FLYING SAFE
FLYING TOGETHER

NEURAL NETWORKS & OPTIMIZING AVIATION SAFETY

AARON O'NEAL | WEI ZHANG | WILLIAM ENGLEHART

AGENDA

THE PROBLEM

GOALS OF ANALYSIS

THE DATA

DATA CLEANING & EDA

PREPROCESSING & MODELING

RECOMMENDATIONS

CONCLUSION

PROBLEM

PROBLEM STATEMENT

The National Transportation Safety Board (NTSB) has collected data on aviation accidents since 1962. Though this information has long helped to guide safety protocols and regulations, the data remains under utilized in the prevention of future incidents.

Guiding Question: What human, environmental, and engineered factors most influence the likelihood of fatalities or injuries when an in-flight incident occurs?

THE DATA PROBLEM

The NTSB maintains a massive and continuous collection of aviation accident data. The depth and complexity presented in the information presents both a challenge and opportunity in the ability for predictive models to promote industry-wide safety guidance.

GOALS

ADDRESSING THE PROBLEM

Guiding the approach to the problem is the focus on using NLP in partnership with Neural Networks to create a comprehensive and holistic approach.

SOLUTION CHARACTERISTICS

- Natural Language Processing
- Neural Networks
- Random Forest Classifier
- Geo Mapping

THE DATA

NATIONAL TRANSPORTATION SAFETY BOARD

The data that guided the production of this analysis was produced by the National Transportation Safety Board (NTSB) and covers investigations of aviation incidents between January 1, 2010, and January 1, 2021.

The NTSB aviation incident database contains information from 1962 and later about civil aviation accidents and selected incidents within the United States, its territories and possessions, and in international waters.

DATA CLEANING & EXPLORATORY DATA ANALYSIS

DATA CLEANING & EXPLORATORY DATA ANALYSIS

APPROACHING THE DATA

After discovering the vast amount of data that the NTSB had on aviation incidents, we scraped all of the incidents from January 1, 2010 to January 1, 2021.

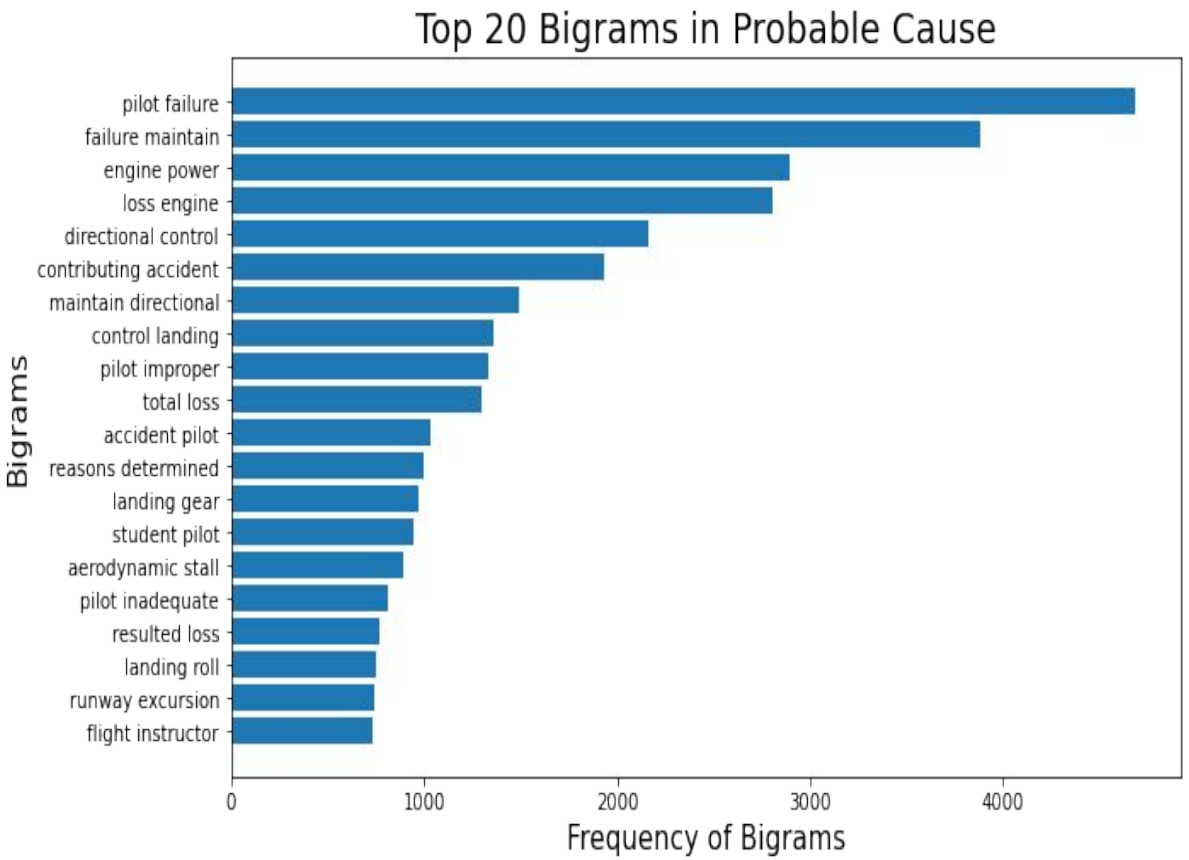
After scraping, we realized there were many issues with how the data was collected and input. There were several values missing and several data points incorrectly input into the dataset. We also narrowed down the incidents to those occurring specifically in the United States

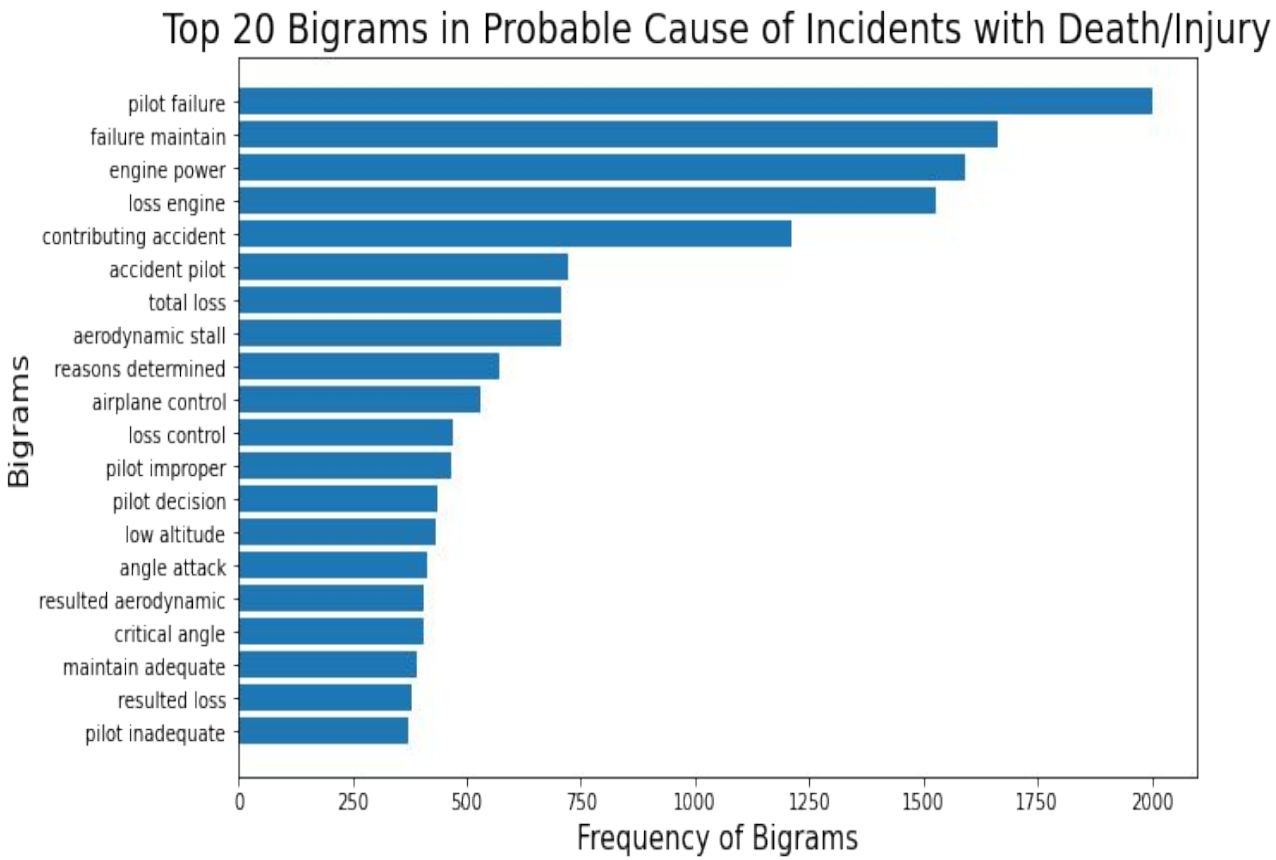
EXPLORATORY DATA ANALYSIS

Returning focus to the problem statement that guides the analysis, the featured EDA works to present and study a variety of statistical perspectives. The primary objective is to identify key trends within the data that will serve both the predictive model and a meaningful solution to our problem of underutilized data.

