BrainStation®

Eyes in the Sky: Predicting Aircraft Damage Caused by Bird Strikes Using Machine Learning

Capstone Project for the BrainStation's Data Science Diploma

Prepared by

Alan Fleck

alansfleck@gmail.com

September 27th, 2021

1. Background and Objectives

A bird strike is defined as a collision between a bird and an aircraft during flight, or on a take-off or landing roll. This term is often expanded to cover other wildlife strikes like bats or ground animals. According to the US Federal Aviation Administration (FAA), there have been about 227,000 wildlife strikes with civil aircraft in the US between 1990 and 2019, with an estimated direct cost as high as $500 million per year [1]. These costs, however, could be potentially higher when considering the resulted disruptions to airlines and airports, and especially, the risks that these strikes pose to people's safety. In this context, there is a need for the identification of the characteristics of these incidents and the factors associated with damaging strikes. Such analysis could be used to inform pilots and operators, as well as to implement more efficient warning systems.

Thus, this project aims to better understand the causes of wildlife strikes and predict damaging incidents based on the characteristics of the aircraft, birds, and flights.

2. The Data

The data for this project was collected and maintained by the FAA Wildlife Strike Database [2]. The file used for this project was downloaded from data.world [3]. The dataset has 25,558 unique observations of bird strikes and 26 columns which represent characteristics of the flights, birds, and aircrafts involved in the incidents. Our target variable is whether the strikes caused damage to the aircraft. The period covered by this dataset is between January 02, 2000 to December 31, 2011.

3. Data Preprocessing

The missing information for the column *"Aircraft: Number of engines?"*, (i.e., around 1% of total observations) was filled by looking at the column of the model of the aircraft. In addition, missing values for *"Origin State"* (i.e., around 1.8% of total observations) represented values for flights that originated outside the US and were filled with the string "International". The column *"Wildlife: Species"* contained information on both the name of the species and its size. Since the information of the size was already represented in the column *"Wildlife: Size"*, only the name of the species was retained in this column. In addition, the columns *"Year", "Month" and "Weekday"* were created from the column *"FlightDate"*, and the US states were grouped according to the US Standard Federal Regions in an attempt to better capture patterns in geographical features and bird migration routes.

The categorical variables were encoded due to the required numeric format for the machine learning algorithms used in this project. For feature selection, only features collected before the strike were retained in the final dataset because these are the variables in the causal path leading to a damaging strike. Finally, since our target variable is divided as 9.7% of damaging strikes and 90.3% of non-damaging strikes, class 0 (i.e. non-damaging strikes) was re-balanced to match the number of observations of class 1 (i.e. damaging strikes) before fitting the machine learning models.

4. Results

The analyses performed in this project were focused on EXPLAINING the features that are associated with damaging strikes, FORECASTING the monthly number of incidents using autoregressive (AR) models, and PREDICTING damaging strikes based on characteristics of the aircraft, birds, and flights using different algorithms for binary classification.

4.1 Explaining Damaging Strikes

We used the odds ratios calculated from the coefficients of logistic regression to determine the features positively and negatively associate with birds strikes. **Figure 1** shows that the factors positively related to damaging strikes (i.e. Odds ratio larger than 1) are the size and number of animals, descent and climb phases of the flight, as well as being in an altitude higher than 1000 feet. On the other hand, factors related to the size of the aircraft as well the landing roll phase of the flight are negatively associated with damaging strikes (i.e. Odds ratio smaller than 1) and could be considered "protective" factors. Interestingly, these were also the important features identified in EDA when looking at the damage rates.

