Business Inferences and Risk Modeling with Machine Learning; The case of Aviation Incidents

Burak Cankaya
Embry-Riddle Aeronautical University
Worldwide, College of Business
bcankaya@erau.edu

Kazim Topuz
The University of Tulsa, Collins
College of Business
kat0141@utulsa.edu

Aaron Glassman
Embry-Riddle Aeronautical University
Worldwide, College of Business
glassf10@erau.edu

Abstract

Machine learning becomes truly valuable only when decision-makers begin to depend on it to optimize decisions. Instilling trust in machine learning is critical for businesses in their efforts to interpret and get insights into data, and to make their analytical choices accessible and subject to accountability. In the field of aviation, the innovative application of machine learning and analytics can facilitate an understanding of the risk of accidents and other incidents. These occur infrequently, generally in an irregular, unpredictable manner, and cause significant disruptions, and hence, they are classified as "high-impact, low-probability" (HILP) events. Aviation incident reports are inspected by experts, but it is also important to have a comprehensive overview of incidents and their holistic effects. This study provides an interpretable machine-learning framework for predicting aircraft damage. In addition, it describes patterns of flight specifications detected through the use of a simulation tool and illuminates the underlying reasons for specific aviation accidents. As a result, we can predict the aircraft damage with 85% accuracy and 84% in-class accuracy. Most important, we simulate a combination of possible flight-type, aircraft-type, and pilot-expertise combinations to arrive at insights, and we recommend actions that can be taken by aviation stakeholders, such as airport managers, airlines, flight training companies, and aviation policy makers. In short, we combine predictive results with simulations to interpret findings and prescribe actions.

Keywords: Business Analytics, Machine Learning, Decision Support Systems, Big Data, Aviation Risk Modeling, Business Inferences with Machine Learning

1. Introduction

It is difficult to learn from aviation incidents since it is difficult to find the same combination of factors – aircraft, flight type, and pilot capabilities – in various incidents. Every incident has distinct characteristics, which makes it challenging to generalize any lessons learned to a broader scope of business decisions. The solution to this business problem calls for a simulation decision support system (DSS) to create scenarios and predict the likelihood of damage. A DSS could provide business insights by evaluating the inferences between variables and providing a what-if analysis to interpret incidents. Aviation is one of the earliest specialized fields because of its high-risk nature. Every incident is documented with detailed reports, which include records from devices, specifications of the nature of an event (such as the weather conditions and airfield data), and expert notes and judgments. We found that the severity of aircraft damage and the chains of events that cause incidents can be predicted; consequently, this study offers a way to apprehend the patterns that cause aviation incidents, by classifying them by flight type. In addition, using a simulation tool, what-if scenarios were analyzed, and actionable insights are offered to aviation decision-makers. The findings of this can help stakeholders by providing business insights mined from aviation incident reports.

For the data, we used incident reports from the Federal Aviation Authority's (FAA's) Aviation Safety Information Analysis and Sharing (ASIAS) Accident and Incident Data System (AIDS) from 2000 to 2020, which include all aviation incidents that happened in the U.S. during that period. Out of 25,527 records, 21,065 of the aircraft damage instances can be categorized as "minor," 3,264 records as "none," (no accidents or incidents), and 1,199 as "substantial/destroyed." We utilized CRISP-DM methodology to prepare the data and conduct the analysis. The process required deep expertise in

URI: https://hdl.handle.net/10125/102783
978-0-9981331-6-4

aviation incidents and data analytics. The initial data could be considered dirty, with many variables in free-text format, missing data, typos, names' changing over time, and mergers and acquisitions that require subject matter expertise (SME) in aviation and analytics. A similar degree of expertise was required to create variables to amplify the model's predictive power; these were Busy Airport, Helicopter, Weekday, Day of Week, and Month. The data preparation included collecting and scrapping data, cleaning and filtering it, treating missing data, conducting a variable selection of the data, and splitting it for training and testing.

After the most time-consuming data understanding and preparation were concluded, we start modelling the data. The original dataset consisted of 28 variables; these consisted of general event location information, aircraft information, environment information, PIC information, and event remarks. (Event remarks are free text information and were not used in this study.) Since the scope of the data spanned twenty years, the data had changed over time, and not all the variables were informed by high-quality data.

The solution to the business prediction problem called for a simulation decision support system (DSS) to create scenarios and predict the likelihood of aircraft damage. The DSS provided business insights by evaluating the relationships among variables and enabling a what-if analysis to assess the likelihood of incidents.

The DSS predicts aircraft damage and provides a better understanding of the patterns that cause aviation incidents by classifying them by flight type. It can also help various stakeholders with business insights mined from aviation incident reports.

