ML in Aviation
How to improve safety

Robert Reynoso

Summary

- Background in Aviation (aeronautics)
- ML in Aviation business cases
    - Forecasting (Ticket sales, profit, etc.)
- What about safety?

Outline

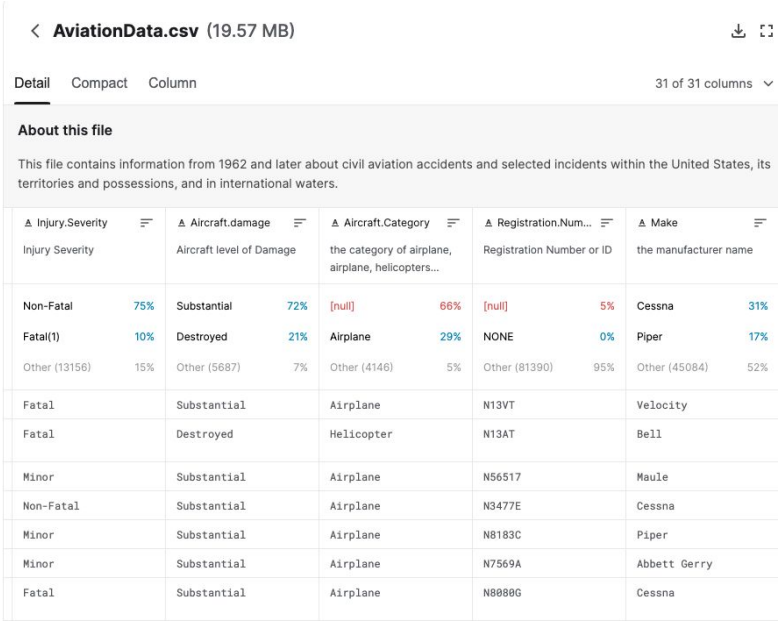
1. Business Problem
2. Data
3. Results
4. Conclusions

Business Problem

Using machine learning (ML) specifically supervised learning.

Can we use classification algorithms to classify past aviation accidents as fatal or non-fatal based on features of that accident?

Data



AviationData.csv (19.57 MB)

Detail Compact Column 31 of 31 columns

About this file

This file 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.

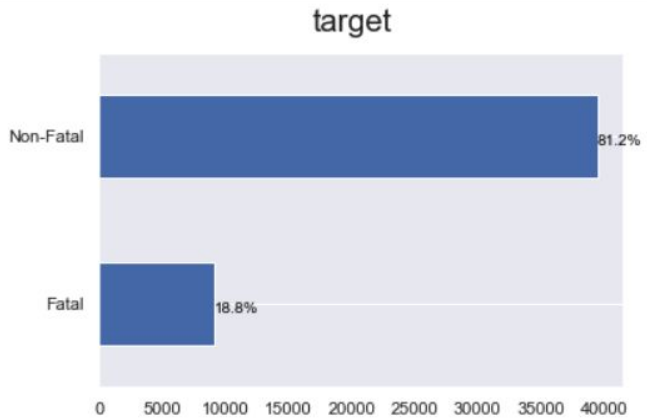
Injury.Severity Injury Severity		Aircraft.damage Aircraft level of Damage		Aircraft.Category the category of airplane, airplane, helicopters...		Registration.Num... Registration Number or ID		Make the manufacturer name	
Non-Fatal	75%	Substantial	72%	[null]	66%	[null]	5%	Cessna	31%
Fatal(1)	10%	Destroyed	21%	Airplane	29%	NONE	0%	Piper	17%
Other (13156)	15%	Other (5687)	7%	Other (4146)	5%	Other (81390)	95%	Other (45084)	52%
Fatal		Substantial		Airplane		N13VT		Velocity	
Fatal		Destroyed		Helicopter		N13AT		Bell	
Minor		Substantial		Airplane		N56517		Maule	
Non-Fatal		Substantial		Airplane		N3477E		Cessna	
Minor		Substantial		Airplane		N8183C		Piper	
Minor		Substantial		Airplane		N7569A		Abbett Gerry	
Fatal		Substantial		Airplane		N8080G		Cessna	

