was based on the ability to use 2001 to 2003 FDR data to train and validate the machine learning-based models and used 2004 FDR data to score the 3 branch DT model. An FDR data set consisting of 2001, 2002, and 2003 data was used to construct (a) decision tree with 2, 3 and 5 branches, (b) logistic regression, (c) neural network, (d) support vector machine, (e) gradient boost machine, and (f) random forest models. A model comparison was used to determine that the 3 branch Decision Tree model had the highest predictive power, with a 98.8% prediction score. Once DT was determined to be the highest scoring model, it was used to run 2004 FDR data in order to (a) determine the predictive probability of the target variable, UARM, and (b) rank input variables in order of importance.

The combined 2001, 2002, and 2003 data sets were partitioned 50/50 percent to train, validate, and compare the performance of all of the models. 2004 data was then used to score the DT (3 branches) model by using the Score node within the SAS® EM™. The 2004 scored data resulted in predictive probabilities of UARM that were then compared with the actual UARM instances that were observed in 2004. These results are included in subsequent sections of this chapter.

Mathews et al. (2013) asserted that once anomalous events are detected using the knowledge discovery algorithm, validation is generally performed in three ways: (a) domain expert input, (b) ASRS pilot narrative reports, and/or (c) pilot interviews. As the

AVSKD process was validated in previous research, domain expertise was represented by using FAA guidelines as the basis of feature selection. One advantage of using FAA unstable approach criteria was that any potential bias was eliminated regarding feature selection.

As part of the model training and validation process, various combinations of input variables were selected for testing. The first variable input combinations used all of the flight data variables available in each of the data sets used. Variables in the data set used to describe aircraft specific FMS software version updates, as well as date and time data, were rejected in early iterations. The iterative process dictated that subsequent model construction was based on feature selection using estimates of variable importance. For example, initial runs resulted in MNS (selected Mach speed) included in the predictive modeling process. Selected Mach was determined to be a variable associated with the cruise (enroute) flight phase, and was not determined to be associated with, or relevant to, the approach and landing phases of flight (a delimitation in the study used to develop the UARM algorithm). Therefore, MNS was rejected in further iterations of the model runs. Ultimately a predictive power score of 98.8% was achieved with the highest scoring model, 3 branch DT. Additionally, variable importance regarding the occurrence of UARM was determined and is reported in tables listed in subsequent sections of this chapter.

Demographics

Information in this section presents data regarding the sample aircraft in the research and the population of commercial transport airplanes in the domestic NAS. Although the recorded flight data made available by NASA were de-identified to specific aircraft type, the flight data variables included in the data sets were sufficient to make a comparison between the sample aircraft in the study and the population of commercial airplanes in the U.S. Demographic information regarding commercial transport aircraft is presented both at the time the sample aircraft flight data was recorded as well as projections based on growth and operational use into the future.

A comparison of the flight data from sample aircraft used in the research to demographic data of the population of commercial transport aircraft was conducted. The comparison results indicate the aircraft used to gather flight data in the research were representative of those in the population of commercial transport aircraft, as described by Vértesy (2017). For example, engine data samples indicate four sets of turbine-powered engine parameter data. The aircraft data indicated retractable landing gear, multi-place flaps and slats, and digital flight controls. Additionally, avionics data are sampled which describe an advanced commercial transport category aircraft. Pilot flight control data indicate a two-pilot configuration with two control yokes, two rudder position indicators, and dual primary flight displays. Aircraft state data representing airspeed, altitude, pressurization, and avionics are typical of advanced commercial category aircraft. Pressure altitude and indicated airspeed data indicate that the sample aircraft operated within the same standard flight envelope as the population of commercial transports aircraft. Although payload and range data were not specifically provided, fuel flow and average flight segment time were used to estimate these data.

In the late 1990s and early 2000s, the industry standard description of regional jets was summarized in FAA AC 150/5325-4B, *Runway Length Requirements for Airport Design* (FAA, 2005). In this document, the FAA described an RJ as "a commercial jet configured for 100 passengers or less" (p. 1). Wong, Pitfield, and Humphreys (2005) agreed that there was no universally accepted definition of RJs, and in their study, refined the FAA position, asserting that RJs should be defined as jet powered aircraft built after 1992 with fewer than 100 passenger seats. Subsequent developments in commercial transport aircraft technologies have resulted in a newer generation of RJs that have grown in (a) size, (b) passenger configuration, (c) maximum takeoff weight, and (d) range (Vértesy, 2017). As RJs evolved, so did the industry standard description. Brueckner and Pai (2009) described RJs as a unique combination of several attributes, including (a) approximately 70 passengers, (b) typical range of approximately 1500 nautical miles, (c) high cruising speed compared to mainline jets (over 500 mph), and (d) passenger amenities and comfort similar to mainline jets (Brueckner & Pai, 2009). As technology and industry leadership changed, so did industry standards on RJ categorization. Curtis, Rhoades, and Waguespack (2013) challenged the traditional definition as larger RJs began to come into service. The researchers asserted that newer aircraft, such as the Embraer E190, E195, and Bombardier CS100/300 series transport aircraft (with passenger configurations of approximately 130 seats) shared similar traits to smaller Boeing and Airbus products, although Airbus and Boeing transport aircraft were not considered to be RJs (Curtis, Rhoades & Waguespack, 2013). Vértesy (2017) supported a broader contemporary definition with the claim that RJ are "a turbofan-engine-powered aircraft carrying typically 30 to 120 passengers at a range up to 2000-2500 nautical miles" (Vértesy, 2017, p. 389).

The FAA (2017b) predicted commercial aircraft in the U.S. would grow from 7,039 in 2016 to 8,270 in 2037, an average annual growth rate of 0.8 percent a year.

These demographic data include narrow body fleet growth at a rate averaging 37 aircraft a year, while the number of wide-body aircraft is forecast to grow by an average of 17 aircraft a year in the same timeframe. The FAA also claimed that the U.S. carrier wide-body fleet will increase by 67 percent over the years 2016 to 2037 (FAA, 2017b. Vértesy (2017) attributes these demographic changes to three factors: (a) windows of opportunity provided by mainline pilot scope contractual issues, (b) strategic responses to rising fuel prices and more efficient engine technology, and (c) leadership changes involving regulatory policies including loans to support new aircraft development and export-financing regimes (Vértesy, 2017).

The FAA (2017b) integrated RJ demographic data into projected numbers of U.S. commercial aircraft with the assertion that the regional aircraft fleet is forecast to decline from 2,156 aircraft in 2016 to 2,027 in 2037. The FAA stated that the one of the factors involved in this decrease is the trend of carriers to remove 50 seat regional jets and retire older small turboprop and piston aircraft, while adding 70-90 seat jets. The FAA continued to claim that by 2037, the number of turbojet powered RJs should total 1,828, up from 1,637 in 2016. Conversely, the turboprop/piston regional aircraft fleet was forecast to decrease by 62% from 519 in 2016 to 199 by 2037. Figures 9 and 10 depict FAA gathered RJ demographic data.

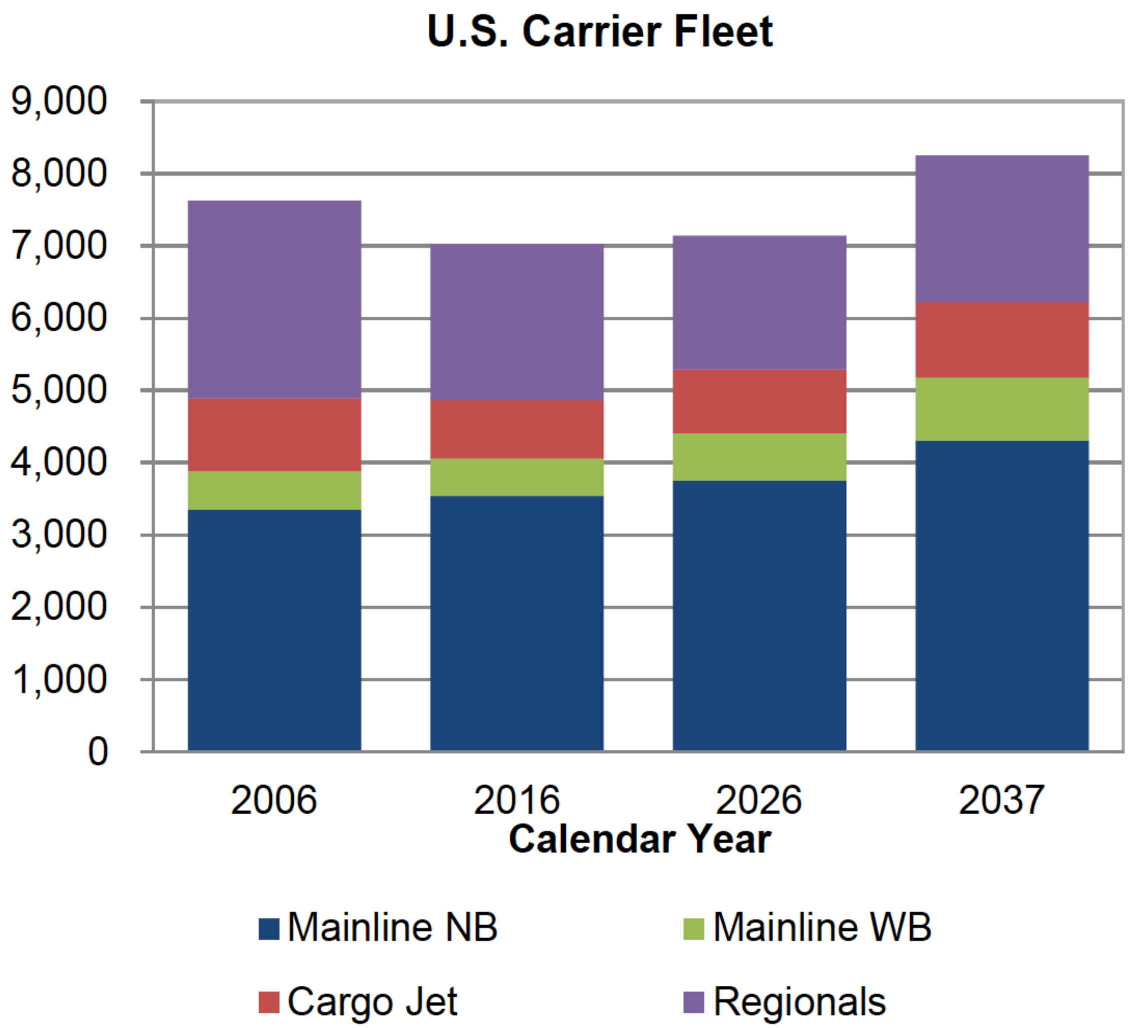

Figure 9. U.S. carrier fleet. Reprinted from "FAA Aerospace Forecast, Fiscal Years 2017-2037," by Federal Aviation Administration, 2017b, p. 29. Retrieved from https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/fy2017-37_faa_aerospace_forecast.pdf

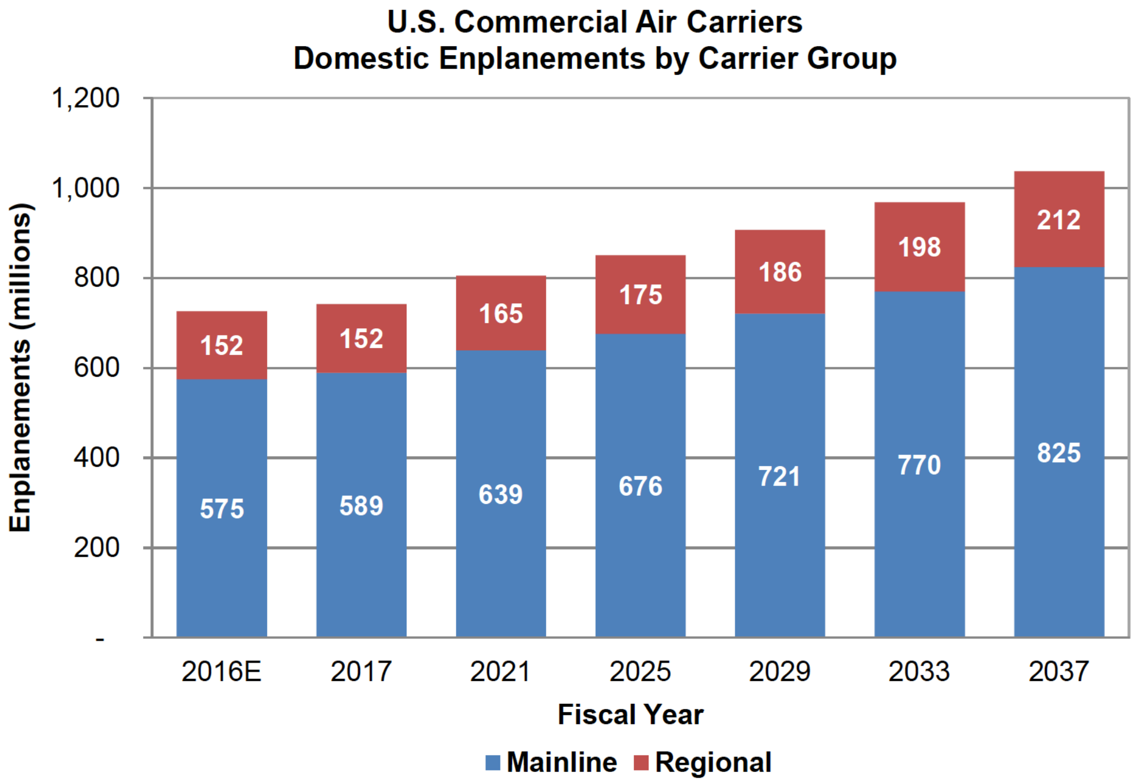

Figure 10. U.S. commercial air carrier domestic enplanements by carrier group. Reprinted from "FAA Aerospace Forecast, Fiscal Years 2017-2037," by Federal Aviation Administration, 2017b, p. 10. Retrieved from https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/fy2017-37_faa_aerospace_forecast.pdf

BAC (2017) and DVBank (2019) include maximum range and maximum takeoff weight in the classification of commercial transport aircraft. Figure 11 provides payload and range demographic information on the population of worldwide regional jet operations in 2018-2019.

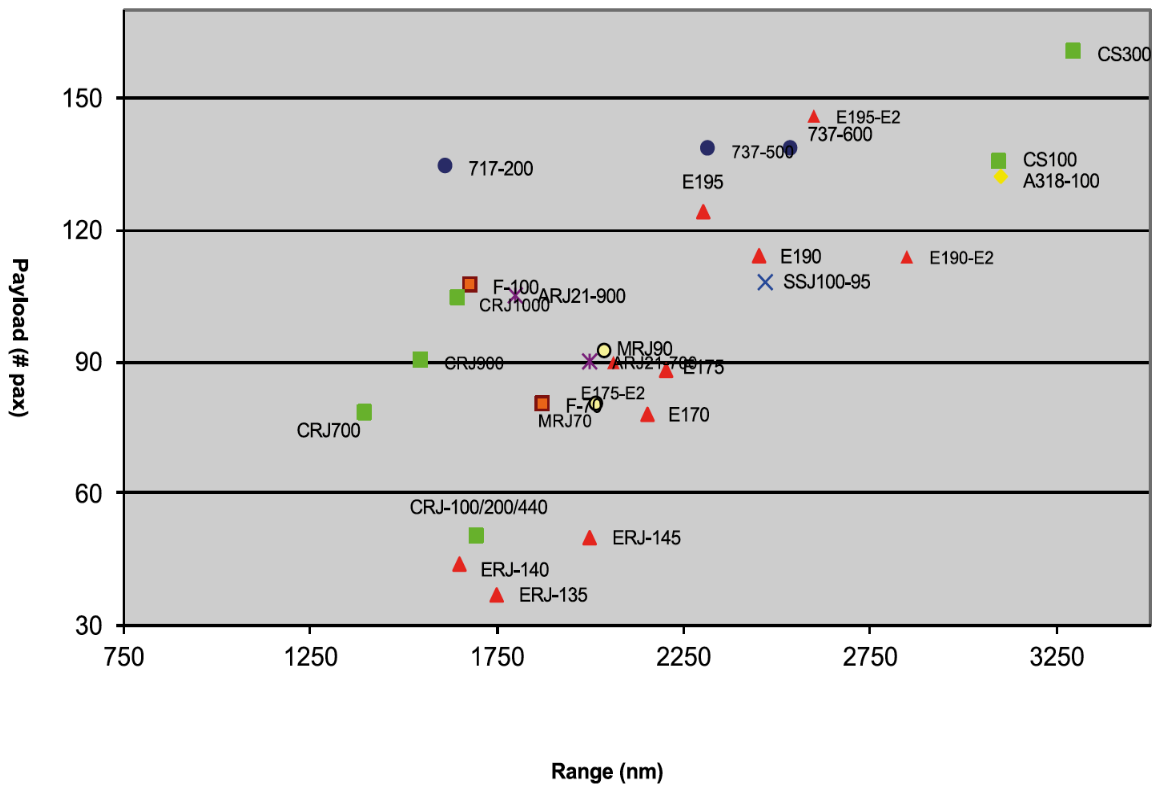

Figure 11. Regional jet payload-range data in 2017. Certified jet airplanes greater than 60,000 pounds maximum gross weight. Reprinted from "An Overview of Commercial Aircraft 2018-2019," Aviation Research, 2019. Copyright 2019 by DVB Bank, p. 163. Source: https://www.dvbbank.com/~/media/Files/D/dvbbank-corp/aviation/dvb-overview-of-commercial-aircraft-2018-2019.pdf

Mozdzanowska and Hansman (2005) provide research describing the economic and operational significance of RJ integration into the NAS in the late 1990's to early 2000's (timeframe of aircraft operations in the research). The researchers analyzed and compared flight operational data between RJs and mainline jets and turboprops. Results of the study indicate that, in 1998, U.S. regional jet operations closely resembled those of turboprops. However, by January 2003, RJs began to fill a gap in the NAS by flying longer routes than turboprops but shorter routes than narrow-body mainline jets. The researchers conclude that RJs' cost per trip was very similar to traditional jet costs when trip length was normalized (Mozdzanowska & Hansman, 2005).

Figure 12 presents a data comparison between the number of unstable approaches observed from sample aircraft flight data and the number of unstable approaches reported by industry sources as presented in previous sections. Based on the comparison, it was concluded that the sample aircraft flight data in the research were representative of the population of U.S. commercial transport aircraft.

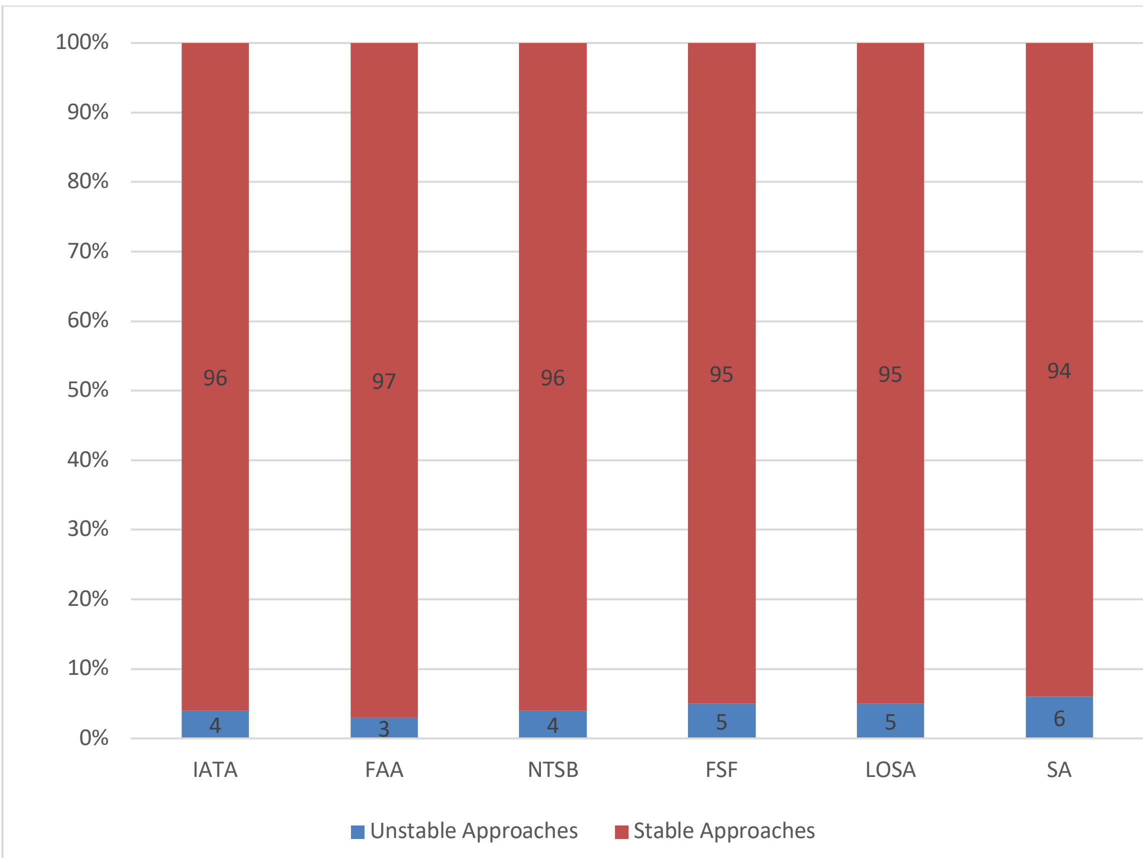

Figure 12. Unstable approach data comparison.
Note. Data for IATA from IATA (2017), for FAA from FAA (2015), for NTSB from NTSB (2019), for FSF from FSF (2009), for LOSA from Moriarity and Jarvis (2014), and for SA from NASA Dashlink (2012). Dashed box highlights Sample Aircraft Data. IATA=International Air Transport Association, FAA=Federal Aviation Administration, NTSB=National Transportation Safety Board, FSF=Flight Safety Foundation, LOSA=Line Observation Safety Audit, SA=Sample Aircraft.

UARM Algorithm Development

One of the most important objectives of the study was to develop an algorithm for the target variable, UARM. In order to successfully accomplish this task, it was first necessary to identify unstable approaches within the data. Initial examination of the data indicated the recorded flight data contained variables that could be used to represent FAA exceedance criteria for unstable approaches. Results of this determination also indicated that flight data variables could be used to develop the algorithm for UARM. Once this determineation was made, a coding process was developed to create a data variable representing landing, rejected landing, and UARM. One important step in this process was the development of assessment criteria limitations as presented in subsequent sections. For example, weight on wheels (WOW>0) was used to determine if a landing (or rejected landing, WOW=0) was accomplished. Exceedance of any of the flight variables used in the construction of stable approach criteria (engergy state, landing configuration, location relative runway) was successfully used to determine if an unstable approach was evident. These variables used to identify unstable approaches were selected based on the measurement of these three constructs used by the FAA to define and describe an unstable approach. A straightforward If/Then decision process was then developed and used to complete the UARM algorithm (See Figure 8). Results of this If/Then assessment process were successful in the discovery that evidence of UARM had occurred or not. For example, once an unstable approach was identified, a determination was made whether or not a rejected landing was performed. *If* evidence of an unstable approach was indicated, and a rejected landing was not performed, *then* UARM resulted.

Results of this UARM algorithm development were then successfully used to construct predictive models as well as the identification of UARM. The rest of the flight data variables, including those representing the FAA stable approach criteria, were utilized as input variables and were anticipated to be continuous or categorical. For example, approach speed was expected be continuous, based on numerical values while landing gear position was expected to be categorical (i.e. either up or down). Sarma (2013) described the process of *binning* to classify continuous variables in a categorical manner.

An important result of the data coding process was the successful transcription of the NASA provided MATLAB data into EXCEL for importation into SAS® EM™. Initial coding of the FDR data was accomplished with the construction of two separate data sets in EXCEL format. The first data set consisted of 2001, 2002 and 2003 flight data. The second data set consisted of 2004 flight data. The three-year data set was partitioned into training and validation samples and subsequently was used to perform a model comparison to determine the best performing model. The 2004 data set was then used to score the model which was determined to be the best preforming model based on the model comparison.

An important aspect of the UARM algorithm development was that of feature selection. The iterative process was used to reject or accept flight data variables based on decision making rules established and described in the UARM algorithm. In order to preclude selection bias, FAA stable approach criteria were used to identify flight data variables which represented stable approach assessment. Results indicated that the *Explore* node was quite useful in the iterative process of feature selection. For example, because unstable approach was a prerequisite for UARM, it was obviously a predictor of the occurrence of UARM. Accordingly, failure to reject the landing during an unstable approach was also a prerequisite for UARM and was therefore anticipated to be present in 100% of UARM occurrences. One key element in the development of the UARM algorithm was the ability to make the important distinction between the identification and analysis of unstable approaches and the ability to successfully predict the probability of occurrence of UARM. For example, exceedance of approach speed criteria at the assessment point would be an indication of an unstable approach, but not necessarily an indication of the probability of the occurrence of UARM if the pilot rejected the landing.