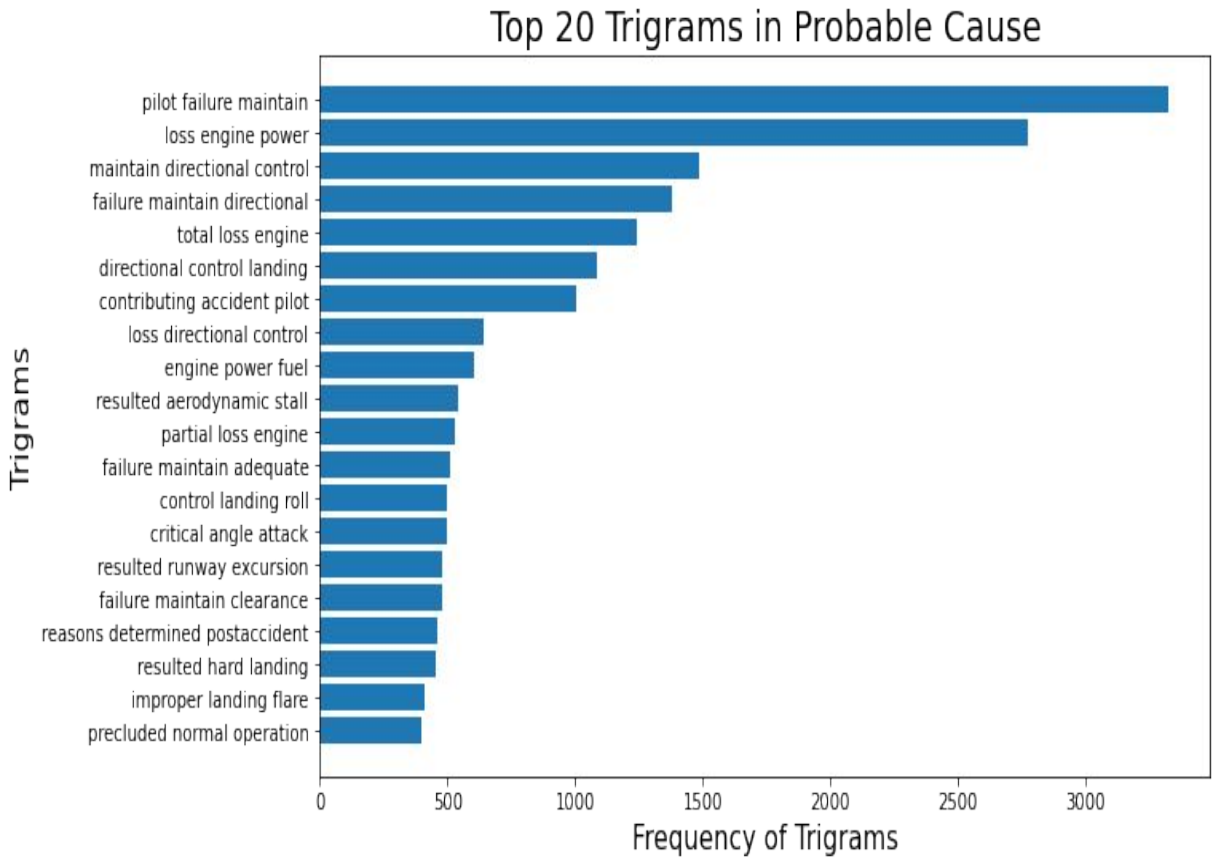
During our EDA, we completed natural language processing on the probable cause, factual narrative, and the analysis narrative parts of each incidents which will be further explained in the following slides.

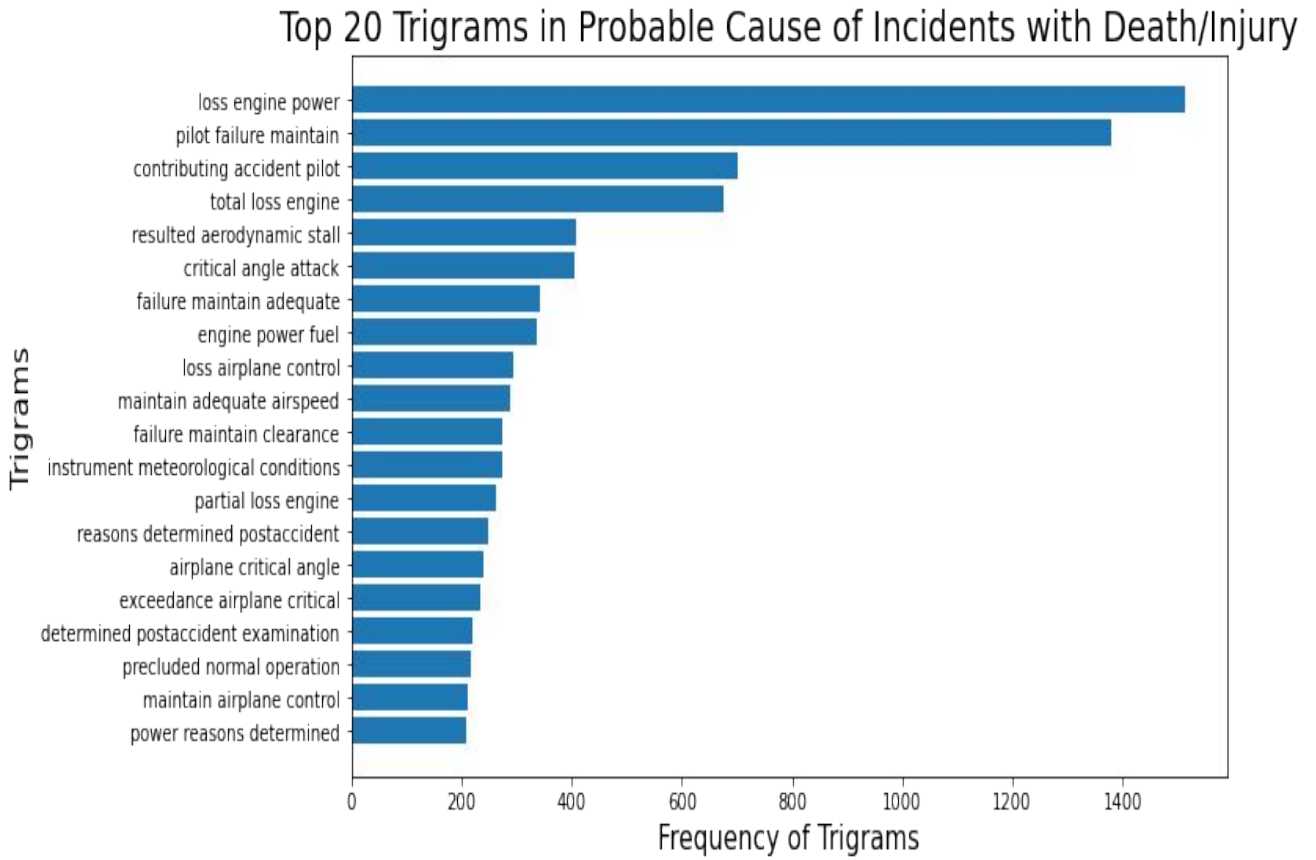
NLP Bigrams
Probable Cause vs Probable Cause w/ Death/Injury



