Proposal for a Modernized Flight Risk Assessment Tool for General Aviation Pre-Flight Planning

Category: Improve Aviation Safety

Hunter Walden

Kathleen Hill

Chi Quinn

Erick Torres

Dr. Isaac Gang, Project Advisor

GEORGE MASON UNIVERSITY

Our Team

Hunter Walden

Data Engineer
Steampunk, Inc.
Former USMC MV-22 Pilot
M.S. Data Analytics Engineering, Student
George Mason University

Kathleen Hill

Data Scientist
Systems Planning and Analysis, Inc.
M.S. Data Analytics Engineering,
George Mason University

Chi Quinn

Business Intelligence Analyst
American Society for Radiation Oncology
M.S. Data Analytics Engineering, Student
George Mason University

Erick Torres

Data Scientist
Serco-NA
M.S. Data Analytics Engineering,
George Mason University

Dr. Isaac Gang

Project Advisor
Associate Professor
College of Engineering and Computing
MS Data Analytics Program
George Mason University

Problem Statement

Pre-flight risk assessment is a key step in the pre-flight planning process to ensure the safety of flight prior to takeoff. Traditional flight risk assessment tools (FRATs) are often inaccessible, time-intensive, and incomplete. This can lead to improper or time-critical risk assessment in flight, rather than complete and deliberate risk mitigation.

The solution is to augment traditional FRATs with automated risk assessment based on pilot and flight profiles to produce a risk score that the pilot can then further evaluate for potential risk mitigation. This speeds up the risk assessment process, potentially reduces pilot bias, and can be incorporated into pre-flight planning applications.

Problem → **Solution**

GEORGE MASON UNIVERSITY

Background

"Poor risk management was a root cause of nearly 50 percent of the fatal business aircraft accidents."

– Wright, 2019

"Civilian aviation operations excluding paid passenger transport, accounts for 94% of civilian aviation fatalities." *– Boyd, 2017*

"60% of fatal accidents occurred while flying in instrument meteorological conditions (IMC), most frequently "between October and April, on weekends, in early morning and evening periods, and along the West Coast, Colorado Rockies, Appalachian Mountains, and the Northeast."

– Fultz & Ashley, 2016

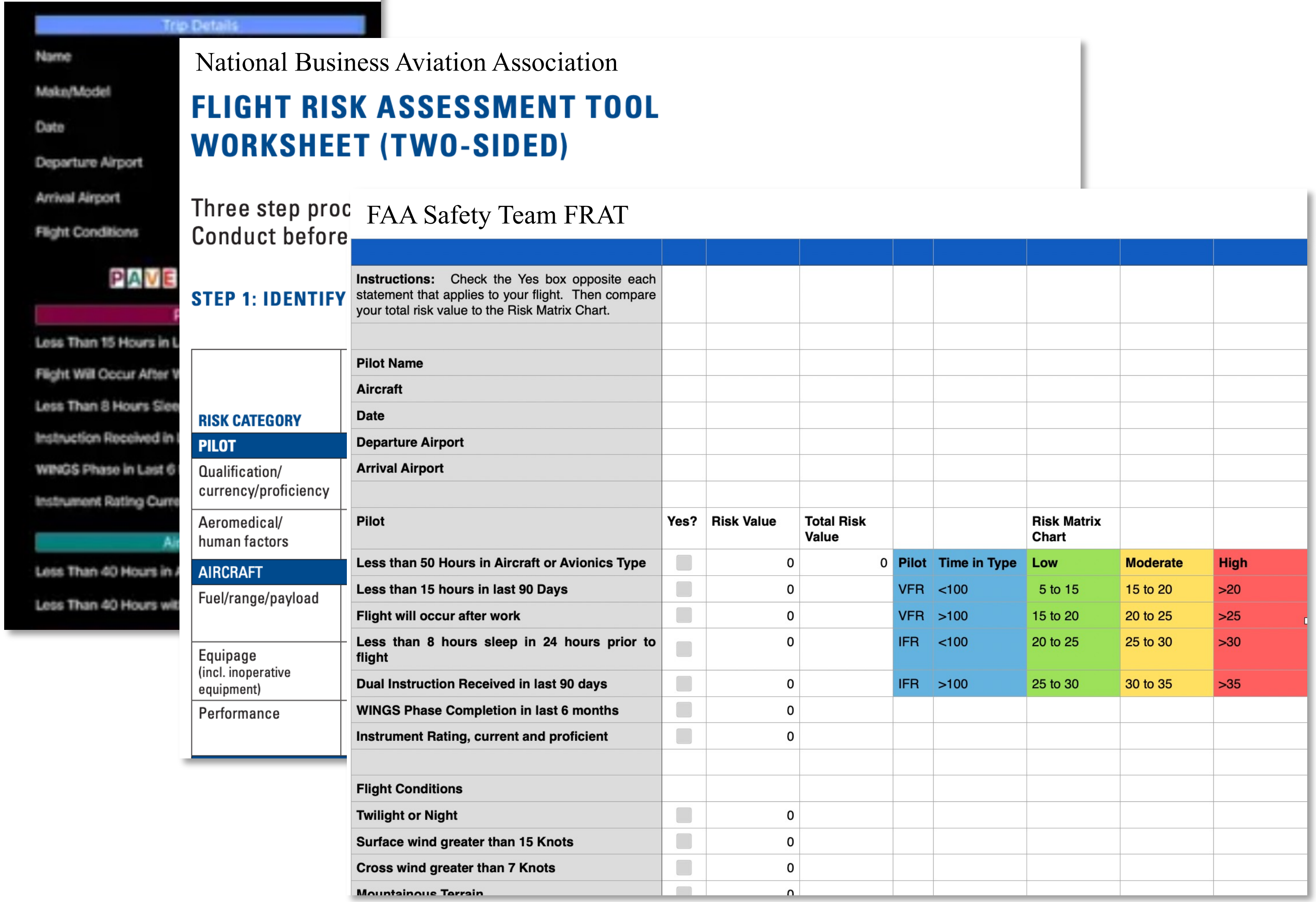
Available Flight Risk Assessment Tools (FRAT)