Results indicated that one advantage of the UARM algorithm was that of scalability. Several different predictive models successfully utilized the UARM algorithm with the application of recorded flight data. Results demonstrated that real world recorded flight data was successfully assessed in the UARM algorithm process to predict the probability of occurrence of the target variable. The UARM algorithm was successfully used repeatedly with consistent results to evaluate large recorded flight data. The algorithm was successfully developed based on initial data coding, subsequent use of an If/Then decision-making process, and ultimately the extremely accurate and precise predictive power regarding the target variable, UARM. The UARM algorithm provided a step by step, repeatable process to the analysis of recorded flight data and allowed for a reliable and valid methodology for the analysis of FDR data to predict UARM. Evidence of successful employment of the UARM algorithm is presented in subsequent sections.

Descriptive Statistics

The recorded flight data variables were initially processed by SAS™ EM® using the STAT EXPLORE node. As part of the feature selection process, variables were examined for relevance to the analysis of unstable approaches and rejected landings. Several variables were rejected for analysis based on relevance to flight analysis. For example, DVER_1 represented the database version that the FMS used and was not included in the predictive model building process. The complete list of flight variables are listed in Table C6.

The recorded flight data was then examined for evidence of unstable approaches. Using FAA exceedance criteria as previously described, unstable approaches were extracted and examined for causal factors (i.e., exceedance of FAA flight parameters). Flight data variables associated with FAA unstable approach exceedance variables were also examined to determine frequency data. Frequency data for approaches and occurrence of UARM is presented in Table 4. Although assumptions of normality and multicolliniarity were not required to be met for the machine learning algorithms utilized in predictive model building, the exceedance criteria variables were examined for general consistency and to better visualize the data. Results of the descriptive analysis are presented in Table 5. Fight data variable exceedances are summarized in Table 6. Exceedance criteria for glideslope deviation was observed most frequently, landing configuration deviations second (power levers at idle, engine thrust not stable) with localizer deviation (lateral distance from extended runway centerline) third most frequent.

Table 4

Summary of Recorded Flight Data Approach Frequencies

Approach Classification	*N*	Percentage
Total Approaches	152,442	
Unstable	11,348	7.44
Stable	141,094	92.6
Rejected Landings	450	0.29
Landings	151,992	99.7
UARM	11,047	7.24

Source: Adapted from "*Sample flight data for 35 aircraft*," by National Aeronautics and Space Administration DASHlink, 2012. Retrieved from https://c3.nasa.gov/DASHlink/projects/85/resources/?type=ds. *Note.* UARM=Unstable Approach Risk Misperception.

Table 5

Summary of Unstable Approach Criteria Descriptive Statistics

Variable	*N*	Missing	*M* (*SD*)	Min	Max	Skewness	Kurtosis
IVV	33	0	-659 (109)	029	1675	0.280	17.4
GS	349	0	119(12.1)	104	179	6.452	180
ALTR	34	0	-670 (204)	005	1685	0.097	10.5
CAS	456	0	124 (7.61)	104	179	1.25	159
LGDN	359	0					
PLA	3124	0					
FLAP	64	0					
DA	383	0	0.34(0.041)	0.02	5.5	-0.517	27.2
LOC	2742	0	0.071(0.020)	0.001	1.7	-2.02	52.4
GLS	4338	0	0.082(0.003)	0.001	2.6	-1.83	30.4

Note. LGDN, PLA, and FLAP are categorical variables, and are presented for informational purposes in order to include the complete list of unstable approach criteria variables. IVV = Inertial Vertical Velocity. FPM = Feet Per Minute. GS = Ground Speed. ALTR = Altitude Rate. CAS = Calibrated Airspeed. LGDN = Landing Gear Down. PLA = Power Lever Angle. DA = Drift Angle. LOC = Localizer. GLS = Glideslope. *SD* = Standard Deviation.

Table 6

Summary of Unstable Approach Flight Data Variable Exceedance Frequency

Unstable Approach Construct	Flight Data Variables	Exceedance Frequency
Energy State	IVV	33
	GS	349
	ALTR	34
	CAS	456
Landing Configuration	LGDN	359
	PLA	3124
	FLAP	1159
Relative Runway Location	DA	383
	LOC	2742
	GLS	4338

Source: Adapted from "*Sample flight data for 35 aircraft*," by National Aeronautics and Space Administration DASHlink, 2012. Retrieved from https://c3.nasa.gov/DASHlink/projects/85/resources/?type=ds. *Note*. IVV = Inertial Vertical Velocity. FPM = Feet Per Minute. GS = Ground Speed. ALTR = Altitude Rate. CAS = Calibrated Airspeed. LGDN = Landing Gear Down. PLA = Power Lever Angle. DA = Drift Angle. LOC = Localizer. GLS = Glideslope.

Data Exploration

The *Explore* component of the SEMMA process initially included the variable importance analysis of the entire data set using the *Stat Explore* node. Chi-square was used to determine the variable importance in this step of the process. Sarma (2013) describes the primary purpose of the *Stat Explore* node as to make a preliminary assessment of the importance of input variables regarding strength of relationship with the target variable. SAS™ EM® software was used to rank flight data variables in order of importance to the occurrence of UARM. SAS™ EM® defined variable importance as the rank order (from 0 to 1) of input variables determined by the Chi-square statistic and described the strength of the relationship between categorical input variables and the target variable. SAS™ EM® used *binning* to derive categorical input variables from continuous input variables (Sarma, 2013).

Variable importance analysis provided an indication of which predictor variables could be considered most useful in the prediction of UARM; conversely, those variables deemed not relevant to the analysis of unstable approaches and/or UARM were rejected. The Explore process was iterative in nature and began with the exploration of the entire 186 variable data set. The purpose of the exploration of the data was to become familiar with which variables were most relevant to the study and to iteratively exclude variables unrelated to, or irrelevant to the prediction of UARM. For example, if flight data variables that defined UARM (unstable approach, landing or rejected landing) were not rejected they would by default be artificially ranked at the top of the variable importance list because they occurred 100% of the time that UARM occurred.

- Initial data exploration indicated that several of the recorded flight data variables were unrelated to flight operations and were rejected on the first iteration. For example, variables associated with nomenclature (date, time, tail number) were rejected. Also, variables associated with FMS software version nomenclature were rejected.
- Subsequent iterations resulted in the rejection of variables which were prerequisites for UARM. For example, prerequisites to UARM include

unstable approach, landing, or rejected landing; therefore, those variables representing unstable approach and rejected/not rejected landing were excluded. The purpose of rejecting these variables was to preclude construct bias from the predictive model construction process.

- The latest iterations resulted in the rejection of the following variables: selected Mach and engine compressor speed. Although these variables indirectly represent data which may or may not be important to unstable approaches, they were determined not to be relevant to phases of flight included in the study (approach and landing) and were therefore excluded from analysis.
- Based on the iterative nature in the *Explore* component of the SEMMA process, three variables were outstanding regarding importance (strength of relationship with the target variable, UARM: (a) auto-throttles off (A_T), (b) flaps not extended (FLAP), and (c) airbrakes deployed (ABRK).

Model Reliability and Validity

Reliability assessment. With the data mining approach used in this research, reliability was strongly determined by the quality of the quantitative input data. Because these data have been collected using previously FAA certified recording devices, the data were considered to be reliable. Because the data source was considered as credible, initial reliability testing consisted of the evaluation of the results in the different machine learning techniques to be utilized. Reliability was demonstrated when the FDR data was analyzed in different models with similar results (Tuffery, 2011). Reliability assessment of the best performing model, DT with 3 branches, was also performed separately using the 2004 FDR test data set to construct ROC and Cumulative lift charts. The reliability analysis with the test data set is presented in subsequent sections.

Model reliability in predictive modeling was also demonstrated with the comparison between training and validation partitioning. In order to perform the reliability assessment of the models, the data samples were partitioned into *training* and *validation* samples using the data partition node in SAS EM. Similar results between the training and validation partitioned samples indicated high reliability of the models. Comparison between training and validation characteristics are presented for the best model, decision tree, in Figure 13, as well as for the models evaluated in the model comparison, in Figures 14 and 15.

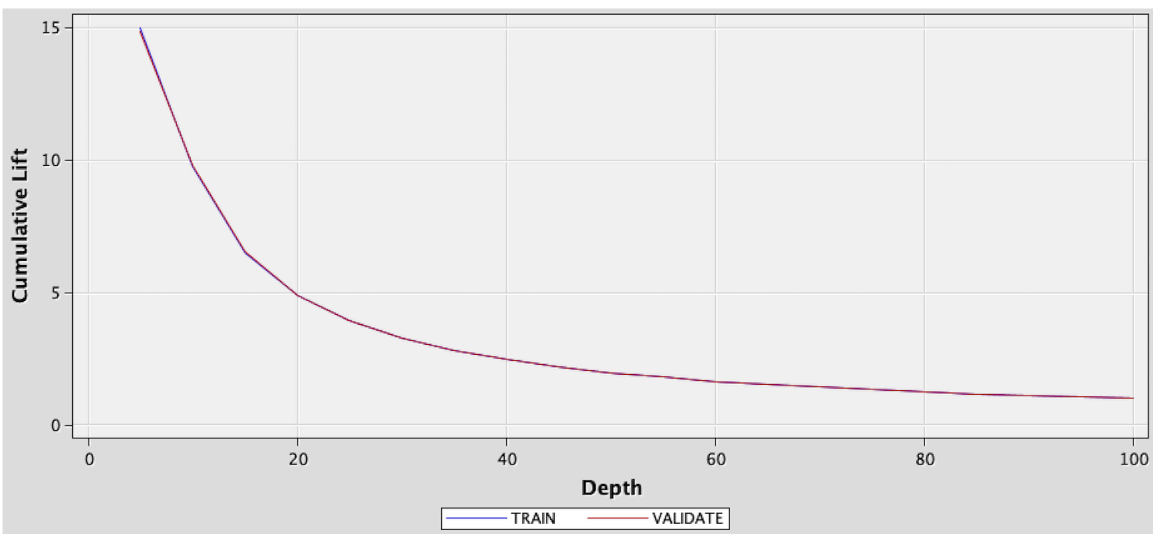

Figure 13. Cumulative Lift Characteristics Chart for DT Train and Validate Comparison. A presentation of relative performance between training and validation data samples.

The close approximation between the training and validation data indicated the validity of the DT model. Additionally, misclassification rate comparison data between training and validation samples for the decision tree model are presented in Table 7. Similar values between training and validation data provided additional indications of the validity of the DT model.

Table 7

Decision Tree Misclassification Rate Training and Validation Comparison

	Train	Valid
MR	0.0028	0.0036
ASE	0.0026	0.0036

Note. MR=Misclassification Rate. ASE=Average Square Error

Validity assessment. The validity of the model was assessed based on model prediction accuracy using MR, accuracy, sensitivity and specificity. Additionally, predictive power assessment was used to demonstrate validity using ROC and CL charts. Examples of the validity assessment are presented in Figures 17 through 21. The *score* assessment function was used to further enhance the validity of the results. Using the *score* assessment function, a SAS® *scripts* technique was created and a different data set was loaded to test and validate the model predictive capability.

In the development of predictive models, validating the models was important for the results to be generalizable and reliable. Further details involving the process of feature selection, data partitioning, and the iterative nature of training and validating the models were described in the SEMMA process application section of the study. The *model comparison* node was used to compare the models. Misclassification rate and Receiver Operating Characteristic (ROC) curves were two assessment techniques used to evaluate the models. Truong et al. (2018) describe how misclassification rate can be used to ensure validation accuracy of prediction. Because the misclassification rate = 1 –

(validation accuracy), lower misclassification rates indicate higher validation accuracy. The researchers continue to describe how a ROC curve indicates validity using the *validation* data ROC curve. For example, the ROC curve indicated the contrast between sensitivity and specificity. *Sensitivity* depicts the probability of UARM, while *specificity* reflects the probability of UARM not occurring. The ROC curve was compared to the baseline straight line, and this comparison provided an assessment of the best predictive model. Figure 14 presents a comparison of the training and validation scores. Similar ROC curve analysis results discovered in the training and validation data suggest the model was valid. Figure 15 provides a graphical description of the Cumulative Lift characteristics validation and train data that also support model validity. Further validity assessment was conducted separately on the champion model, DT with 3 branches, using the 2004 test data set and is presented in the next section describing the champion model performance.

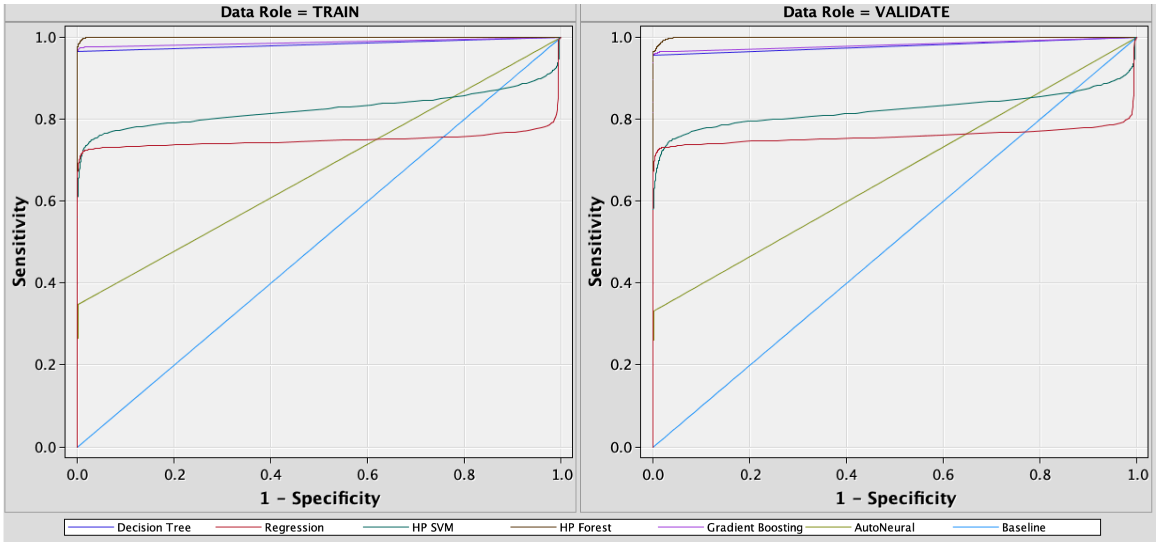

Figure 14. Receiver Operating Characteristics Chart for Train and Validate Comparison. A demonstration of relative performance between models. The straight line was a baseline and larger the area under the curve, the better the performance of the model. Vertical Axis contains *sensitivity,* or true positive fraction values. Horizontal axis contains *specificity,* or false positive fraction values for training and validation data. Similar results between training and validation data charts indicated consistency of model performance.

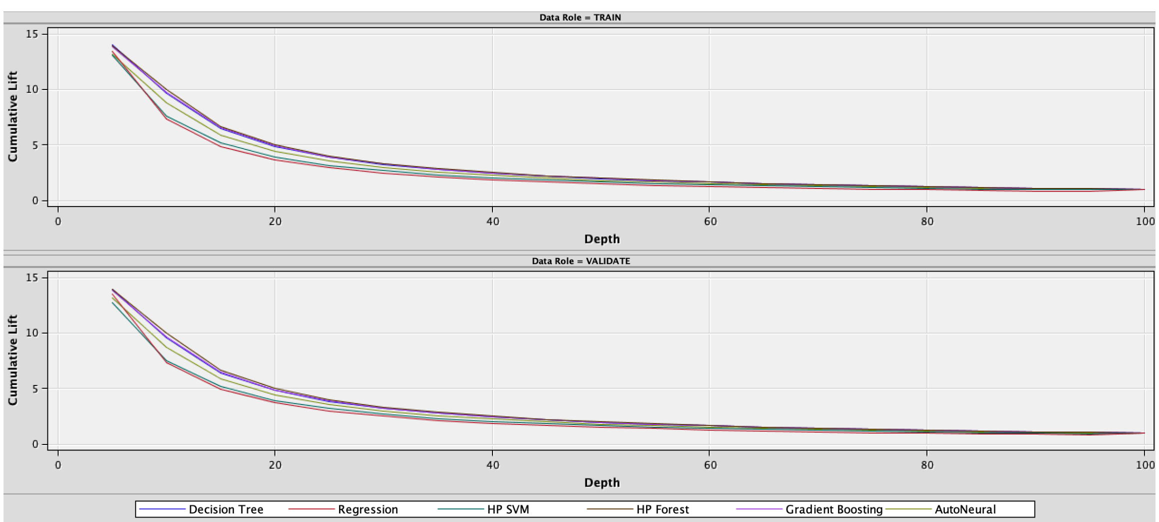

Figure 15. Cumulative Lift Chart for Train and Validate Comparison. A demonstration of relative performance between models.

Model Building and Evaluation

Six machine learning-based prediction models were constructed to predict the probability of UARM occurrence: (a) support vector machine, (b) gradient boost machine, (c) random forest, (d) neural network, (e) decision tree with 2, 3 and 5 branches, and (f) logistic regression. SAS® EM™ model diagrams were constructed containing the machine learning-based models and is depicted in Figure 16. The diagram constructed to train and validate the models contained: (a) the data set with variables representing the recorded flight data from years 2001-2003, (b) the Stat Explore node, (c) the Data Partition node, (d) the six predictive models, and (e) a model comparison node. Model performance was compared using the Model Comparison node to rank models based on the misclassification rate, ROC and Cumulative Lift Charts. Predictive Power data is provided for the best model based on model comparison. A confusion matrix was constructed to provide analysis information regarding true positive, true negative, false negative and false positive predictive data of the model. These data were presented in Tables 8 and 9.

Table 8

Decision Tree Confusion Matrix

	Predicted UARM (1)	Predicted UARM (0)
Actual (1)	4736(TP)	27(FP)
Actual (0)	228(FN)	64376(TN)

Note. 1=occurrence of UARM, 0=nonoccurrence of UARM. N = 69,367. TP=True Negative. TN=True Negative. FP=False Positive. FN=False Negative.

Sensitivity was calculated by dividing true positive predictive values of UARM by the combination of true positive and false negative prediction values of UARM. Specificity was calculated by dividing true negative values of UARM by the combination of false positive and true negative values of UARM. These results were presented in Table 9.

Table 9

Decision Tree Sensitivity/Specificity

	UARM (1)	UARM (0)
Sensitivity	0.954	0.960
Specificity	0.961	0.999

Note. 1=occurrence of UARM, 0=nonoccurrence of UARM.

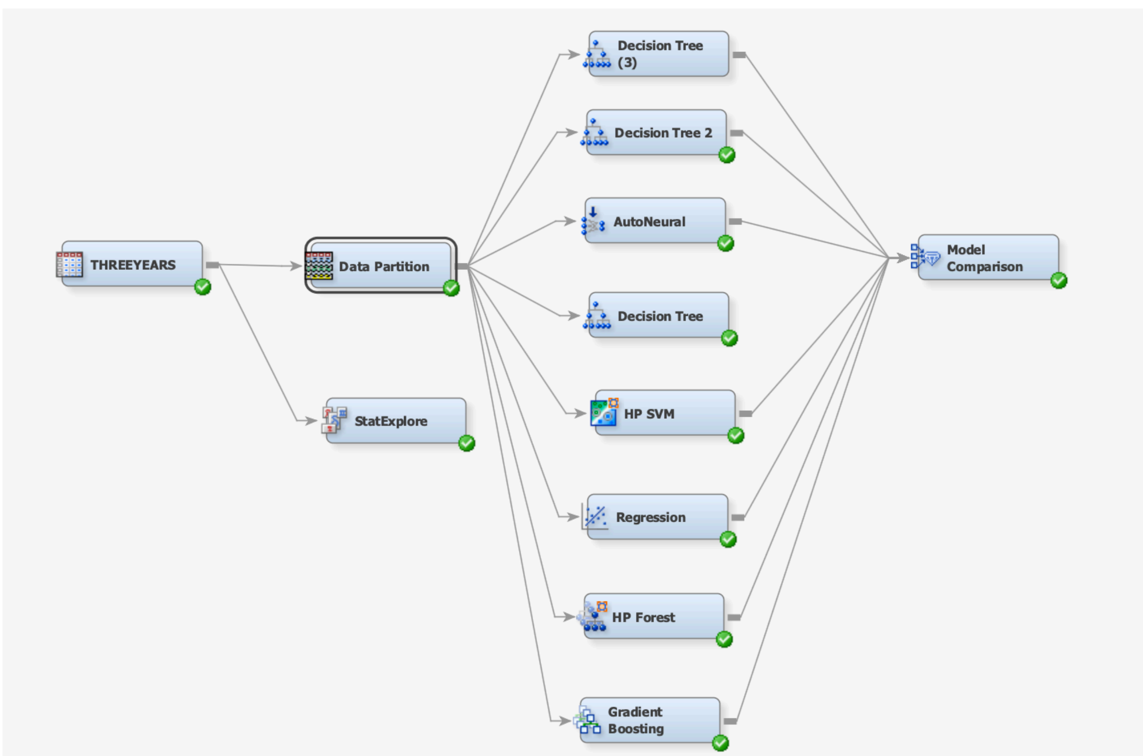

Figure16. SAS EM model diagram example. A visual presentation of the SAS modeling process.

Table 10 presents the results of the comparison of misclassification rates for validation sample data (sample size was 138,733 cases) between the six models and were ranked using the lowest to highest values. The model comparison results indicate that the decision tree (2, 3, and 5 branches) and gradient boost machine had the lowest misclassification rates compared to the other models. An examination of the ROC chart results indicated a performance differential between the DT model and other models. Thus, the DT model was determined to be the model with the highest accuracy for prediction of UARM.

Table 10

Summary of Model Comparison Misclassification Rate

Model Description	Misclassification Rate
Decision Tree, 3 branch	0.003
Decision Tree, 2 branch	0.004
Decision Tree, 5 branch	0.004
Gradient Boost Machine	0.006
Random Forest	0.009
Support Vector Machine	0.033
Logistic Regression	0.040
Neural Network	0.048

Note. Sample size was 138,733 cases for years 2001-2003 data.

The Lift Chart was used to determine the predictive power of the models in more detail. Truong et al. (2018) asserted that if Lift curves for training and validation samples are very close, model validity is strong. Figure 17 presents the ROC results of the model comparison. The ROC chart indicates a performance advantage for the DT model. Figure 18 presents results for the Cumulative Lift Chart for the 6 models, which was also used to compare model predictive performance. Because the Lift curve of the model is in close proximity to that of the best cumulative Lift curve, results were confirmed with additional evidence of the predictive power of the DT model. An examination of the Lift curve substantiates the results in with the top 20% of the responses generating a Lift value of 1.5. This indicates that approximately 30% of UARM occurrences could be predicted with this model compared with prediction of UARM without the model.

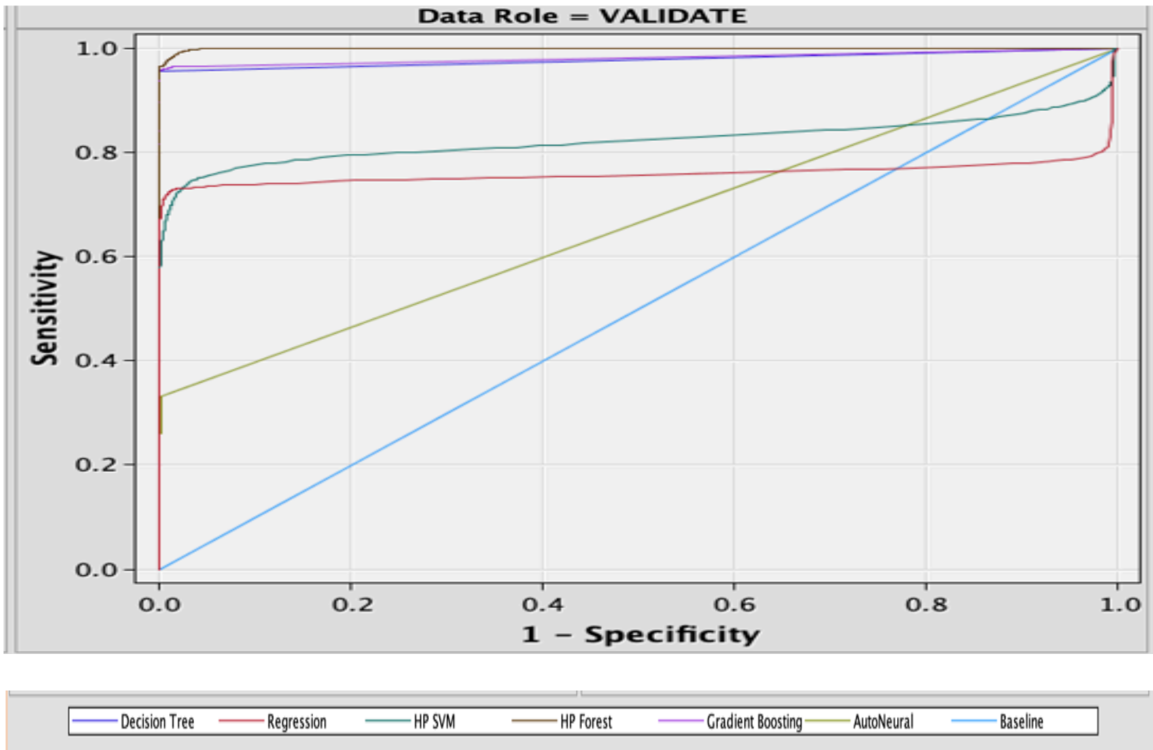

Figure 17. Receiver Operating Characteristics chart for the model comparison. Validation data used to compare model performance indicated DT model performed the best.