- The NTSB has a database from accidents since 1962
- Pulled from Kaggle

Final shape (4867,17)

```
data_2.columns
```

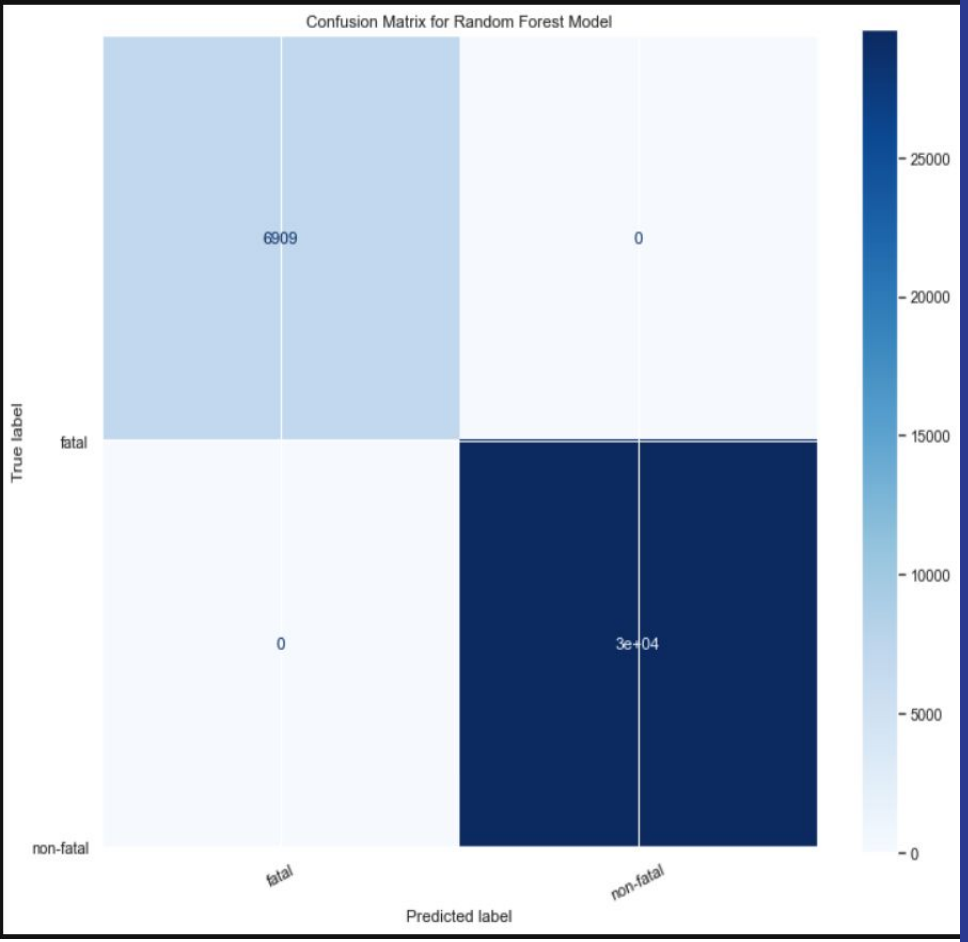
```
Index(['Unnamed: 0', 'target', 'location', 'aircraft_damage', 'make', 'model',
       'number_of_engines', 'engine_type', 'weather_conditions',
       'phase_of_flight', 'Year', 'Month', 'Day', 'injuries', 'pax_onboard',
       'fatality_percentage', 'survived', 'amateur_built'],
      dtype='object')
```