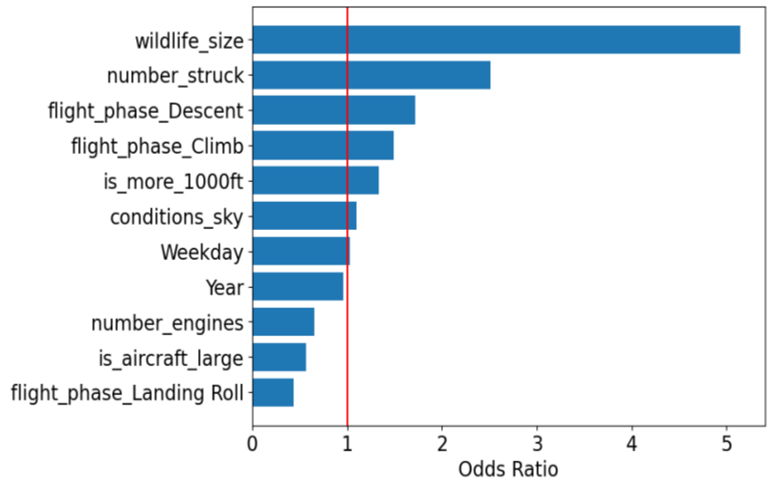

Figure 1: Features Associated with Damaging Strikes

4.2. Forecasting the Monthly Number of Strikes with AR Models

SARIMAX and ARIMA models were fitted with seasonally differenced data because our original series was not stationary. The parameter *"p"* was set as 2 in both models based on the autocorrelation plots; the parameter *"q"* was set to 3 in the ARIMA model based on the partial autocorrelation plot; and *"d"* was set to 0 because seasonal differencing was already applied. As the evaluation metric, the *Mean Absolute Percentage Error* (MAPE) of the models were compared against a baseline model based on the prediction of the average monthly number of strikes. The MAPE of the test set was 3.2%, 2.7%, and 2.8% for the baseline, SARIMAX, and ARIMA models, respectively, indicating that both AR models performed better than the prediction of the mean value. **Figure 2** shows our forecast reconstructed on the original series. As we can see in the figure below, our model seemed to accurately predict the trend and capture the monthly patterns almost perfectly. However, we can see an overestimation in the predicted number of strikes, especially for the data between February 2011 and December 2011 (i.e. the last peak).

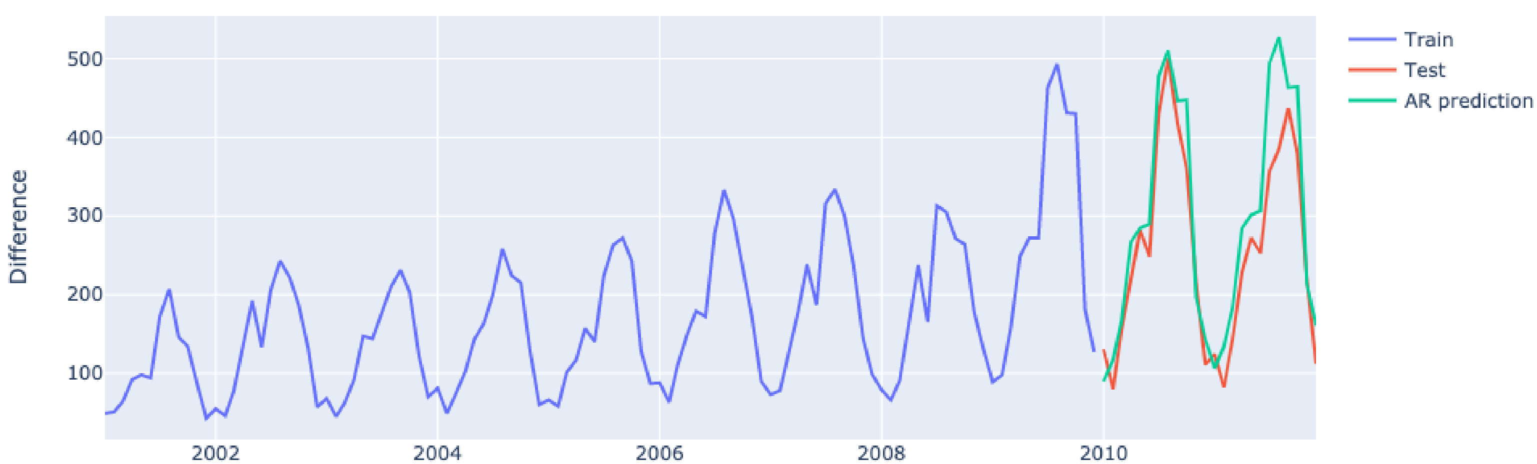

Figure 2: Forecast of Number of Strikes Over a Monthly Period (Reconstructed Series).