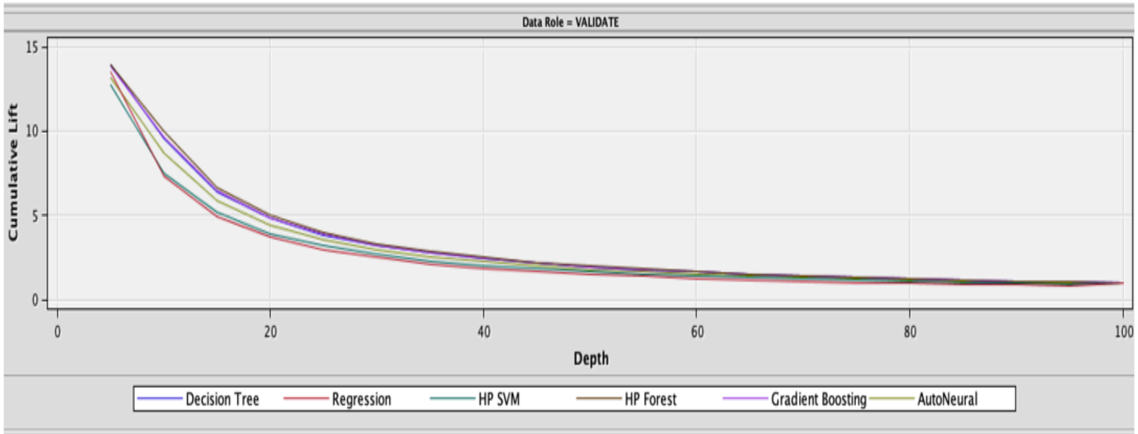

Figure 18. Cumulative Lift chart for the model comparison. Lift charts provided a model comparison and depicted similar performance between the models.

Additional reliability assessment methods available included a comparison of model performance between different models using the same data set. Figures 19, 20 and 21 presented model comparison information between examples used with DT, LR, and NN models. These similarities between these results of different models using misclassification rates indicated good consistency in the data, hence good reliability.

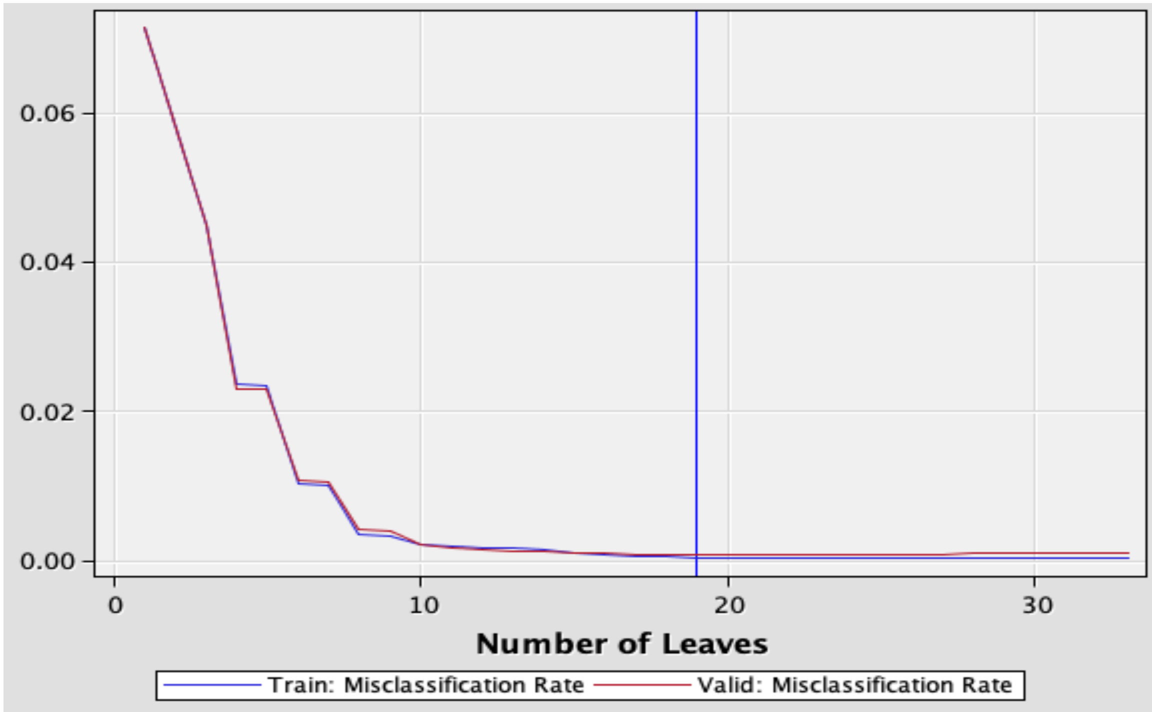

Figure 19. Subtree Assessment Plot for DT model. Provided a basis for comparison between DT model and other predictive models. A favorable comparison between different models with similar results in misclassification rate indicated good model reliability.

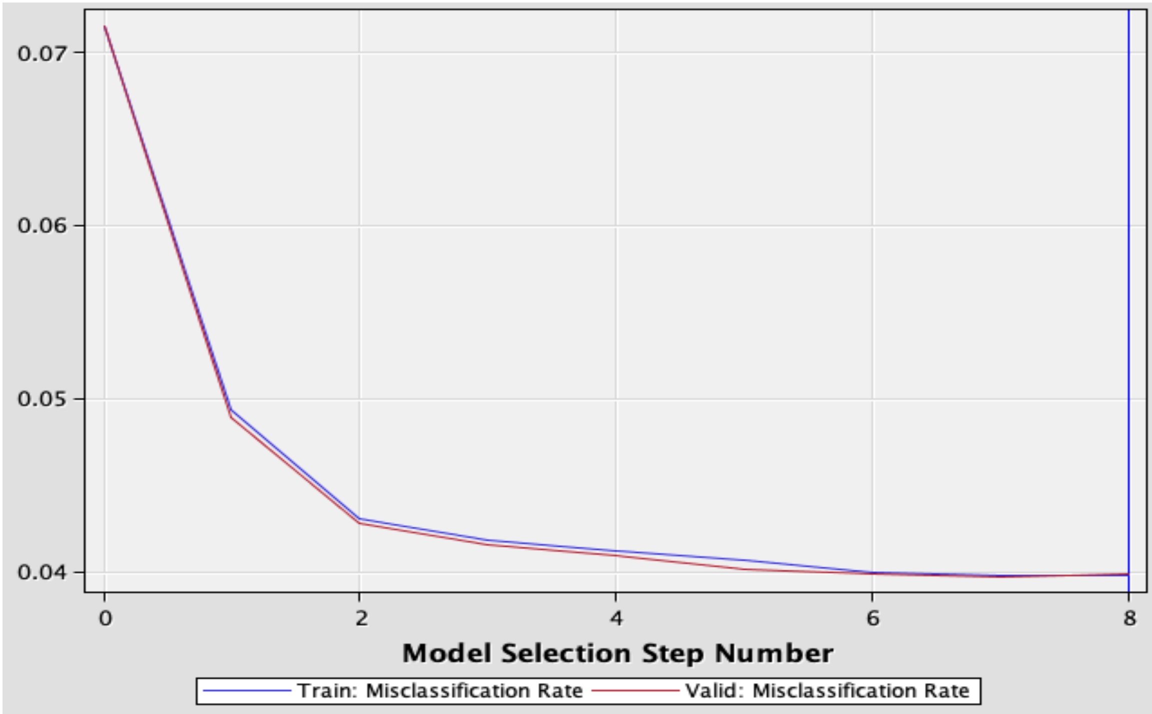

Figure 20. Assessment Plot for logistic regression model. Provided a basis for comparison between LR model and other predictive models. A favorable comparison between different models with similar results in misclassification rate indicated good model reliability.

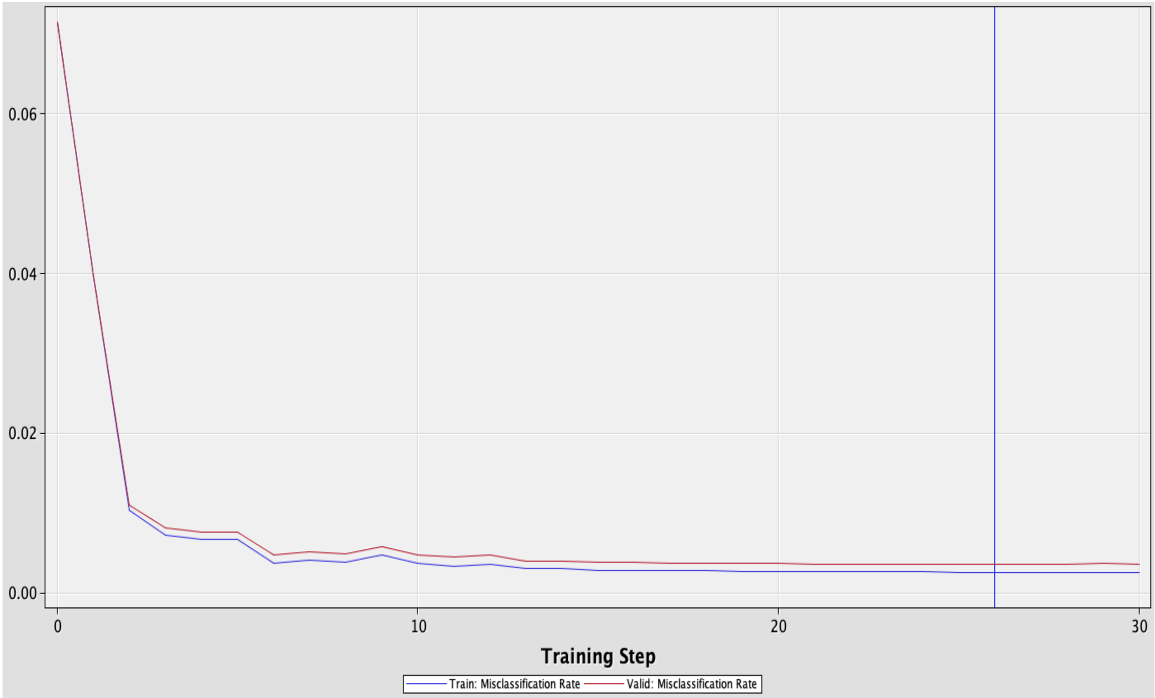

Figure 21. Assessment plot for neural network model. Provided a basis for comparison between NN model and other predictive models. A favorable comparison between different models with similar results in misclassification rate indicated good model reliability.

Variable Importance

Once the DT model was determined to be the best performing model, variable worth was determined using the Score node using only the 2004 data set. Sarma (2013) asserts that relative importance is measured between 0 and 1. The Variable Worth Plot used the Gini split statistic, which Sarma (2013) describes as a total leaf impurity. For example, a variable worth was calculated with a decision tree with a depth level of six. Variable importance was determined using the best performing model, DT with three branches. Based on the analysis of the 2004 flight data with a sample size of 13,492 cases, findings indicate that six important variables stood out in the prediction of UARM. Glideslope deviation (GLS) was the most important variable in the prediction model.

The other important predictor variables in ranked order were: (2) selected airspeed, (3) localizer deviation, (4) flaps not extended, (5) drift angle, and (6) approach speed deviation. Figure 22 presents results of variable importance regarding the value of each predictor variable in the prediction of UARM.

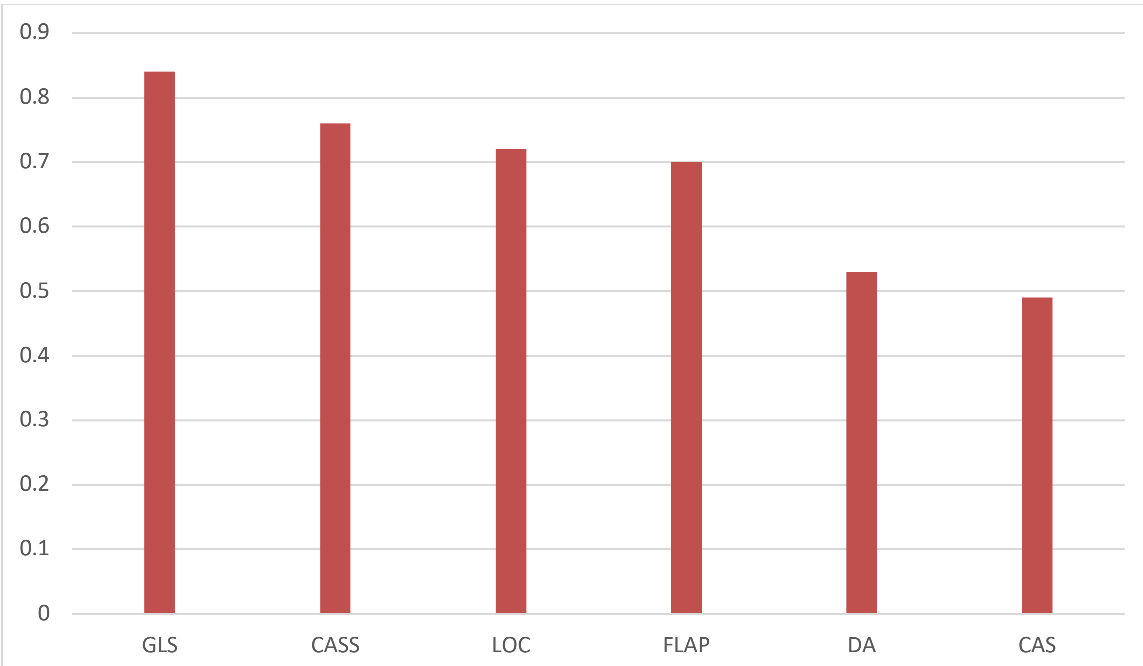

Figure 22. Variable importance plot for predictor variables. Horizontal axis presented important predictors. Vertical axis indicated relative worth. *Note.* GLS=Glideslope Deviation, CASS=Selected Airspeed, LOC=Localizer Deviation, FLAP=flap position, DA=Drift Angle, CAS=Calibrated Airspeed.

Additional details regarding the relative impact of these variables can be analyzed using the best performing model, DT. Figure 23 shows an overview example of the DT model including the validation results used to analyze and interpret the DT model. Figure B3 provides an enlarged presentation of the DT model. The parent node (UARM) indicates that 92.84 percent of the flights do not contain evidence of UARM whereas 7.86

percent are classified as containing evidence of UARM. Probability of UARM occurrence and the effects of impact variables are analyzed and interpreted as follows:

- In the parent node, glideslope deviation (GLS, vertical deviation) was considered an important predictor of occurrence of UARM or non-occurrence of UARM. If the glideslope deviation was greater than 0.17 (GS > 0.17), then there was approximately 99 percent likelihood of the occurrence of UARM. If the glideslope deviation was less than -0.17 (GS <-0.17), there was approximately 98 percent likelihood of occurrence of UARM.
- Distance to Waypoint (DWPT) was determined to be a predictor used in splitting criterion for the DT. If the distance to waypoint was less than 18610 then there was approximately 99 percent likelihood of occurrence of UARM.
- The next important factor was selected calibrated airspeed (CASS). If CASS indicated less than 104 knots indicated airspeed, then 98 percent likelihood of UARM occurred. If CASS was greater than 104, then 2 percent occurrence of UARM was likely.
- Altitude Rate (ALTR) was determined to be a predictor in splitting criterion for the DT. If ALTR was greater than 1704 feet per minute there was a likelihood of approximately 99 percent occurrence of UARM. Consequently, if ALTR was less than 120 feet per minute, then there was also approximately 99 percent likelihood of occurrence of UARM.

- LOC deviation was indicated to be an important predictor of UARM. When LOC (lateral deviation) exceeded 0.12 (LOC > 0.12), there was a likelihood of 98 percent occurrence of UARM. LOC values of less than – 0.12 (LOC < - 0.12) indicated a likelihood of occurrence of UARM of approximately 98 percent.
- CASS was again used in splitting criteria in the next child node in the tree. If CASS indicated greater than 130 knots indicated airspeed, then 99 percent likelihood of UARM occurred. If CASS was less than 130, then only a 2 percent occurrence of UARM was likely.
- Flaps not extended (FLAP) was used in the following node tree decision criteria: flaps not extended indications contributed to the occurrence of UARM 99 percent of the time, whereas flaps extended indicated the occurrence of UARM 0.33 percent likelihood.
- Distance to Waypoint (DWPT) was again used determined to be a predictor used in splitting criterion for the DT. If the distance to waypoint was greater than 2488 then there was approximately 99 percent likelihood of occurrence of UARM. Conversely, if the distance to waypoint was less than 2488, then likelihood of UARM occurrence dropped to approximately 5 percent.
- Approach speed (CAS) was used in the next branch of the tree. If CAS indicated between 150 knots indicated airspeed and 211 knots indicated airspeed, then there was approximately a 90 percent likelihood of UARM occurrence. If CAS was less than 150 KIAS, then 4 percent occurrence of

UARM was likely. Consequently, if CAS was greater than 211 KIAS, then UARM was approximately 0.3 percent.

- Drift Angle (DA) was used in the next branch. If DA indicated greater than 11.8, then 86 percent likelihood of UARM occurred. However, if DA values increased to 12.6 or greater, then likelihood of occurrence of UARM dropped to nearly 0 percent.
- The terminal leaf in the tree used CAS in the splitting criteria. If CAS indicated greater than 150 knots indicated airspeed, then 79 percent likelihood of UARM occurred. If CAS was less than 150, then likelihood of occurrence of UARM dropped to nearly 0 percent.
- Drift Angle (DA) was used in the terminal leaf. If DA indicated greater than 11.9, then 98 percent likelihood of UARM occurred. If DA was less than 11.9, then likelihood of occurrence of UARM dropped to nearly 0 percent.

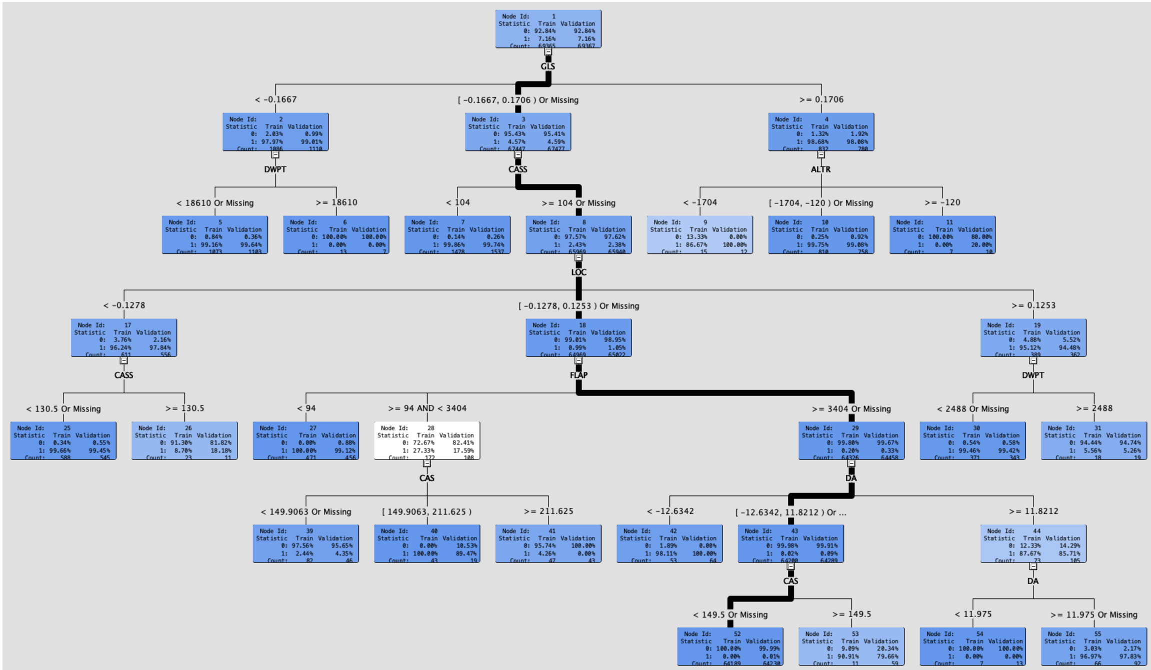

Figure 23. Decision tree model of UARM. Provides a visual representation of the best performing model, Decision Tree model with three branches.

Champion model misclassification rate analysis was used in conjunction with the Cumulative Lift and ROC charts presented in Figures 24 and 25 to demonstrate model validity. The Cumulative Lift chart in Figure 24 provided a model comparison with the use of no model and indicated a lift of 100% with both 10% and 20% of the population sampled, which compared favorably to the no Lift conditions without the use of the model. Additionally, the ROC chart in Figure 25 provided a demonstration of model validity with the very high value of area under the curve between model sensitivity/specificity and the diagonal line of discrimination.

High values of model validity were also demonstrated based on assessment of model accuracy. Model accuracy was a measure of how well the champion model correctly predicted UARM. Table 12 confusion matrix data (TP = True Positive, TN =

True Negative, FP = False Positive, and FN = False Negative) were used to calculate model accuracy as presented in Equation 1.

$$\text{Accuracy} = (TP + TN)/(TP + TN + FP + FN) = .997. \quad (1)$$

Truong et al. (2018) asserted that model predictive accuracy above 80% was considered acceptable performance, therefore the champion DT model demonstrated very high validity.

Predictive power of the champion model was presented using Cumulative Lift and ROC charts in Figures 24 and 25, respectively. In the Cumulative Lift Chart, lift was the ratio of the percent of captured responses within each percentile bin to the average percent of responses for the model. *Cumulative Lift* was calculated by including all data up to the percentile bin. The Cumulative Lift chart provided a depiction of the advantage of using the predictive model regarding probability of UARM over the random prediction of UARM without the use of the predictive model. Perfect prediction models representing 100% sensitivity (no false positives) and 100% specificity (no false negatives) would be represented by sensitivity/specificity of (0,1) on the ROC chart with the AUC (area under curve) between the model results and the line of discrimination maximized. The AUC represented the probability randomly chosen positive results of UARM prediction occurred higher than a randomly chosen negative instance. The substantial AUC indicated the strength of the predictive power of the model.

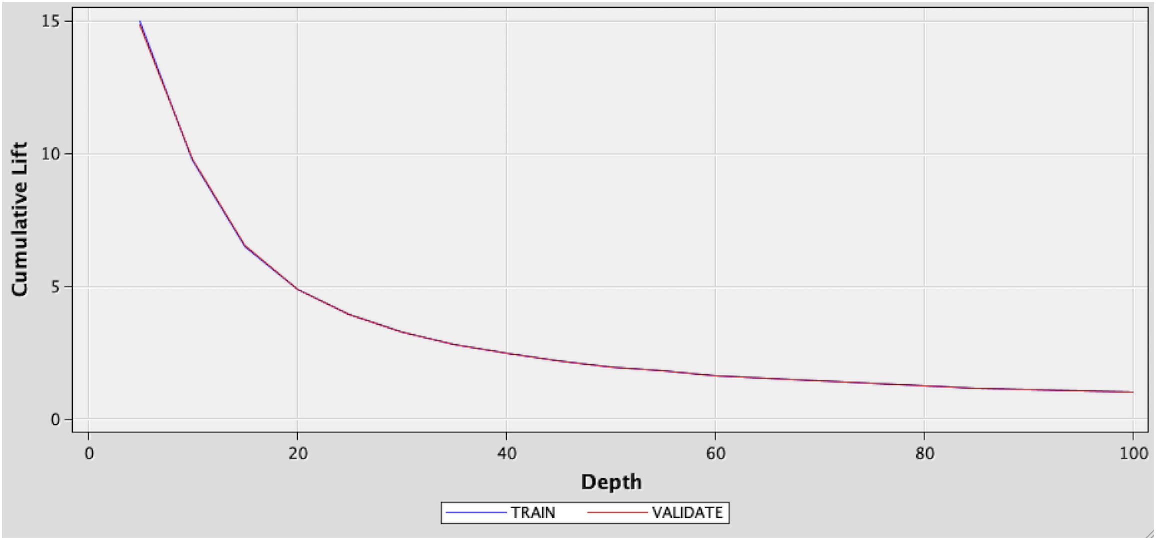

Figure 24. Cumulative lift chart for DT model. Provides a visual representation of the lift of the best performing model, Decision Tree model with three branches. The Chart provided information comparing model performance of UARM prediction using a random process rather than the use of the predictive model.