Table 1: Data Overview

Variable	Explanation	Data Type	Descriptive Statistics*	**Prc Mss
DOW	Day Week	N	Sunday(28.25), Saturday(22.46)	0
Weekday	Weekday/ Weekend	N	Weekday(68.2), Weekend(31.7)	0
Month	Month of Year	N	July(10.41), August(9.53)	0
Day	Day of Month	N	29th(3.47)-17th(3.45)	0
State	State the Incident Happened	B	CA(11.14), FL(10.78)	3.29
Busy	If the Airport is one of the Busy Airports	N	Yes (JFK, MCO)(6.0), No(94.0)	0
Aircraft Damage	Type of Damage	N	Minor (82.1),None (13.3), Substantial_Destroyed (4.61)	0
Flight_Phase	Flight Phase	N	Landing(26.86), Touchdown (17.59)	0
Helicopter	If the Aircraft is a Helicopter	B	Yes (2.81), No (97.19)	0.9
Flight_Type	Flight Phase	N	Personal(61.69, Instruction(14.34)	23.3
MajorCarrier	If the fight is managed by a major carrier	B	Yes (8.33), No (91.67)	4.14
FlightConductCode	Flight Conduct Code	N	General(78.99), AirCarrier(8.36)	0.39
Engine_Make	Maker of Engine	N	Lyoming(39.98), Continental(29.31)	40.98
Aircraft_Engine_Model	Engine Model	N	IOSeries(65.59), PW6(5.43)	41.78
Nbr_of_Engines	Number of Engines	N	1(70.14), 2(27.67)	26.33
PIC_Certificate_Type	PIC Certificate Type	N	Private Pilot(38.81), Commercial (16.92)	14.77
PIC_Category	PIC Category Hours	N	3955.3 (6134.7)	16.38
PIC_Model	PIC Model Hours	N	738.0 (1657.2)	18.95
90	Flight hours in last 90 days	N	54.0 (72.5)	18.95

* Descriptive Statistics: Binary- % of each category; Nominal- % of most common two categories; Numeric- mean (standard deviation) ** Percent Missing

A machine-learning (ML) algorithm was chosen for its fast performance, short run-time, and interpretability. The ML algorithms were compared and evaluated on a confusion matrix for their mean receiver operating characteristic curve (ROC), summarized in terms of area under the curve (AUC), recall, reliability, and in-class accuracy. The summary of the weights of the variables and their relative importance referred to the flight-related findings. The results of the simulation model belong to the Multinomial Logistic Regression algorithm because of its interpretability. The predictions and their deployment were tested randomly.

Visualizations to depict the cumulative story of the flight types were created. The simulations, incorporating flight types, flight times, pilot expertise, and aircraft-type combinations, were storified with a decision-support system. The business insights derived from complex relations of multiple variables and categories of variables for flight type were documented and reported to the relevant aviation stakeholders. The findings of the study are that scheduled, rather than a commuter, air taxis are at significantly less risk of aircraft damage. The highest likelihood of destroyed aircraft occurs with flights involving illegal or stolen drugs. Cargo flights have a greater likelihood of being destroyed than normal flights (excluding illegal flights). Aerial applicators and executive flights have a lower risk of aircraft damage during cruise phases.

These claims are explained in more depth in the discussion section. In sum, we have shown that aircraft damage can be accurately predicted before a plane leaves the ground. In addition, the relation between the variables and specific types of categorical variables can reveal patterns that might lead to an incident. The evidence that points to patterns of liability for business flights, training flights, scheduled flights, and airline flights were reported to the relevant aviation authorities. These results could re-shape the aviation landscape. Stakeholders can utilize the reported findings, and other scenarios can be simulated using the DSS created for this study.

In conclusion, the study contributes significantly to the aviation business literature by introducing a DSS that can predict aircraft damage by flight type using machine-learning algorithms. In addition, and more importantly, the DSS which is a simulation tool throws light on the big picture of aviation incidents, identifies the interactions between complex factors and the risks associated, and interpret them as managerial strategies. It furnishes aviation authorities with insights for strategic policy decisions on aviation and aerospace management procedures. It should be highlighted that due to the summarizing approach in this study the predominant aim of this study is not only to predict the aircraft damage severity category, but also more of mapping the complex conditional dependencies amongst aircraft, flight, and pilot related variables and bring interpretability to the fault structure that create broad term of aviation incidents.

2. Literature Review

Comprehensible data for making inferences and business decisions were important even before the Information Age (Teng et al., 1994). After that, organizations began to gather data about their areas of interest area to make smarter decisions. In attempts to summarize and understand what had happened in the past, business analysts started applying descriptive statistics (Hardoon & Shmueli, 2013).The next step was to make predictions based on historical data. Traditional statistical methods were employed to make basic predictions, and the first uses were for making scientific predictions (Waljee et al., 2014). As the amount of data stored in databases began to expand, the computational power of distributed storing and computing systems enabled machine-learning algorithms to be used for scientific tasks (Cankaya et al., 2021).

Machine-learning applications effectively perform predictive tasks, and they are widely used for many business and scientific applications (Delen et al., 2020). Machine learning has produced promising predictions in a variety of business contexts, ranging from healthcare (Almeda et al., 2019) and stochasticity problems in transportation scheduling (Cankaya et al., 2019) to stock market fluctuations (Shen et al., 2012) and the analysis of incident reports (Topuz & Delen, 2021)

One problem with the primitive machine-learning application algorithms is that they are black-box algorithms; namely, you can make accurate predictions but cannot understand how the system works, so you cannot explain the root processes for making the predictions (Papernot et al., 2017). Recently, models have been developed that merge predictions with relational inferences and are being used to understand business inferences hidden in the data (Topuz et al., 2018).