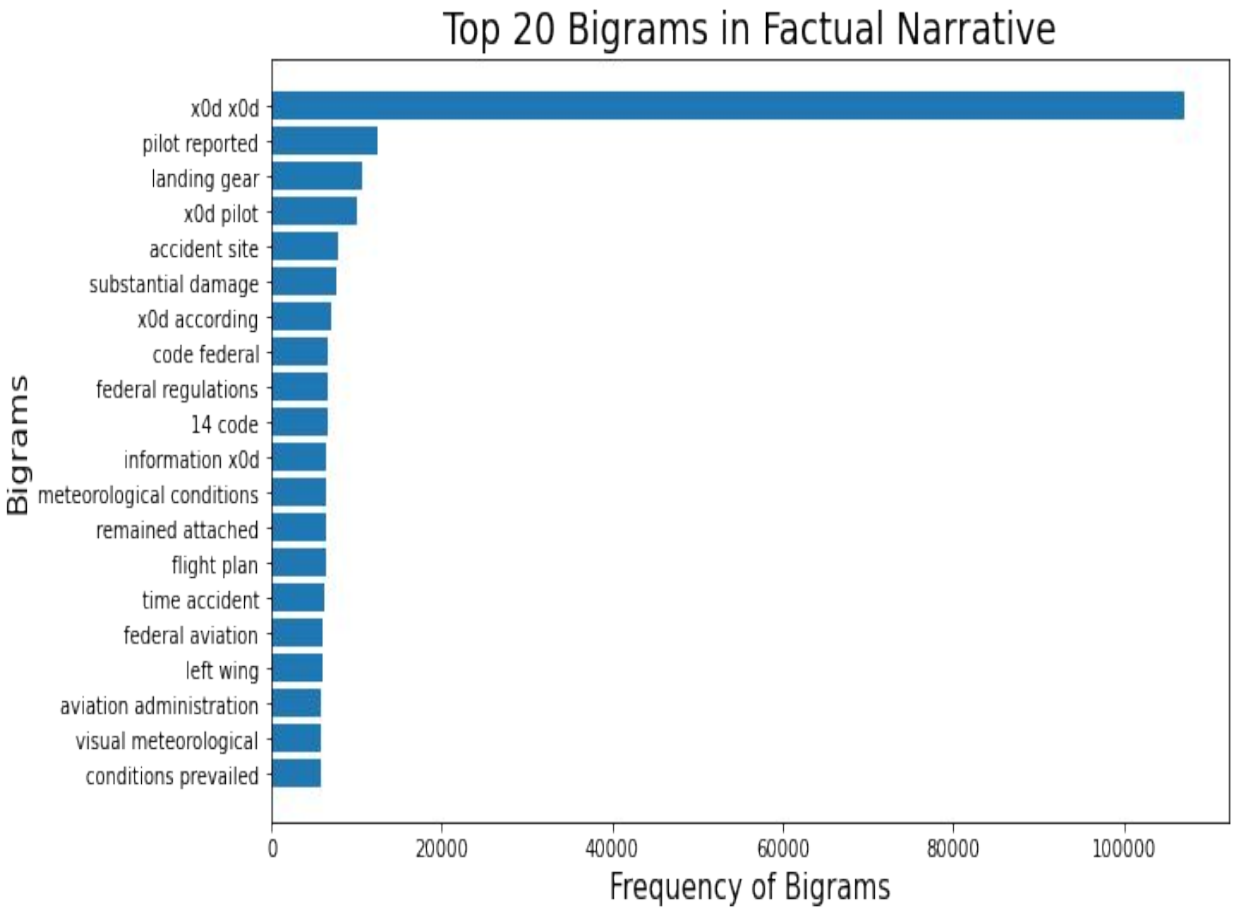


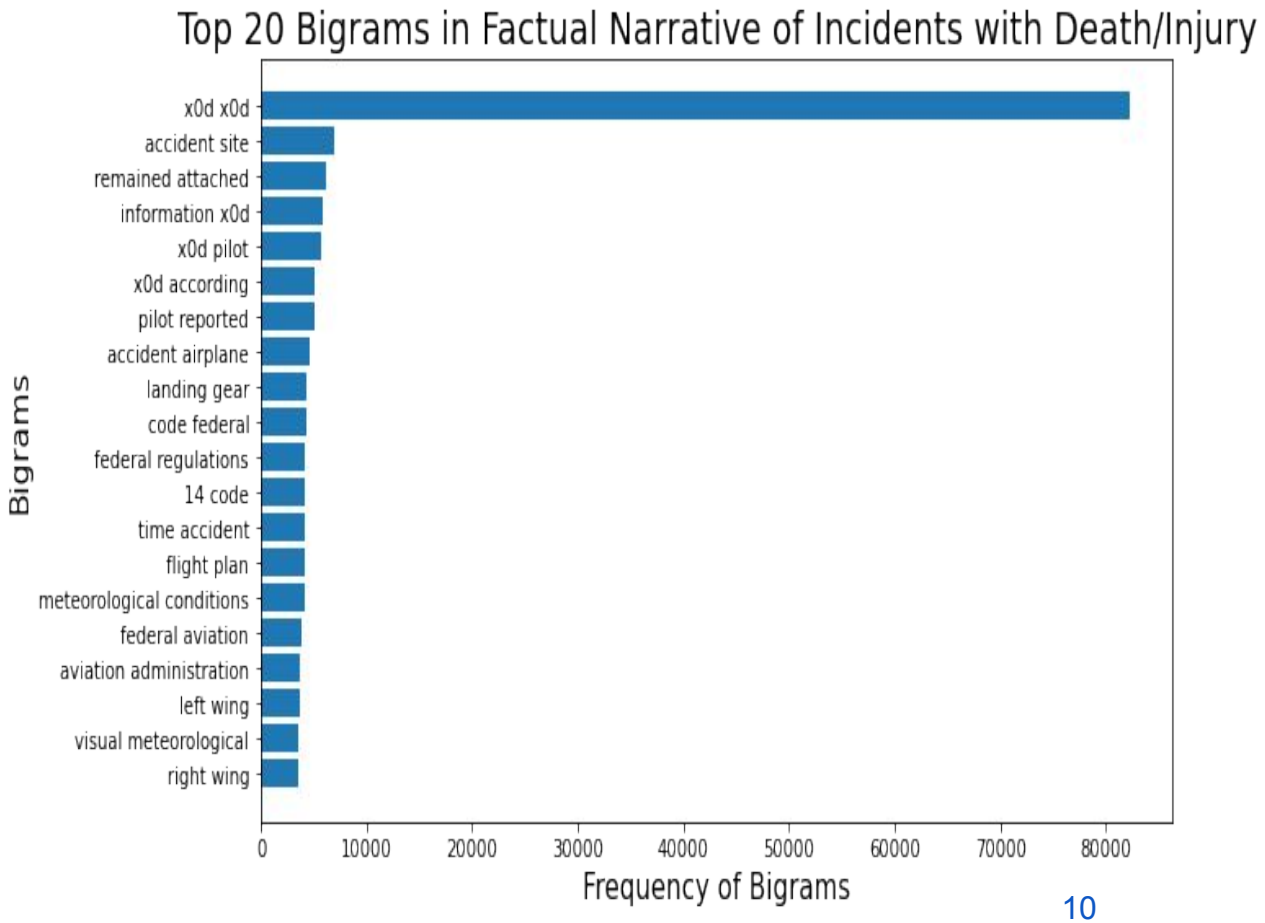
NLP Trigrams
Probable Cause vs Probable Cause w/ Death/Injury