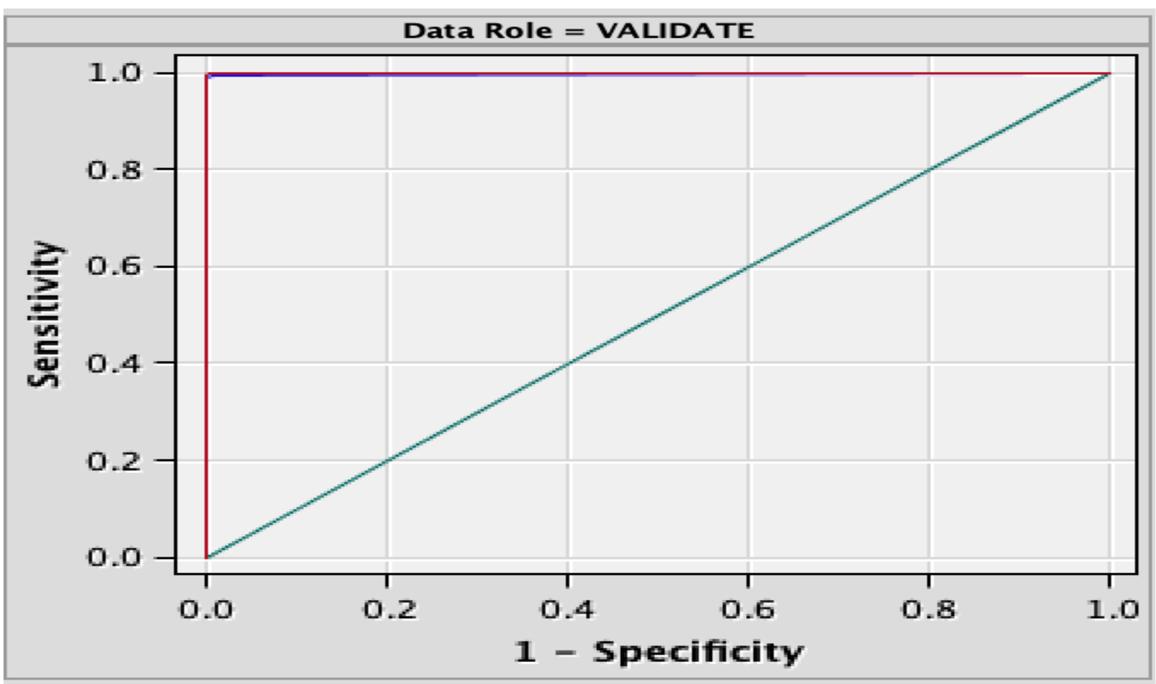

Figure 25. Receiver Operating Characteristics chart for DT model. Provides a visual representation of the predictive power of the best performing model, Decision Tree model with three branches.

Figures 26 and 27 provide information on the leaf statistics and subtree assessment plot of the DT model. The leaf statistics plot provided summary statistics for the best performing model, DT with 3 branches. The subtree assessment plot provided a visual depiction of the number of subtrees in the DT model including a reference line that indicated the number of leaves in the final model. Tuffery (2011) asserted that for classification DT models, the Leaf Statistics chart displayed the percentage of frequency of class levels (UARM or no UARM) within each leaf node. The Subtree Assessment chart in Figure 27 provided a visual representation of how many observations were correctly and incorrectly classified for each value of the target variable, UARM. A low number of misclassifications indicated that the model fit the data. Misclassification rate was determined as depicted in Equation 2.

$$MR = (FP + FN)/(TP + TN + FN + FP). \tag{2}$$

Where MR = misclassification rate, FP = False Positive, FN = False Negative, TP = True Positive, and TN = True Negative.

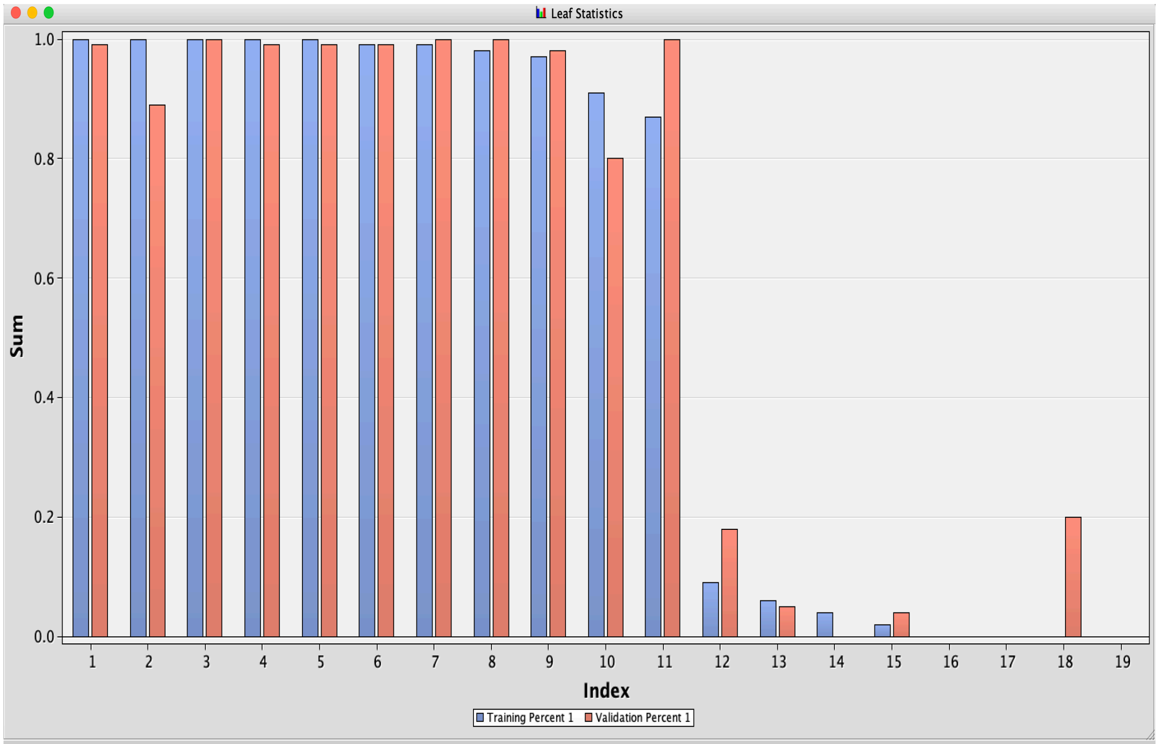

Figure 26. Leaf statistics plot for DT model. Provides a visual representation of the leaf statistics of the best performing model, Decision Tree model with three branches. Leaf statistics contain information about frequency percentages for each class level of the target variable between training and validation data.

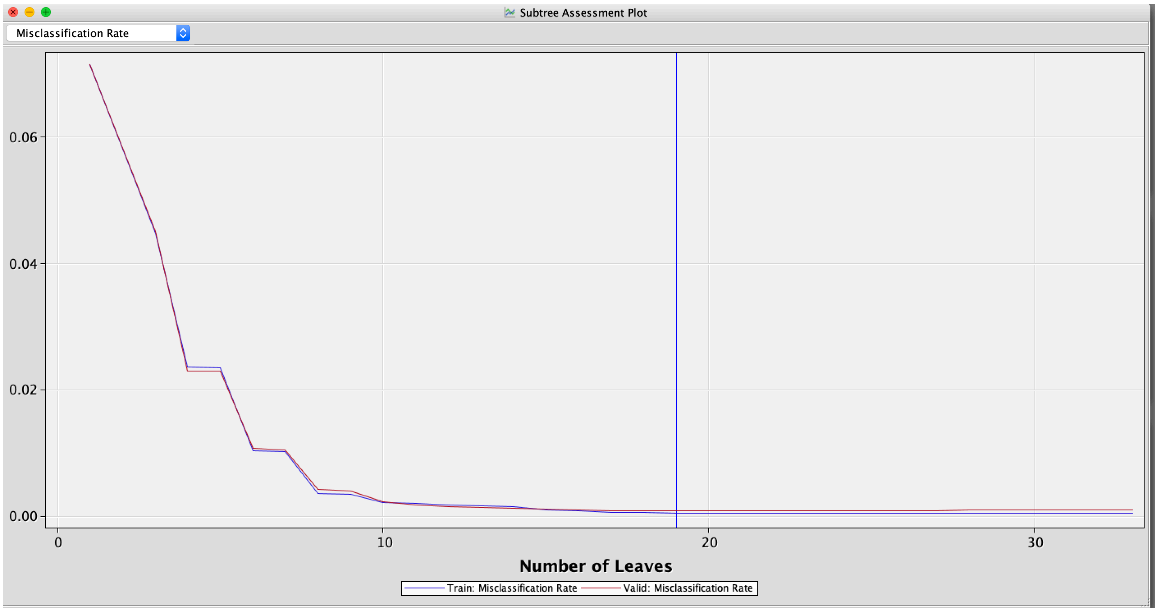

Figure 27. Subtree assessment plot for DT model. The reference line indicated optimum number of leaves in the final model, 19. The chart demonstrated similar performance between training and validation data samples using misclassification rate.

Table 11 shows the variable importance for the best performing model, DT. Relative importance score from 0 to 1. The results indicate Glideslope deviation (GLS) was the most important predictor, followed by (2) selected approach speed (CASS), (3) localizer deviation (LOC), (4) flaps not extended (FLAP), (5) Drift Angle (DA), and (6) approach speed deviation (CAS). These were predictors that contributed most to predicting *Unstable Approach Risk Misperception*.

Table 11

Variable Importance for UARM

Variable	Number of Splitting Rules	Importance
GLS	67	0.84
CASS	48	0.76
LOC	135	0.72
FLAP	56	0.70
DA	45	0.53
CAS	13	0.49

Note. GLS=Glideslope Deviation. CASS=Selected Airspeed. LOC=Localizer. FLAP=Flaps. DA=Drift Angle. CAS=Calibrated Speed.

Scoring

The Score node in SAS® EM™ was used to score and code the DT model. A separate data set, Year 2004 Recorded Flight Data, was used to score the best performing model, DT. The purpose of the coding was to illustrate the relationships between the predictor variables and the target variable. The first task in the scoring process was to evaluate variable importance using the 2004 flight data set.

Oehling and Barry (2019) assert that discovering anomalous flight events is a classification task. An evaluation of this task is commonly performed using a confusion matrix. The purpose of a confusion matrix is to evaluate the number of correct (true) and incorrect (false) results that occur. Table 12 presents the number of true positives, true negatives, false positives, and false negatives. Sensitivity, specificity, precision, and recall can be determined from these results. In the research, the purpose of the confusion matrix was to better understand the predictive accuracy of the DT model for the target variable, UARM. Sarma (2013) asserts that prediction accuracy can be described based on sensitivity, specificity, and overall accuracy. *Sensitivity* measures the true positive fraction, and *specificity* measures the true negative fraction. For example, *sensitivity* describes the ability of the model to correctly predict UARM, while *specificity* describes the ability of the model to correctly predict the non-occurrence of UARM. Truong et al. (2018) describe the overall prediction accuracy as the total accurate prediction number divided by total number, or one minus misclassification rate.

Table 12

UARM Confusion Matrix for Champion DT Model

	Predicted UARM (1)	Predicted UARM (0)
Actual (1)	483(TP)	3(FP)
Actual (0)	17(FN)	6224(TN)

Note. 1=occurrence of UARM, 0=nonoccurrence of UARM. N = 6727. TP=True Negative. TN=True Negative. FP=False Positive. FN=False Negative.

Table 13 presents details of the sensitivity analysis. *Sensitivity* and *specificity* values of UARM prediction are presented. The results indicate that the probability of correctly detecting UARM was very high, with 99% accuracy. The sensitivity for the occurrence of UARM was very high at the 96% level, given the low rate of frequency of the occurrence of UARM as compared to the total number of approaches evaluated. The specificity value showed the probability of correctly detecting the non-occurrence of UARM, at over 99%. Finally, the overall prediction accuracy was 99%, indicating a very strong predictive performance. Truong et al. (2018) assert that prediction accuracies of this level (performance > 80%) indicate very high predictive power.

Table 13

Specificity and Sensitivity of Champion DT Model

	UARM (1)	UARM (0)
Sensitivity	0.966	0.960
Specificity	0.961	0.999

Note. 1=occurrence of UARM, 0=nonoccurrence of UARM.

These SAS codes could be used in conjunction with the DT model to make actual predictions using recorded flight data. Figure 28 presented an example of the SAS scoring code for the DT model. A complete presentation of the SAS codes was included in Figure B2.

```


******      TEMPORARY VARIABLES FOR FORMATTED VALUES      ******;
LENGTH _ARBFMT_12 $     12; DROP _ARBFMT_12;
_ARBFMT_12 = ' '; /* Initialize to avoid warning. */


******              ASSIGN OBSERVATION TO NODE                ******;
IF  NOT MISSING(GLS ) AND
  GLS  <    -0.16633498886108 THEN DO;
  _NODE_  =                    2;
  _LEAF_  =                    1;
  P_UARM0  =                    0;
  P_UARM1  =                    1;
  Q_UARM0  =                    0;
  Q_UARM1  =                    1;
  V_UARM0  =     0.00581395348837;
  V_UARM1  =     0.99418604651162;
  I_UARM  = '1' ;
  U_UARM  =                    1;
  END;
ELSE IF  NOT MISSING(GLS ) AND
```

Figure 28. Example of scoring codes for DT model. Provided a representation of the scoring code in the final model for the DT model with GLS depicted.

Summary

NASA has provided public access to FDR data, which was gathered from 35 regional jets operating in the NAS from 2001-2004. Data mining techniques were utilized to address the nature of the research. Specifically, the successful development and deployment of the UARM algorithm and the procedural guidelines of AVSKD process provided the framework for the treatment of the data as well as the discovery of unstable approaches and evidence of the occurrence or the absence of UARM. The SEMMA process was then successfully applied to the data, with the construction and evaluation of predictive models. These NASA FDR data were modeled using: (a) decision trees with 2, 3, and 5 branches, (b) neural networks, (c) logistic regression, (d) support vector machine, (e) random forest and, (f) gradient boost machine predictive models. Model performance was compared and tested, and achieved the goal of predicting the probability of UARM in the event an unstable approach occurs at 98%. Once the models were analyzed, a successful determination of how DM techniques could be utilized to predict UARM, and which were the most important predictors to the probability of UARM, was determined.

CHAPTER V

DISCUSSION, CONCLUSIONS, AND RECOMMENDATIONS

The study addressed two areas of aviation safety: (a) pilot aeronautical decison making lapses regarding unstable approaches and (b) the ability to predict *Unstable Approach Risk Misperception*. A new algorithm, based on UARM, was successfully developed and deployed to augment other advanced machine learning algorithms used to explore large recorded flight data. Data mining techniques were also successfully applied to recorded flight data in order to develop predictive models. Federal Aviation Administration stable approach exceedance criteria were employed to investigate unstable approaches, their impact factors, and suggested lapses in pilot aeronautical decison making regarding landing or rejected landing. Results of the study demonstrated that large recorded flight data could be used to discover new knowledge in flight operations.

Results of the study successfully demonstated the deployment of a new algorithm which was used as a predictive tool for UARM. The scalability of the new algorithm was employed with the adaptation, application, and comparison of six advanced machine learning algorithms to recorded flight data. The UARM algorithm was then used to predict conditions when pilots continued an unstable approach, rather than executing a rejected landing, which indicated a lapse in pilot aeronautical decision making. Because of this evidence obtained in the demonstration, decision makers should be able to utilize this predictive capability to mitigate the hazards associated with pilot misperecption of runway excursions.

Discussion

A knowledge discovery process was used to facilitate the prediction of a known aviation hazard, runway excursion caused by continuation of an unstable approach to landing. The research questions were addressed regarding the application of machine learning alorithms and predictive modeling of recorded flight data; as well as the prediction of the probability of UARM and how to identify important predictors of UARM. The study supported the findings of Oehling and Barry (2019), whose research showed that ML techniques could be applied to large recorded flight data for purposes of knowledge discovery. Another similarity in the findings of the study and that of Oehling and Barry (2019) was the observation and recommendation that the application of the knowldege discovery process should consider other phases of flight, not only the approach and landing phase. One key difference with Oehling and Barry (2019) was the disagreement in results indicating that the NN model was among the ML techniques with the highest predictive power. Somewhat unexpected was that the Decision Tree model was the model that had the highest predictive power. The literature had indicated that advancements in ensemble learning algorithms such as Random Forest and Gradient Boost Machine should provide modeling capacities that outperform traditional ML algorithms such as Decision Tree-based models. Additionally, this finding did not support the assertion of Paul and Dupont (2015) that Random Forest should perform the best among ML techniques regarding the discovery embedded variables selected in the feature selection process.

The flight data were recorded by FDRs on a fleet of 35 regional jets over a period of four years (2001-2004). NASA had de-identified these data and made them available to the public. These data were analyzed to identify unstable approaches and to construct prediction models. Although the age of the data was in excess of 15 years, several factors ensured the validity of not only the AVSKD process, but also the validity of the data used in the research. At the time of the study, the FAA stable approach exceedance criteria had not been modified or changed since originally developed. The stable approach criteria listed in Appendix 2 of FAA AC 120-71A were still used in SOP guidance for Part 121 US air carriers at both the time of data collection and of the study. The scalability of the UARM algorithm allowed for the implementation of these flight data, and flight data variables were successfully identified to represent the stable approach constructs. Additionally, unstable approaches were successfully identified based on the assessment of flight data variables. The identification and extraction of unstable approach occurrence was critical to the deployment of the UARM algorithm. Additionally, algorithm scalability ensured that no performance fallibility would be expected with the application of flight data gathered on other airframe types or models or timeframe of data collection. For example, the AVSKD and UARM algorithm process would be expected to demonstrate both reliability and validity independent on the data source (i.e., Boeing 747, Embraer 145, Airbus 320, etc.). Scalability of the UARM algorithm included the timeframe of data collection would not have been expected to affect the reliability and/or validity of the model as well. For example, results indicated that the data collected for the study (2001-2004) did not affect model performance, hence, no effects would be expected if the opportunity for a more recent data source became available for public research (i.e., 2017-2019).

The purpose of the research was to utilize machine learning techniques to explore large flight data in order to predict the target variable, *Unstable Approach Risk Misperception.* Machine learning algorithms were used to develop a prediction model for *Unstable Approach Risk Misperception* and to determine important variables that contributed to the prediction of *Unstable Approach Risk Misperception*. Predictive models were constructed based on advanced machine learning algorithms using 186 recorded flight data variables. Specific machine learning techniques applied to the flight data included: (a) decision tree with 2, 3 and 5 branches, (b) logistic regression, (c) neural network, (d) support vector machine, (e) random forest, and (f) gradient boost machine algorithms. Once the models were built and validated, the model with the highest predictive score was used to predict the probability of UARM, which could be used to identify runway excursion hazard. Additionally, SAS® EM™ software was used to rank flight data variables considered to be most important to the occurrence of UARM.

The research was exploratory and data-driven in nature, based on answering two research questions:

Research Question 1. How can the application of data-mining and machine learning techniques to recorded flight data be used to predict the probability of *Unstable Approach Risk Misperception* by the pilot? The research question was addressed with the development of a new algorithm based on unstable approach identification and subsequent occurrence (or not) of UARM. Additionally, predictive models using advanced machine learning algorithms were successfully employed with the application of these models to recorded flight data. Data mining techniques were successfully used to explore a large-volume FDR data from commercial flight operations and to predict *Unstable Approach Risk Misperception*. Data coding was successfully developed using FAA stable approach exceedance criteria, which was applied to FDR flight data. A new algorithm was successfully developed using the *Aviation Safety Knowledge Discovery* model developed and validated by Mathews et al. (2013) at the NASA Ames Research Center. The successful development and deployment of the UARM algorithm was a key accomplishment of the study. The UARM algorithm successfully demonstrated how large flight data could be used to predict the probability of pilot risk misperception regarding the hazard of runway excursion. Evidence of pilot risk misperception was represented by the decision the pilot to continue to a landing even when evidence existed of exceedance in any one or more of the flight data variables from the FAA stabilized approach criteria. For purposes of this research, this target variable was defined as *Unstable Approach Risk Misperception* (UARM). Data mining techniques were used to populate and compare various predictive models and to determine the most accurate model, decision tree with three branches, which was then used to make predictions of the target variable.

In order to predict the probability that UARM would occur during an unstable approach, the following ML algorithms were used to build the models: (a) logistic regression, (b) decision tree with 2, 3, and 5 branches, (c) neural network, (d) support vector machine, (e) gradient boost machine, and (f) random forest. Results indicated flight-related variables representing: (1) glideslope deviation, (2) selected approach speed (3) localizer deviation, (6) flaps not extended, (7) excessive drift angle, and (8) approach speed deviation were the most important predictors of probability of UARM occurrence.

Additionally, the occurrence of a rejected landing, or continued approach to landing, when confronted with evidence of an unstable approach, was also included.

Findings indicated that the DT model performed with the highest predictive power, 96%. Once the DT model was determined to be the highest scoring model, a separate data set (2004 FDR data) was used to: (a) determine the predictive probability of the target variable, UARM, and (b) rank input variables in order of importance. Results of this analysis described the predictive accuracy of UARM as 98%. A sensitivity and specificity analysis was conducted which indicated a true positive prediction of 96% and a true negative prediction of 92%. Thus, the model was acceptable to answer the question of how to predict the probability of UARM.

Research Question 2. What flight data variables are the most important predictors of pilot misperception of a runway excursion hazard as evidenced by continuing an unstable approach to a landing? The research question was successfully addressed by analyzing the recorded flight data, specifically the approach and landing phases of flight which indicated exceedance of FAA stable approach criteria, and using these criteria to develop and compare several different predictive models. The successful development and deployment of the new UARM algorithm provided an appropriate tool to accomplish this task. These data were extracted using snapshots at two assessment windows, once again using FAA stable approach assessment criteria. Data was sampled at a 500 ft AGL assessment window and also at a point of either landing or a rejected landing (WOW either greater than or equal to 0). Assessment was conducted using the flight data variables with FAA AC-120-71A (FAA, 2003) providing guidance for variable selection for the UARM algorithm. For example, (a) target approach speed deviation, (b) flap position, (c) landing gear position, (d) engine speed, (e) altitude above ground level (AGL), and (f) glide path deviation were variables stated in the FAA stable approach criteria categories. Adherence to stable approach criteria was determined based on the data, including: (a) the vertical and lateral position of the aircraft with reference to the landing runway, (b) energy state, and (c) landing configuration. The information gathered in the data analysis was then used to develop models to predict the probability of the pilot misperceiving the runway excursion risk of continuing an unstable approach to landing. Pilot risk misperception was suggested by the decision to continue to a landing even when evidence exists of exceedance in any one or more of the flight data variables from the stabilized approach criteria.

An advantage regarding internal validity of the study was the restriction to FDR data from one air carrier, one type of aircraft, and one FAA certified FDR. This limitation favorably decreased the likelihood of selection bias regarding feature selection and the application of FAA stable approach criteria. For example, the three constructs (energy state, landing configuration, and aircraft location relative the landing runway) were based on exceedance criteria to flight data variables from only one aircraft type. These findings were not unexpected as the literature indicated that deviations in energy management were frequently found to be contributing factors in runway excursions (FAA, 2014).

Variable importance was determined using the best performing model, DT. Findings indicate that six important variables stood out in the prediction of UARM. Glideslope Deviation (GLS) was the most important variable in the prediction model. The other important predictor variables in order of worth: (2) selected airspeed, (3) localizer deviation, (4) flaps not extended, (5) drift angle, and (6) approach speed deviation. Interpretation and effects of these important predictors to the probability of the occurrence of UARM are as follows:

- ***Glideslope deviation.*** Exceedance of glideslope deviation limits was also interpreted to be evidence of energy mismanagement. High deviations would support the exceedance of other high ranked variables regarding excessive energy and was also interpreted to support indications by other exceedance variable importance rankings that indicated energy mismanagement. This variable supports the high approach path that demonstrated *high and fast* at the assessment point.
- ***Approach speed deviation.*** The inclusion of airspeed deviations also supported the interpretation of energy mismanagement. The exceedance of approach speed limitations provided additional evidence of *high and fast* energy mismanagement. When combined with other important predictors of UARM, this flight variable provided support when combined with other important predictors regarding hazard of runway excursion.
- ***Selected Calibrated Approach Speed.*** A new finding that indicated the pilot selection of approach speed was an important predictor in UARM. The literature did not provide information of pilot selection of airspeed as an important predictor involved in either unstable approaches or runway excursions. The inclusion of this flight data variable represented a new discovery in the analysis of unstable approaches and pilot risk misperception of runway excursions. Significance of pilot selected

approach in the prediction of UARM was a key indicator in approach speed deviation contributory factors to unstable approaches and the suggested lapse in ADM. Deviations between approach speed actually flown and selected approach speed were determined to be important in the prediction of UARM.