Combining subject matter expertise (SME) and machine learning to simulate various business scenarios helps to provide an understanding of the details of cases that are expensive to replicate in a controllable environment (Topuz et al., 2018). In addition to cost scenarios, many other business and scientific applications are possible. In situations where events can be risky for people, computational simulations of possible scenarios are indispensable (Čokorilo, 2013)

Simulations have long been a part of aviation culture, and pilot training programs have been using simulation tools for a long time, which enable them to test pilots-in-training and assess their reactions to difficult scenarios (Hight et al., 2022). Similarly, simulating various scenarios and testing for the likelihood of aircraft damage can enable managers to make inferences and business decisions based on incident and accident reports (Madeira et al., 2021). These aviation reports are released individually; each report is unique and is evaluated by a current for a future pilot, as a lesson learned. however, it is also essential for aviation authorities to combine numbers of similar events and simulate all the factors that can cause an event (Skorupski, 2016). Predictions of aviation accidents have been made in the literature but in summary they aimed on either a specific accident scenario combination not a summarizing study or only focused on prediction not expansibility (Srinivasan et al., 2019; Mehta et al., 2021).

Studies even combined text-based reports with tabular incident datasets. That type of meta-studies are used to investigate the deeper reasoning of specific types of incidents. These meta-studies are not for testing conditional dependencies and creating a visual network illustration of the main fault lines that create incidents (Sarkar et al., 2020; Srinivasan et al., 2019).

Recent studies have focused on various accident scenario combinations such as; accidents involving fatalities and serious injuries for Part 91 manufactured aircrafts (Burnett &Si., 2017). It was also relevant to create a unique risk metric for aviation incidents. Still, as expected, the variation reasoning of different aviation accidents focused on airport surface environment with incidents and close calls (Bati & Withington, 2019). Similarly, others only focused on predicting fatalities in FAA incidents which is only possible by limiting the flight type to high-risk combinations but did not aim at the highly frequent Minor damage category events (Lukáčová et al., 2014). In contrast, others aim to quantify and categorize risk with an ensemble model and designate event outcomes for risk levels (Zhang & Mahadevan, 2019)

Nevertheless, prediction simulations are important, so that aviation-related businesses can anticipate direct and indirect financial and vital risks (Tulechki, 2015). Insurance agencies, government regulators, flight training schools, airport managers, airlines, and other aviation business actors desperately need a simulation tool to visualize the risks of an aviation incident or accident and its possible damages (Oster et al., 2013) In this study, we introduce an incident and accident simulation and prediction tool that all aviation stakeholders can use to test the likelihood of damage from an event. The event might happen in the exact combination, or the event never happened on the tested combination, but the simulator is calculating a likelihood by summarizing similar events.

3. Methodology and Data

The analysis of the aviation data required solid knowledge and experience in aviation operations and a reliable, explainable analytics methodology. With the increasing popularity of data analytics in industry and academia, projects using machine learning to make business predictions and provide insights have increased rapidly. However, many of these studies lack a set of assumptions and procedures for validating and applying their methods. Their generalizations may overstate the findings, and their robustness might be questioned.

This study employs CRISP-DM, a well-known data analytics procedure that incorporates an information loop equipped with a feedback mechanism. It was essential throughout the data preparation and comprehension phases to have a thorough grasp of the aviation business and the FAA data to generate new variables and validate the data. The algorithm selection process also required an understanding of the problem and the business context. Particular algorithms act better on related types of data, and since this study valued interpretability as much as predictability, knowledge of the aviation business was needed in the modeling phase to interpret the model. In addition, predictions and simulation results were evaluated with feedback mechanisms to create related variable combinations and business inferences. Deployment of the model means both deploying the predictions and deploying the simulation. Furthermore, simulation results need to be evaluated to arrive at findings useful for business.

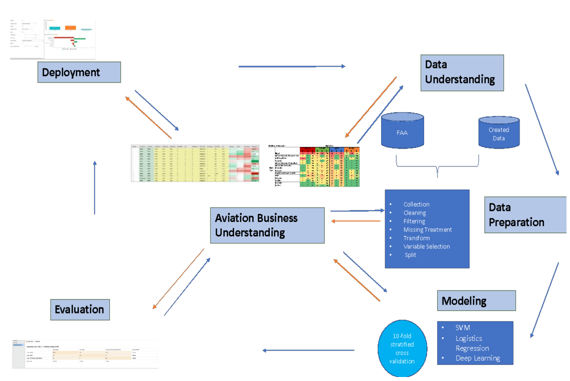

Figure 1: CRISP-DM Methodology Visualization

As it is visualized in Figure 1, the six steps of the CRISP-DM methodology as a loop of information flow with additional feedback mechanisms to aviation business understanding on each step due to the nature of the deepness of the aviation data. Aviation Business Understanding is essential in the data understanding and preparation phase to understand FAA data, create new variables, and validate the data. The algorithm selection process is also requiring aviation business understanding. Different algorithms act better on related data types, and since this study values interpretability as much as predictability, aviation business understanding is involved in the modelling phase. In addition, at the evaluation step, the predictions and simulation results are evaluated with feedback mechanisms to create related variable combinations and business inferences. The deployment of the model means deploying the prediction and deploying the simulation. Simulation results need to be evaluated to create the business findings.