Though there are FRATs on the market, most include a score-based system based on pilot input which might be prone to bias.

ZDigitalPro, LLC FRAT App

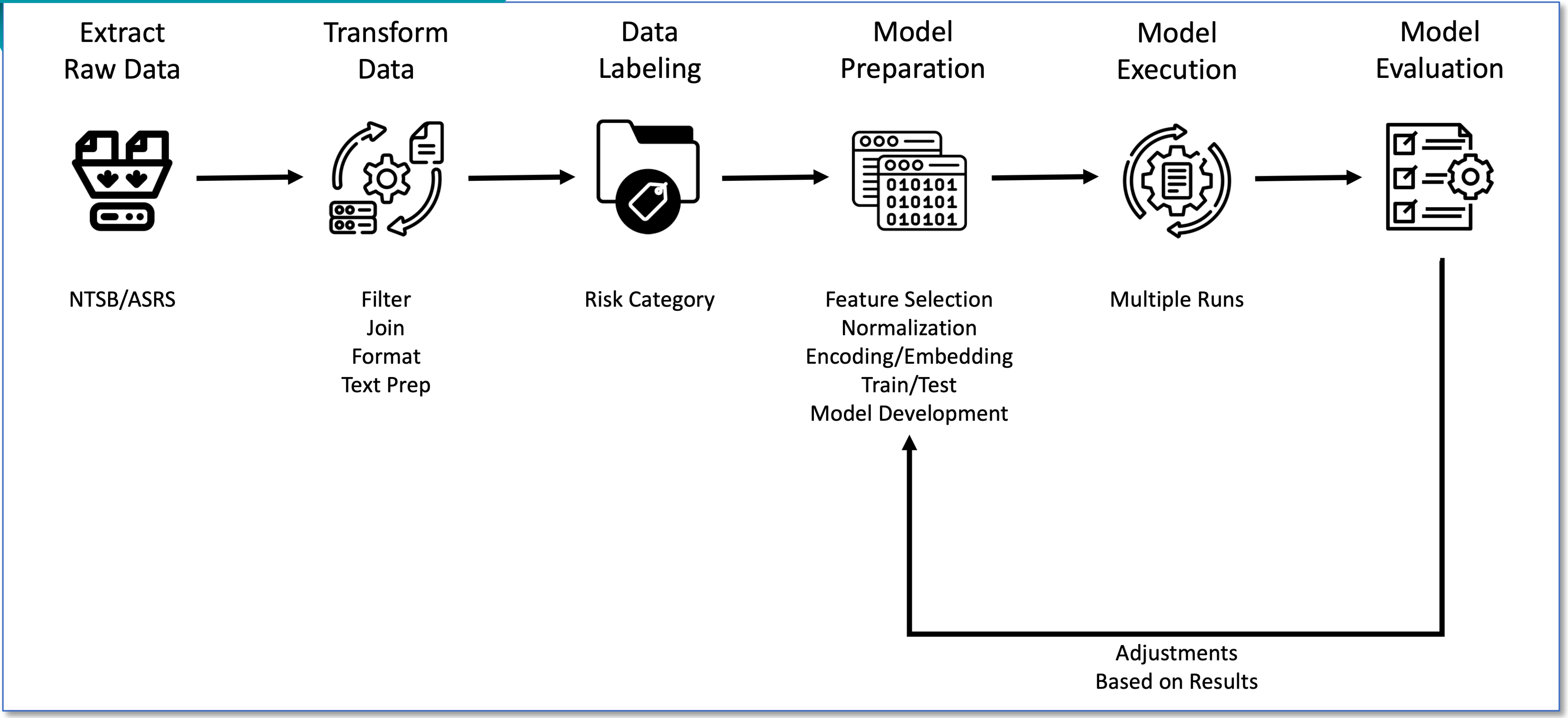
National Business Aviation Association

FAA Safety Team FRAT



Methodology

FRAT Project Modeling Methodology



GEORGE MASON UNIVERSITY

Data

NTSB National Transportation Safety Board

Number of Records: 46,052

Years: 1983-2022

Total Flight Hours
Actual Instrument Time
Flight Hours in Make
Flight Hours at Night
Rotary Wing Flight Hours
Single Engine Flight Hours
Multi-Engine Flight Hours
Simulated Instrument Hours

Dept/Dest Airport
Airport Location
Airport Elevation
Departure Time
Runway length
Runway width
Day of week
Month

Pilot | Aircraft

Feature Selection

Environment | Weather

Aircraft Make
Aircraft Model
Airframe Hours
Certified Max Gross Weight
Airframe Inspection Timeline
Engine hours
No. Engines
Engine Manufacturer

Temperature / Dewpoint
Wind Direction / Speed
Meteorological conditions
Ceilings
Visibility