- ***Localizer deviation and drift angle.*** These two variables indicated exceedance in lateral relative position with the landing runway. Although not a direct indicator of energy mismanagement, lateral exceedances were interpreted to indicate risk misperception of runway excursion (veer off) that could have resulted from inaccurate runway alignment for landing.
- ***Flaps.*** This variable was of key importance supporting the energy mismanagement at the assessment point. If a pilot has the landing flaps at a 0 setting (flaps not deployed) at 500 feet, difficulty in maintaining approach speed and rate of descent would be expected. This variable exceedance supported the interpretation that a combination of flaps up and speed brakes deployed indicated a risk misperception in ability to reduce energy prior to landing.

Exceedances in GS, runway alignment (LOC and DA) and excessive airspeed were noted as off nominal (i.e., FAA exceedance criteria) energy management. As such, analysis of those variables indicating a flight path trajectory of *high and fast* supported the assertion by the NTSB that energy mismanagement issues were contributory factors in many accidents/incidents involving runway excursions (NTSB, 2008; 2016). Findings supported those of NTSB in several accidents and incidents described in previous sections. Pilots attempting to reduce energy with incorrect lift device deployment (speed brakes) was determined to be a contributory factor in American Airlines 1420 as well as several other incidents including runway excursions (NTSB, 2001).

The findings fill the gaps in the literature for: (a) federal guidelines and oversight of hazards associated with unstable approaches and runway excursions, (b) aviation research conducted on pilot risk perception and risk tolerance, and (c) aviation research using predictive modelling based on advanced ML techniques applied to large FDR data. While the literature review described many examples of aviation research in each of these topics, the research filled the gaps as follows:

Unstable approach and runway excursion hazards. Findings of the study supported those provided by the FAA, NTSB, and FSF. These entities described in detail oversight, guidance, and/or recommendations to operators regarding the hazards associated with mitigating the risk of runway excursions. The FAA had listed unstable approaches as one of the most common causal factors in runway excursions (FAA, 2014). The FSF and NTSB corroborated this assertion that stable approaches (and safe landings) begin early in the approach planning phase of flight (FSF, 2009; NTSB, 2016, 2019b). Findings in the study concurred with these organizations that exceedances in stable approach criteria had been demonstrated with the analysis of large volumes of recorded flight data. Interpretation of the results of important flight variable predictors of UARM supported the recommendations by the NTSB for the necessity of more focused data-driven training in pilot ADM regarding risk perception. Results of the study supported the FAA and NSTB call for improved pilot training initiatives, enhanced CRM training, as well as research into risk mitigation strategies for operators to avoid the hazards associated with unstable approaches (FAA, 2017a; NTSB, 2016, 2019b). Although the recorded flight data used in the research was insufficient to determine any actual occurrences of REs, results of important predictors indicated consistent energy mismanagement (*high and fast*) which were listed as contributory factors in runway excursions (runway overruns) in several incidents/accidents by the NTSB. This interpretation also supported the discovery in recent aviation accidents that had demonstrated unstable approaches continue to be causal factors. Results of the study also supported the NTSB recommendation that the aviation industry should respond to the hazard of unstable approaches with improvements in pilot training, as well as the development of CRM techniques to enhance pilot risk assessment and perception in flight operations (NTSB, 2013, 2019b).

Pilot risk perception and risk tolerance. Findings of the study successfully addressed the recommendation for future research by You and Han (2013), who recommended that potential factors affecting pilot lapses in ADM should be investigated. The researchers had also concluded that the safe operational behavior of pilots could be affected by HF characteristics such as ADM, HIP, SA, interpersonal communications and teamwork attitudes. Identification of potential indications of lapses in pilot ADM, particularly those involving risk misperception associated with energy mismanagement on approach were demonstrated. Results corroborated those of Hunter (2005) regarding the correlation between pilot risk perception and hazardous events. Actions by pilots regarding energy management risk misperception were interpreted to have indicated the potential overestimation of ability to reduce energy for landing. Excessive airspeed, rate of descent, and high on glidepath with a continuance to landing, rather than a rejected landing, suggested risk misperception. Results supported the conclusions of Hunter (2009), who asserted that pilot attitudes associated with perception of risk were strongly related to the relative levels of safety inherent in airline operations. Findings also addressed recommendations from Hunter (2009) and addressed gaps in the literature regarding the need for future research in pilot risk perception. Hunter (2009) specifically called for future research to be conducted to identify key factors that contribute to inaccurate perceptions of risk, which was one of the successful accomplishments of the research.

Predictive modeling using recorded flight data. Results of the study addressed gaps in the literature that were identified regarding anomaly detection in large flight data. Findings were in agreement with those of Bharadwaj et al. (2013), who detailed a multifaceted process of discovering and describing unusual events as Anomaly Detection. An area of agreement with Bharadwaj et al. (2013) indicated that anomaly detection was achievable in large flight data with the successful investigation to detect unusual events. One area of difference was that results did not support the use of cluster analysis to identify anomalies, but rather used standardized exceedance criteria. For example, an unstable approach was considered an anomaly in the context of this research, as defined by any exceedance of limitations presented in FAA AC 120-71A. However, results of the research did support the notion that anomaly detection could be described using events that did not fall into normal regions of expectations or standards.

One key difference with results provided by Li et al. (2015) and Aslaner, Unal, and Iyigun (2016) was that the current study demonstrated how SMEs should not necessarily be needed for interpretation and classification tasks. The successful demonstration of the use of standardized FAA exceedance criteria and the scalability of the UARM algorithm precluded the need for SME interpretation of exceedance criteria used in anomaly detection. Other differences in substance with the works of Li et al. (2015) and Aslaner, Unal, and Iyigun (2016) were noted. These researchers applied clustering techniques, rather than standardized exceedance criteria, to flight data to identify anomalies in the takeoff and landing phases of flight. Findings addressed the gap demonstrated in these works that included a vague description of what constitutes abnormal flight events, variables of interest in the clustering analysis, and the lack of a clearly defined target variable. Findings also addressed deficiencies concerning the lack of a clearly described coding process. Another result was the successful demonstration of the use of a repeatable coding process, that was then used to build the models used in the evaluation of the algorithms. Conversely, results corroborated findings in the literature with the successful investigation into the application of advanced ML techniques to large FDR data to discover previously unknown anomalies, as recommended by Li et al. (2015) and Aslaner, Unal, and Iyigun (2016),

Results supported those of Gera and Goel (2015), who suggested that data mining was part of a more general process based on the discovery of knowledge pertaining to large data. Results also successfully addressed recommendations by Tong et al. (2018), who suggested that future research should use large FDM data to explore and discover anomalous events in the NAS. One area of disagreement was the assertion by Tong et al. (2018) that random forest, rather than decision tree models, were best suited to extract the most important features in predicting a target variable, landing speed in their case.

The demonstration of the capability of knowledge discovery process research model to discover important predictors of flight trajectory prediction was an important result. Additional results concurred with and supported the conclusions of Gallego et al. (2018), who produced research with the objective of investigating the effects of operational input variables on the vertical flight path trajectory prediction. Conclusions also supported findings with Kang and Hansen (2018) and Achenbach and Spinler (2018) that future research should incorporate weather-related data for improving accuracy of the predictive models.

Findings of the study built on those provided by Oehling and Barry (2019), who presented the use of ML techniques to detect unknown occurrences in flight data, generated by approximately 300 aircraft, from six different Airbus A320 fleets and sub-fleets, for over 1000 flights per day, from March 2013 to March 2016. Results support assertions by the researchers that methods enhancing the safety knowledge discovery process could be applied to large flight data. The research also built of the results obtained by Oehling and Barry (2019), who described ML in terms of algorithms which learn from the data.

Results of analysis described the strong predictive probability of UARM by the DT model as 98%. A sensitivity and specificity analysis was conducted that indicated a true positive prediction of 95% and a true negative prediction of 99%. These findings supported those of Maxson (2018), Truong et al. (2018), Oehling and Barry (2019), who asserted that models with predictive power above 90% indicated a high level of predictive performance. Considering the high predictive power of the best model, findings indicate that the AVSKD research model was acceptable to address the objectives of the study.

Findings demonstrating the ability to predict hazards supported that FAA assertion that effective SRM strategies should incorporate predictive risk identification and mitigation. For example, the ability to predict the probability of future occurrences of UARM could be useful in the successful safety risk mitigation strategies regarding the risk of runway excursions. Traditional SMS strategies have focused on *reactive* and *proactive* mitigation strategies. FAA guidelines suggested the development of *predictive* techniques in the SRM component of an organization's SMS. The FAA recommended that operators should be able to identify safety issues and spot trends *before they result* in an incident or accident. The evolution of SMS strategies has resulted in the requirement for carriers to develop and implement predictive risk management (FAA, 2007a). A SMS strategy favoring predictive methods rather than reactive is depicted in Figure 24.

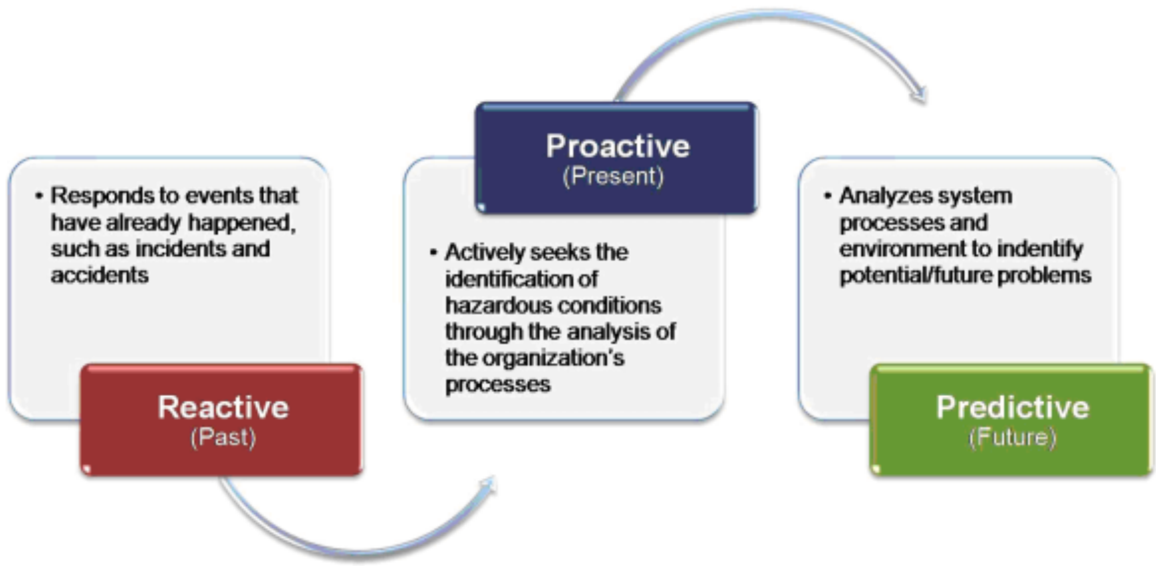

Figure 29. FAA predictive Safety Risk Mitigation strategy. Reprinted from "Safety Management System," by Federal Aviation Administration, 2016. Retrieved from https://www.faa.gov/about/initiatives/sms/explained/basis/

Conclusions

Continuation of an unstable approach to a landing had been identified by civil aviation authorities and the airline industry as one of the primary contributory factors to runway excursion hazards (FAA, 2014; NTSB, 2019b). Although the FAA had stipulated that operators adhere to criteria defining stable approaches, results corroborated industry data that unstable approaches still occur with some pilots not often following the FAA guidance to perform a rejected landing (FSF, 2009). Results indicated that 6% of unstable approaches resulted in a rejected landing, also supported this assertion. Additionally, agreement was noted with FAA-provided LOSA data, which indicated that 97% of unstable approaches resulted in a safe landing and that 3% of unstable approaches resulted in rejected landings (FAA, 2013). With the advent and deployment of advanced digital data recording devices, required under 14CFR §91.609 for all air carriers with an operating certificate, opportunities existed to analyze recorded flight data. Concurrently, recent developments in complex mathematical machine learning algorithms have improved research capability regarding the analysis of these flight data. The successful development and deployment of the UARM algorithm used in the research demonstrated the power and precision that such capabilities could achieve. Data mining techniques, both exploratory and predictive, can provide aviation researchers with the tools necessary to both analyze these large flight data and also to predict abnormal flight occurrences.

Results also successfully demonstrated how the UARM algorithm could be used to identify models that accurately and precisely predicted the probability of pilot misperception of runway excursion risk based on stabilized approach criteria, as well as the identification of important flight variables associated with frequent non-compliance of rejected landing guidelines. The UARM algorithm was also successfully used to identify important predictors of the occurrence of pilot risk misperception of runway excursion risk.

Theoretical contributions. Gaps in the literature were addressed through the demonstration of a reliable and valid methodology to predict pilot *Unstable Approach Risk Misperception*. Although hazards have been identified with the continued occurrences of runway excursions, the literature indicated that a reliable and valid representation of rejected landing decision making based on unstable approach criteria had not been fully investigated. Results addressed this gap with the investigation of potential pilot lapses in aeronautical decision making, specifically in the rejected landing following an unstable approach. Previous studies that have applied advanced data mining techniques to investigate and analyze large flight data have focused primarily on the validation and evaluation of advanced mathematical algorithms. Another important accomplishment was the demonstration of a predictive capability that could be used to mitigate the risk of *Unstable Approach Risk Misperception*.

Two factors indicating opportunities for research became evident with a review of the literature. The first factor was the large amount of data that are being recorded by advanced digital flight recorders on every commercial airline flight in the NAS. Airlines are encouraged by the FAA to voluntarily participate in the FOQA program. FOQA was designed to improve safety in commercial aviation by allowing airlines and pilots to share de-identified aggregate information with the FAA who can then monitor national trends in aircraft operations and focus its resources to address risk issues (e.g., flight operations, air traffic control (ATC), airports). Although voluntary (in the United States), the FOQA program has resulted in very large amounts of flight data that have not been accessed on a scale appropriate for these data. Even though pilot safety reports, accident reports, and safety debrief narratives constitute a large amount of data, the literature indicated that these data have only been explored with the use of text mining and qualitative methods. The research successfully addressed this gap with the demonstration of FDR data analysis and predictive model building. Previous studies relied on cluster analysis or exceedance criteria to discover abnormal flight events, but none offered a reliable and valid predictive model to predict *Unstable Approach Risk Misperception*.

The second factor that indicated opportunities for research was that although various statistical analytical methods have revealed clear patterns in the prediction of pilot performance, the literature indicated these data have not been exploited in order to fully investigate significant relationships of the predictors. An examination of the extant literature indicated a gap in research providing evidence of a relationship between pilot performance and flight anomaly variables. Previous studies described in the literature relied on results provided by the evaluation of subject matter experts, who were required to analyze the results and apply them to aviation problems. The research demonstrated the ability to predict *Unstable Approach Risk Misperception* without the necessity of subject matter expert analysis.

Algorithm development for predictive modeling. A key finding was the demonstration of the successful development and deployment of the UARM algorithm used in the predictive modeling process. The successful application of the AVSKD knowledge discovery process model to large recorded flight data using the UARM algorithm was also a significant finding. The scalability of the UARM algorithm allowed for use of multiple sets of flight data variables to determine pilot risk misperception of runway excursion risk. Additional findings were the discovery of the important predictors of UARM. Results of the knowledge discovery process suggested that pilot risk misperception, specifically regarding energy mismanagement, had a strong relationship with the target variable, UARM. New predictors were flight variables specifically associated with energy management exceedances. Although airbrakes deployed and excessive approach speed were previously reviewed in the literature, key findings of the research included the discovery of additional new important predictors of UARM based on energy mismanagement of being *high and fast* (flaps up and excessive approach speed deviations). The discoveries of new important predictors of pilot risk misperception of runway excursion risk was a key result of the research.

These results provided the successful demonstration that flight data variables could be used to develop the algorithm for UARM. Once this determination was realized, a coding process was developed to create a data variable representing landing, rejected landing, and UARM. An important development in the study was that of a straightforward If/Then decision process used to construct the UARM algorithm (See Figure 8). Results of this If/Then assessment process were successful in the identification that evidence of UARM had occurred or not. For example, once an unstable approach was identified, a determination was made whether or not a rejected landing was performed. *If* evidence of an unstable approach was indicated, and a rejected landing was not performed, *then* UARM resulted. Results of this UARM algorithm development were then successfully used to construct predictive models as well as the identification of UARM. The rest of the flight data variables, including those representing the FAA stable approach criteria, were utilized as input variables and were anticipated to be continuous or categorical. For example, approach speed was expected be continuous, based on numerical values while landing gear position was expected to be categorical (i.e. either up or down).

Results indicated that one advantage of the UARM algorithm was that of scalability. Several different predictive models successfully utilized the UARM algorithm with the application of recorded flight data. Results demonstrated that real world recorded flight data was successfully assessed in the UARM algorithm process to predict the probability of occurrence of the target variable. The UARM algorithm was successfully and repeatedly used with consistent results to evaluate large recorded flight data. The algorithm was successfully developed based on initial data coding, subsequent use of an If/Then decision-making process, and ultimately the extremely accurate and precise predictive power regarding the target variable, UARM. The UARM algorithm provided a step-by-step, repeatable process to the analysis of recorded flight data and allowed for a reliable and valid methodology for the analysis of FDR data to predict UARM.

Practical contributions. Key beneficiaries of the research are airline pilot simulator training programs and airline Safety Management System managers. The ability of airline pilot training managers to not only predict UARM but also identify hazardous trends in aircraft state variables involved in ADM could have a positive impact on airline safety risk mitigation strategies inherent in pilot simulator training programs, such as the development of realistic runway excursion scenarios. Results of the study could be used to further refine not only FAA (2014) stabilized approach criteria but also in the FAA oversight of air carrier pilot training programs.

Safety Management Systems managers could use the results to improve SRM effectiveness, as required under 14CFR Part 5. Because SMS programs have traditionally relied on hazard identification using accident and incident reports rather than predictive measures, predictive capabilities could be beneficial (FAA, 2007a). The ability to predict UARM could provide SMS managers with a predictive tool that would enhance safety risk mitigation effectiveness.

Unstable approaches and runway excursion hazards. Results support positions taken by the FAA, NTSB, and FSF, who have provided oversight, guidance, and/or recommendations to operators regarding the hazards associated with mitigating the risk of runway excursions. Results indicated that unstable approaches, followed by failure of the pilot to execute a rejected landing, continued to occur in the NAS. Interpretation of these results leads one to conclude that efforts to improve training and awareness of the runway excursion hazard have been ineffective. Recommendations for improved pilot training initiatives, enhanced CRM training, as well as research into risk mitigation strategies for operators to avoid the hazards associated with unstable approaches have seemingly not addressed the critical factors (FAA, 2017a; NTSB, 2016, 2019b). Results support evidence of pilot energy mismanagement, which had been discovered to be among contributory factors in runway excursions in recent aviation accidents and incidents (Campbell et al., 2018). Results concur with and support the position taken by the NTSB, which had recommended that the aviation industry respond to the hazard of unstable approaches with improvements in pilot training, as well as the development of CRM techniques to enhance pilot risk assessment and perception in flight operations (NTSB, 2013; 2019b).

Limitations of the findings

The findings should be interpreted for external validity in the context of the following: (a) normal cockpit procedures; (b) calm weather; (c) uncontaminated runway with proper approach and runway lighting; (d) no pilot physiological anomalies; (e) no failures or degradation of aircraft equipment/systems; (f) fully operational navigational aids, (g) crews exercising proper CRM, and (h) proper use of automation, as supported by the 2013 FAA report *Operational Use of Flight Deck Automation Systems* (FAA, 2013).

These limitations provided stimulation for recommendations for future research as well as recommendations to the target population.

Recommendations

The ability of airline pilot training managers to not only predict UARM but also identify hazardous trends in aircraft state variables involved in ADM could have a positive impact on airline safety risk mitigation strategies inherent in pilot simulator training programs, such as developing realistic scenarios for an unstable approach resulting in a runway excursion. Additionally, Safety Management Systems managers should use the predictive capabilities to support the FAA mandate for SRM mandate to be proactive in identifying and mitigating hazards. SMS programs have traditionally relied on hazard identification using accident and incident reports rather than predictive measures.

Recommendations for the target population. Results suggested that advances in airline pilot simulator training programs were necessary. Findings indicate that unstable approaches were followed by a landing occurred at a rate of 94%. Regulatory guidance has been provided and the hazard identified, yet the risk misperception of unstable approaches has not been successfully mitigated. The ability of airline pilot training managers to not only predict UARM but also identify hazardous trends in aircraft state variables involved in ADM could have a positive impact on airline safety risk mitigation strategies inherent in pilot simulator training programs, such as developing realistic runway excursion scenarios.

Recommendation 1. Airline pilot training managers should develop simulator training scenarios to address potential pilot lapses in aeronautical decision making regarding *Unstable Approach Risk Misperception* based on an analysis of recorded flight data. For example, results suggested that pilots facing evidence of unstable approach regarding energy mismanagement attempted to reduce energy with deployement of speed brakes and idle power settings, rather than reject the landing. Improved simulator training could assist pilots in the recognition of energy mismanagement, and the associated hazard of runway excurion based on being *high and fast* (Campbell et al., 2018; FAA, 2013) A goal of this enhanced training should be the goal of improving ADM resulting in increased likelihood of rejected landing when faced with evidence of unstable approach.

Recommendation 2. Airline industry Safety Management System managers should develop strategies, based on an analysis of recorded flight data, to better mitigate unstable approaches by identifying conditions where pilots misperceive the risk. SMS managers should use safety awareness initiatives regarding pilot risk misperception of runway excursion risk. Organizational safety culture should be enhanced to alleviate potential lapses in ADM regarding *Unstable Approach Risk Misperception*. Pilots should be made more aware of the notion that a rejected landing is a successful outcome regarding unstable approach, which should also be supported by air carrier management.

Recommendations for future research. Future research involving data mining, machine learning algorithms, and predictive modeling should focus on more comprehensive flight data sets, not only aircraft state variables, to better represent the complex operational environment. Based on the unexpected result of the identification of selected Mach early in the iterative variable importance investigation, factors related to other phases of flight could be important predictors of UARM. This result was unexpected as selected Mach was not observed to have contributed to the literature on unstable approaches or runway excursion contributory factors. Mach speed is a flight variable associated with the cruise phase of flight, rather than approach and landing variables. Therefore, it is recommended that *all* phases of flight should be investigated, not only the approach and landing phase, to determine if other factors and co-variates contribute to UARM. For example, analysis of pilot descent planning and instrument approach briefings could reveal additonal factors contributing to *Unstable Approach Risk Misperception*. Safety risk mitigation strategies should be included in the predictive model. Furthermore, existing pilot alerting technologies such as Terrain Awareness and Warning System (TAWS) and other technologies that alert the pilot of incorrect aircraft configuration or excessive rate of descent, should be incorporated into the model in order to investigate potential risk reduction strategies.

Recommendation 3. Develop more thorough and comprehensive models that extend beyond recorded flight data sets, to other data sets such as cockpit voice recorder data (e.g. crew/ATC coordination information) affecting decision-making, and weather data for external conditions (e.g. turbulence, wind-shear) affecting aircraft state variables. Predictive modeling of flight events should be enhanced with more thorough replication of those factors which could influence pilot ADM. For example, weather conditions, including runway condition (ice or wet) should be included in flight data in order to enhance and improve predictive capability of models.

Results of the study supported the literaature provided by FAA, NTSB, FSF, and IATA in terms of low SOP compliance rates regarding the unstable approach rejected landing decision-making process. Future research should include the investigation of the effects of not only CRM issues, but automation interface. Findings provided by Giles (2013) were supported by the results of this study with the suggestion that low SOP compliance rates could indicate pilot to machine interface issues. Giles also suggested that safety culture should be investigated for effects on low SOP compliance rates regarding the rejeced landing ADM process.

Recommendations for future testing support those of Achenbach and Spinler (2018), who stated that significant limitations to their study were the lack of real-time weather data, the omission of crew resource management considerations as well as ATC flow control in the construction of predictive models. For example, weather data (e.g. turbulence/convective activity) and ADM aspects of CRM could enhance the knowledge discovery process in the prediction of UARM and should be included in studies predicting pilot risk misperception.

Recommendation 4. Develop and enhance predictive pilot alerting technologies regarding unstable approaches to mitigate runway excursions. Pilot alerting technologies have been previously developed to mitigate various risks, such as controlled flight into terrain, landing with gear up, and landing with evidence of windshear. Similar technological developments should be pursued that would alert pilots when unstable approach conditions are evident and a rejected landing should be executed. For example, AI should be further developed and utilized to enhance current certified technology such as Honeywell's SmartLanding™ and other systems currently under development. AI technology improvements to current systems should be further developed to comply with FAA recommendations shifting current SRM strategies from reactive and proactive to predictive (FAA, 2016).

