Semester Project [rename]

Carlos Rojas, San José State University

October 2021

Abstract

Introduction

You hear the pilot yelling "Mayday Mayday!" over the intercom. What do you do? Brace? Hug your loved ones? Turn off airplane mode and text your family? We hope you never have to experience anything remotely close to this but the recent 20 year memorial of 9/11 got us thinking about what causes a plane to crash. Though we cannot completely eradicate the mishaps of airplanes, we will study the causes of fatalities and use various data mining techniques to determine the risk associated with flying.

Our dataset comes from the National Transportation and Safety Board and consists of over 74 thousand accidents and incidents from 1948 to 2013 (https://data.ntsb.gov/avdata). The problem that we set out was to determine whether aviation is becoming safer as time moves forward. This is quantified by analyzing the number of fatal accidents with respect to relevant crash data as well as quantifying the total number of accidents that have occured overtime. Not only that but we set out to determine if a crash can be predicted given various factors/features. If a crash can be predicted then we will be able to evaluate the likelihood that a crash will occur and would eventually be able to step in and stop the flight. In our analysis, we have considered specific parameters that are crucial for predicting (or can lead up to) a crash, such as no. of engines, weather, etc. With these pointers the manufacturers and air crew members can take extra caution before taking off.

This was performed with techniques such as dimensionality reduction to cluster whether certain types of crashes are fatal, linear regression to predict the total number of fatalities, logistic regression to determine the strongest flight predictors to lead to a fatal crash, random forest to binary classify if a crash would be fatal or not. More in-depth details of these processes are seen in the Methods section.

Methods

Comparisons

Example Analysis

Conclusions

References