Weather Observation Time
Light Conditions
Wind Gust Speed

Data Transformation

Labeling the Dataset: Risk Categories

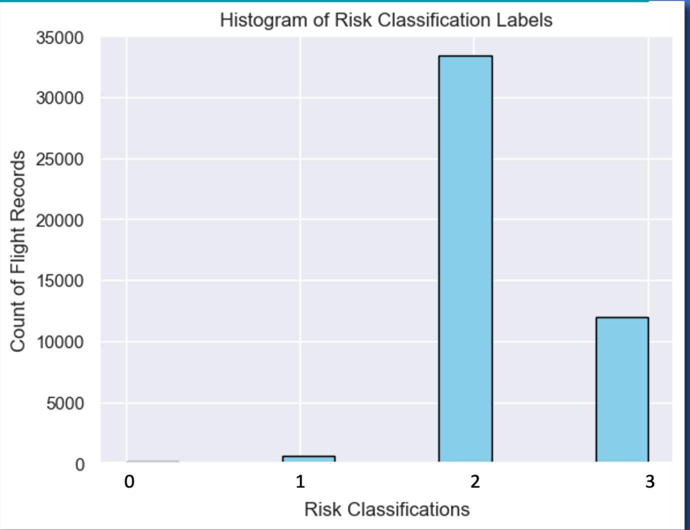
Proposed Risk Categories

Risk Score	Consequence Level	Description
0	Low	Flight resulted in an incident or accident with no aircraft damage AND no minor injury.
1	Medium	Flight resulted in an incident or accident with either minor aircraft damage OR minor injury, AND no fatality.
2	High	Flight resulted in an incident or accident with either severe aircraft damage OR severe injury, AND no fatality.
3	Catastrophic	Flight resulted in an incident or accident with either fatality or destruction of the aircraft.

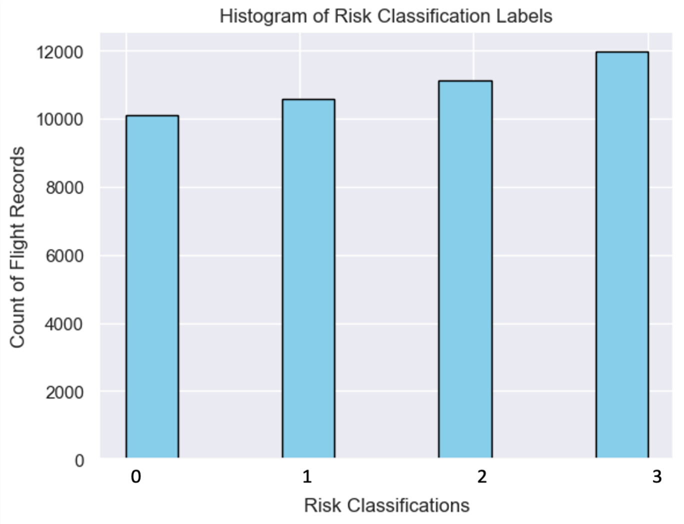
Note: When comparing aircraft damage and injury, the highest of the two dictates the classification. Numerical risk scores are only utilized for modeling purposes, and it is preferred that an end user see an output of “Low” or “High” rather than 0 or 2. The risk score of 0 does not imply a no-risk flight.