REFERENCES

Achenbach, A., & Spinler, S. (2018). Prescriptive analytics in airline operations: Arrival time prediction and cost index optimization for short-haul flights. *Operations Research Perspectives, 5*, 265-279. doi:10.1016/j.orp.2018.08.004

Alligier, R., & Gianazza, D. (2018). Learning aircraft operational factors to improve aircraft climb prediction: A large scale multi-airport study. *Transportation Research Part C, 96*, 72-95. doi:10.1016/j.trc.2018.08.012

Aslaner, H. E., Unal, C., & Iyigun, C. (2016). Applying data mining techniques to detect abnormal flight characteristics. *Proceedings of the SPIE, Vol. 9850* (pp. 1-19). doi:10.1117/12.2224061

Benbassat, D., & Abramson, C. I. (2002). Landing flare accident reports and pilot perception analysis. *The International Journal of Aviation Psychology, 12*(2), 137-152. doi:10.1207/S15327108IJAP1202_3

Biswas, G., Mack, D., Mylaraswamy, D., & Bharadwaj, R. (2013). *Data mining for anomaly detection*. Hampton: NASA/Langley Research Center. Retrieved from Advanced Technologies & Aerospace Collection; ProQuest Central (20130011182).

Blair, B., Lee, H. K. H., & Davies, M. (2017). Real-time detection of in-flight aircraft damage. *Journal of Classification, 34*(3), 494-513. https://doi.org/10.1007/s0035

Boeing Commercial Airlines, Aviation Safety. (2017). *Statistical summary of commercial jet airplane accidents worldwide operations 1959 – 2016*. Seattle, WA: Author. Retrieved from https://www.skybrary.aero/bookshelf/books/4239.pdf

Breiman, L. (1997). *Arcing the edge* (Technical Report 486). Department of Statistics, University of California, Berkeley. Retrieved from https://statistics.berkeley.edu/tech-reports/486

Breiman, L. (2001). Random forests. *Machine Learning. 45(*1), pp. 5-32. doi:10.1023/A:101933404324

Brueckner, J. K., & Pai, V. (2009). Technological innovation in the airline industry: The impact of regional jets. *International Journal of Industrial Organization, 27*(1), 110-120. doi:10.1016/j.ijindorg.2008.05.003

Budalakoti, S., Budalakoti, S., Srivastava, A. N., & Otey, M. E. (2009). Anomaly detection and diagnosis algorithms for discrete symbol sequences with applications to airline safety. *IEEE Transactions on Systems, Man, and*

Cybernetics, Part C (Applications and Reviews), 39(1), 101-113. doi:10.1109/TSMCC.2008.2007248

Campbell, A., Zaal, P., Schroeder, J. & Shah, S. (2018). *Development of possible go-around criteria for transport aircraft.* Paper presented at 2018 Aviation Technology, Integration, and Operations Conference. doi:10.2514/6.2018-3198

Chang, Y., Yang, H., & Hsiao, Y. (2016). Human risk factors associated with pilots in runway excursions. *Accident Analysis and Prevention, 94*, 227-237. doi:10.1016/j.aap.2016.06.007

Chen, J., Zhang, X., Zhao, M., & Xia, Y. (2017). Research on extraction of QAR key parameters during approach phase of civil aviation. *Advances in Materials, Machinery, Electrical Engineering (AMMEE 2017)*. Atlantis Press.

Chidambaram, S., & Srinivasagan, K. G. (2018). Performance evaluation of support vector machine classification approaches in data mining. *Cluster Computing.* Springer. doi:10.1007/s10586-018-2036-z

Christopher, A. B. A., Vivekanandam, V. S., Anderson, A. B. A., Markkandeyan, S., & Sivakumar, V. (2016). Large-scale data analysis on aviation accident database using different data mining techniques. *The Aeronautical Journal, 120*(1234), 1849–1866. doi:10.1017/aer.2016.107

Curtis, T., Rhoades, D. L., & Waguespack, B. P., Jr. (2013). *Regional jet aircraft competitiveness: Challenges and opportunities.* Retrieved from https://commons.erau.edu/cgi/viewcontent.cgi?article=1076&context=publication

Das, S., Matthews, B. L., & Lawrence, R. (2011). Fleet level anomaly detection of aviation safety data. *IEEE Conference on Prognostics and Health Management* (pp. 1-10). doi:10.1109/ICPHM.2011.6024356

Das, S., Li, L., Srivastava, A., & Hansman, R. J. (2010). Comparison of algorithms for anomaly detection in flight recorder data of airline operations. *12th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference* (pp. 1-26). doi:10.2514/6.2012-5593

Diallo, O. (2012). A predictive aircraft landing speed model using neural network. *IEEE 31st Digital Avionics Systems Conference* (pp. 1-33). doi:10.1109/DASC.2012.638215

Dismukes, R. K. (2010). Understanding and analyzing human error in real-world operations. In *Human factors in aviation* (pp. 335-374). Academic Press.

Dubey, A., Kamath, S., & Kanakia, D. (2016). Learning data mining techniques. *International Journal of Computer Applications 136*(11), 5-8.

DVB Bank SE. (2019). *An overview of commercial aircraft 2018-2019.* Retrieved from https://www.dvbbank.com/~/media/Files/D/dvbbank-corp/aviation/dvb-overview-of-commercial-aircraft-2018-2019.pdf

Federal Aviation Administration. [FAA]. (1991). *Aeronautical decision making* (Advisory Circular 60-22). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_60-22.pdf

Federal Aviation Administration. [FAA]. (2003). *Standard operating procedures for flight deck crewmembers* (Advisory Circular 120-71A [Cancelled]). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC120-71A.pdf

Federal Aviation Administration. [FAA]. (2004). *Flight operational quality assurance* (Advisory Circular 120-82). https://www.faa.gov/documentLibrary/media/Advisory_Circular/AC_120-82.pdf

Federal Aviation Administration. [FAA]. (2005). *Runway length requirements for airport design* (Advisory Circular 150/5325-4B). https://www.faa.gov/documentLibrary/media/advisory_circular/150-5325-4B/150_5325_4b.pdf

Federal Aviation Administration. [FAA]. (2007a). *Introduction to safety management systems (SMS) for airport operators.* Advisory Circular 150/5200-37. Washington, DC: Author.

Federal Aviation Administration. [FAA]. (2007b). *Flight standards information management system (FSIMS)* (Order 8900.1). Washington, DC: Author. Retrieved from https://www.faa.gov/documentLibrary/media/Order/8900.1.pdf

Federal Aviation Administration. [FAA]. (2011). *Continuous descent final approach.* Advisory Circular 120-108. Washington, DC: Author.

Federal Aviation Administration. [FAA]. (2012). *Aircraft approach category.* Safety Alert for Operators: SAFO 12005. Washington, DC: Author.

Federal Aviation Administration. [FAA]. (2013). *Operational use of flight path management systems.* Final Report of the Performance-Based operations Aviation Rulemaking Committee/Commercial Aviation Safety Team Flight Deck Automation Working Group. Washington, DC: Author. Retrieved from https://www.faa.gov/aircraft/air_cert/design_approvals/human_factors/media/oufpms_report.pdf

Federal Aviation Administration. [FAA]. (2014a). *Mitigating the risks of a runway overrun upon landing.* Advisory Circular 91-79A. Washington, DC: Author.

Federal Aviation Administration. [FAA]. (2014b). *FAA approval of aviation training devices and their use for training and experience.* Advisory Circular 61-136A. Washington, DC: Author.

Federal Aviation Administration. [FAA]. (2015). *National runway safety report.* Washington, DC: Author. Retrieved from https://www.faa.gov/airports/runway_safety/publications/media/Runway-Safety-Report-2013-14.pdf

Federal Aviation Administration. [FAA]. (2016). *Pilot's handbook of aeronautical knowledge* (FAA-H-8083-25). Washington, DC: Author. Retrieved from https://www.faa.gov/regulations_policies/handbooks_manuals/aviation/phak/media/pilot_handbook.pdf

Federal Aviation Administration. [FAA]. (2017a). *Upset prevention and recovery training.* Advisory Circular 120-111. Washington, DC: Author.

Federal Aviation Administration. [FAA]. (2017b). *FAA aerospace forecast, fiscal years 2017-2037.* Washington, DC; Author. Retrieved from https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/fy2017-37_faa_aerospace_forecast.pdf

Federal Aviation Administration. [FAA]. (2018). *FAA stabilized approach and go-around concept.* FAA Safety Brief 18-09. (FAA-AFS-920-18-09). Washington, DC: Author. Retrieved from https://www.faa.gov/news/safety_briefing/2018/media/SE_Topic_18-09.pdf

Federal Aviation Administration. [FAA]. (2019). *Air traffic procedures air traffic bulletin procedures* (ATB 2019-1). Washington, DC: Author. Retrieved from https://www.faa.gov/air_traffic/publications/media/atb_april_2019.pdf

Flight Safety Foundation. [FSF]. (2009). FSF ALAR Briefing Note 7.1: Stabilized approach. *Flight Safety Foundation Approach and Landing Accident Reduction tool kit.* Flight Safety Foundation: Alexandria, VA. Retrieved from https://www.skybrary.aero/bookshelf/books/864.pdf

Forlizzi, L., Güting, R. H., Nardelli, E., & Schneider, M. (2000). *A data model and data structures for moving objects databases*, *29*(2), 319-330. ACM.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics, 29*(5), 1189-1232. doi:10.1214/aos/1013203451

Friso, H. F., Richard, C., Visser, H. G., Vincent, T., & Bruno, D. (2018). Predicting abnormal runway occupancy times and observing related precursors. *Journal of Aerospace Information Systems, 15*(1), 10-21. doi:10.2514/1.I010548

Gallego, C. E., Gómez Comendador, V. F., Sáez Nieto, F. J., Orenga Imaz, G., & Arnaldo Valdés, R. M. (2018). Analysis of air traffic control operational impact on aircraft vertical profiles supported by machine learning. *Transportation Research Part C, (95)*, 883-903. doi:10.1016/j.trc.2018.03.017

Genuer, R., Poggi, J., Tuleau-Malot, C., & Villa-Vialaneix, N. (2017). Random forests for big data. *Big Data Research, 9*, 28-46. doi:10.1016/j.bdr.2017.07.003

Gera, M., & Goel, S. (2015). Data mining-techniques, methods and algorithms: A review on tools and their validity. *International Journal of Computer Algorithms 113*(18)*,* 22-29.

Ghoson, A. M. A. (2011). Decision tree induction & clustering techniques in SAS Enterprise Miner, SPSS Clemintine, and IBM intelligent miner-a comparative analysis. *International Journal of Management & Information Systems 14*(3)*),* 57-70. doi:10.19030/ijmis.v14i3.841

Giles, C. N. (2013). Modern airline pilots' quandary: Standard operating procedures--to comply or not to comply. *Journal of Aviation Technology and Engineering, 2*(2), 1. doi:10.7771/2159-6670.1070

Gorinevsky, D., Mathews, B., & Martin, R. (2012). Aircraft anomaly detection using performance models trained on fleet data. *2012 Conference on Intelligent Data Understanding* (pp. 17-23). doi:10.1109/CIDU.2012.6382

Gregorutti, B., Michel, B., & Saint-Pierre, P. (2015; 2014). Grouped variable importance with random forests and application to multiple functional data analysis. *Computational Statistics and Data Analysis, 90*, 15-35. doi:10.1016/j.csda.2015.04.002

Hapfelmeier, A., & Ulm, K. (2014). Variable selection by random forests using data with missing values. *Computational Statistics and Data Analysis, 80*, 129-139. doi:10.1016/j.csda.2014.06.017

Hu, C., Zhou, S., Xie, Y., & Chang, W. (2016). The study on hard landing prediction model with optimized parameter SVM method. *Proceedings of the 35th Chinese Control Conference* (pp. 4283-4287). doi:10.1109/ChiCC.2016.7554018

Hui, Y., & Fanxing, M. (2012). Application of improved SAX algorithm to QAR flight data. *Physics Procedia (24)* 1406-1413. 2012 International Conference on Applied Physics and Industrial Engineering.

Hunter, D. R. (2005). Measurement of hazardous attitudes among pilots. *The International Journal of Aviation Psychology, 15*(1), 23-43. doi:10.1207/s15327108ijap1501_2

International Air Transport Association [IATA] (2017). *Unstable approach:Risk mitigation, policies, procedures and best practices, 3rd edition.* Montreal: Author. Retrieved from https://www.canso.org/system/files/Unstable%20Approaches%203rd%20Edition.pdf

Ju, C., Ji, M., Lan, J., & You, X. (2017). Narcissistic personality and risk perception among Chinese aviators: The mediating role of promotion focus. *International Journal of Psychology, 52*(S1), 1-8. doi:10.1002/ijop.12243

Ju, X. & Tian, Y. (2012). Knowledge-based support vector machine classifiers via nearest points. *Procedia Computer Science, (9)*, 1240-1248.

Kang, L., & Hansen, M. (2018). Improving airline fuel efficiency via fuel burn prediction and uncertainty estimation. *Transportation Research Part C, 97*, 128-146. doi:10.1016/j.trc.2018.10.002

Koteeswaran, S., Malarvizhi, N., Kannan, E., Sasikala, S., & Geetha, S. (2017). Data mining application on aviation accident data for predicting topmost causes for accidents. *Cluster Computing.* doi:10.1007/s10586-017-1394-2

Kumar, A., & Ghosh, A. K. (2019). Decision tree and random forest–based novel unsteady aerodynamics modeling using flight data. *Journal of Aircraft, 56*(1), 403-409. doi:10.2514/1.C035034

Kumar, V., Pang-Ning, T, & Steinbach, M. (2018). Data mining. In D. Mehta & S Sahni (Eds.), *Handbook of data structures and applications (2nd ed.), (pp. 997-1012).* Portland: CRC Press

Kommadi, B. (2019). *Learn data structures and algorithms with golang.* Birmingham, UK: Packt Publishing

Lauer, F., & Bloch, G. (2008). Incorporating prior knowledge in support vector machines for classification: A review. *Neurocomputing, 71*(7), 1578-1594. doi:10.1016/j.neucom.2007.04.010

Leading Regional Jet Operators. (2005). *Market Share Reporter*. Ed. R. Lazich and V. Burton. Detroit, MI: Gale. Retrieved from https://bi-gale-com.ezproxy.libproxy.db.erau.edu/essentials/article/GALE%7CI2502015786?u=embry&sid=summon

Lee, S., Park, W., & Jung, S. (2014). Fault detection of aircraft system with random forest algorithm and similarity measure. *The Scientific World Journal,* 727359-7. doi:10.1155/2014/727359

Lee, H., & Park, J. (2016). Comparative study on hazardous attitudes and safe operational behavior in airline pilots. *Journal of Air Transport Management, 54*, 70-79. doi:10.1016/j.jairtraman.2016.03.024

Li, L., Das, S., Hansman, R. J., Palacios, R., & Srivastava, A. N. (2015). Analysis of flight data using clustering techniques for detecting abnormal operations. *Journal of Aerospace Information Systems, 12*(9), 587-598. doi:10.2514/1.I010329

Li, L., Hansman, R.J., Palacios, R., & Welsh, R. (2016). Anomaly detection via a Gaussian mixture model for flight operation and safety monitoring. *Transportation Research Part C,* 45-57. doi.10.1016/j.trc.2016.01.007

Lv, H., Yu, J., & Zhu, T. (2018). A novel method of overrun risk measurement and assessment using large scale QAR data. *2018 IEEE Fourth International Conference on Big Data Computing Service and Applications* (pp. 213-220). doi:10.1109/BigDataService.2018.00039

Martinussen, D., & Hunter, D. (2010). *Aviation psychology and human factors.* Boca Raton, FL: CRC Press.

Matthews, B., Das, S., Bhaduri, K., Das, K., Martin, R., & Oza, N. (2013). Discovering anomalous aviation safety events using scalable data mining algorithms. *Journal of Aerospace Information Systems, 10*(10), 467-475. doi:10.2514/1.I010080

Maxson, R. W. (2018). *Prediction of airport arrival rates using data mining methods* (Doctoral Dissertation). Retrieved from https://commons.erau.edu/edt/419/

Mehta, D. P., & Sahni, S. (2018; 2017;). *Handbook of data structures and applications* (2nd ed.). Portland: Chapman and Hall/CRC. doi:10.1201/9781315119335

Mendes, H. (2012). *Study of mathematical algorithms to identify abnormal patterns in aircraft flight data* (Doctoral Dissertation). Institute of Superior Technology, University of Lisbon, Portugal. Retrieved from https://fenix.tecnico.ulisboa.pt/downloadFile/395144228216/TeseHM12Jun2012.pdf

Moriarty, D., & Jarvis, S. (2014). A systems perspective on the unstable approach in commercial aviation. *Reliability Engineering and System Safety,131*, 197-202. doi:10.1016/j.ress.2014.06.019

Mozdzanowska, A., & Hansman, J. (2005). Growth and operating patterns of regional jets in the United States. *Journal of Aircraft, 42*(4), 858-864. doi:10.2514/1.11748

Mugtussidis, I. (2000). *Flight data processing techniques to identify unusual events* (Doctoral Dissertation). Retrieved from https://vtechworks.lib.vt.edu/bitstream/handle/10919/28095/phd.pdf?sequence=1&isAllowed=y

Moretti, L., Di Mascio, P., Nichele, S., & Cokorilo, O. (2018). Runway veer-off accidents: Quantitative risk assessment and risk reduction measures. *Safety Science, 104*, 157-163.doi:10.1016/j.ssci.2018.01.010

National Aeronautics and Space Administration [NASA]. (2012). *Sample flight data for 35 aircraft*. Retrieved from the NASA DASHlink website at https://c3.nasa.gov/DASHlink/projects/85/resources/?type=ds

National Transportation Safety Board [NTSB]. (2000). *Crash during landing, Federal Express Flight 14, McDonnell Douglas MD-11, N611FE, Newark, New Jersey, July 31, 1997* (Aircraft Accident Report NTSB/AAR-00/02). Washington, DC: Author. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR0002.pdf

National Transportation Safety Board [NTSB]. (2001). *Runway overrun during landing, American Airlines Flight 1420, McDonnell Douglas MD-82, N215AA, Little Rock, Arkansas, June 1, 1999* (Aircraft Accident Report NTSB/AAR-01/02). Washington, DC: Author. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR0102.pdf

National Transportation Safety Board [NTSB]. (2008). *Runway overrun during landing, Delta Connection Flight 6448, Embraer ERJ-170, N862RW, Cleveland, Ohio, February 18, 2007* (Aircraft Accident Report AAR-08/01). Washington, DC: Author. Retrieved from http://www.ntsb.gov/investigations/accidentreports/pages/aar0801.aspx

National Transportation Safety Board [NTSB]. (2014a). *Crash during a nighttime nonprecision instrument approach to landing, UPS Flight 1354, Airbus A300-600, N155UP, Birmingham, Alabama, August 14, 2013* (Aircraft Accident Report NTSB/AAR-14/02). Washington, DC: Author. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1402.pdf

National Transportation Safety Board [NTSB]. (2014b). *Descent below visual glide-path and impact with seawall, Asiana Airlines Flight 214, Boeing 777-200ER, HL7742, San Francisco, California, July 6, 2013* (Aircraft Accident Report NTSB/AAR-14/01). Washington, DC: Author. Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/AAR1401.pdf

National Transportation Safety Board [NTSB]. (2016). *Runway overrun due to unstable approach, General Aviation, Embraer S A EMB-505, N32FL, Eden Prairie, Minnesota, August 5, 2013* (Aircraft Accident Report CEN13LA462). Washington, DC: Author. Retrieved from https://reports.aviation-safety.net/2013/20130805-1_E55P_N327FL.pdf

National Transportation Safety Board [NTSB]. (2019a). *Safety advocacy group: most wanted list of safety recommendations.* Washington, DC: Author. Retrieved from https://www.ntsb.gov/safety/mwl/Pages/mwl-4.aspx

National Transportation Safety Board [NTSB]. (2019b). *Stable approaches lead to safe landings.* (NTSTB Safety Alert 077, March 2019). Retrieved from https://www.ntsb.gov/safety/safety-alerts/Documents/SA-077.pdf

Nanduri, A. & Sherry, L. (2016). Anomaly detection in aircraft data using recurrent neural networks (RNN). *2016 Integrated Communications Navigation and Surveillance Conference* (pp.1-8). doi:10.1109/ICNSURV.2016.7486356

Oehling, J. & Barry, D. J. (2019). Using machine learning methods in airline flight data monitoring to generate new operational safety knowledge from existing data. *Safety Science, 114*, 89-104.doi:10.1016/j.ssci.2018.12.018

O'Hare, D. (1990). Pilots' perception of risks and hazards in general aviation. *Aviation, Space, and Environmental Medicine, 61*(7), 599.

Orasanu, J., Davison, J., Ciavarelli, A., Cohen, M., Fischer, U., & Slovic, P. (2001). The many faces of risk in aviation decision making. *Human Factors and Ergonomics Society Annual Meeting Proceedings, 45*(4), 307-310. doi:10.1177/154193120104500409

Patel, T., & Thompson, W. (2013). *Data mining from A to Z: Better insights, new opportunities* (White Paper). Retrieved from SAS® website at http://www.datascienceassn.org/sites/default/files/Data%20Mining%20from%20A%20to%20Z.pdf

Paul, J., & Dupont, P. (2015). Inferring statistically significant features from random forests. *Neurocomputing, 150*, 471-480. doi:10.1016/j.neucom.2014.07.067

Puranik, T. G., & Mavris, D. N. (2018). Anomaly detection in general-aviation operations using energy metrics and flight-data records. *Journal of Aerospace Information Systems, 15*(1), 22-35. doi:10.2514/1.I010582

Ravindran, A., & Meht, D. (2018). Data structures for big data stores. In D. Mehta & S Sahni (Eds.), *Handbook of data structures and applications (2nd ed.), (pp. 983-995).* Portland: CRC Press.

Rogerson, E. C., & Lambert, J. H. (2012). Prioritizing risks via several expert perspectives with application to runway safety. *Reliability Engineering and System Safety, 103*, 22-34. doi:10.1016/j.ress.2012.03.001

Sarma, K. S. (2013). *Predictive modeling with SAS enterprise miner: Practical solutions for business applications* (2nd ed.). Cary, NC: SAS Institute.

Sherry, L., Wang, Z., Kourdali, H. K., & Shortle, J. (2013). *Big data analysis of irregular operations: Aborted approaches and their underlying factors. 2013 Integrated Communications, Navigation and Surveillance Conference* (pp. 1-10). doi:10.1109/ICNSurv.2013.6548548

Shi, D., Guan, J., Zurada, J., & Manikas, A. (2017). A data-mining approach to identification of risk factors in safety management systems. *Journal of Management Information Systems, 34*(4), 1054-1081. doi:10.1080/07421222.2017.1394056

Shish, K., Kaneshige, J., Acosta, D., Schuet, S., Lombaerts, T., Martin, L., & Madavan, A. N. (2017). Aircraft mode and energy-state prediction, assessment, and alerting. *Journal of Guidance, Control, and Dynamics, 40*(4), 804-816. doi:10.2514/1.G001765

Singh, V., Gupta, R. K., Sevakula, R. K., & Verma, N. K. (2016). Comparative analysis of Gaussian mixture model, logistic regression and random forest for big data classification using map reduce. *11th International Conference on Industrial and Information Systems* (pp. 333-338). doi:10.1109/ICIINFS.2016.8262961

Smart, E. (2011). *Detecting abnormalities in aircraft flight data and ranking their impact on the flight.* (Doctoral Dissertation). University of Portsmouth, United Kingdom). Retrieved from https://researchportal.port.ac.uk/portal/en/theses/detecting-abnormalities-in-aircraft-flight-data-and-ranking-their-impact-on-the-flight(d9678b70-41e6-459a-82fb-ba2d12a0f971).html

Takahashi, T. T., & Delisle, M. (2018). (Un) stabilized approach-an introduction to dynamic flight conditions during takeoff and landing climb. In *2018 Aviation Technology, Integration, and Operations Conference* (p. 3500). doi:10.2514/6.2018-35

Thompson, M. (2017). Using ADS-B big data analysis to create stable approach models. *Journal of Air Traffic Control.* Retrieved from: https://trid.trb.org/view/1473370

Tong, C., Yin, X., Wang, S., & Zheng, Z. (2018). A novel deep learning method for aircraft landing speed prediction based on cloud-based sensor data. *Future Generation Computer Systems, 88*, 552-558. doi:10.1016/j.future.2018.06.023

Treder, B., & Crane, B. (2004). *Application of insightful corporation data mining algorithms to FOQA data at JetBlue airways.* Flight Safety Foundation Report. Retrieved from https://flightsafety.org/files/FOQA_data_mining_report.pdf.

Truong, D., Friend, M. A., & Chen, H. (2018). Applications of business analytics in predicting flight on-time performance in a complex and dynamic system. *Transportation Journal, 57*(1), 24-52. doi:10.5325/transportationj.57.1.0024

Tufféry, S. (2011). *Data mining and statistics for decision making.* West Sussex, United Kingdom: John Wiley & Sons.

Vértesy, D. (2017). Preconditions, windows of opportunity and innovation strategies: Successive leadership changes in the regional jet industry. *Research Policy, 46*(2), 388-403. doi:10.1016/j.respol.2016.09.011

Vogt, W. P., Gardner, D. C., Haeffele, L. M., & Vogt, E. R. (2014). *Selecting the right analyses for your data: Quantitative, qualitative, and mixed methods*. Guilford Publications.