3.1 Methods

For this study, we utilized some of the most common machine-learning techniques. Supervised machine-learning methods were chosen based on the problem, data, and labelling settings. In this case, the data came labelled by FAA experts to fit the aircraft damage categories, so we chose one of the supervised machine-learning methods because the data is already labelled. We applied several algorithms to the dataset, compared their performance, and chose the most interpretable model with acceptable performance. The supervised machine-learning methods we used were multinomial logistic regression (LR), support vector machine (SVM), and deep learning.

3.1.1 Multinomial Logistic Regression

Multinomial Logistic regression is a machine-learning algorithm that identifies the pattern differences between two categories of target variables with a logistic function. When the category of the target variable has more than two discrete output categories, multinomial logistic regression can identify predictive patterns where the relationship between input and output variables is not linear. Multinomial logistic regression uses maximum likelihood estimation as a conditional probability for the prediction. When the probability is over 0.33, the prediction falls into the True category; when it does not, it falls into the False category (Belyadi &Haghighat, 2021)

3.1.2 Support Vector Machine

A support vector machine (SVM) is a supervised machine-learning algorithm that can produce promising prediction results for both regression and classification tasks by breaking the search space into hyperplanes, named support vectors. Data instances that have a minimum distance to the hyperplane surface are known as hyperplanes. The shapes of the support vectors are defined by the kernels, and there are different types of kernels, such as linear kernel functions, polynomial kernel functions, and sigmoid kernel functions. One factor that distinguishes SVMs are the margins between classes, where it is possible to linearly separate data. Soft margins are complex so that it is impossible to linearly separate classes of data (Rani et al., 2022)

3.1.3 Deep Learning-H2O

Deep-learning algorithms are multi-layer, artificial neural network algorithms that can find more complex patterns than ANN in multiple layers of pattern recognition. The information from the data enters a hidden layer with related weights. The information is processed in the neuron and transferred to the next hidden layer, depending on the activation function. Information flowing from input to output is named *feed-forward*. The information flows back from the output to the input in the next round, and the pattern of the prediction gets written again. This rewriting process is called *back-propagation* (Cullell-Dalmau et al., 2020)

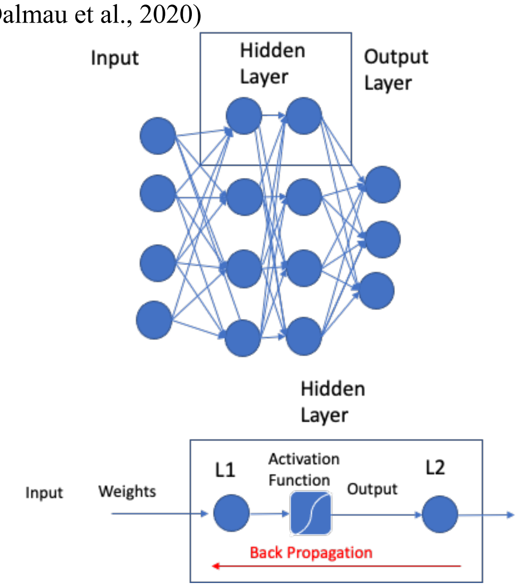

Figure 2: Representation of a simple Deep Learning H2O

An H2O algorithm is a type of deep-learning algorithm with single layers and a feed-forward ANN algorithm that uses a stochastic, gradient-descent, back-propagation algorithm (Candel et al., 2016)

3.2 Model Evaluation

In our case, the target variables are of three types: None, Minor, and Destroyed/Substantial. The evaluation of the models will be demonstrated with the confusion matrix below.

In Table 1, the Minor class is Class A, The None class is Class B, and the Destroyed/Substantial class is Class C.

3.2.1 Confusion Matrix

Table 1: Confusion Matrix

		Predictions		
		Class A	Class B	Class C
Actual Values	Class A	1	2	3
	Class B	4	5	6
	Class C	7	8	9

A confusion matrix with three categories is shown in Table 1. The predicted values of the instances that each algorithm predicts versus the actual values of the target variables layed-out in the matrix; by the ratios of true positives (TPs), true negatives (TNs), false positives (FPs), and false negatives (FNs), we define the performance indicators.

TP = Cell 1
FN = Cell 2+Cell 3
FP = Cell 4+ Cell 7
TN= Cell 5+ Cell 6+ Cell 8+Cell 9

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

$$Recall = \frac{TP}{TP+FN} \quad (3)$$

High accuracy (1), precision (2), and recall (3) values indicate when there is a greater likelihood of a prediction's being correct. However, due to the nature of the aviation data set, the distribution of the categories in the data was not balanced; the Minor category equaled 82%, the None category 12%, and the Destroyed/Substantial category only 4%. Because of this imbalance, the training and testing of the prediction had to be repeated to confirm the accuracy of the predictions.