Balancing the Dataset: Resampling

Risk Classification Labels before resampling



Risk Classification Labels after resampling



Cleaning the Dataset

- Fixing random errors, binning, and formatting
- Standardizing Datatypes
- Label Encoding
- Handling Missing Data

Modeling

XGBoost

Random Forest

Feed-Forward Neural Network

BERT

Modeling

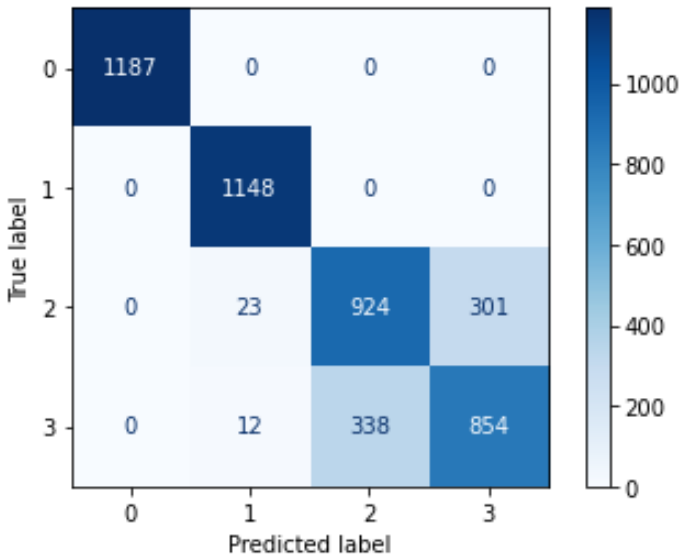
XGBoost

XGBoost Modeling Results

Test Iteration	Configuration	F1 Score Results
Baseline	Imbalanced	0.76
1	Resampled	0.85
2	Category reduction, imbalanced	0.77
3	Category reduction, resampled	0.79
4	Normalized, resampled	0.84
5	Category reduction, normalized, resampled	0.79
6	10-fold cross-validation, resampled*	0.85

Note: * indicates the preferred model configuration

Preferred Model Configuration Confusion Matrix



Process

Dataset Balancing

Model Configuration

Label Effects

10-Fold Cross Validation

Results

Overall, class 0 and 1 seemed overfit due to resampling techniques, but it improved the ability to more accurately predict class 2 and 3.

Modeling

Random Forest

Process

Dataset Balancing

Model Configuration

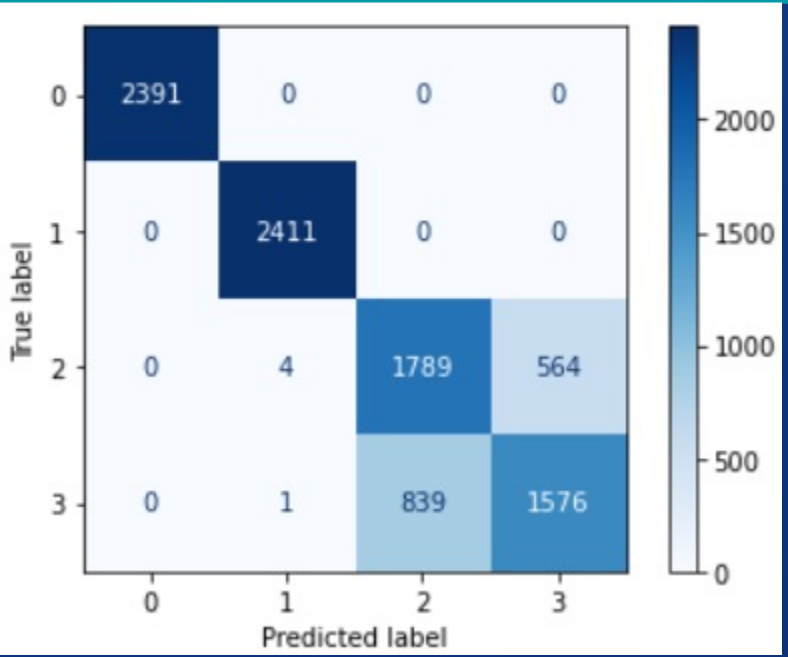
10-Fold Cross Validation

Random Forest Modeling Results

Test Iteration	Configuration	F1 Score Results
Baseline	Imbalanced	0.75
1	Resampled	0.85
2	Resampled, 10-fold cross validation*	0.86

