Algorithims and Analyses to Predict Trends in General Aviation Incidents

Jeffrey Wu | Francis Zamora | Ricardo Ballesteros

Abstract

The federal government keeps a close eye on commercial aircraft data to maintain safety for passengers; however, recreational flight are loosely regulated, and, as a result, accidents occur much more frequently. This project focuses on three different objectives: clustering reports based on hidden trends, binary classification on likelihood of injuries, and discovering general trends in the data.

Datasets

Federal Aviation Administration

- Accident and Incidident Data
  MA,NH,VT,ME

NASA

- Aviation Safety Reporting data

Distance Formula for K-means clustering

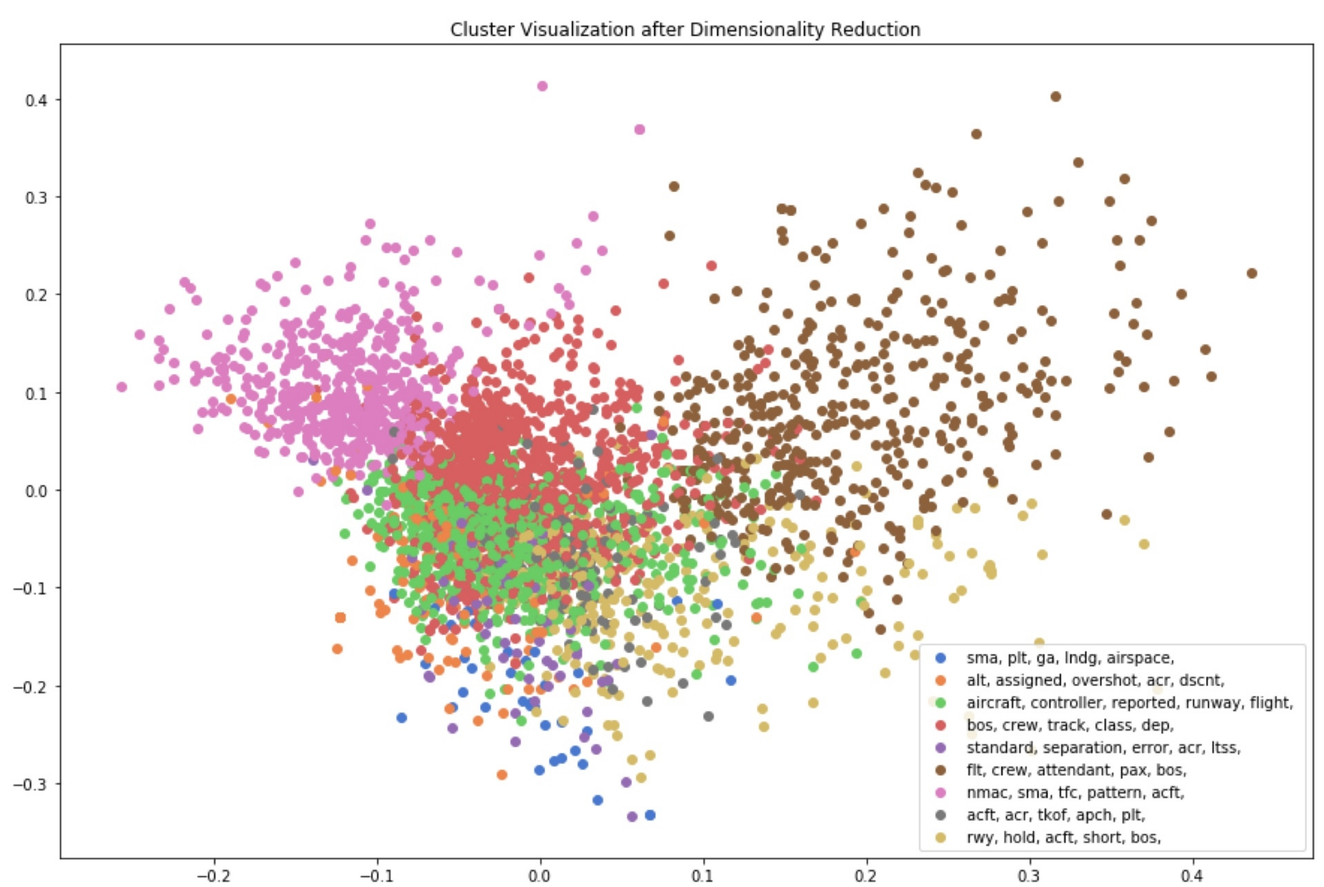
Objective Function: Minimize the distance between the cluster centers and cluster points for report scores

$$J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - c_j \right\|^2$$

Visualising K-means

Objective: Understand and find different causes of aviation incidents and accidents based off of reports provided by the FAA & NASA.