3.2.2 Cross-Validation

Formula 4 represents the cross-validation process, and k represents the number of times the experiment is repeated. The data were broken into k parts, then training and testing ratios were defined. We chose 90/10 % as the training and testing ratio. Each piece was chosen for testing at least once, and the other remaining 90% of the data was retained as the training set. Each repetition result was marked as a fold formulation (4). We considered 10 folds to be sufficient for testing the randomness of the data. The accuracy, precision, and recall values of each fold were compared to the initial non-cross-validated data to check for randomness. Then, for simulation purposes, the fold closest to the mean accuracy value was used in the simulation to arrive at the business findings.

$$Performance = \frac{1}{k}\left(\sum_{i=1}^{k} P_i\right) \quad (4)$$

5. Results

The model uses deep learning, multinomial logistic regression, and SVM models. The predictive power of the models was tested with the confusion matrix, and the results are summarized in the model performance results in Table 2. In summary, the performance results for all the models were sufficient and very close to each other. In this study, interpretability, explained in Figures 2 and 3, showed the superiority of multinomial logistic regression because its accuracy, recall, and precision were high. Also, the variables chosen for the multinomial logistic regression were important for building business relationships and yielding interpretable results.

Table 2: Model Performance Results

		Method Accuracy	
	Deep Learning	Logistic Regression	SVM
Accuracy	84.2	**84.6**	82.9

Class Recall	99.2	**98.4**	99.6
AUC	77.6	**77.9**	79.3
“Minor” Damage In Class Precision	84.4	**85.4**	83.12

Due to the nature of the aviation incidents, the Destroyed/Substantial cases are vital to claim predictability. This study also checks False Negatives (i.e., allowing an airplane destined to crash to take off-Destroyed/Substantial cases) to compare the performance of different models, including Multinomial Logistic Regression, SVM, Deep Learning-H2O, and Decision Tree Based models. Due to the nature of the problem, these Destroyed/Substantial cases have low- frequency in the data, and none of the ML models have successfully reduced the FN ratio even though 5,10, 20, 50, and 100 x weight balancing have been tested on the system. We then balanced the data using Synthetic Minority Oversampling Technique (SMOTE) and compared the results. Again, our overreaching goal is to use the simulation tool for risk inferences and make business decisions with more than a predictive decision-making tool on all of the flights in the national aerospace.

The results direct us to use the simulation tool for risk inferences and make business decisions more than a predictive decision-making tool on all of the flights in the national aerospace. To increase predictability, certain types of flights should be isolated, and a more predictable outcome can be achieved. The scope of this study is to summarize all flights that prevent us from analyzing isolated cases.

5.1 Sensitivity Analysis

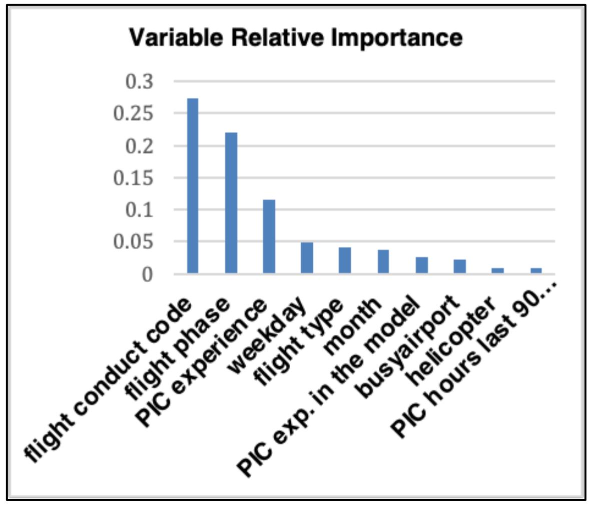

Figure. 2: Variable Importance-Weight by Correlation

The variable importance chart visualizes the relative importance of the variables. It shows how important each variable is relative to the others. The results are interesting, with flight type specifications such as the flight conduct code (0.274) and flight phase (0.221) being more important than pilot experience (0.115). The chart also shows the importance of creating variables with an SME, who in this case created variables such as weekday (0.049), month (0.037), busyairport (0.023), and helicopter (0.009). The other variables chosen were flight type (0.042), PIC experience in the model (0.027), and PIC hours in the last 90 days (0.009).

We also experimented with automatically computed variables as combinations of other variables; however, since interpretability was essential for this model and since the results were not significantly improved by the auto-created variables, we stuck with the original variables.

