Data Mining Techniques to Predict Aircraft Damage Levels for Wildlife Strikes in United States

Researchers : Shlok Misra and Ila Toppo
Faculty Advisor: Dr. Flavio A. C. Mendonca

Abstract

Wildlife strikes pose a major economic and safety threat to aviation all around the world. Approximately eight percent of those strikes caused damage to aircraft. A primary method to understand the magnitude of this economic and safety hazard is through data collection and analyses. Data mining methods can be used to predict the likelihood of events and significant factors of contribution based on historical datasets. Researchers will collect and analyze data (2011-2020) from the National Wildlife Strike Database. The purpose of this study is twofold:

1. Create a prediction model to predict the probability of damage to an aircraft during a wild life strike.
2. 2. To identify the potential predictors of substantial and minor damaging wildlife strikes to aviation in the U.S.

Findings of the current study can help determine the nature and magnitude of this problem as well as provide the ground work for the development and implementation of integrated safety management and research efforts to improve aviation safety.

Introduction

Wildlife strike pose a major hazard to aviation. There were more than 227,000 incidents of wildlife strikes in the United States during 1990-2019. The rate of wildlife strikes per one million aircraft is currently estimated to 37.30 strikes/ million aircraft. A wildlife strike that attracted global attention was that of U.S Airways Flight 1549 in 2009 involving an Airbus-319. Those strikes resulted in total failure of both engines. The flight crew managed to ditch the aircraft in the Hudson River in New York. Wildlife strikes pose a major risk to aviation due to increasing population of certain high risk birds, such as the New World Vultures. This study will utilize data mining classification techniques to create prediction models to predict the probability of damage to an aircraft during a wildlife strike and detect significant factors that increase the probability of damage to an aircraft during a wild strike.

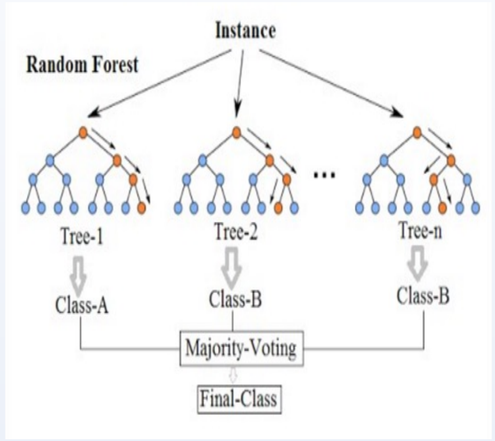
Method

Data Collection:
Data has been retrieved from the FAA Wildlife Strike Database from the year 2011- 2020. Considering the available data, the study will analyse the following eleven factors: altitude above ground level (AGL), bird size, number of birds struck, daylight conditions, aircraft mass, number of engines, phase of flight, precipitation, season, sky conditions, and airspeed. The goal of the study is to estimate the statistical relationship between these factors and the probability of aircraft damage in the event of a bird strike.

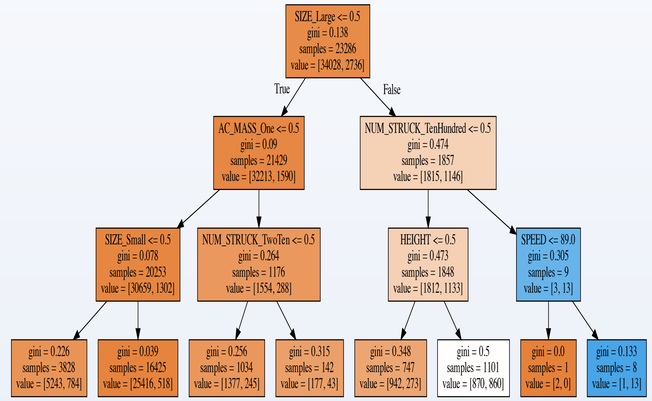
Process Data:
Since the data in the FAA Wildlife database are collected through a voluntary reporting system, some reports may contain unanswered questions resulting in missing values. Multiple Imputation algorithm of Approximate Bayesian Bootstrap methodology through Laplace library in R, will be used to replacing missing values with estimates.
A combination of Random Under-sampling and Synthetic Minority Oversampling Technique (SMOTE) through Python 3 script will be used to rectify the data imbalance for the minority class.

Results

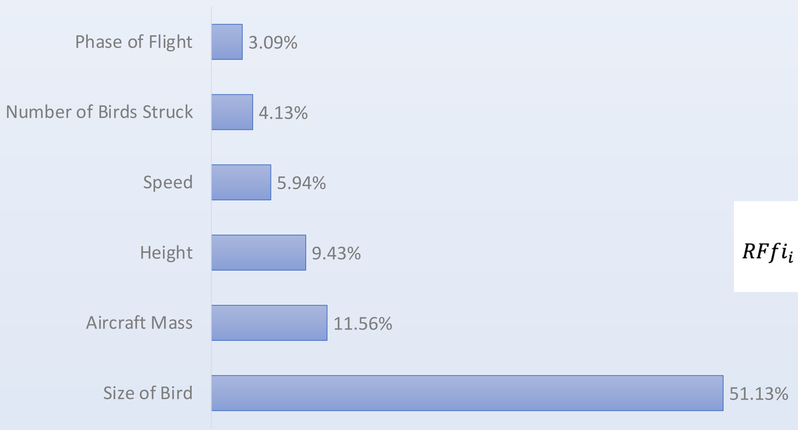
Random Forest Classifier Methodology



Decision Tree



Significant Variable Contribution



Gini Index Impurity Criterion

$$I_G = 1 - \sum_{j=1}^{c} p_j^2$$

Feature Importance Criterion

$$RFfi_i = \frac{\sum_j normfi_{ij}}{\sum_{j \in all\ features, k \in all\ trees} normfi_{jk}}$$

Expected Results

At the conclusion of the study, we expect to build a model with the following performance parameters:

Metric	Threshold
Accuracy	0.93
Precision	0.93
Recall	0.98
F1 Score	0.96

References

- Federal Aviation Administration. (2021). Wildlife strikes to civil aircraft in the United States in 1990-2019. National Wildlife Strike Database Serial Report Number 26. https://www.faa.gov/airports/airport_safety/wildlife/media/Wildlife-Strike-Report-1990-2019.pdf