Technique: The data was given in the form of narratives reported by pilots after an accident or incident. To cluster the report, we first converted the narratives into their word frequency scores, and then their tf-idf scores. Afterwards, we applied SVD to reduce the dimensionality of the dataset. Finally, we clustered the truncated scores using the k-Means algorithm to discover hidden clusters in the accident reports as shown below with the labels reporting the top 5 words in each cluster.



Binary Classification on Injuries

Objective: Predict whether a report will occur or not based on weather conditions, plane model, flight plan, number of engines, and number of passengers

Technique: Logisitic Regression

Event City	Aircraft Make	Aircraft Model	Primary Flight Type	Flight Conduct Code	Flight Plan Filed Code	Total Fataliti	Total Injurie
MARTHAS VINEYARD	CESSNA	CE-150	PERSONAL	GENERAL OPERATING RULES	NONE	0	0
BEVERLY	BEECH	BE-23	PERSONAL	GENERAL OPERATING RULES	NONE	0	0
LAWRENCE	BLANCA	BL-8	PERSONAL	GENERAL OPERATING RULES	NONE	0	0
MARTHAS VINEYARD	CESSNA	CE-172	PERSONAL	GENERAL OPERATING RULES	VISUAL FLIGHT RULES	0	0
SOUTH WEYMOUTH	MIKOYAN	MIG-17	OTHER	GENERAL OPERATING RULES	VISUAL FLIGHT RULES	0	0
BOSTON	BOEING	737	SCHEDULED AIR CARRIER	AIR CARRIER/COMMERCIAL	INSTRUMENT FLIGHT RULES	0	1
NEW BEDFORD	GULSTM	GA-1159	BUSINESS	GENERAL OPERATING RULES	INSTRUMENT FLIGHT RULES	0	0
MANSFIELD	THNDCT	THNDCT-AX8	FOR HIRE	GENERAL OPERATING RULES	VISUAL FLIGHT RULES	0	3
BEVERLY	PIPER	PA-32	PERSONAL	GENERAL OPERATING RULES	NONE	0	0
BOSTON	CESSNA	CE-310	AIR TAXI COMMUTER (SC	AIR TAXI/COMMUTER	INSTRUMENT FLIGHT RULES	0	0
MONTAGUE	PIPER	PA-28	PERSONAL	GENERAL OPERATING RULES	VISUAL FLIGHT RULES	0	0
LAWRENCE	BEECH	BE-23	INSTRUCTION	GENERAL OPERATING RULES	INSTRUMENT FLIGHT RULES	0	0
MARTHAS VINEYARD	PIPER	PA-31	EXECUTIVE	GENERAL OPERATING RULES	INSTRUMENT FLIGHT RULES	0	0
BEDFORD	PIPER	PA-34	EXECUTIVE	GENERAL OPERATING RULES	INSTRUMENT FLIGHT RULES	0	0
BOSTON	BOEING	727	SCHEDULED AIR CARRIER	AIR CARRIER/COMMERCIAL	INSTRUMENT FLIGHT RULES	0	4
LAWRENCE	PIPER	PA-28	PERSONAL	GENERAL OPERATING RULES	NONE	0	0
BOSTON	SHORTS-BOMBARDIER	360	AIR TAXI (SCHEDULED- N(	AIR CARRIER/COMMERCIAL	VISUAL FLIGHT RULES	0	1
WESTFIELD	CESSNA	CE-172	PERSONAL	GENERAL OPERATING RULES	NONE	0	0
MARTHAS VINEYARD	PIPER	PA-24	PERSONAL	GENERAL OPERATING RULES	INSTRUMENT FLIGHT RULES	0	0

Accuracy: 97.098%
Recall: 11.079%

Future Work

Discover trends nationwide based on accident and incidents using binary classification and statistical analysis.

Report findings to FAA and NASA to influence government regulations on recreational pilots.

Compile flight plans filed by recreational pilots before their flight and predict likelihood of accident after consolidating datasets and flight plans.