6. Discussions

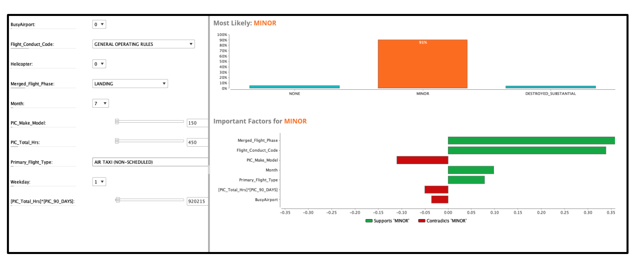

Figure 3. The Simulation Tool

The simulation tool is one of the main contributions of this study. All the methods – deep-learning-H2O, SVM, and multinomial logistic regression – gave acceptable results in terms of predictive power, but the multinomial logistic regression model gave the best interpretable simulation results. The simulator tool is a graphical user interface (GUI) that we can change the combination of both categorical and continuous input variables and visualize the prediction likelihood for a specific category. This likelihood diagram can be seen as the most likely category on the top right. We can also see the contributing and contradicting factors with their relative importance in the chosen target variable category on the bottom right.

By the support of this simulation tool, we can understand the fault network that creates incidents for different types of flights. It used all the main flight-related variables evaluated in Figure 2. By changing the categories for flight type, PIC experience, and flight phase combinations, the simulator produced a summary of likelihood percentages and substantial supporting and contradicting variables, ranking the variables relative to the selected target variable class.

In addition to the summary values, the simulator also creates predictive results for missing data. Even when a combination of flights did not happen, it can create a novel case by referring to relative results and can create a simulation of likely outcomes. Moreover, the simulator can be used as a prescriptive tool to create optimized results for maximum or minimum possible variable combinations by keeping some variables constant. For example, we can keep flight type as scheduled air carrier and flight phase as landing, and it will create the contributing and contradicting variable combinations and give us the likelihood of event ending as minor incident.

The screenshot given in Figure 3 demonstrates how simulation tools can be used. The image comes from an instance of a DSS tool predicting and explaining an event: a non-scheduled air-taxi flight in July, where the PIC had 150 hours of experience with the model and 450 hours of total flight expertise. The tool gives a 91% chance that the event will result in minor damage, a 5% chance that there will be no damage, and 4% chance that the aircraft will be destroyed. In addition, the right corner of the figure explains which factors are important and whether they support or contradict a minor damage scenario. The flight phase and flight-conduct code significantly support the chance of a minor incident, while the month and flight type offer less support. On the other hand, the PICs' experience with the model, the PIC's total experience, and the amount of airport traffic suggest that the case might be more than minor.

Table 3: Flight Type and Flight Phase Event Likelihood,

Likelihood of Event		Flight Phase											
		Ground			TakeOff			Cruise			Landing		
		N	D	M	N	D	M	N	D	M	N	D	M
Flight Type	Illegal	6	19	75	20	10	70	40	9	**51**	7	10	83
	Air Taxi-Scheduled-Not Commuter	5	5	90	14	4	82	34	4	**62**	6	4	90
	Aerial Applicator	6	5	89	9	5	86	23	4	**73**	5	5	90
	Executive	5	5	90	9	4	87	22	5	**73**	5	4	91
	Air Taxi-Scheduled-5Trips/Week	5	5	90	8	5	87	18	5	**77**	5	4	91
	Air Taxi-Non Scheduled	4	6	90	7	6	87	19	4	**77**	5	4	91
	Instruction	4	6	90	8	5	87	19	4	**77**	5	4	91
	Personal	4	6	90	7	6	87	18	5	**77**	5	5	90
	Supplemental or Commercial	5	7	88	8	5	87	18	5	**77**	5	5	90
	Cargo	4	10	86	7	7	86	16	6	**78**	5	6	89
	Scheduled	5	6	89	7	6	87	17	5	**78**	5	5	90
	For Hire	5	5	90	7	5	88	16	4	**80**	5	4	91
	Industrial	4	5	91	7	5	88	16	4	**80**	5	4	91
	Business	4	5	91	6	4	90	15	4	**81**	5	4	91

Some of the leading business inferences for different flight phases and flight types can be seen in Table 3, which evaluates cruise, take-off, landing, and ground flight scenarios. Even though the predictive metrics for each category are not high, the general patterns lead us to relevant risk inferences. The greatest likelihood of a predicted destroyed aircraft is found for illegal flights, such as flights carrying drugs or stolen goods. Among them, 19% of the incidents in the ground phase end up classified as Destroyed and 75% as Minor Damage. Illegal flights are likely to have 40% no damage and about 51% are likely to have minor damage in the cruise phase. Similarly, events in the Take-Off phase show a 20% likelihood of no damage to the aircraft and a 70% likelihood of minor damage. In summary, if an illegal flight has an incident on the ground, there is a 19% likelihood it will destroy/substantial the aircraft. If the aircraft manages to take off, cruise, or land, the chances of being substantial/destroyed destruction are about 10%. It can be inferred from this comparison that a significant part of the illegal flights end up at a chase on the ground and have major accidents that total the aircraft. One other significant finding is that cargo flights have a greater likelihood of being destroyed than normal flights (excluding illegal flights), with a 10% likelihood on the ground, 7% at take-off, and 6% when cruising or landing.