Note: * indicates the preferred model configuration

Preferred Model Configuration Confusion Matrix



Results

- 10-fold cross validation returned the highest F1 score – 0.86
- Model correctly predicted highest risk classifications

Modeling

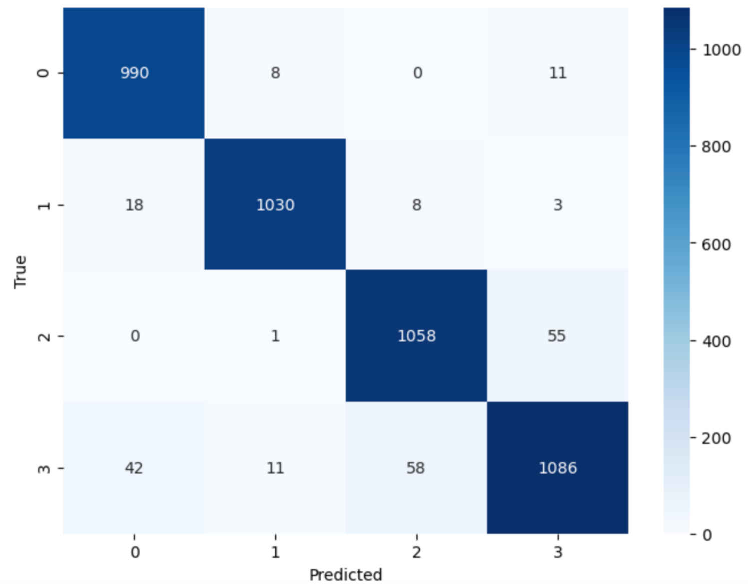
Feed-Forward Neural Network

Neural Network Modeling Results

Test Iteration	Configuration	F1 Score Results
Baseline	Imbalanced, whole data set imputation	0.67
1	Imbalanced, class-based imputation*	0.60
2	Resampled, class-based imputation*	0.60
3	Resampled, whole data set imputation	0.89
4	Resampled, whole data set imputation, 10-fold cross validated using KerasClassifier	0.90
5	Resampled, whole data set imputation, 10-fold cross validated using StratifiedKFold**	0.98

Notes: * This method for imputation introduced error and was not used moving forward
** indicates the preferred model configuration

Preferred Model Configuration Confusion Matrix



Results

- While imputation by class label seems ideal, the process utilized introduced error.
- The resampled dataset utilizing whole dataset imputation, validated using StratifiedKFold 10-fold cross validation resulted in the most accurate predictions.

Process

Dataset Balancing

Missing Value Imputation

Model Configuration

10-Fold Cross Validation

Modeling

BERT

Process

Label Resampling

Data Concatenation

Embeddings

Hyper-parameter Tuning

Evaluation

Bert Modeling Results

Test Iteration	Configuration	F1 Score Results
1	Data with context	1.0
2	Data without context	1.0