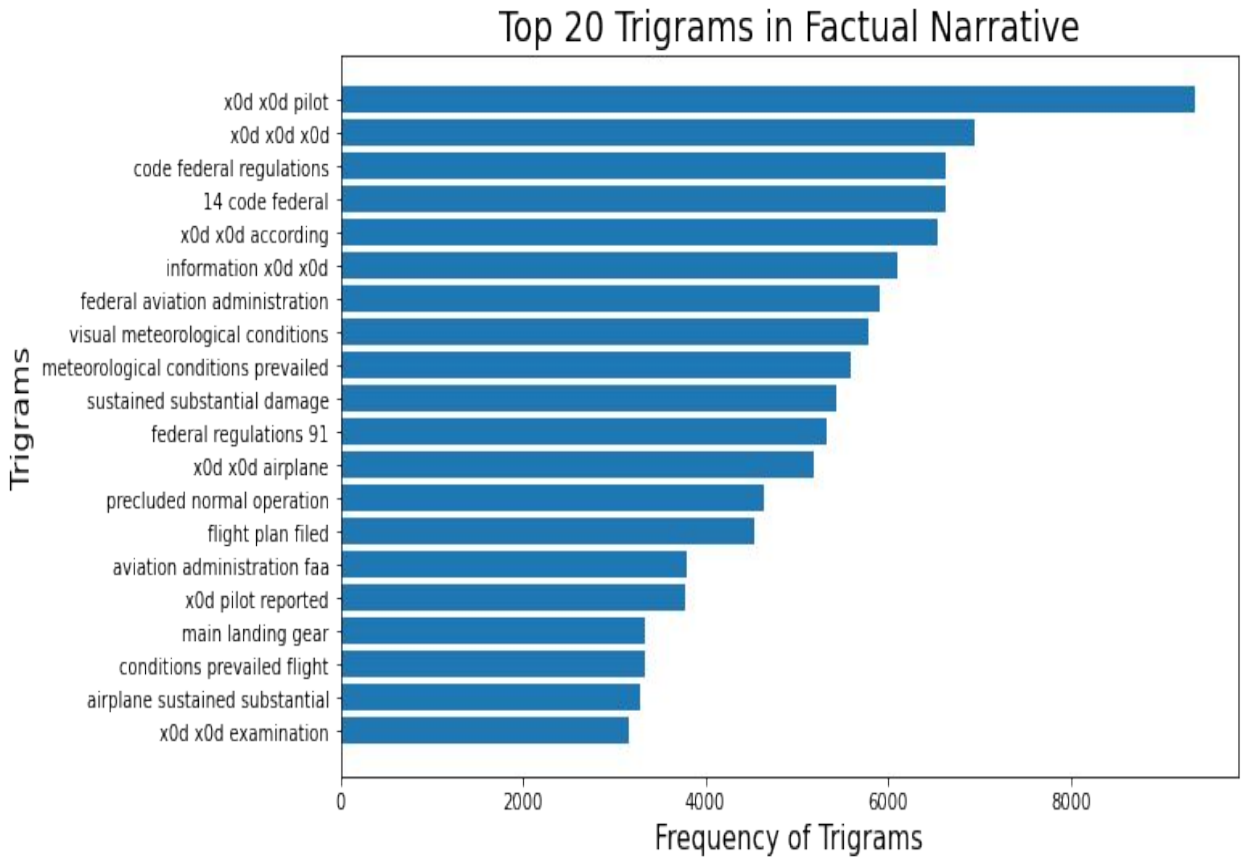


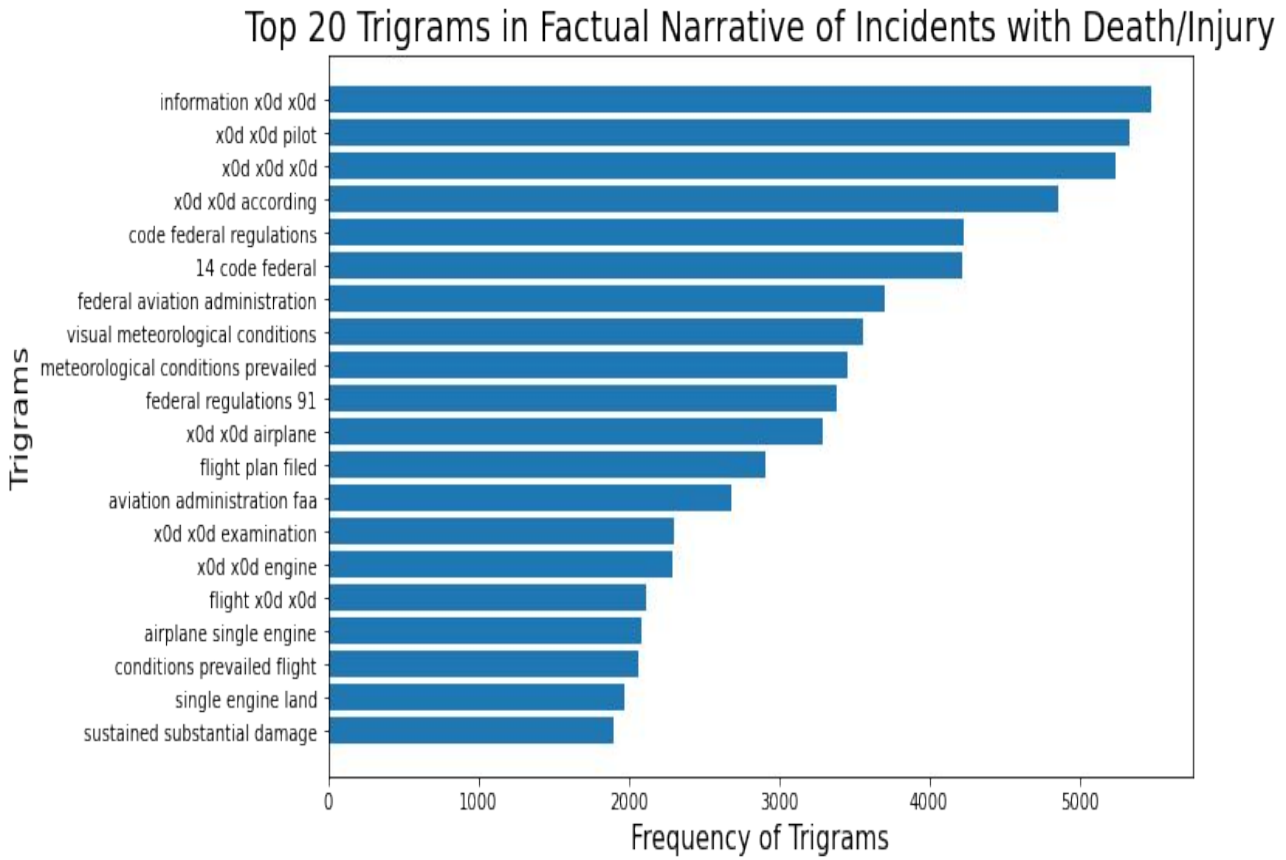
NLP Bigrams
Factual Narrative vs Factual Narrative w/ Death/Injury