The main reason cargo flights are more likely to end up destroyed is that the weight load distribution and set-up in the aircraft need expertise, and many cargo flights fail to engineer the weight distribution and damage the aircraft. These findings lead the aircraft insurance companies to check airport security and illegal flight risk to define higher premiums and prices. Even though most cargo flights do not carry passengers and avoid the risk of potential damage to more humans, they have a higher risk of damaging the aircraft, and their insurance prices are high to be protected from these incidents.

Air taxi flights are categorized as Scheduled Not-Commuter, Scheduled with at least five flights per Week, and Non-Scheduled On-Demand Air Taxi Flights. Scheduled Not-Commuter flights have a significantly greater likelihood of avoiding damage in the take-off phase (14%) and in the cruise phase (34%). In comparison of the phase, take-off presents an 82% likelihood of minor damage, and the cruise phase a 62% likelihood. These numbers are some of the highest of all the flight types. Scheduled and non-commuter air taxis are at a significantly lower risk of aircraft damage, injuries to people, or non-flight-related objects in the flight. None damage to aircraft category events means there is an event, but it does not have damage to the aircraft. So, for these None category events for Scheduled Not-Commuter flights, the crew and passengers have a high risk of being hurt than on other flights at takeoff and cruise. The frequency of these flights are less than five times a week. The reason for these events is that the cabin crew typically are not as professional as a commuter. In cases like turbulence or other disruptions at takeoff and cruise, the people in the aircraft are more likely to be hurt. Insurance agencies can raise the premiums for body injury for passengers and non-flight-related objects in the aircraft for these types of low-frequency flights in comparison to the other scheduled flights. We recommend paying extra attention to the preparedness of Scheduled Not-Commuter flights, and governing agencies can advance the cabin crew requirements for these flights.

Furthermore, aerial applicators (23%) and executive flights (22%) have similar behavior in running fewer risks of damage to the aircraft in the cruise phase. Still, they have None damage to aircraft category an incident. Aerial applicators are mostly agricultural flights, and their numbers for aircraft damage at the cruise phase are very similar to those of industrial flights (16%) and personal flights (18%), albeit there is less likelihood of damage to the aircraft but more None category events. That also means these Aerial applications and agricultural flights have a higher risk of having an incident that injures the pilot and damages the non-flight-related accessories. These Aerial application flights can be specifically insured for these non-flight-related accessories.

Also, summarizing all flight types, there is a significantly more likelihood of None damage to the aircraft category event if an event occurs during the cruise phase (20% on average) than in other phases (6.2% on average). These events are primarily due to turbulence, or anything happening after the take-off will likely be recorded as the cruise phase. For any passenger flights, these incidents can be reduced by improving the cabin crew resources and educating them about the likelihood of these events.

6.1 Conclusion Limitation and Future Research

This study clearly shows that machine learning can effectively predict aircraft damage, interpret, and explain patterns in aircraft-related data, and provide business insights to shape future industry policy. Stakeholders can use the DSS developed here to provide a summary of related events and estimate the likelihood of an aircraft incident involving minor, no, and serious damage. They can also simulate important supporting and contradicting categories of variables. The methodologies used here are effective with different combinations of datasets, and the results visualized in Table 3 show an ordinance only in the context of the study. In addition, in this study, making business inferences is prioritized over making predictions.

Due to lack of space, the discussion in this paper is limited to analyses based on different flight types. We had to modify some assumptions, such as adjusting for 150-hours PIC flight experience in the model and 450-hours of total PIC flight experience; however, the number of actual flight hours may differ significantly for instructional and executive flights. To provide solutions for each aviation stakeholder, the DSS assumptions should be set up to meet the usage requirements of a specific business.

Another limitation of the study is rooted in the complexity of the data. New variables from the AIDS dataset and from various other datasets in free-text formats, such as those containing weather data, NTSB accident data, accident remarks, and reports can be combined, and a natural language processing (NLP) analysis can be done on them. Other sub-categories of minor accidents can be added so that financial analyses can be done since minor accidents occur way more frequently, at a cost, but rarely involve injuries or major aircraft repairs.

Future researchers can consider gathering datasets from other sources and making predictions based on other variables. In addition, factors such as injuries, weather conditions, and pilot expertise levels can be considered to enrich the business scenarios and make predictions. Furthermore, the DSS can be designed to select other simulation scenarios and offer solutions. Instructional flights, personal flights, business flights, air-taxi flights, agricultural flights, and airline flights can be considered, along with pilot expertise, airport specifications, flight phases, and timelines to create unique scenarios for a variety of flight-type combinations.

References

Almeda, García-Alonso, C. R., Salinas-Pérez, J. A., Gutiérrez-Colosía, M. R., & Salvador-Carulla, L. (2019). Causal Modelling for Supporting Planning and Management of Mental Health Services and Systems: A Systematic Review. International Journal of Environmental Research and Public Health, 16(3), 332

Bati, F., & Withington, L. (2019, September). Application of Machine Learning for Aviation Safety Risk Metric. In *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)* (pp. 1-9). IEEE.

Belyadi, H., & Haghighat, A. (2021). machine learning Guide for Oil and Gas Using Python: A Step-by-Step Breakdown with Data, Algorithms, Codes, and Applications. Gulf Professional Publishing.

