Prediction of Animal Strike on US Commercial Flights

Gábor Horváth

Business objectives

- Create a statistical analysis to identify those reasons (based on the data available), which are determining the the risk of an animal strike for an airport.
- Create a prediction model, which can be used to predict the risk of an animal strike for a given flight.

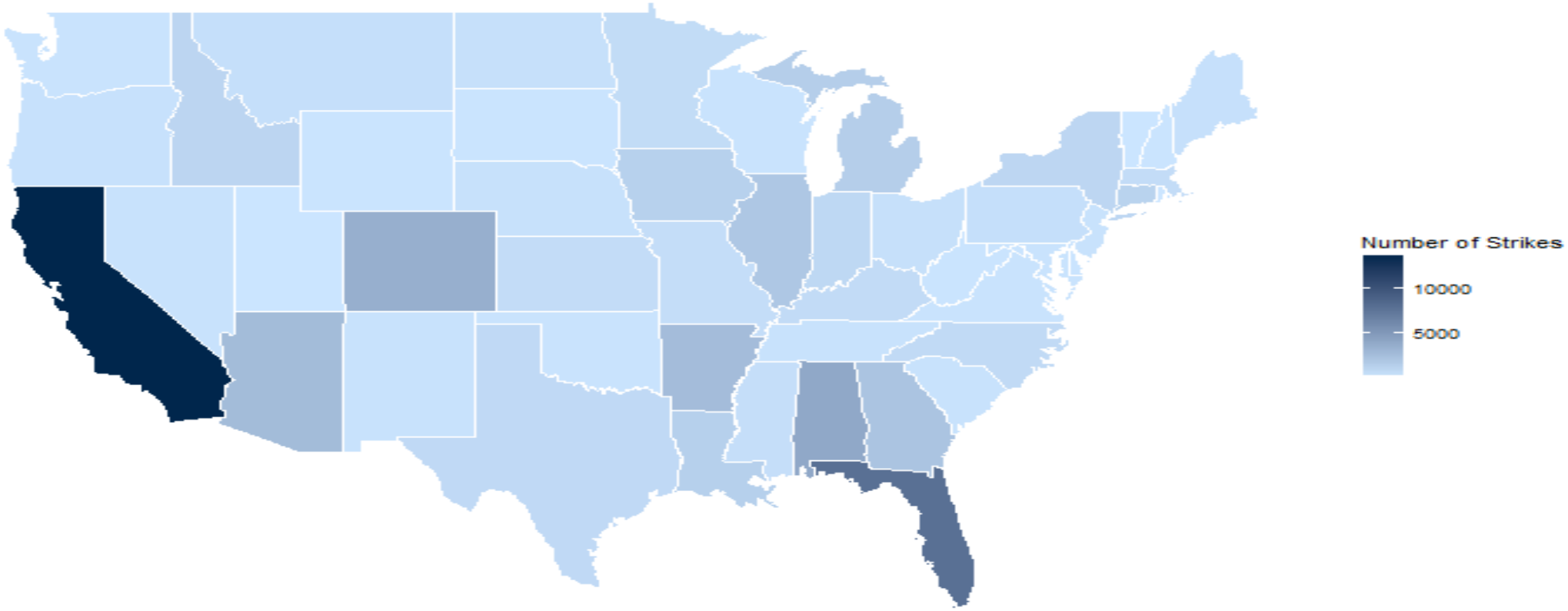
Data sources

- Federal Aviation Administration - Wildlife Strike Database
- United States Department of Transportation - Bureau of Transportation Statistics - Flight performance
- Federal Aviation Administration - Airport Data & Contact Information

Tools used

- R
- R Studio
- knitr, MiKTeX
- ggplot2
- H2O
- CRIPS-DM process model
- buckets, github

Strike distribution



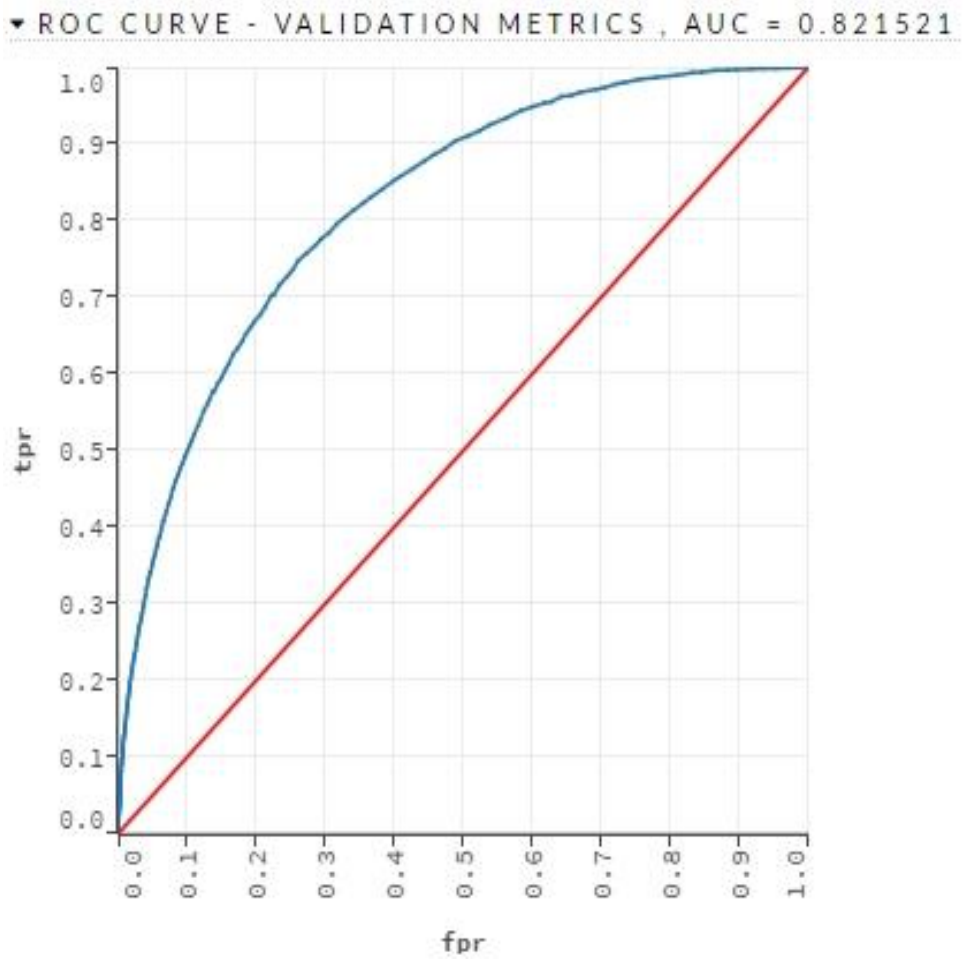
Model 01 - details

- log-log linear regression model
- categorical and continous predictor variables from airport point of view
- outcome variable: number of animal strike per airport
- location and traffic of the airport influence the outcome

Model 02 - details

- H2O platform - Java
- categorical and continous predictor variables from flight point of view
- outcome variable: flight got animal strike or not
- most influental predictors were the airports (origin, destination), the arrival and departure time blocks and the month

Receiver Operating Characteristic (ROC) curve



Confusion matrix

	0	1	Error
0	3082474	25926	0.008341
1	14791	2224	0.869292
Totals	3097265	28150	0.013028

Thank you