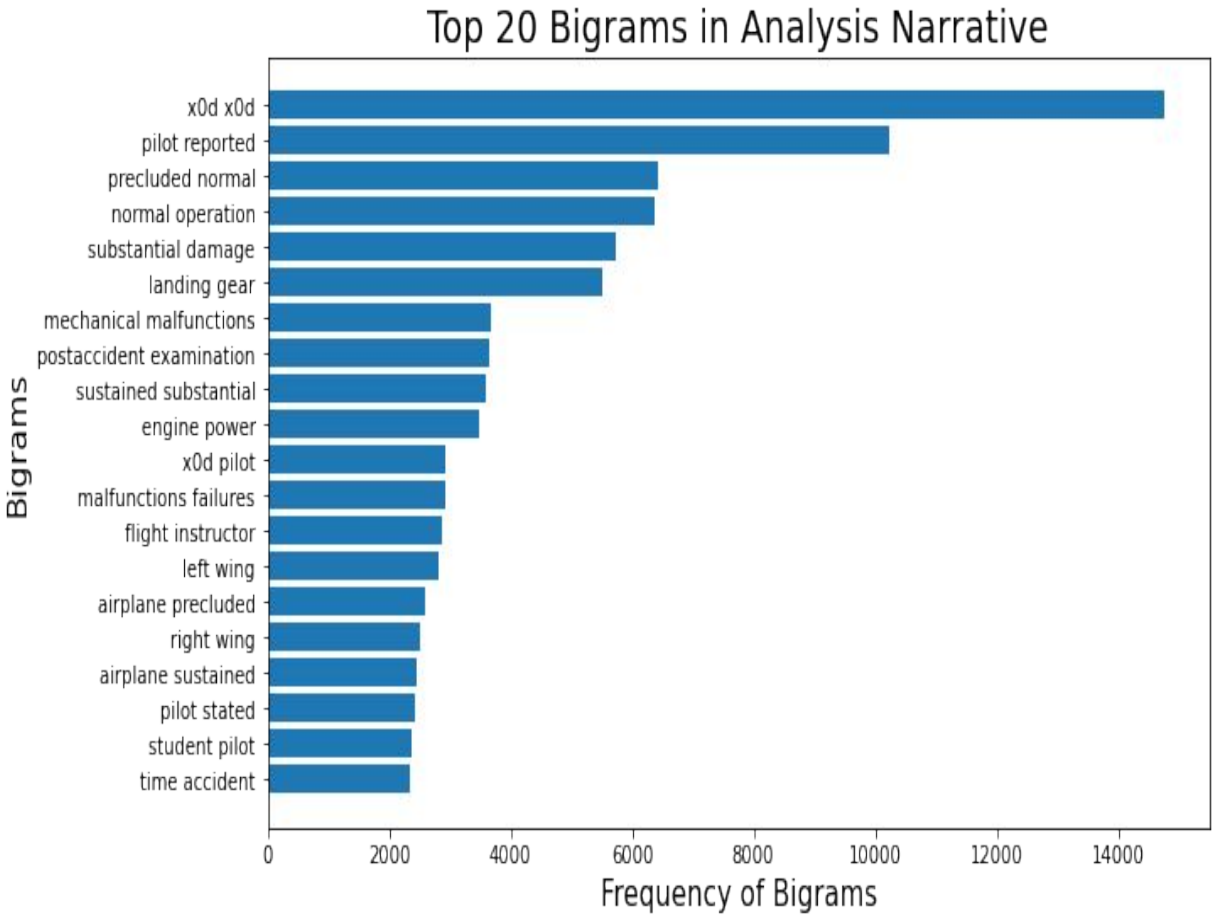


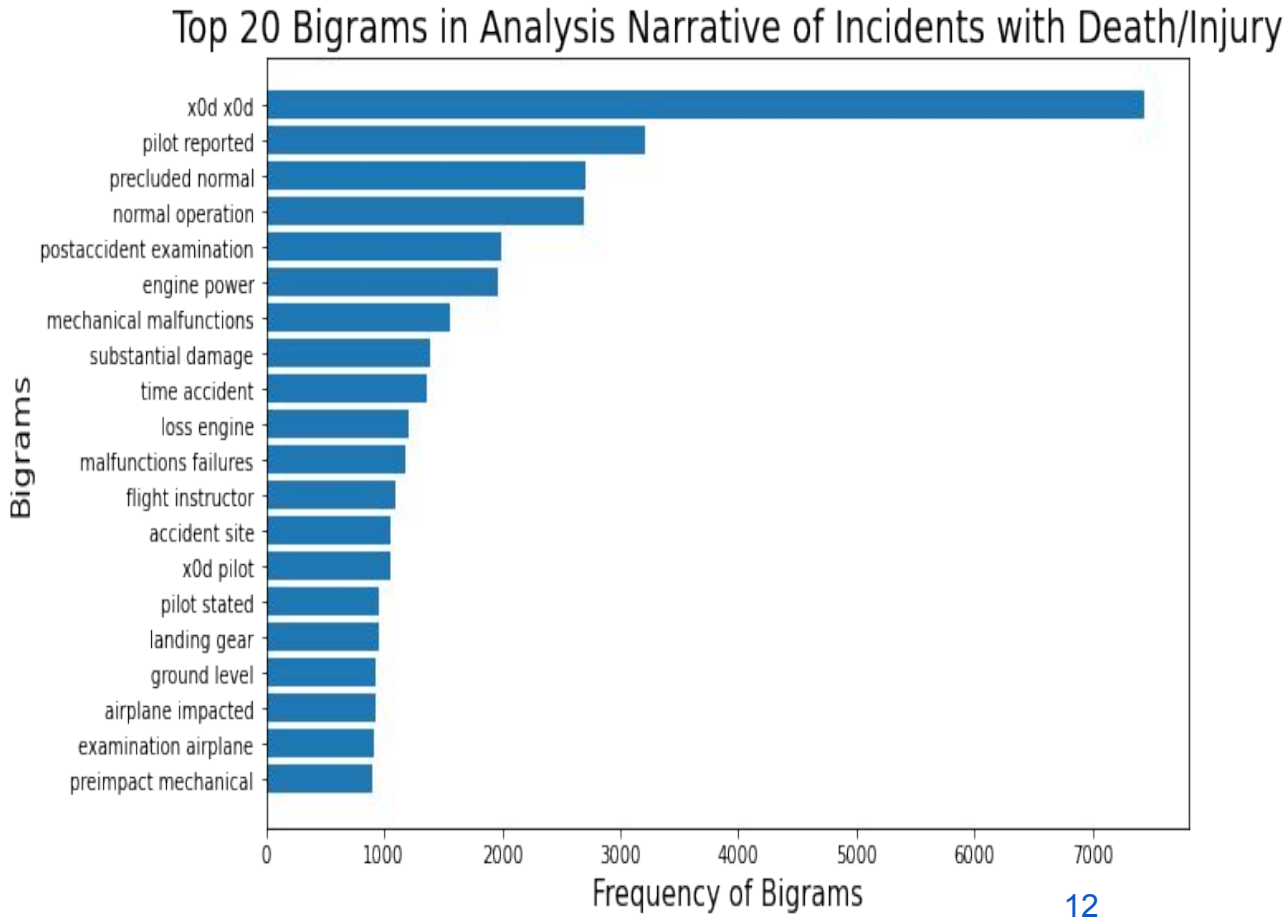
NLP Trigrams
Factual Narrative vs Factual Narrative w/ Death/Injury



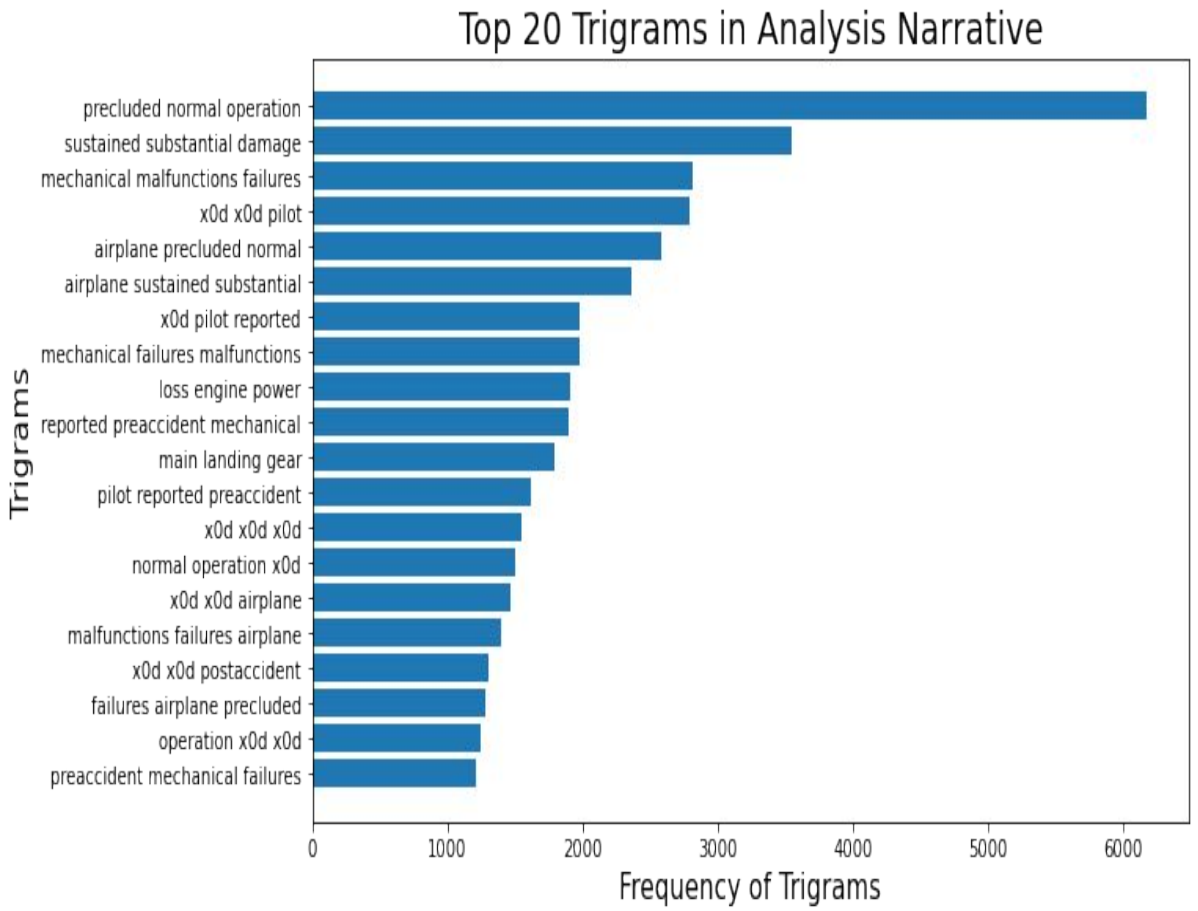


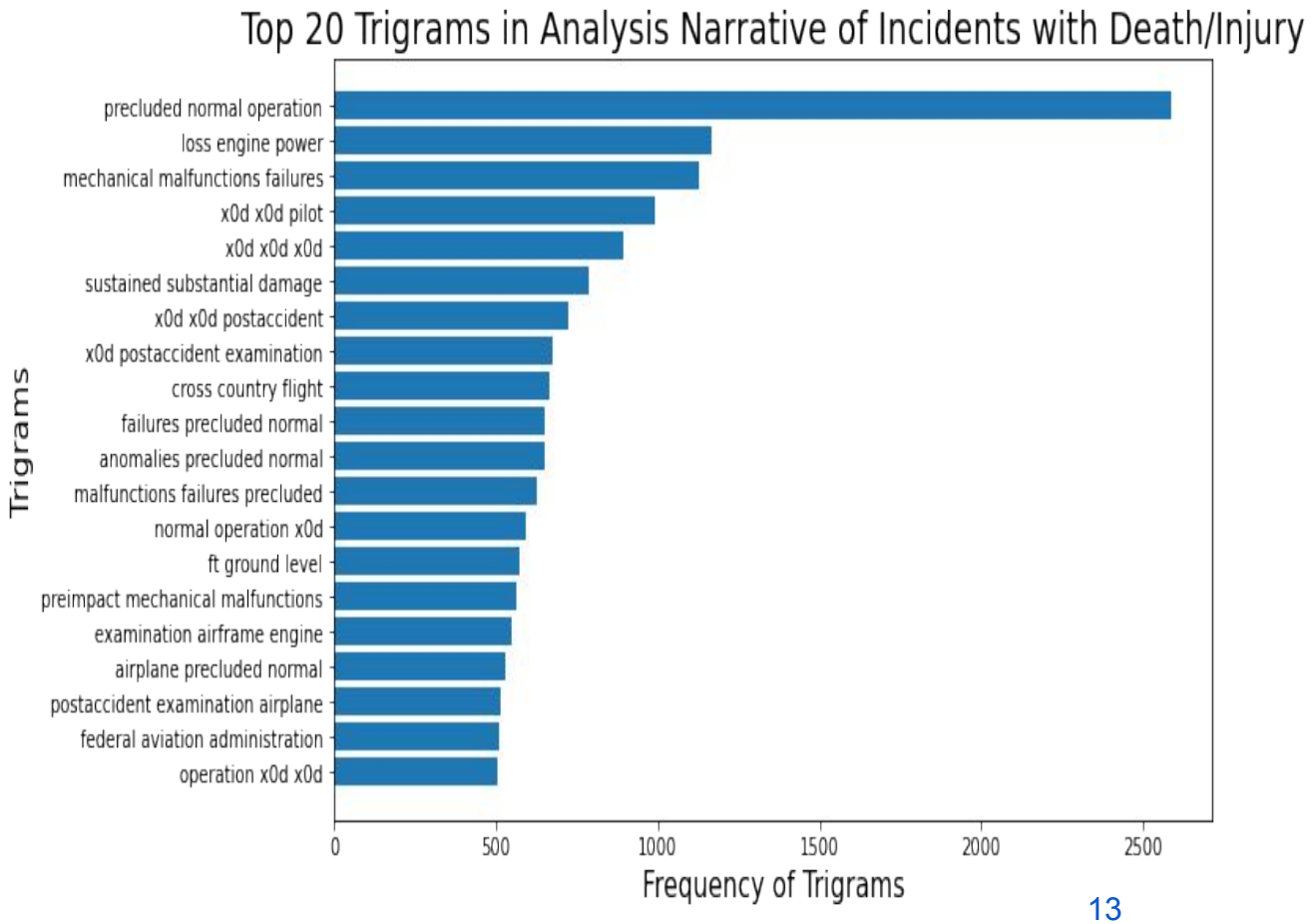
NLP Bigrams
Analysis Narrative vs Analysis Narrative w/ Death/Injury





NLP Trigrams
Analysis Narrative vs Analysis Narrative w/ Death/Injury





PREPROCESSING & MODELING

UNDERSTANDING NEURAL NETWORKS

1

Comprised of a node layers, containing an input layer, one or more hidden layers, and an output layer.