Bert Modeling Training & Validation Loss

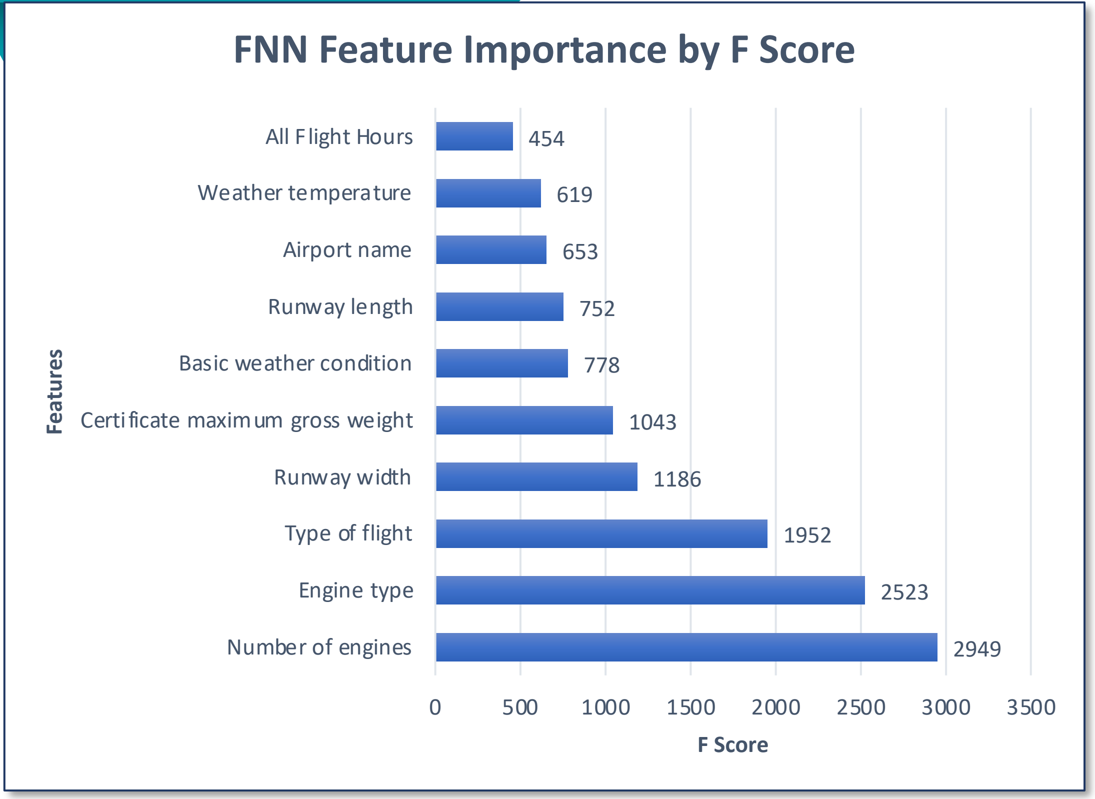
Results

- Although the training and validation loss curves converge nicely, such extremely high accuracy and F1 scores suggest overfitting and the inability for the model to generalize.
- Further research is necessary to determine the cause of these results.