Walker, G. (2017). Redefining the incidents to learn from: Safety science insights acquired on the journey from black boxes to flight data monitoring. *Safety Science, 99*, 14-22. doi:10.1016/j.ssci.2017.05.010

Wang, Z. (2016). *A methodology for nowcasting unstable approaches* (Doctoral dissertation). Retrieved from ProQuest Dissertations Publishing. (10135073)

Wang, L., Ren, Y., & Wu, C. (2018). Effects of flare operation on landing safety: A study based on ANOVA of real flight data. *Safety Science, 102*, 14-25. doi:10.1016/j.ssci.2017.09.027

Wang, L., Wu, C., & Sun, R. (2014). An analysis of flight quick access recorder (QAR) data and its applications in preventing landing incidents. *Reliability Engineering and System Safety, 127*, 86-96. doi:10.1016/j.ress.2014.03.013

Wong, D. K. Y., Pitfield, D. E., & Humphreys, I. M. (2005). The impact of regional jets on air service at selected US airports and markets. *Journal of Transport Geography, 13*(2), 151-163. doi:10.1016/j.jtrangeo.2004.04.012

You, X., Ji, M., & Han, H. (2013). The effects of risk perception and flight experience on airline pilots' locus of control with regard to safety operation behaviors. *Accident Analysis and Prevention, 57*, 131-139. doi:10.1016/j.aap.2013.03.036

Zhao, X. B., Bin, L. I., & Wang, C. G. (2017). There is a gold mine in flight data: A framework of data Mining in Civil Aviation. *DEStech Transactions on Social Science, Education and Human Science*. Retrieved from http://www.dpi-proceedings.com/index.php/dtssehs/article/view/12896/12429

APPENDIX A

Data Source Authorization

The archived data utilized in the research were made available to the public by NASA on its DASHlink website and did not require permission for either access or use. The link for the data is: https://c3.nasa.gov/dashlink/projects/85/

APPENDIX B

Flight Data Code

```
close all; clear; clc;
write_values = false;
if write_values
%mat_files_struct = dir('data\*.mat');
mat_files_struct = dir('data\652*.mat');
mat_files = {};
for it = 1:length(mat_files_struct)
if it == 1
prev_percent = 0;
fprintf('%d percent complete.\n', prev_percent);
end
mat_files{end+1} = mat_files_struct(it).name;
S = load(['data\' mat_files{it}]);
if it == 1
% Construct headers and initialize.
data2write = cell(1, 1);
fields = fieldnames(S)';
data2write{1, 1} = 'Tail Number';
data2write{1, 2} = 'Identifier';
for f = 1:length(fields)
data2write{1, f+2} = fields{f};
end
data2write{1, end+1} = 'Landed';
end
% Run checks.
WS_RALT = 5; % Window size to take average of radio
altimeter
(seconds).
MA = movmean(S.RALT.data, S.RALT.Rate*WS_RALT);
if ~any(MA > 500)
%fprintf([mat_files{it} ' is BAD.\n']);
continue;
end
pks_dat = MA - 500;
pks_dat(pks_dat > 0 | pks_dat < -50) = -50;
[~, locs] = findpeaks(pks_dat);
desc = movmean(diff(S.RALT.data), S.RALT.Rate*WS_RALT);
locs(desc(locs - 1) > 0) = [];
% locs is now the location where it is closest to 500ft
altitude on
% descent, ideally there should be only 1 element in locs,
unless a
% go around was executed.
```

```
t = unique(4.*ceil(locs./32));
if ~isempty(t)
% Delete peaks too close (must be greater than 60 seconds
apart).
while any(diff(t) <= 60)
t(logical([1 (diff(t) <= 60)'])) = [];
1
end
else
%fprintf([mat_files{it} ' is BAD.\n']);
continue;
end
if length(t) > 4
%fprintf([mat_files{it} ' is BAD.\n']);
continue;
else
for L = 1:length(t)
if L == 1
r = size(data2write, 1);
end
data2write{r + L, 1} = num2str(mat_files{it}(1:3));
data2write{r + L, 2} = num2str(mat_files{it}(3:
(end-5)));
for f = 1:length(fields)
temp_dat = S.(fields{f}).data;
idx = S.(fields{f}).Rate*t(L);
data2write{r + L, f+2} = S.(fields{f}).data(idx);
end
if length(t) == 1
data2write{r + L, end} = 1;
elseif length(t) ~= 1 && length(t) ~= L
data2write{r + L, end} = 0;
else
data2write{r + L, end} = 1;
end
end
end
% More percentage stuff...
if 100*it/length(mat_files_struct) >= prev_percent + 1
clc;
prev_percent = prev_percent + 1;
fprintf('%d percent complete.\n', prev_percent);
end
end
xlswrite('data500.xls', data2write);
end
vars_for_STD = {...
```

```
'LOC';
'GLS';
'DA';
};
vars_for_STD = sort(vars_for_STD);
data = readcell('data500.xls');
headers = data(1, :);
data(1, :) = [];
% [Lia, Locb] = ismember(A, B)
2
[Lia, Locb] = ismember(vars_for_STD, headers);
for it = 1:length(Locb)
col_mat = cell2mat(data(:, Locb(it)));
pd(it) = fitdist(col_mat, 'Normal');
end
%{
Energy-state
IVV
GS
ALTR
CAS
Landing-config
LGDN - must be 0.
PLA - if all 4 0, unstable approach.
FLAPS - what unique values? histogram. ON HOLD
Physical location state.
LOC
GLS
DA
%}
UNSTABLE = cell(length(data), 1);
UNSTABLE_EXCEEDED_VARS = cell(length(data), 1);
UARM = cell(length(data), 1);
num_unstable = 0;
for r = 1:size(data, 1)
exceeded = {};
[~, idx] = ismember('IVV', headers);
if data{r, idx} > 1000
exceeded = [exceeded {'IVV'}];
end
[~, idx] = ismember('GS', headers);
if data{r, idx} < 70
exceeded = [exceeded {'GS'}];
end
[~, idx] = ismember('ALTR', headers);
if data{r, idx} > 1000
exceeded = [exceeded {'ALTR'}];
```

```
end
[~, idx] = ismember('CAS', headers);
if data{r, idx} < 100 || data{r, idx} > 150
exceeded = [exceeded {'CAS'}];
end
[~, idx] = ismember('LGDN', headers);
if data{r, idx} ~= 0
exceeded = [exceeded {'LGDN'}];
end
```
3
```
[~, idx] = ismember({'PLA_1', 'PLA_2', 'PLA_3', 'PLA_4'},
headers);
if all(cell2mat(data(r, idx)) == 0)
exceeded = [exceeded {'PLA'}];
end
for it = 1:length(Locb)
bounds = [(pd(it).mu - 3*pd(it).sigma), (pd(it).mu +
3*pd(it).sigma)];
if data{r, Locb(it)} < bounds(1) || data{r, Locb(it)} >
bounds(2)
exceeded = [exceeded vars_for_STD(it)];
end
end
if isempty(exceeded)
UNSTABLE{r} = 0;
else
UNSTABLE{r} = 1;
UNSTABLE_EXCEEDED_VARS{r} = strjoin(exceeded, ', ');
num_unstable = num_unstable + 1;
end
if UNSTABLE{r} == 1 && data{r, 189} == 1
UARM{r} = 1;
else
UARM{r} = 0;
end
end
data2write2 = [headers, {'UNSTABLE', 'UNSTABLE EXCEEDED
VARS', 'UARM'};
data, UNSTABLE, UNSTABLE_EXCEEDED_VARS, UARM];
xlswrite('data500_w_EXCEEDANCE.xls', data2write2);
```
Published with MATLAB® R2019a

Figure B1. Flight variable data code.

```
*****************************************************************
******               DECISION TREE SCORING CODE               ******
*****************************************************************

******          LENGTHS OF NEW CHARACTER VARIABLES           ******
LENGTH I_UARM  $   12;
LENGTH _WARN_  $    4;

******               LABELS FOR NEW VARIABLES                 ******
label _NODE_ = 'Node' ;
label _LEAF_ = 'Leaf' ;
label P_UARM0 = 'Predicted: UARM=0' ;
label P_UARM1 = 'Predicted: UARM=1' ;
label Q_UARM0 = 'Unadjusted P: UARM=0' ;
label Q_UARM1 = 'Unadjusted P: UARM=1' ;
label V_UARM0 = 'Validated: UARM=0' ;
label V_UARM1 = 'Validated: UARM=1' ;
label I_UARM = 'Into: UARM' ;
label U_UARM = 'Unnormalized Into: UARM' ;
label _WARN_ = 'Warnings' ;
```

```
******      TEMPORARY VARIABLES FOR FORMATTED VALUES      ******;
LENGTH _ARBFMT_12 $     12; DROP _ARBFMT_12;
_ARBFMT_12 = ' '; /* Initialize to avoid warning. */


******              ASSIGN OBSERVATION TO NODE              ******;
IF  NOT MISSING(GLS ) AND
  GLS  <    -0.16633498886108 THEN DO;
  _NODE_  =                    2;
  _LEAF_  =                    1;
  P_UARM0  =                    0;
  P_UARM1  =                    1;
  Q_UARM0  =                    0;
  Q_UARM1  =                    1;
  V_UARM0  =     0.00581395348837;
  V_UARM1  =     0.99418604651162;
  I_UARM  = '1' ;
  U_UARM  =                    1;
  END;
ELSE IF  NOT MISSING(GLS ) AND
```

```
        0.17062500119209 <= GLS   THEN DO;
    _NODE_  =                    4;
    _LEAF_  =                    9;
    P_UARM0  =     0.01052631578947;
    P_UARM1  =     0.98947368421052;
    Q_UARM0  =     0.01052631578947;
    Q_UARM1  =     0.98947368421052;
    V_UARM0  =                    0;
    V_UARM1  =                    1;
    I_UARM  = '1' ;
    U_UARM  =                    1;
    END;
  ELSE DO;
    IF  NOT MISSING(FLAP ) AND
      FLAP  <                 58.5 THEN DO;
      _NODE_  =                    5;
      _LEAF_  =                    2;
      P_UARM0  =                    0;
      P_UARM1  =                    1;
      Q_UARM0  =                    0;
      Q_UARM1  =                    1;
      V_UARM0  =                    0;
```

```
    V_UARM1   =                    1;
    I_UARM   = '1' ;
    U_UARM   =                    1;
    END;
  ELSE IF  NOT MISSING(FLAP ) AND
                  58.5 <= FLAP  AND
    FLAP  <                3575 THEN DO;
    IF  NOT MISSING(CAS ) AND
                149.6875 <= CAS  AND
      CAS  <           209.0625 THEN DO;
      _NODE_  =                 11;
      _LEAF_  =                  4;
      P_UARM0  =                  0;
      P_UARM1  =                  1;
      Q_UARM0  =                  0;
      Q_UARM1  =                  1;
      V_UARM0  =     0.33333333333333;
      V_UARM1  =     0.66666666666666;
      I_UARM  = '1' ;
      U_UARM  =                  1;
      END;
    ELSE IF  NOT MISSING(CAS ) AND
```

```
                    209.0625 <= CAS  THEN DO;
      _NODE_  =                 12;
      _LEAF_  =                  5;
      P_UARM0  =                  1;
      P_UARM1  =                  0;
      Q_UARM0  =                  1;
      Q_UARM1  =                  0;
      V_UARM0  =                  1;
      V_UARM1  =                  0;
      I_UARM  = '0' ;
      U_UARM  =                  0;
      END;
    ELSE DO;
      _NODE_  =                 10;
      _LEAF_  =                  3;
      P_UARM0  =     0.96153846153846;
      P_UARM1  =     0.03846153846153;
      Q_UARM0  =     0.96153846153846;
      Q_UARM1  =     0.03846153846153;
      V_UARM0  =                0.875;
      V_UARM1  =                0.125;
      I_UARM  = '0' ;
```

```
      U_UARM   =                    0;
      END;
    END;
  ELSE DO;
    IF  NOT MISSING(LOC ) AND
      LOC  <    -0.12377397894096 THEN DO;
      _NODE_  =                  13;
      _LEAF_  =                   6;
      P_UARM0  =                   0;
      P_UARM1  =                   1;
      Q_UARM0  =                   0;
      Q_UARM1  =                   1;
      V_UARM0  =     0.02439024390243;
      V_UARM1  =     0.97560975609756;
      I_UARM  = '1' ;
      U_UARM  =                    1;
      END;
    ELSE IF  NOT MISSING(LOC ) AND
          0.12426399812102 <= LOC  THEN DO;
      _NODE_  =                  15;
      _LEAF_  =                   8;
      P_UARM0  =                   0;
```

```
         P_UARM1   =                     1;
         Q_UARM0   =                     0;
         Q_UARM1   =                     1;
         V_UARM0   =                     0;
         V_UARM1   =                     1;
         I_UARM   = '1' ;
         U_UARM   =                     1;
         END;
       ELSE DO;
         _NODE_   =                    14;
         _LEAF_   =                     7;
         P_UARM0   =     0.99824673254701;
         P_UARM1   =     0.00175326745298;
         Q_UARM0   =     0.99824673254701;
         Q_UARM1   =     0.00175326745298;
         V_UARM0   =     0.99840840362883;
         V_UARM1   =     0.00159159637116;
         I_UARM   = '0' ;
         U_UARM   =                     0;
         END;
       END;
     END;
```

Figure B2. SAS DT data code.

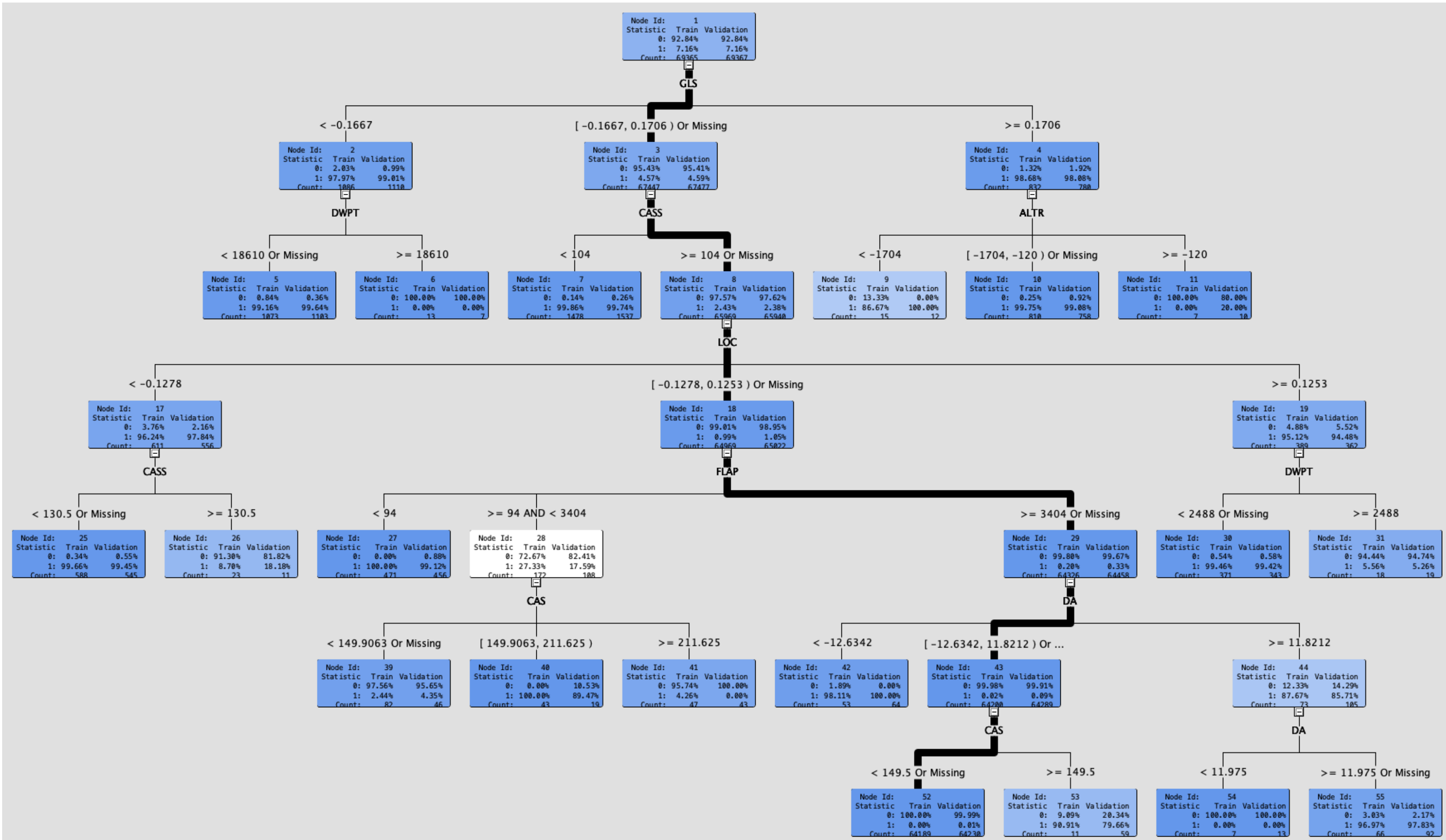

Figure B3. Final decision tree model.

APPENDIX C

Tables

Table C1

Decision Tree Node Properties

Function	Definitions
Node ID	The first DT node in a diagram. Subsequent Decision Tree nodes that are added to a diagram was identified as Tree2, etc.
Imported Data	Consists of a list of data sources to the DT node. The achieved data source, NASA FDM data variables, was used to select the row of data using the following menu options: • Browse – Opens a window to observe the data set. • Explore – Allows sampling and plotting of the data.
Properties	
Exported Data	List of the output that the DT node creates. Was used for selecting the table row using the following: • Browse – Enables observation of the data set. • Explore – Allows sampling and plotting of the data.
Properties	Describes flight data variables.

Property	Definitions of the Properties to Train the Model
Variables	Describes each variable data. Can also be used to generate a report or generate a variable table. The Use Status feature allows selection of input variables and determines the target variable. The Explore option allows viewing of the distribution of a variable.
Interactive	Commences an interactive training session.
Import Tree Model	Determines if the DT node was used to import a model generated using any other DT node. Tree Model Data Set: Enables the selection of the DT model from an earlier iteration of the DT node. Changes to the Score and Report properties for the imported DT can be modified using this node.
Use Frozen Tree	Allows for the iterative process of the creation of a new tree or a frozen tree in the training process.

(continued)

Table C1

Decision Tree Node Properties (Continued)

Property	Definitions of the DT Splitting Rule Node Train Properties
Interval Target Criterion	Notes the constraints considered in potential interval variable splitting rules and to select variables. Criteria options available include: • ProbF – Node variance F-test and p-values. • Variance – Node mean reduction of the square of the error.
Nominal Target Criterion	Evaluates potential splitting rules and to select variables. Criteria options available include: • ProbChisq – Generates Chi-square p-values for the branch node. • Entropy measure • Gini index
Ordinal Target Criterion	Considers rules for splitting and selecting of variables. The following splitting criteria options include: • Entropy measure • Gini index
Significance Level	A *p*-value which represents the worth of potential splitting rules.
Missing Values	Determines how rules consider missing values among the following options: • Use in search – Considers missing data and generates a model with maximum worth. • Most correlated branch – Assigns missing data to branches containing least sums of the squares of residuals • Largest branch – Assigns missing data to the largest branch
Use Input Once	Determines if a flight data variable can be used iteratively or once.
Maximum Branch	Determines how many branches that were used to construct the DT model, from a range of 2 – 100.
Maximum Depth	Specifies the maximum number of generations of nodes that limited the DT model between 1 and 50.
Minimum Categorical Size	Determines the least number of training observations required to a split search.

(continued)

Table C1

Decision Tree Node Properties (Continued)

Property	Definitions of the DT Node Train Properties
Leaf Size	The least number of training iterations contained in a leaf.
Number of Rules	The number of the splitting rules contained in a node. Only one rule was saved and the remaining were used.
Number of Surrogate Rules	Defines how many rules were used by the decision tree when missing values are used in the main splitting rule.
Split Size	The least quantity training observations contained in a node prior to a split.
Property	Definitions of the DT Split Search Node Train Properties
Use Decisions	Uses selected information in the split search.
Use Priors	Prior probabilities are used in the search.
Exhaustive	Indicates how many splits were used in an exhaustive search. Applies to binary splits up to 2,000,000,000.
Node Sample:	Determines node sample size n.
Property	Definitions of the DT Subtree Node Train Properties
Method	Selects a subtree from a mature tree based on the number of leaves:
Assessment	Highest assessment with the smallest subtree.
Largest	The largest tree.
N	Used to determine N, the number of leaves on the largest tree.
Number of Leaves	The highest value of N, which identifies largest number of leaves that were selected in a subtree of n leaves.
Assessment Measure	Used to select the performing DT:
Decision	Was set to Misclassification, using UARM.
Average Square Error	Least average square error.
Misclassification	Lowest misclassification rate.
Lift	Ranks top N% of the observations, based on the prediction of UARM and evaluated using top n% of the ranked observations.
Assessment Fraction	Indicates the proportion of the top n% of observations in the model assessment.

(continued)

Table C1

Decision Tree Node Properties (Continued)

Property	Definitions of the DT Cross Validation Node Train Properties
Perform Cross-Validation	Performs cross-validation of subtrees in the sequence.
Number of Subsets	Calculates cross-validation subsets, up to 20.
Number of Repeats	Calculates the number of iterations in the cross-validation process, up to 100.
Seed	For generating validation sets, generated randomly.
Property	**Definitions of the DT Observation-Based Node Train Properties**
Observation Based Importance	Determines if observation-based importance statistics should be calculated.
Number Single Variable Importance	Determines the number of variables for which statistics should be calculated.
Leaf Variable	Suppresses NODE variables in the output data, when set to NO.
Property	**Definitions of the DT Interactive Sample Node Train Properties**
Create Sample	Uses all of the data or only a sample.
Sample Method	Determines the sampling method. Options are Stratify, First N, or Random. When User is selected to Create Sample, Stratify is utilized.
Sample Size	Determines sample size.
Sample Seed	Indicates the random number generator seed used to sample the data.
Performance	Determines where to store the training data.
Disk	Uses a disk utility file.
RAM	Uses memory.
Property	**Definitions of the DT Score Properties Node**
Variable Selection	Uses value of importance to select variables.
Leaf Role	The Segment, Input, or Rejected properties specify the variable role.

(continued)

Table C1

Decision Tree Node Properties(Continued)

Property	Definitions of the DT Report Properties Node
Precision	To display the number of decimals in the Observation Based Importance, Variable Importance, and Subtree Assessment tables and plots.
Tree Precision	Displays the average values of the nodes and the splitting value in decimal places.
Class Target Node Color	Selects color as display options:
Percent Correctly Classified	Correct classification percentage of observations.
Percent of Event	Colors correspond to the UARM event level.
Single Color	Enables selection of the same color for all nodes.
Interval Target Node Color	Color selection for target variable:
Average	Average target value.
Root Average Square Error	Square root of the average square error.
Single Color	All nodes have a same color.
Node Text	Selects node text to be included in the DT: • Node ID • Show Validation Statistics • Count • Class Targets • Predicted Value • Percent Correct
Property	**Definitions of the DT Status Property Node**
Create Time	When the creation of the node occurred.
Run ID	Node run identifier.
Last Error	Last run error message.
Last Status	Node status.
Last Run Time	Node run time.
Run Duration	Duration of last node run.

Table C2

Results of the Decision Tree

Property	Description of Node Results
Settings	Properties of the DT node.
Run Status	DT node run status, to include: start time, duration, ID, and success of the run.
Variables	Training set data variables
Fit Statistics Table	Presented on Leaf Statistics Plot
Classification Chart	Correct classification a bar chart
Score Rankings Overlay Chart	Presented Data on DT model Scoring
Score Rankings Matrix	Used to demonstrate model accuracy
Score Distribution Plot	Presented results for DT model score
Variable Importance	Importance of each input variable
Observation Based Importance Statistics	Used to present important statistics for DT model
Table	Chart construction variables
Plot	Used to modify an existing Results plot or create a Results plot

Table C3

Data Mining Tasks

Term	Description
Sample the Source of Input Data	Appropriate for large sources of data to decrease model training time.
Create Partitioned Data Sets	Splits the data sample into training, validation, and test data sets. The training data set is used to calculate the regression equation. The validation data set can be used to fine-tune stepwise regression models (prevent the models from over-fitting the training data). The validation data set is also used by default for model assessment. The test data set can be used to obtain an unbiased estimate of the generalization error of a model.
Select Important Variables	Although prior knowledge is valuable in the selection of important variables to unstable approaches, the exploratory nature of the study encourages the use of the entire data set to train the regression model. This may however, increase training time for the regression model as well as negatively affect the prediction results. The Multiplot node was used to generate exploratory plots to help identify important predictors. This iterative process will also help in the rejection of unimportant predictors using the Variable Selection node.
Transform Data and Filter Outliers	Used to stabilize predictor values using appropriate transformation options such as log, exponential, inverse and square root.
Property	**Regression Node Functions**
Node ID	The first DT node in a diagram. Subsequent Regression nodes that are added to a diagram were identified as R2, etc.
Imported Data	Consists of a list of data sources to the REG node. The achieved data source, NASA FDM data variables, were used to select the row of data using the following menu options: • Browse – opens a window to observe the data set. • Explore – allows sampling and plotting of the data.
Exported Data	Lists output data created by the REG node. Was used to select the table row using the following: • Browse – Enables observation of the data set. • Explore – allows sampling and plotting of the data.
Properties	Describes flight data variables.