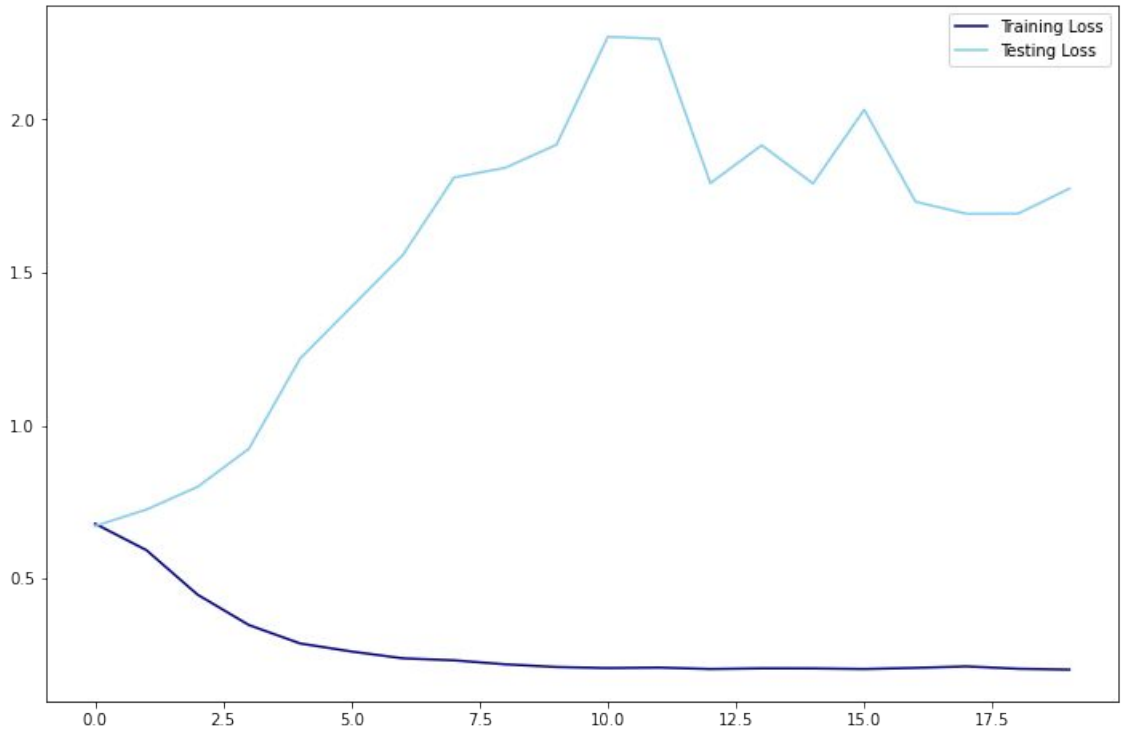
2

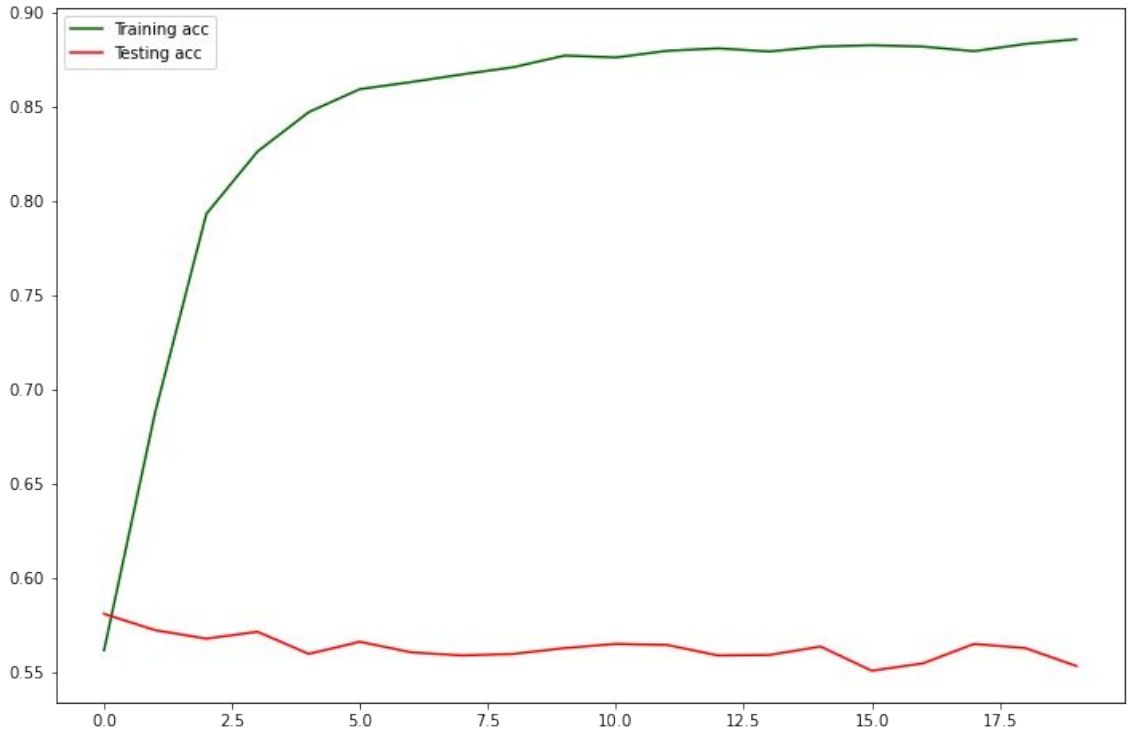
Neural networks rely on training data to learn and improve their accuracy over time.

3

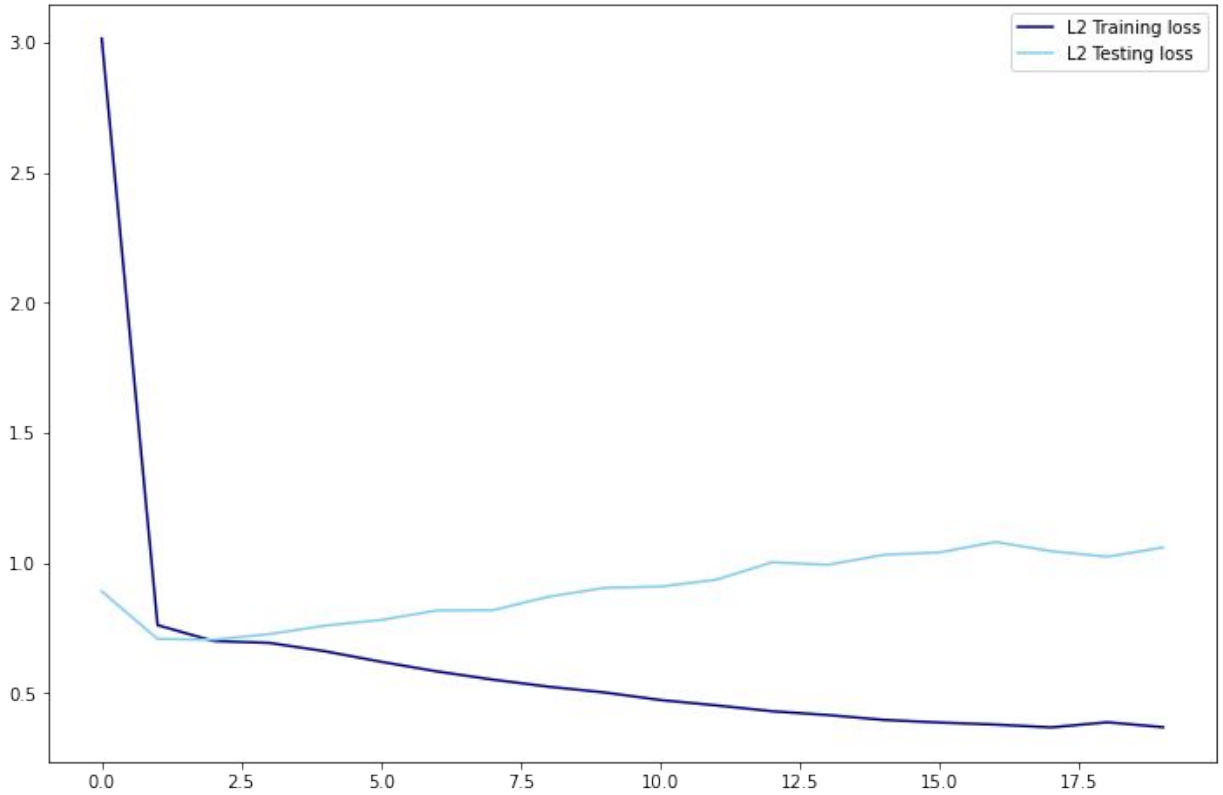
Classify and cluster data at a high velocity.