Initial Results

results

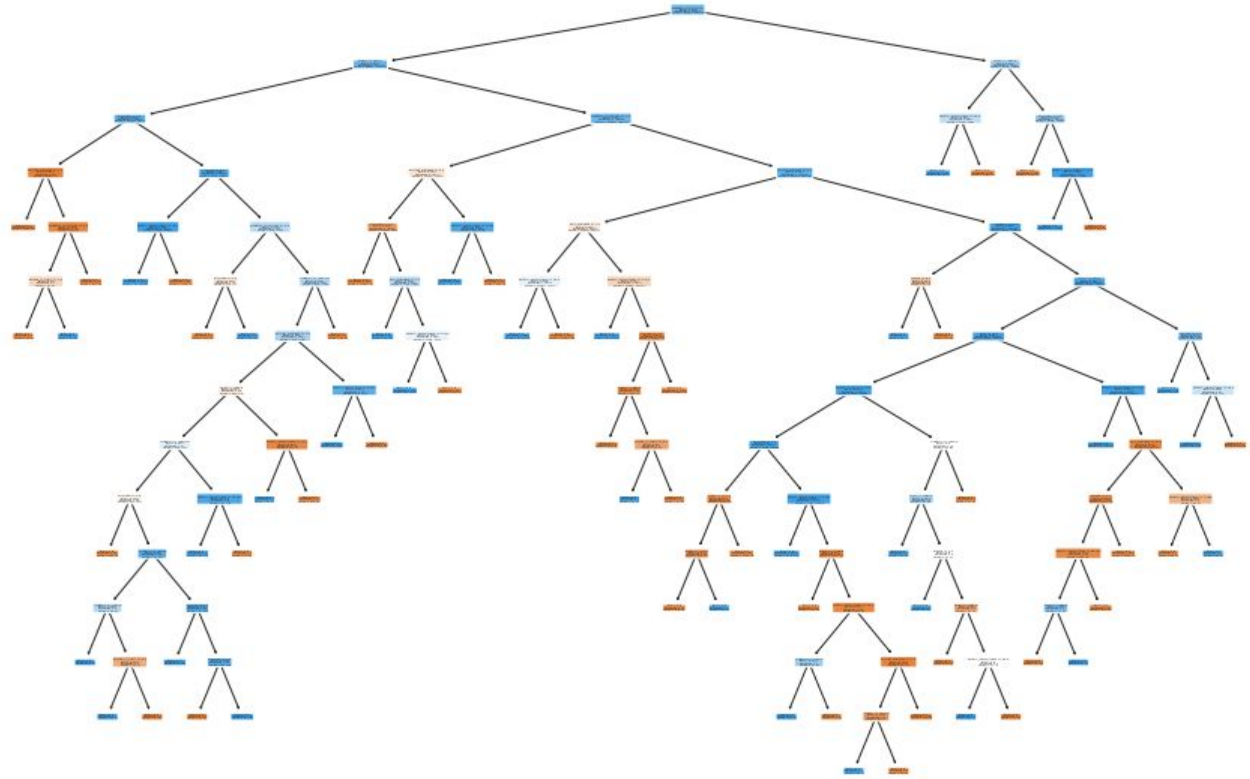
	model	train_precision	train_recall	train_accuracy	train_f1	train_time	test_precision	test_recall	test_accuracy	test_f1	test_time
0	knn	1.000000	1.000000	1.000000	1.000000	0.023560	0.923077	0.957389	0.900148	0.939920	4.634440
1	logistic_regression	0.998683	0.999662	0.998658	0.999173	2.625757	0.998591	0.999597	0.998521	0.999094	0.001482
2	decision_tree	1.000000	1.000000	1.000000	1.000000	0.058102	0.999597	0.999799	0.999507	0.999698	0.002410
3	random_forest	1.000000	1.000000	1.000000	1.000000	0.937998	1.000000	0.999799	0.999836	0.999899	0.043773
4	naive_bayes	0.987394	0.659053	0.716763	0.790484	0.023357	0.988198	0.657903	0.714497	0.789913	0.004797
5	adaboost	1.000000	1.000000	1.000000	1.000000	0.900672	0.999799	0.999799	0.999671	0.999799	0.041151
6	gradient_boosting	1.000000	0.999899	0.999918	0.999949	3.212887	0.999899	0.999799	0.999753	0.999849	0.011706



