AVIATION DATA ANALYSIS AND PREDICTION

Sai Krishna Emmadishetty | Advisor Dr.. Charles Hoot

Computer Science and Information Systems | North West Missouri State University

Introduction

- Predicted ratio of fatal injuries in Aircraft crashes for Aviation Dataset obtained from kaggle.
- Classified aviation related events as accident/incident using meta-data related to the flight and the event.

Materials

Features:

1. Data related to Aircraft

2. Details about the flight and injuries count if any.

- Performed multiple Regression techniques including RF, gradient boosting and linear regression to train the model for regression problem.
- Used SVM, decision tree and RF with grid search for event classification.

Results

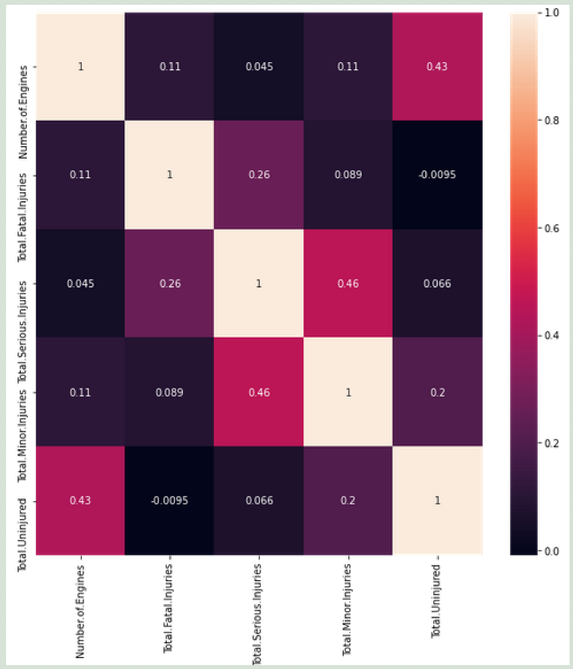

Fig1: Correlation between features along with label

We find a correlation between Number of air craft engines and fatal injuries which explains that large air crafts if crashed would cause more loss.

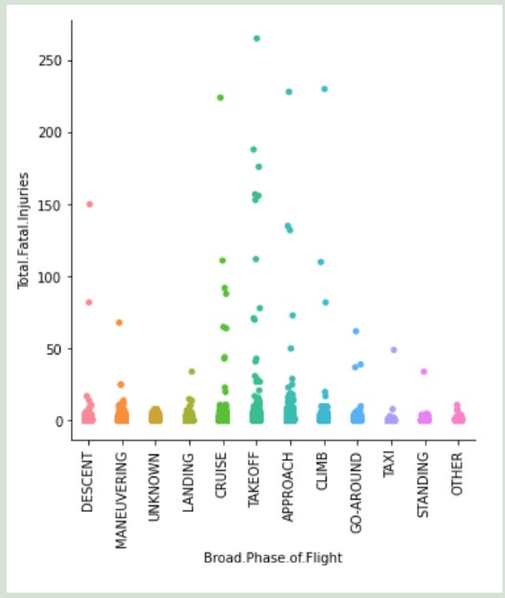

Fig2: relation between fatal injuries count and flight phase

Aircraft crashes happened during takeoff account for the highest number of fatal Injuries

Conclusion

- Evaluate the minimum number nights of Aviation dataset.
- The Graphical representation show that highest number of the fatal injuries occur in accidents at takeoff time
- Using Random Forest Classifier accuracy of 99.44% is obtained.
- Model score of 0.936 is achieved with gradient boosting regression for the fatal injuries prediction
- We can avoid over-fitting by regularizing data and using ensemble technique.

Additional Recourses

- https://scikit-learn.org/stable/modules/svm.html
- www.kaggle.com/khsamaha/aviation-accident-database-synopses

Acknowledgment

Dr.Charles Hoot

Further Information

https://github.com/44-599-MachineLearning-S21/project-machine-learning-s21-Saikrishna1545