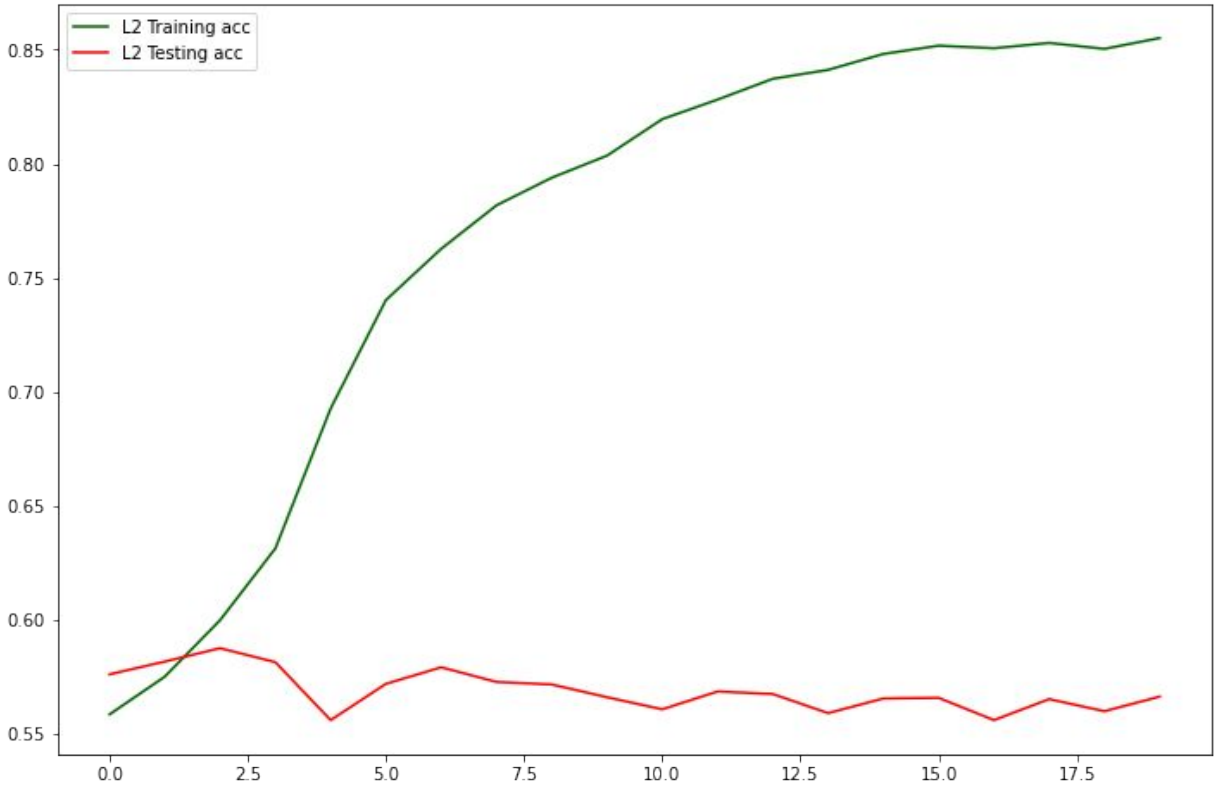
NN Baseline Model



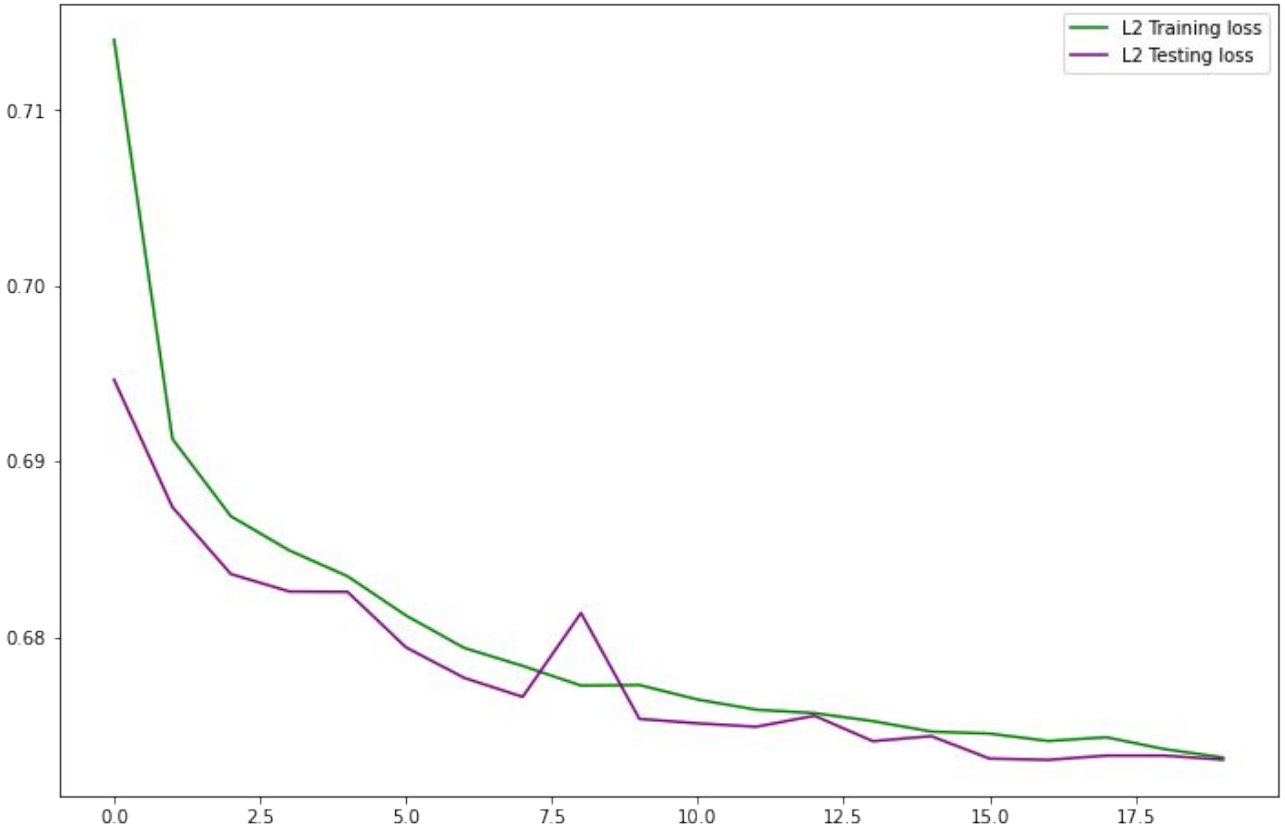


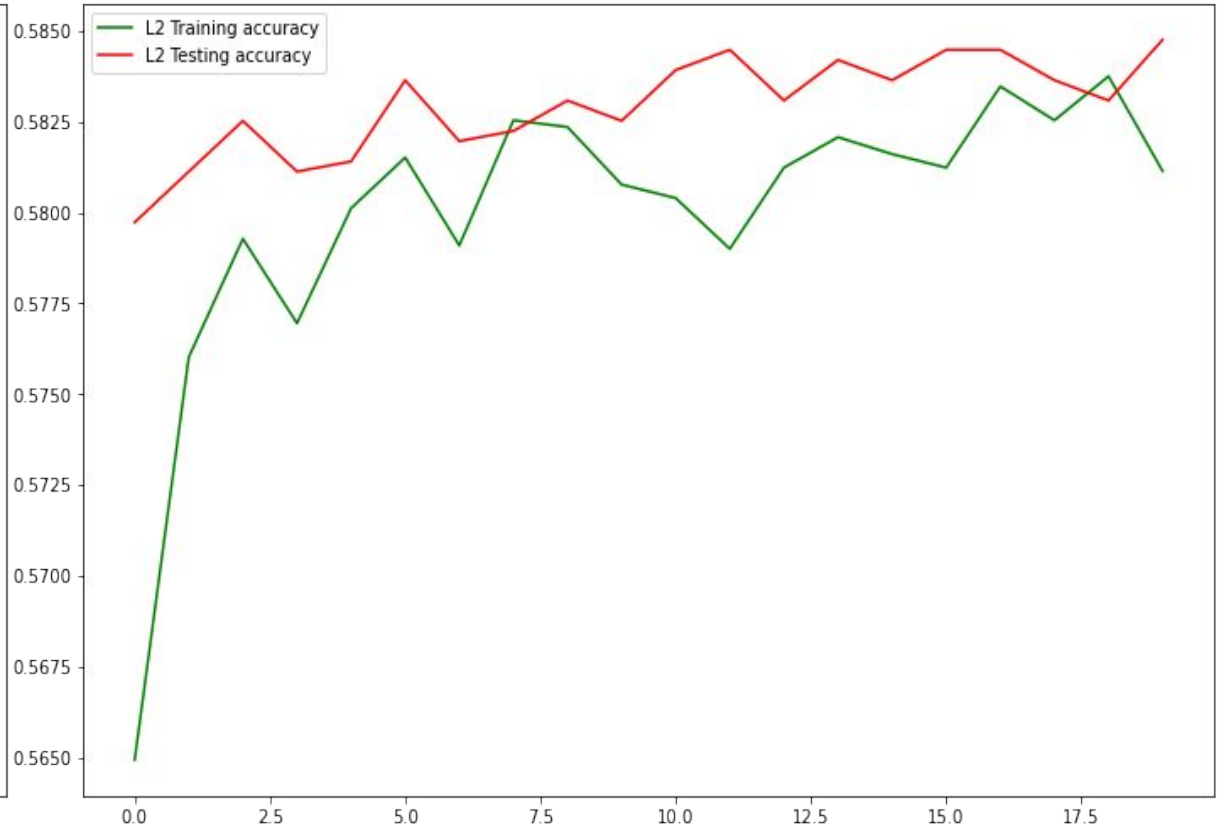
NN Model Improved



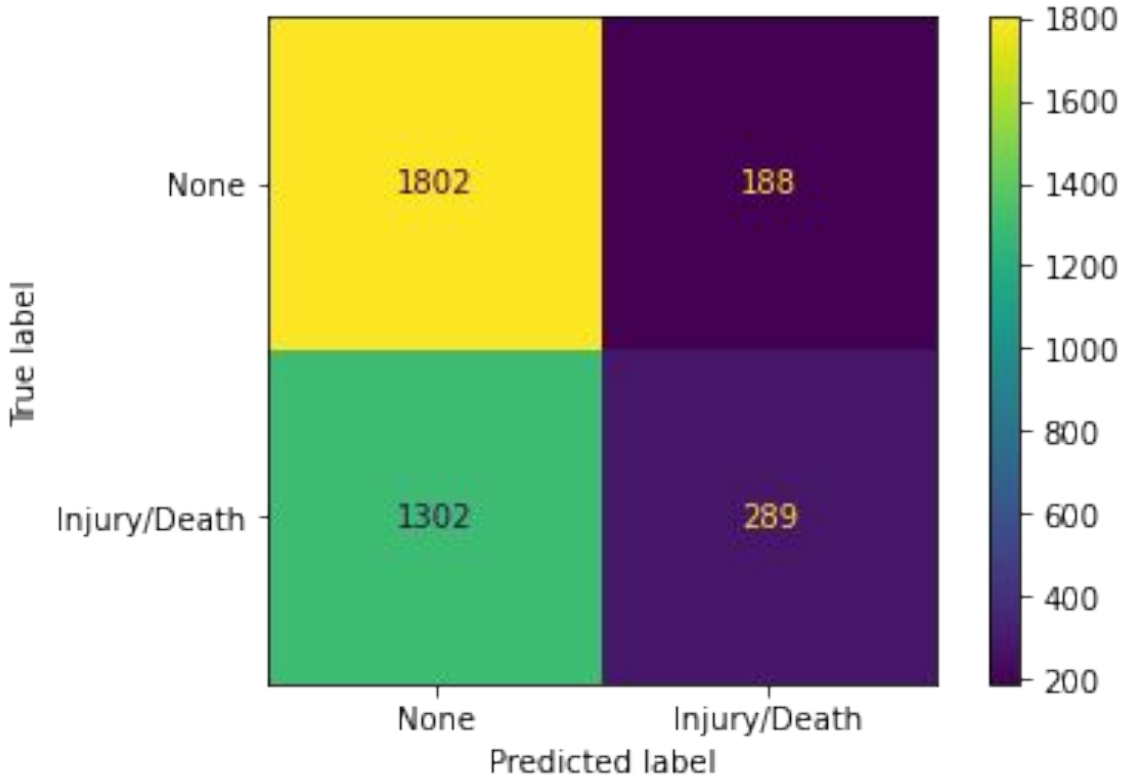


NN Reduced Features





Random Forest



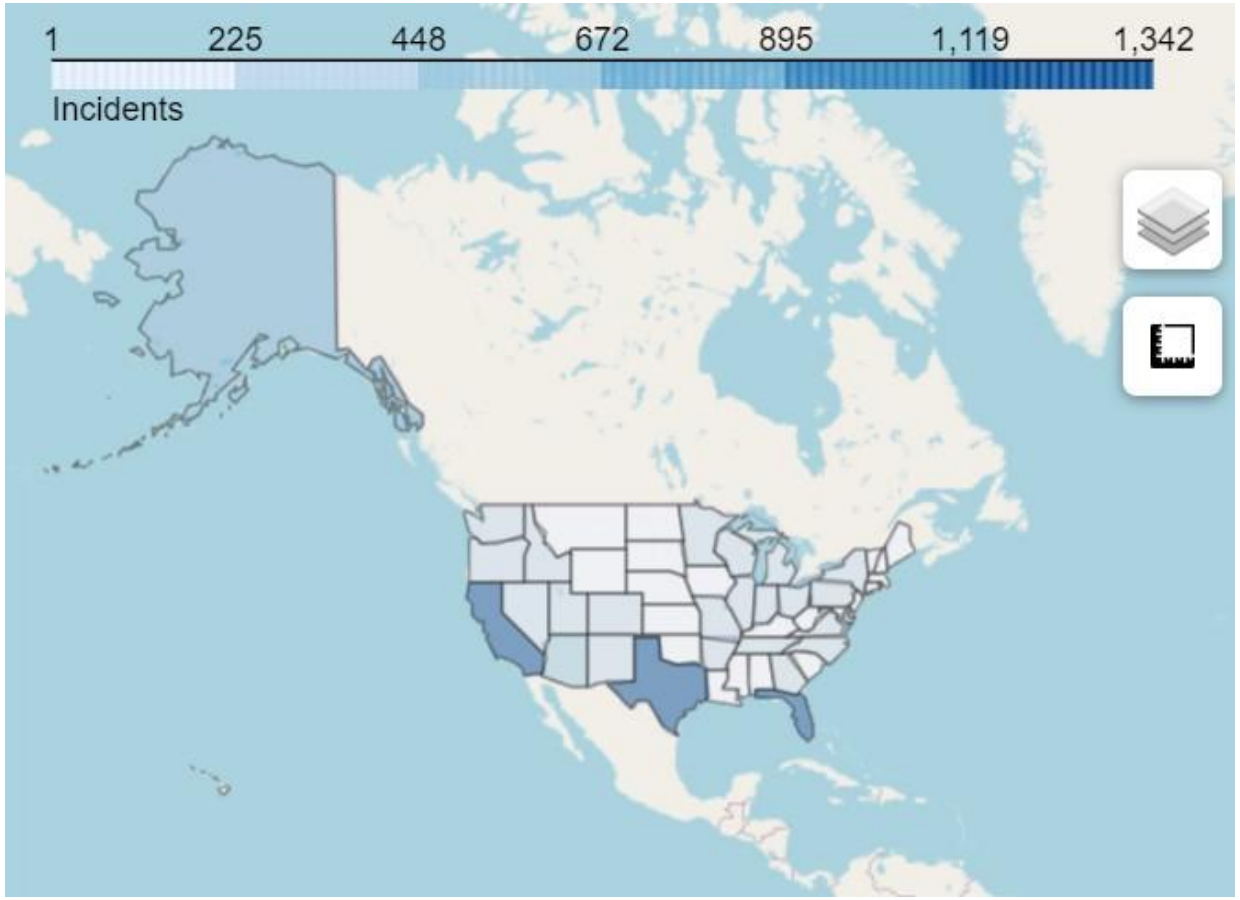
Baseline: 0.5497

Model Score: 0.5787

GOING FORWARD

INTERACTIVE MAP

Map of Incidents



Recommendations & Conclusions

How we can ensure the safety of flight

The Data	Improve the quality and collection of data to promote greater research and understanding.
The Pilot	Understand the significance of human error and the educational commitments that can be made to address this
The Focus	Recognize the complexity of aviation accidents and the creativity required to produce meaningful solutions

THANK YOU

Aaron O’Neal

Wei Zhang

William Englehart