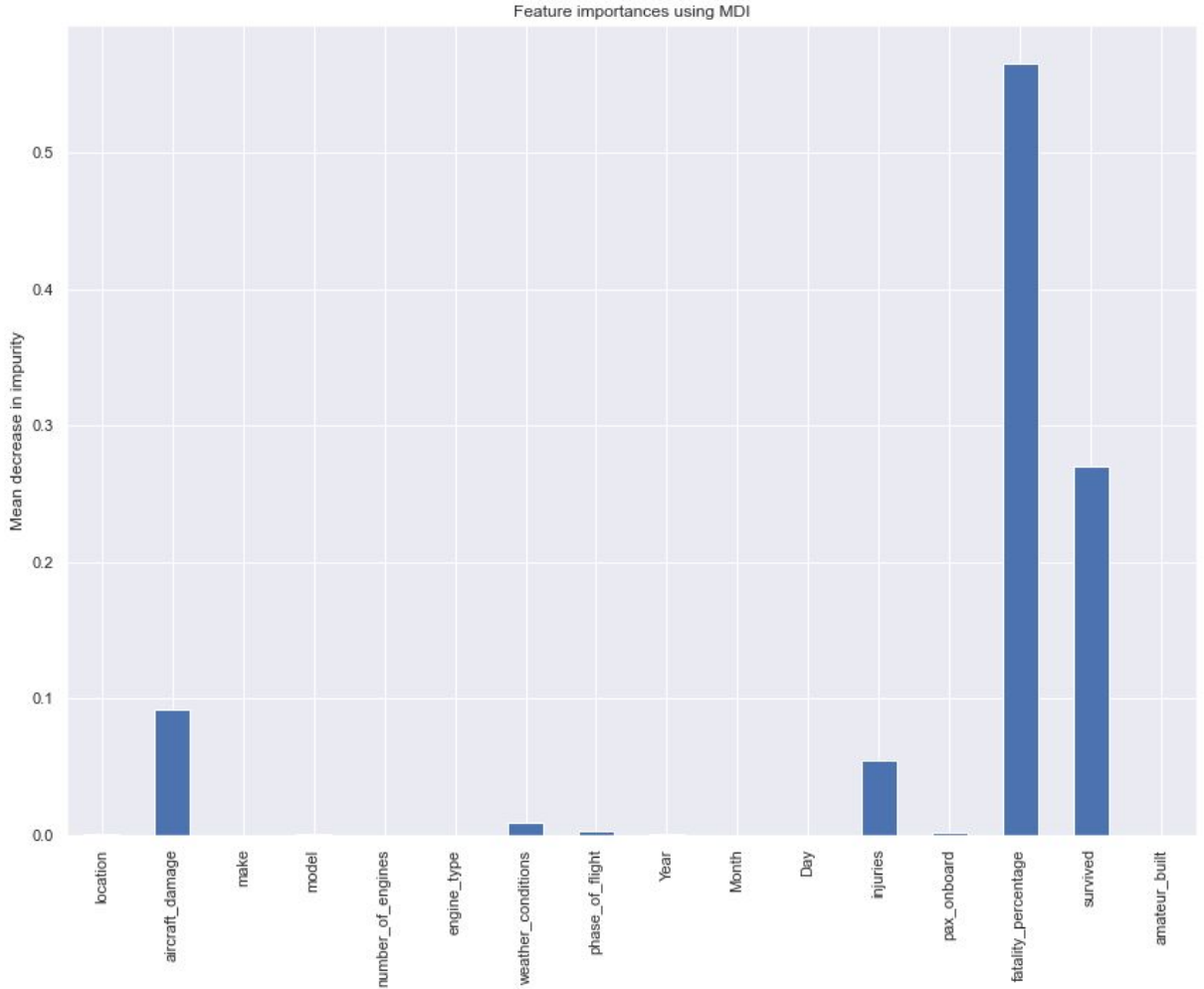
Analysis Random forest

```
{'model': 'random_forest',
 'train_precision': 1.0,
 'train_recall': 1.0,
 'train_accuracy': 1.0,
 'train_f1': 1.0,
 'train_time': 0.9320237636566162,
 'test_precision': 1.0,
 'test_recall': 0.9997985292636244,
 'test_accuracy': 0.9998356344510191,
 'test_f1': 0.9998992544831755,
 'test_time': 0.045164823532104 49}
```