Table C4

Equation Modeling Properties

Property	Description
Main Effects	Used to suppress, or not to suppress input and rejected variables.
Two-Factor Interactions	Used to include, or not, two-factor interaction of class variables.
Polynomial Terms	Used to include, or not, polynomial terms for interval variables.
Polynomial Degree	Includes, or not, the highest degree polynomial terms to be included in the regression analysis.
Term Editor	Was used to specify the variable interaction terms.
Property	RN Model Options Train Properties
Suppress Intercept	Used for classification of variables, not used for ordinal target variables.
Input Coding	Enables method of coding class variables.
General Linear Models (GLM)	Uses dummy coding, to calculate differences between levels. Coding of 1 is used for the dummy indicator variables, except for the terminal level, which is represented by 1.
Deviation	Effects coding, calculates differences between specific levels and the average value.
Property	RN Model Selection Train Properties
Selection Model	Identifies the type of training model.
Backward	Initial iteration includes all predictors and filters with each iteration until Significance Level or the Stop Criterion has been reached.
Forward	Initial iteration includes no predictors and adds predictor values begins until the Entry Significance Level or the Stop Criterion has been reached.
Stepwise	Initial iteration commences in Forward but continues until Stay Significance Level or Stepwise Stopping Criteria have been reached.
None	Initiates modelling with all input variables.
Selection Criterion	Available options to select final criteria: • Validation Error – selects the model with the lowest error rate based on the validation data set. • Validation Misclassification – Selects the model with the lowest misclassification rate. • Cross-Validation Error – Selects the model with the lowest error rate value of negative log-likelihood for logistic regression. • Cross-Validation Misclassification – Selects the model with the lowest misclassification rate.
Use Selection Defaults	Used to set select model selection criteria or can be set to criteria properties based on defined values, such as exceedance of flight data variables.

(continued)

Table C4

Equation Modeling Properties (Continued)

Property	Selection Options
Sequential Order	Adds or removes variables based on the model statement.
Entry Significance Level	Used in forward and stepwise regression to add variables.
Stay Significance Level	Used in forward or stepwise regression to remove variables.
Start Variable Number	Used to specify initial the number of predictors.
Force Candidate Effects	Used to enter the number of variables that Were used in all candidate models.
Hierarchy Effects	Applies hierarchy rules in the selection process, using the following options: • All – Applies to all variables • Class – Applies to only class variables. • Moving Effect Rule – Options allow either the application of hierarchy or model effects are removed in the iterative process, using the following options: • None – The application of hierarchy is not utilized. • Single – Single predictors are used in the iterative process and hierarchy applications are utilized. • Multiple – Multiple predictors are used in the iterative process with the application of hierarchy utilized. • Maximum Number of Steps – Used to, n, the maximum number of steps that Were used in the stepwise model effect selection process.
Property	**RN Optimization Options Train Properties**
Technique	• Default • Congra – For modelling with more than 500 variables. • Dbldog – Double Dogleg optimization technique. • Newrap – Newton-Raphson with Line Search optimization technique. • Nrridg – Newton-Raphson with Ridging optimization technique. • Quanew – Quasi-Newton optimization technique. • Trureg – For modelling with less than 40 variables. • Default Optimization – Used to select model default optimization. • Max Function Calls – Maximum allowed in the optimization technique. • Maximum Time – Limits processing time.

(continued)

Table C4

Equation Modeling Properties (Continued)

Property	RN Convergence Criteria Train Properties
Uses Defaults	Sets default convergence criterion values.
Options	• Absolute - Absolute convergence criterion. • Absolute Function - Used to specify an absolute function convergence criterion. • Absolute Function Times – Constraints are applied requiring a specific number of successful iterations. • Absolute Gradient - Used to specify the absolute gradient convergence criterion that Was used. • Absolute Gradient Times – Limits termination to that of successful convergence criterion achievement. • Absolute Parameter - Used to specify the absolute parameter convergence criterion. • Absolute Parameter Times – Sets limits to that of satisfaction of parameter convergence. • Relative Function – Used to specify the relative function convergence criterion. • Relative Function – Used to specify the relative function convergence criterion. • Relative Function Times – Sets limits to that of satisfaction of relative convergence. • Relative Gradient – Used to specify the relative gradient convergence criterion. • Relative Gradient Times – Sets limits to that of satisfaction of relative gradient criterion.
Property	**RN Output Options Train Properties**
Confidence Limits	Used to set the Confidence Limits property of the RN.
Save Covariance	Used to set the Save Covariance property.
Covariance	Used to set the Covariance property of the RN.
Correlation	Used to set the Correlation property of the RN.
Statistics	Used to set the Statistics property of the RN.
Suppress Output	Used to set the Suppress Output property of the RN.
Details	Used to set the Details property of the RN.
Design Matrix	Used to set the Design Matrix property of the RN.
Property	**RN Score Properties**
Excluded Variables	Specifies processes for excluded variables.
None	Role unchanged.
Hide	Are removed from the node.
Reject	Role is rejected.

(continued)

Table C4

Equation Modeling Properties (Continued)

Property	RN Status Properties
Create Time	Node creation time.
Run ID	Node run identifier.
Last Error	Any error message displayed.
Last Status	Status of the node is displayed.
Last Run Time	Time of last run displayed.
Run Duration	Last node run amount of time displayed.
Grid Host	Node run grid server displayed.
User-Added Node	Extension node of SAS Enterprise Miner.
Property	**RN Model Selection Methods**
	Allows the selection of specific effects in the regression modeling process. An iterative process Was used in the effort to optimize selected criterion among the following options:
Backward	Not recommended when the target is binary, will not be used in the study.
Forward	Iteratively adds effects until termination criteria are met.
Stepwise	Similar to Forward, but can also remove items not significantly associated with the target variable.
None	Includes all effects in the model.
Property	**RN Model Selection Criteria**
	Based on the Entry and/or Stay Significance Levels using the following options:
None	Uses the last model produced as the final model.
Validation Error	In logistic regression modeling, the error is the negative log-likelihood.
Validation Misclassification	Uses model with the lowest misclassification rate.
Cross-Validation Error	Uses the model with the lowest cross-validation error rate. For logistic regression models, the error is the negative log-likelihood.
Cross-Validation Misclassification	Uses the model with the lowest cross-validation misclassification rate.

(continued)

Table C4

Equation Modeling Properties (Continued)

Property	Input Coding Categorical Variables with the RN
Input Coding Node	Used to choose the coding method that Was used with class variables with the following options:
None	Does not maintain hierarchy (default).
Single	Hierarchy is used to remove or allow to remain, only one variable per iteration.
Multiple	Hierarchy is used to remove or allow to remain, more than one effect variable per iteration.
Results Window	After a successful iteration, this window displays the following: • Properties o Settings o Run Status - displays start time, run duration, and completion status. o Variables –Table of the training data variables. o Train Code – Training code is displayed. o Notes • SAS Results o Log – Regression run log. o Output – Regression run output. o Flow Code – Flow diagram code. • Scoring o SAS Code - Node score code. o PMML Code —Node PMML code • Assessment o Fit Statistics – In table format. o Classification Chart - The Classification chart will display results for UARM, the target variable. The horizontal axis Was used to display the target levels with colors being used to identify the classification of observations. The percentage of total observations Was represented by the height of the bar. o Decision Chart – Was used to the percentage of correct classification observations to misclassified observations for the training and validation iterations. o Score Rankings Overlay – The vertical axis will display statistics for each observational grouping. Best measures Were represented using the model that correctly predicts UARM for all observations. o Score Rankings Matrix - Overlays statistics for standard, baseline and best models using data from the training and validation sets. Best measures Were represented using the model that correctly predicts UARM correctly for all cases. o Score Distribution – Vertical axis contains nonevents. The horizontal axis is the model score of a bin.
Model	Used to display descriptive information about the variables, among the available options: • Effects Plot – Displays a bar graph of the coefficients. • Estimates Selection and Iteration Plots. Displays a graph of statistics of the variables used in the modeling. process.
Table	Displays tables of variables applicable to the steps in the modeling process.

Table C4

Equation Modeling Properties

Property	Output from Logistic Regression Runs
Response Profile	Displays the order the levels of UARM, using Ordered Values of 1to indicate event observations and Ordered Values of 2 to indicate nonevent observations.
Input Class Level Information	Creates a table listing the values of the design matrix.
Model Fitting Information	Used to compare models.
Type III Analysis of Effects	Were used to provide overall tests for the model effects.
Analysis of Maximum Likelihood Estimates	Was used to display significance tests of individual model parameters.
Odds Ratio Estimates	Were used to display odds ratio estimates for each main effect in the model that is not involved in an interaction.
Property	Regression Node Output Data Sets
	The Exported Data Was used to list the two types of data sets that the Regression node outputs.
Scored Data Sets	Will contain the scored test, validation, and training data sets inputs and scores.
Parameter Estimates Data Set	Provides fit statistic information.

Table C5

HP SVM Node Functions

Node	Function
Node ID	Assigns ID codes HPSVM1, HPSVM2, etc.
Imported Data	Inputs data sources into the HP SVM node with the following options: • Browse • Explore • Properties – Contains Table and Variables tabs.
Exported Data	Lists output data with the same options as Imported data.
Notes	Allows the storage of notes of interest.
Variables	Allows the selection of menu options including: “Explore” to view data sampling information, “use” and “report” for further variable functional options in addition to the following: • Apply – opens a submenu to allow the following options: • Reset • Label • Mining – Opens a submenu for the following options: Order, Lower Limit, Upper Limit, Creator, Comment, and Format Type for each variable. • Basic – opens submenu for the following: Type, Format, Informat, and Length of each variable. • Statistics – statistics for each variable. • Explore – Used to view attributes of a data variable, to include: sampling information, observation values, and distribution.
Maximum Iterations	Selects the maximum number of iterations for optimization.
Use Missing as Level	Used to select whether missing values are used.
Tolerance	Used to select termination tolerance for optimization.
Penalty	Assigns penalty value.

(continued)

Table C5

HP SVM Node Functions (Continued)

Node	Function
Optimization Method	Selects method, either "interior point" or "active set."
Interior Point Options	Allows the following submenu options: • Kernel – Used to select the desired kernel function, either "linear" or "polynomial." • Polynomial Degree – Used to select the degree of the polynomial function.
Active Set Options	Opens a submenu allowing for the following options: • Kernel – Linear or polynomial. • Polynomial Degree • RBF Parameter – Used to select the radial basis function.
HD SVM Node	The node contains the following properties: • Create Time • Run ID • Last Error • Last Status • Last Run Time • Run Duration • Grid Host – Used to display the grid server used in the run. • User-Added Node – Used to indicate if extension code was created.
"Results" node	Used to open a submenu containing the following options: • Properties • Settings – Displays configuration information. • Run Status – Displays Start Time, Run Duration, and run success information. • Variables – Displays variable name, use, report, role, and level. • Train Code – Displays the training code.

(continued)

Table C5

HP SVM Node Functions (Continued)

Node	Function
Notes	SAS Results • Log • Output • Flow Code
SAS Code	SAS created Score
Assessment	Used to allow the display of the following information: • Fit Statistics • Classification Chart – Used only in logistic regression to display classification results for the categorical target variable, UARM. • Score Rankings Overlay – Was used to display all observations in descending order of rank, for the binary target variable, UARM. Used to display the following information about training and validation statistics as well as “best” measures related to correct prediction of the target variable for all observations: o Cumulative Lift o Lift o Gain o % Response o Cumulative % Response o % Captured Response

(continued)

Table C5

HP SVM Node Functions (Continued)

Node	Function
Score Rankings Matrix	Used to overlay statistics for standard, baseline, and best models as defined by and validation data sets and displays the following: • Cumulative Lift • Lift • Gain • % Response • Cumulative % Response • % Captured Response • Cumulative % Captured Response
Score Distribution	Plots the proportion of events to the model score. Once again, bins Were used to group categorical variables. Choice of chart options Were as follows: • Percentage of Events • Number of Events • Cumulative Percentage of Events • Report Variables
Model	SVM Fit Statistics Model Information
Table	Used to display a table containing pertinent data.
Plot	Used to open a chart selection menu which can be used to customize chart display.

Table C6

Recorded Flight Data Variables

Variable	Description	Unit	Scale
A_T	Thrust automatic on		Binary
AB	Airbrake	DEG	Binary
ACID	Aircraft Number		
ACMS	Aircraft Maint System		
AIL_1	Left Aileron Position	DEG	Cont
AIL_2	Right Aileron Position	DEG	Cont
ALT	Pressure Altitude	FEET	Cont
ALTR	Altitude Rate	FT/MIN	Cont
ALTS	Selected Altitude	FEET	Cont
AOA1	Angle of Attack 1	DEG	Cont
AOA2	Angle of Attack 2	DEG	Cont
AOAC	Calibrated AOA	DEG	Cont
APFD	Flight Director		
ATEN	Autothrust Engaged		binary
BAL1	Baro Correct Altitude LSP	feet	cont
BAL2	Baro Correct Altitude LSP	feet	cont
BLAC	Body Longitudinal Acceleration	G	cont
BLV	Bleed Air All Valves		binary
BPGR_1	Brake Pressure LH Green	psi	cont
BPGR_2	Brake Pressure RH Green	psi	cont
BPYR.1	Brake Pressure LH yellow	psi	Cont
BPYR_2	Brake pressure RH Yellow	psi	cont
CALT	Cabin High Altitude		binary
CAS	Calibrated Airspeed	Knots	cont
CASM	Max Allowable Airspeed	Knots	cont
CASS	Selected Airspeed	knots	cont
CCPC	Control Column Position Captain	counts	cont
CCPF	Control Column Position FO	counts	cont
Variable	Description	Unit	Scale
CRSS	Selected Course	deg	cont
CTAC	Cross Track Acceleration	G	cont
CWPF	Control Wheel Position FO	counts	cont
CWPC	Control Wheel Position Captain	counts	cont
DA	Drift Angle	deg	cont
DATE-DA	date-day		
DATE-MO	date-month		
DATE-YR	date-year		
DFGS	DFGS 1 & 2 Master		

DVER-1	Database Version Char 1		
DVER-2	Database Version Char 2		
DWPT	Distance to Waypoint		
EAI	engine anti-ice		binary
ECYC_1	Engine Cycle 1	hours	
ECYC_2	Engine Cycle 2	hours	
ECYC_3	Engine Cycle 3	hours	
ECYC_4	Engine Cycle 4	hours	
EGT_1	engine exhaust gas temperature	degrees	cont
EGT_2	engine exhaust gas temperature	degrees	cont
EGT_3	engine exhaust gas temperature	degrees	cont
EGT_4	engine exhaust gas temperature	degrees	cont
EGRS_1	Engine Hours 1		
EGRS_2	Engine Hours 2		
EGRS_3	Engine Hours 3		
EGRS_4	Engine Hours 4		
EVNT	Event Marker		
FADF	FADEC FAIL All engines		binary
FADS	FADEC Status All engines		binary
FF_1	fuel flow	pounds per hour	cont
FF_2	fuel flow	pounds per hour	
FF_3	fuel flow	pounds per hour	
FF_4	fuel flow	pounds per hour	

Variable	Description	Unit	Scale
FGC3	DFGS Status 3		
FIRE_1	fire loop 1		
FIRE_2	fire loop 2		
FIRE_3	fire loop 3		
FIRE_4	fire loop 4		
FLAP	flap position	deg	cat
FPAC	flap augmentation computer		
FQTY_1	fuel quantity 1		
FQTY_2	fuel quantity 2		
FQTY_4	fuel quantity 3		
FRMC	Frame Counter		
GLS	Glideslope Deviation	DDM	cont
GMT_HOUR	Greenwich Mean Time	hours	
GMT_MIN	Greenwich Mean Time	minutes	
GMT_SEC	Greenwich Mean Time	seconds	
GPWS	ground proximity warning system 1-5		cat
GS	ground speed	knots	cont
HDGS	Selected Heading	deg	cont
HF1	high frequency radio 1		
HF2	high frequency radio 2		
HYDG	Low Hydraulic Pressure Green		binary

HYDY	Low Hydraulic Pressure Yellow		binary
ILSF	ILS Frequency LSP		
IVV	Inertial Vertical speed	feet per minute	cont
LATG	Lateral Acceleration	G	cont
LATP	Latitude Position LSP		
LGDN	Gear Left and Right Down and locked		binary
LGUP	Gear Left and Right Up and Locked		binary
LMOD	Lateral Engage Modes	counts	
LOC	localizer deviation	DDM	cont
LONG	Longitudinal Acceleration	G	cont
LONP	Longitude position LSP	deg	cont

Variable	Description	Unit	Scale
MACH	Mach speed		cont
MH	magnetic heading	deg	cont
MNS	Selected Mach	Mach	cont
MRK	Markers-Inner, Middle, outer		
MSQT_1	Squat Switch LH main gear		binary
MSQT_2	Squat Switch RH main gear		binary
MW	Master Warning		binary
N1C	engine speed	PRPM	cont
N1CO	engine core speed	rpm	cont
N1T	engine compressor temp	deg	
N1_1	fan speed 1	% RPM	cont
Variable	Description	Unit	Scale
N1_2	fan speed 2	% RPM	cont
N1_3	fan speed 3	% RPM	cont
N1_4	fan speed 4	% RPM	cont
N2_1	core speed 1	% RPM	cont
N2_2	core speed 2	% RPM	cont
N2_3	core speed 3	% RPM	cont
N2_4	core speed 4	% RPM	cont
NSQT	Squat Swith nose gear		binary
OIPL	low oil pressure all engines		binary
OIP_1	oil pressure	psi	cont
OIP_2	oil pressure	psi	cont
OIP_3	oil pressure	psi	cont
OIP_4	oil pressure	psi	cont
QIT_1	oil temperature	degrees	cont
QIT_2	oil temperature	degrees	cont
QIT_3	oil temperature	degrees	cont
QIT_4	oil temperature	degrees	cont
PACK	air conditioning packs all		binary
PH	Flight Phase from ACMS		
PI	Impact Pressure LSP	mb	cont

Variable	Description	Unit	Scale

PLA_1	Power Lever Angle 1	deg	cont
PLA_2	Power Lever Angle 2	deg	cont
PLA_3	Power Lever Angle 3	deg	cont
PLA_4	Power Lever Angle 4	deg	cont
POVT	Pylon Overheat all engines		binary
PS	static pressure LSP	in.	cont
PSA	average static pressure LSP	in.	cont
PT	total pressure	MB	cont
PTCH	Pitch angle	deg	cont
PTRM	pitch trim	deg	cont
PUSH	stick pusher		binary
RALT	radio altimeter	feet	cont
ROLL	roll angle	deg	cont
RUDD	rudder position	deg	cont
RUDP	rudder pedal position	cont	cont
SAT	static air temperature	deg	cont
SHKR	stick shaker		binary
SMKB	animal bay smoke		binary
SMOK	smoke warning		binary
SNAP	manual snapshot switch	binary	
SPLG	hydraulic system green	gallons	cont
SPLY	hydraulic system yellow	gallons	cont
SPL_1	roll spoiler left	deg	cont
SPLY	spoiler deploy yellow		binary
SPLG	spoiler deploy green		binary
SPL_2	roll spoiler right	deg	cont
TAI	total air temperature	degrees	cont
TAS	true airspeed	knots	cont
TCAS	traffic collision avoidance system		
TH	true heading		
TMAG	true/mag heading select	deg	cont
TMODE	thrust mode		
Variable	Description	Unit	Scale
TOCW	takeoff configuration warning		binary
TRK	track	degrees	cont
TRKM	track angle mag LSP	deg	cont
VAR_1107	synch word for subframe 1		
VAR_26701	synch word for subframe 2		
VAR_5107	synch word for subframe 3		
VAR_6670	synch word for subframe 4		
VHF1	radio 1		
VHF2	radio 2		
VHF3	radio 3		
VIB_1	engine vibration 1	in./sec	cont
VIB_2	engine vibration 2	in./sec	cont

VIB_3	engine vibration 3	in./sec	cont
VIB_3	engine vibration 4	in./sec	cont
VMODE	vertical engage modes		
VRTG	vertical acceleration	G	cont
VSPS	selected vertical speed	ft/min	cont
WAI_1	inner wing deice		binary
WAI_2	outer wing deice		binary
WOW	weight on wheels		binary
WSHR	windshear		binary

Table C7

Random Forest

Property	Function
Node ID	Was used to indicate the ID of the node in the diagram. For example, HPDMForest2, HPDMForest3, etc.
Imported Data	Used to display the following importing options: • Browse – Was used to browse the data. • Explore – Was used to sample and plot the data. • Properties – Was used to open the Table and Variables options.
Exported Data	Was used to open the same submenu options as the “input” menu: • Browse - Was used to browse the data. • Explore - Was used to sample and plot the data • Properties - Was used to open the Table and Variables options. Train properties are also available as follows:
Variables	Was used to open a variables table, which will then be used to view an “Explore” option to observe sampling information, values, or a distribution plot of the variable of interest.
Maximum Number of Trees	Was used to select the number of DT models in a forest. specifies the number of trees in the forest.
Seed	Was used to select a random number generator for sampling.
Type of Sample	Was used to select the number of observations in the training sample.
Proportion of Obs in Each Sample	When Type of Sample is selected to Proportions, Was used to select percentage of observations for each tree in the forest.
Number of Obs in Each Sample	Was used to select the number of observations in each tree when Count is used to select Type of Sample. The following splitting options are available in the HP Forest Node Train Properties menu:
Maximum Depth	Was used to select the maximum depth of a tree node.
Missing Values	Was used to select how missing values are treated. Options include: “Use in Search” Was use missing values as a separate value for splitting. “Distribute” Was used to omit missing values in the splitting of a particular node.

(continued)

Table C7

Random Forest (Continued)

Property	Function
Minimum Use in Search	Was used to select the minimum number of missing value observations Was used in a splitting rule.
Number of Variables to Consider in Split Search	Was used to select the number of input variables in a particular node.
Significance Level	Was used to select a p-value to test association of an input variable with the target variable, UARM.
Max Categories in Split Search	Was used to select the maximum number of categories in a candidate variable in an association test.
Minimum Category Size	Was used to select the minimum number of input variables that a given nominal category will use in a split search.
Exhaustive	Used to select the maximum number of splits when a target variable contains more than two categories.
Method for Leaf Size	Was used to select the value of the leaf size. Options include "default", "count", and "proportion."
Smallest Percentage of Obs in Node	Was used to select the smallest number of training observations a new branch can have.
Smallest Number of Obs in Node	Was used to select the smallest number of training observations a new branch can have.
Split Size	Was used to select the number of observations within a node prior to splitting. The HP Forest node contains the following score properties:
Variable Selection	Was used to select the automatic selection function of the node.
Variable Importance Method	Was used to select which option Was used regarding variable importance among the following options: Loss Reduction (default) or Random Branch Assignments.
Number of Variables to Consider	Was used when "Random Branch Assignments" in the "Number of Variables to Consider" setting is used.
Cutoff Fraction	Was used with when the "Variable Importance Method" property is set to "Random Branch Assignments" to select the threshold setting for largest random branch assignment measure.

(continued)

Table C7

Random Forest (Continued)

Property	Function
"Results" Options	
Settings	Was used to display HP Forest node properties.
Run Status	Was used to display run start time, duration, and status.
Variables	Was used to present a table of variables in the training data set.
Train Code	Was used to present the training code.
Notes	Was used to present data information.
SAS Results	
Log	Displays the HP Forest run log.
Output	Was used to display the HP Forest run output.
Flow Code	Was used to display the HP Forest node flow code.
Scoring	
SAS Code	Was used to present the HP Forest SAS code.
Assessment	
Fit Statistics	Was used to present model fit statistics.
Classification Chart	Was used to present classification results for the categorical target variable, UARM.
Score Rankings Overlay	Was used to present statistics for groups of observations, sorted by the posterior probabilities. Was used to present both training and validation statistics.
Score Distribution	Was used to plot nonevents and proportions of events and the model score of a bin.
Model	Was used to present tables and graphs pertaining to the variables as follows:
Baseline Fit Statistics	Was used to present information about the baseline model.
Iteration History	Was used to present information about the iterative modeling process.
Iteration Plot	Was used to present goodness-of-fit statistics plotted against the number of trees in the forest.

(continued)

Table C7

Random Forest (Continued)

Property	Function
Leaf Plot	Was used to present the total number of leaves in the forest plotted against the number of trees in the forest. See Figure 20.
Leaf Statistics	Was used to leave distribution in the form of a histogram.
Variable Importance	Was used to present information pertaining to each variable's worth.
Table	Was used to present data used to construct charts.
Plot	Was used to modify or create charts.