Burnett, R. A., & Si, D. (2017, May). Prediction of injuries and fatalities in aviation accidents through machine learning. In *Proceedings of the International Conference on Compute and Data Analysis* (pp. 60-68).

Cankaya, B., Wari, E., & Tokgoz, B. E. (2019). A chemical tanker scheduling problem: Port of Houston case study. International Journal of Planning and Scheduling, 3(1), 47-67.

Cankaya, B., Eren Tokgoz, B., Dag, A. and Santosh, K.C. (2021), "Development of a machine-learning-based decision support mechanism for predicting chemical tanker cleaning activity", Journal of Modelling in Management, Vol. 16 No. 4, pp. 1138-1165.

Candel, A., Parmar, V., LeDell, E., & Arora, A. (2016). Deep learning with H2O. H2O. ai Inc, 1-21.

Cullell-Dalmau, M., Otero-Viñas, M., & Manzo, C. (2020). Research techniques made simple: deep learning for the classification of dermatological images. Journal of Investigative Dermatology, 140(3), 507-514.

Čokorilo, O. (2013). Human factor modeling for fasttime simulations in aviation. Aircraft Engineering and Aerospace Technology.

Delen, Topuz, K., & Eryarsoy, E. (2020). Development of a Bayesian Belief Network-based DSS for predicting and understanding freshmen student attrition. European Journal of Operational Research, 281(3), 575–587.

Hardoon, D. R., & Shmueli, G. (2013). Getting started with business analytics: insightful decision-making. CRC Press.

Hight, M. P., Fussell, S. G., Kurkchubasche, M. A., & Hummell, I. J. (2022). Effectiveness of Virtual Reality Simulations for Civilian, Ab Initio Pilot Training. Journal of Aviation/Aerospace Education & Research, 31(1), 1.

Lukáčová, A., Babič, F., & Paralič, J. (2014, January). Building the prediction model from the aviation incident data. In *2014 IEEE 12th International Symposium on Applied Machine Intelligence and Informatics (SAMI)* (pp. 365-369). IEEE.

Madeira, T., Melício, R., Valério, D., & Santos, L. (2021). machine learning and natural language processing for prediction of human factors in aviation incident reports. Aerospace, 8(2), 47.

Mehta, J., Vatsaraj, V., Shah, J., & Godbole, A. (2021, July). Airplane Crash Severity Prediction Using Machine Learning. In *2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT)* (pp. 1-6). IEEE.

Oster Jr, C. V., Strong, J. S., & Zorn, C. K. (2013). Analyzing aviation safety: Problems, challenges, opportunities. Research in transportation economics, 43(1), 148-164.

Papernot, N., McDaniel, P., Goodfellow, I., Jha, S., Celik, Z. B., & Swami, A. (2017, April). Practical black-box attacks against machine learning. In Proceedings of the 2017 ACM on Asia conference on computer and communications security (pp. 506-519).

Rani, A., Kumar, N., Kumar, J., & Sinha, N. K. (2022). machine learning for soil moisture assessment. In Deep Learning for Sustainable Agriculture (pp. 143-168). Academic Press.

Teng, J. T., Grover, V., & Fiedler, K. D. (1994). Business process reengineering: charting a strategic path for the information age. California Management Review, 36(3), 9-31.

Topuz, K., Zengul, F. D., Dag, A., Almehmi, A., & Yildirim, M. B. (2018). Predicting graft survival among kidney transplant recipients: A Bayesian decision support model. Decision Support Systems, 106, 97-109.

Topuz, K., & Delen, D. (2021). A probabilistic Bayesian inference model to investigate injury severity in automobile crashes. Decision Support Systems, 150, 113557.

Sarkar, S., Pramanik, A., Maiti, J., & Reniers, G. (2020). Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Safety science*, *125*, 104616.

Shen, S., Jiang, H., & Zhang, T. (2012). Stock market forecasting using machine learning algorithms. Department of Electrical Engineering, Stanford University, Stanford, CA, 1-5.

Skorupski, J. (2016). The simulation-fuzzy method of assessing the risk of air traffic accidents using the fuzzy risk matrix. Safety Science, 88, 76-87.

Srinivasan, P., Nagarajan, V., & Mahadevan, S. (2019). Mining and classifying aviation accident reports. In AIAA aviation 2019 forum (p. 2938).

Srinivasan, P., Nagarajan, V., & Mahadevan, S. (2019). Mining and classifying aviation accident reports. In *AIAA aviation 2019 forum* (p. 2938).

Tulechki, N. (2015). Natural language processing of incident and accident reports: application to risk management in

civil aviation (Doctoral dissertation, Université Toulouse le Mirail-Toulouse II).
Waljee, A. K., Higgins, P. D., & Singal, A. G. (2014). A primer on predictive models. Clinical and translational gastroenterology, 5(1), e44.
Zhang, X., & Mahadevan, S. (2019). Ensemble machine learning models for aviation incident risk prediction. *Decision Support Systems*, *116*, 48-63.

* The data used in the study have been gathered from FAA ASIAS system and used in a series of research including this study. It can be available for research upon request depending on the availability.