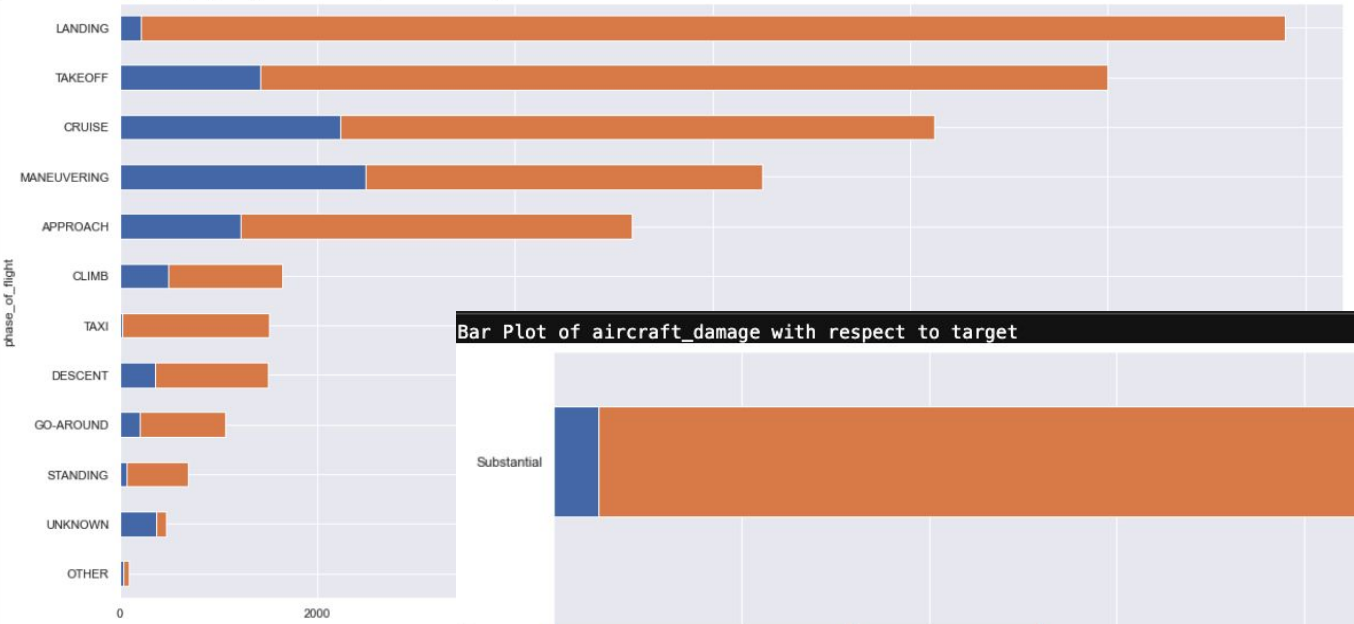
Visualization of my RF model



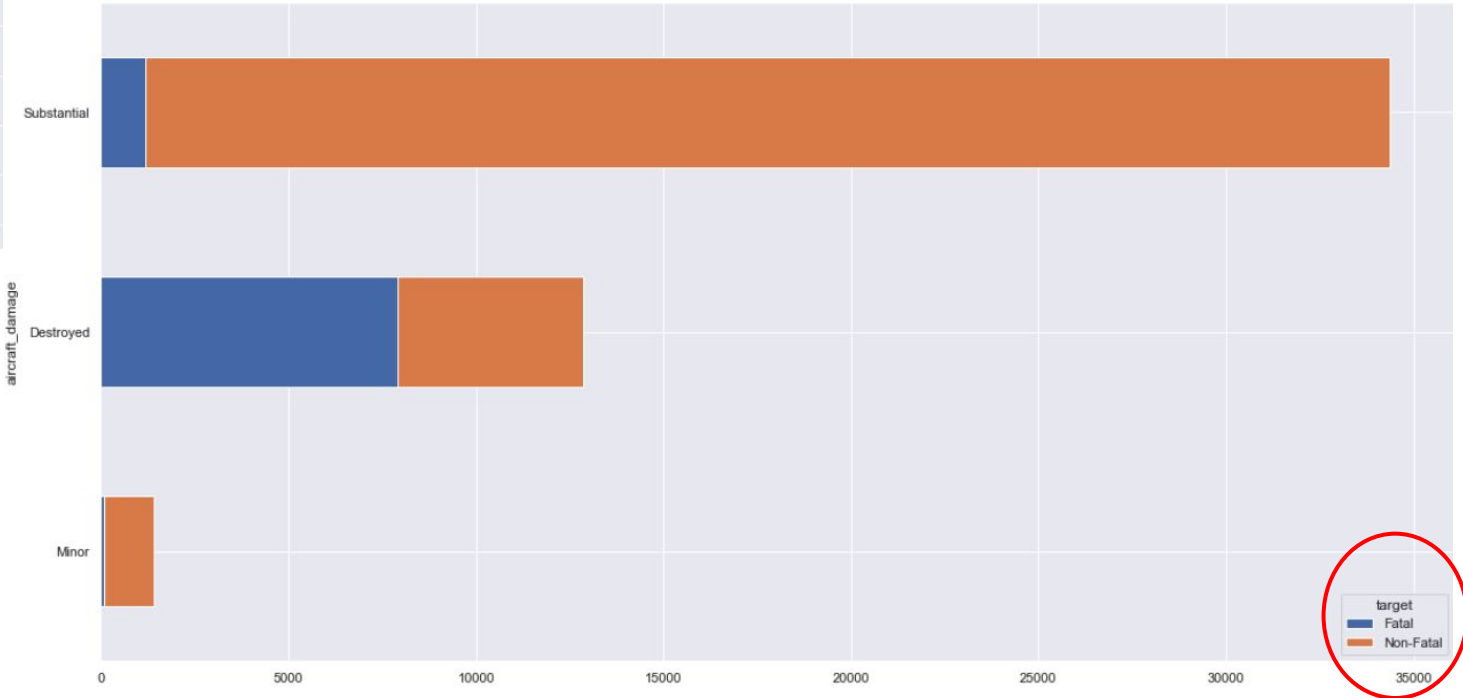
Model evaluation & improving



Bar Plot of phase_of_flight with respect to target



Bar Plot of aircraft_damage with respect to target



Conclusions

1. Successfully ran 7 vanilla ML models to learn how we can improve aviation safety.

2. The above models returned acceptable model performance.

3. Through initial classification modeling, we learned which features are important in classifying a fatal or non-fatal aviation accident.

4. Although we can classify a accident with good model performance. Further investigation and feature engineering is required on the fatality_percentage feature.

Next Steps

1. Find or create more data specifically in the aircraft_damage category
2. Use imputation to replace any unknown data
3. Productionize model with prediction function
4. Look into multiclass classification

Email: robert@birdstop.io

GitHub: @reynoso811

LinkedIn:
linkedin.com/in/rreynoso8

Thank You!