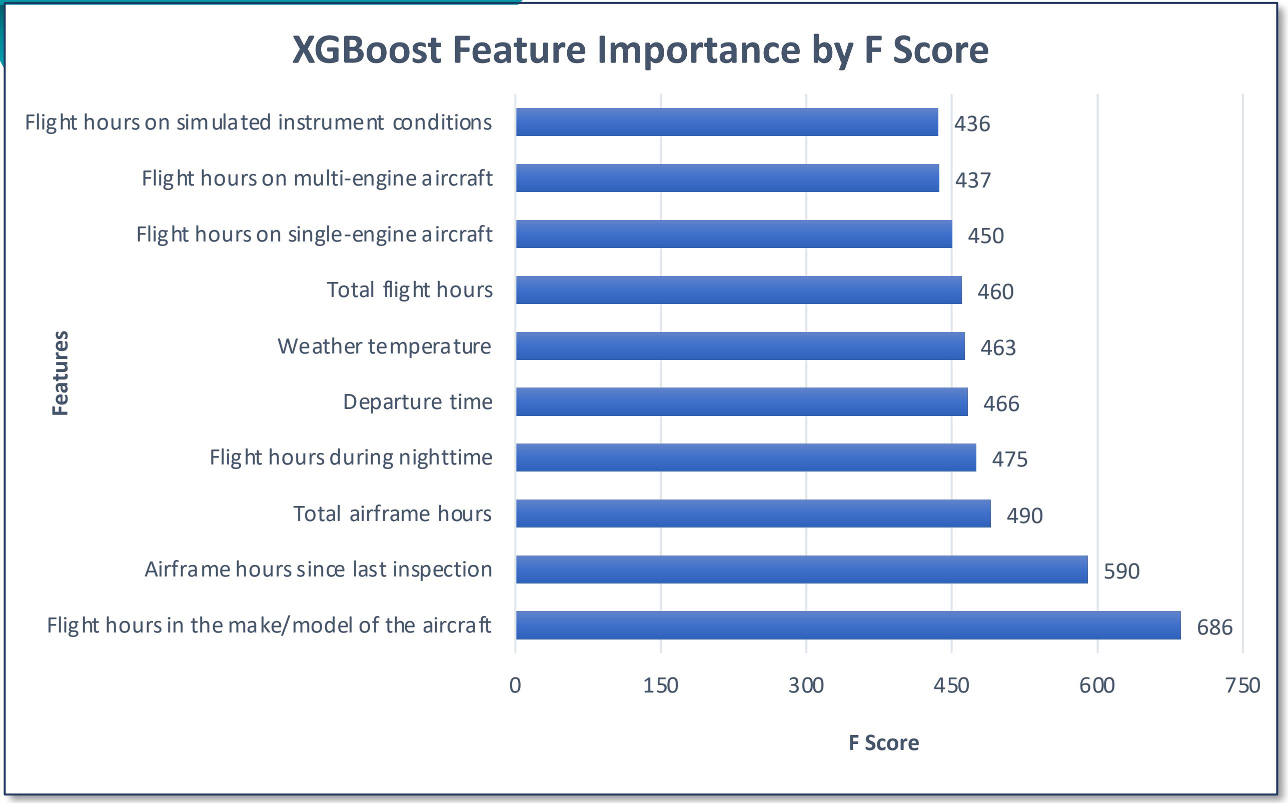
Modeling

Feature Importance

Feed-Forward Neural Network



XGBoost



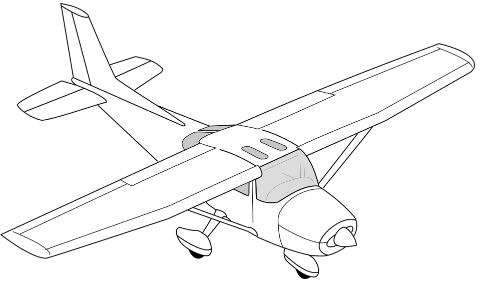
Model Performance by F1 Scores

- FNN
- XGBoost / Random Forest
- BERT

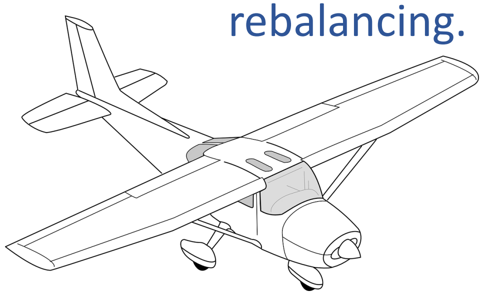
Future Work

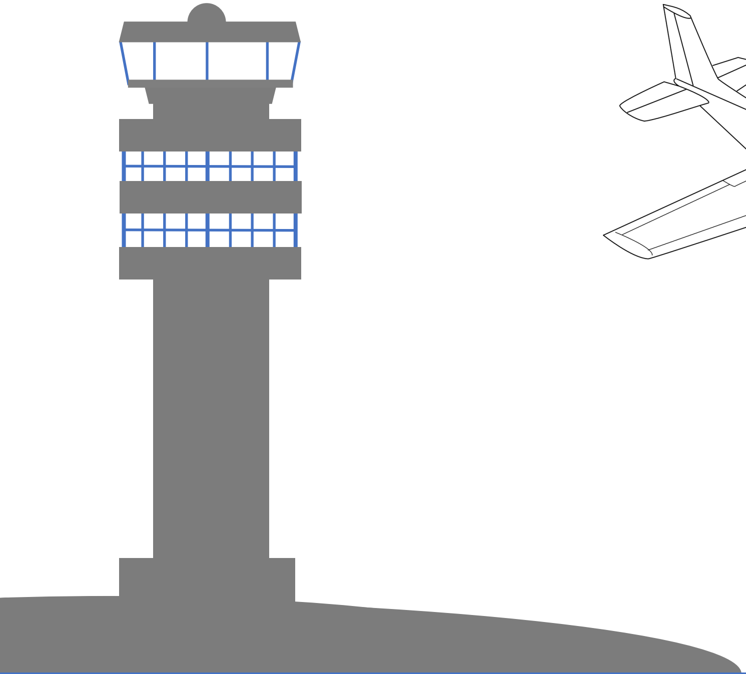
Way Forward

1 | Further development, testing, and evaluation of methods and high performing models, along with the incorporation of improved missing value imputation methodology for neural networks.

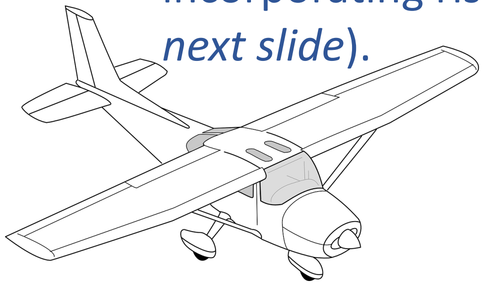


2 | Expansion of training dataset to include low risk flight records based on voluntary general aviation pilot participation. Reduces bias in model by reducing need for rebalancing.