4.3. Predicting Damaging Strikes with Supervised Machine Learning

The following models were used to predict damaging strikes: Logistic regression, K-Nearest Neighbors (KNN), Support Vector Machines (SVM), Decision Tree, Random Forest, XGBoost, and Neural Network. For each model (except for Neural Networks), first a base model was fitted and then hyperparameter optimization was performed using 5-fold cross-validation. Although *accuracy*, *recall, precision, and ROC AUC* scores were reported, model evaluation was primarily focused on *accuracy* (to estimate the proportion of correctly predicted observations to the total observations) and *recall* (to estimate the proportion of the damaging strikes correctly identified out of the total damaging strikes).

Table 1 shows the evaluation metrics of the final models selected after hyperparameter optimization. Among these models, the Neural Network had the highest *accuracy* of the test set (82.3%). Additionally, the highest *recall* was observed for the Decision Tree (77.6%). It is also noticeable that, in general, gains in *accuracy* seemed to come with a cost in the *recall*, and vice versa. When considering this fact, the most balanced model in terms of these two metrics was XGBoost (Accuracy: 75.3%; Recall: 74.7%).

Table 1: Evaluation Metrics of the Models.

	Validation Accuracy	Test Accuracy	Test Recall	Test Precision	Test ROC AUC
Logistic Regression	76.6%	77.6%	73.5%	26.4%	75.8%
KNN	72.7%	76.3%	62.0%	23.1%	70%
SVM	75.2%	77.5%	68.8%	25.5%	73.6%
Decision Tree	75.4%	68.9%	***77.6%***	20.6%	72.8%
Random Forest	76.3%	73.1%	74.3%	22.8%	73.7%
XGBoost	76.5%	***75.3%***	***74.7%***	24.5%	75%
Neural Network	81.6%	***82.3%***	63.2%	30.2%	73.8%

Conclusions and Future Works

In conclusion, the likelihood of a bird strike to cause damage is higher for incidents involving small one-engine airplanes, animals of large size and in large groups, during descent or climb, and at altitudes higher than 1000 feet. Thus, pilots should be warned when these factors are involved.

Our forecasting model correctly captured the seasonal pattern of the data and may be used to predict future trends, as well as to guide plans of action in a monthly time frame. Future analyses should focus on data before and after the pandemic to capture the patterns that emerged from the disruptions in the aviation industry during this period.

The Neural Network had the highest accuracy in predicting the outcome of bird strikes compared to the other models, and it could be used for prediction and warning systems for pilots and other operators if this metric is prioritized. If the recall is preferred as a score, then the Decision Tree model is recommended. Finally, XGBoost is suggested for a model balanced between recall and accuracy. As discussed before, the learning gain between these models seemed to be limited, with gains in accuracy being counter-balanced by loss in the recall and vice versa. Thus, it is suggested to include other explanatory features that may be missing from the present dataset. Such features could improve the predictive power of the models and, consequently, increase both accuracy and recall. Examples of additional features are the speed of the aircraft, the angle of the aircraft to the soil and the bird at the moment of the strike, as well as conditions of the aircraft such as age and miles/years in service.

References

[1] Federal Aviation Administration. Wildlife Hazard Mitigation. https://www.faa.gov/airports/airport_safety/wildlife/faq/. Accessed on September 19, 2021.

[2] Federal Aviation Administration. FAA Wildlife Strike Database. https://wildlife.faa.gov/home. Accessed on September 19, 2021.

[3] data.world. 2000-2011 Birds Strikes Planes. https://data.world/shihzy/2000-2011-birds-strikes-planes. Accessed on August 10, 2021.