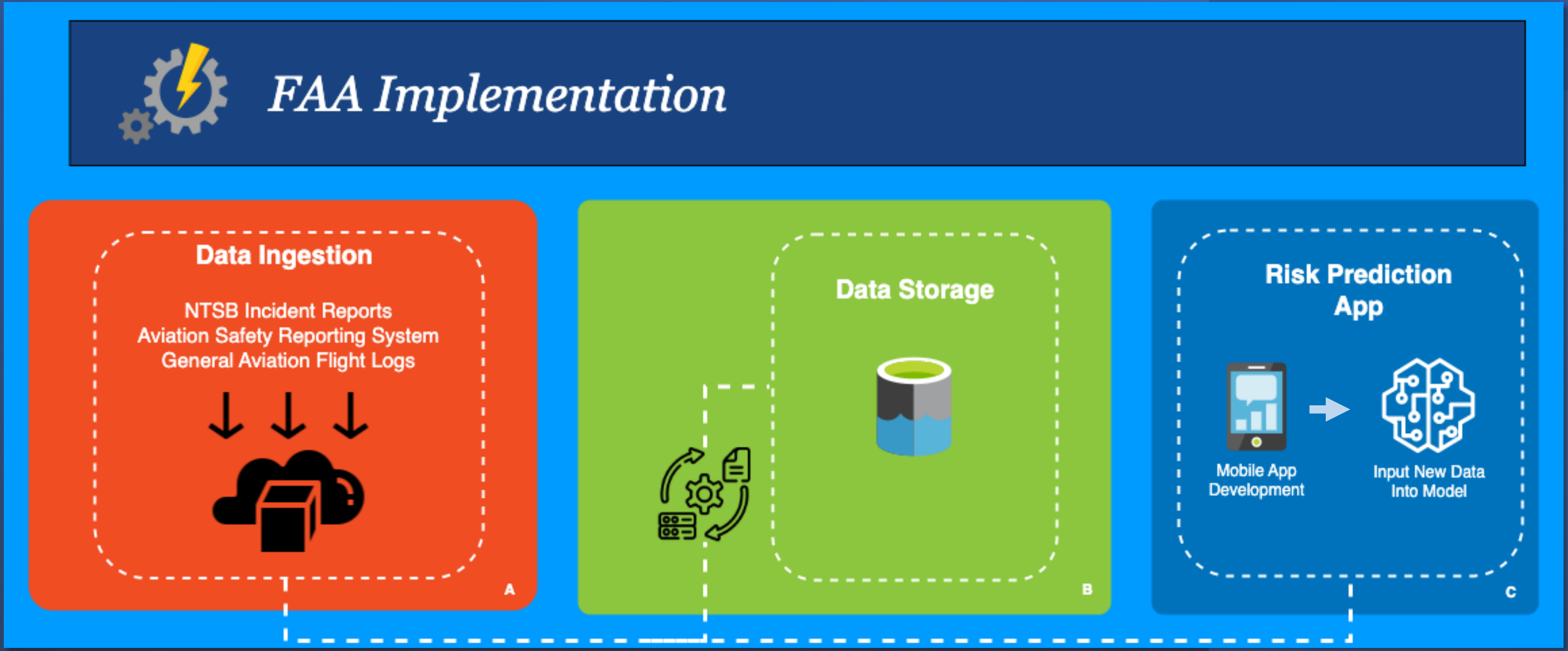


3 | Deploy the optimal model through development of an application for the ingestion of new flight records, incorporating risk assessment class assignment (*see next slide*).



Next Steps

The Way Forward



Questions

References

1. Boyd, D. D. (2017). A Review of General Aviation Safety (1984–2017). Aerospace Medicine and Human Performance, 88(7), 657–664. https://doi.org/10.3357/amhp.4862.2017
2. Flaticon. (2023a, May 30). *Airplane Icon - 561698*. https://www.flaticon.com/free-icons/airport
3. Flaticon. (2023b, May 30). *Data modelling Icon - 8329485*. https://www.flaticon.com/free-icons/predictive-modeling
4. Flaticon. (2023c, May 30). *Digital Transformation Icon - 9768191*. https://www.flaticon.com/free-icons/digital-transformation
5. Flaticon. (2023d, May 30). *Etl Icon - 5708950*. https://www.flaticon.com/free-icons/etl
6. Flaticon. (2023e, May 30). *Evaluation Icon - 6757112*. https://www.flaticon.com/free-icons/risk-management
7. Flaticon. (2023f, May 30). *Plane Icon - 723955*. https://www.flaticon.com/free-icons/plane
8. Fultz, A. J., & Ashley, W. S. (2016). Fatal weather-related general aviation accidents in the United States. Physical Geography, 37(5), 291–312. https://doi.org/10.1080/ 02723646.2016.1211854
9. National Business Aviation Association. (n.d.). *Flight Risk Assessment Tool*. https://nbaa.org/wp-content/uploads/2018/06/flight-risk-assessment-tool.pdf
10. *Resources - Library Contents - FAA - FAASTeam - FAASafety.gov*. (n.d.). https://www.faasafety.gov/gslac/ALC/libview_normal.aspx?id=352396
11. Wright, R. (2019, October 29). Risk assessment tools. Aviation Safety. Retrieved January 11, 2023, from https://www.aviationsafetymagazine.com/features/risk-assessment-tools/
12. ZDigitalPro,LLC. (2017, April 18). *Flight Risk Assessment Tool*. https://apps.apple.com/us/app/flight-risk-assessment-tool/id1080401103?platform=iphone