Analysis of SpaceX Rocket Launch

Sumit Kaware

Agenda

- Executive Summary
- Table of Contents
- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix

Executive Summary

We will be working with SpaceX launch data that is gathered from an API, specifically the SpaceX REST API. This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome. Our goal is to use this data to predict whether SpaceX will attempt to land a rocket or not.

Preparing the data for a machine learning model that will predict if the first stage will successfully land.

Using the best hyperparameter values, we will determine the model with the best accuracy using the training data. We will test Logistic Regression, Support Vector machines, Decision Tree Classifier, and K-nearest neighbors.

Introduction

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage.

Problem:

Space Y wants to compete with Space X

Space Y tasks us to train a machine learning model to predict successful Stage 1 recovery

Methodology

Data Collection And Wrangling

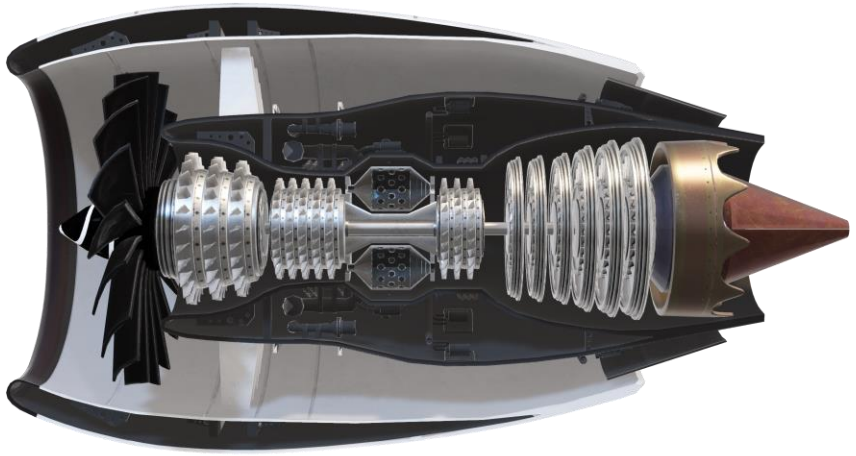
EDA And Interactive Visual Analytics

Predictive Analysis

Data Collection And Wrangling

SPACEX API

- Request to the SpaceX API
- Clean the requested data



WEB SCRAPING WIKIPEDIA

- Web scraping to collect falcon 9 historical launch records from a Wikipedia page titled `List of Falcon 9 and Falcon Heavy launches`

Data Collection And Wrangling

DATA WRANGLING

- There are several different cases where the booster did not land successfully. Sometimes a landing was attempted but failed due to an accident
- True Ocean means successfully landed to a specific region of the ocean
- False Ocean means unsuccessfully landed to a specific region of the ocean.
- True RTLS means successfully landed to a ground pad
- False RTLS means unsuccessfully landed to a ground pad.
- True ASDS means successfully landed on a drone ship
- False ASDS means unsuccessfully landed on a drone ship.
- Convert those outcomes into training labels with `1` means the booster successfully landed `0` means it was unsuccessful.

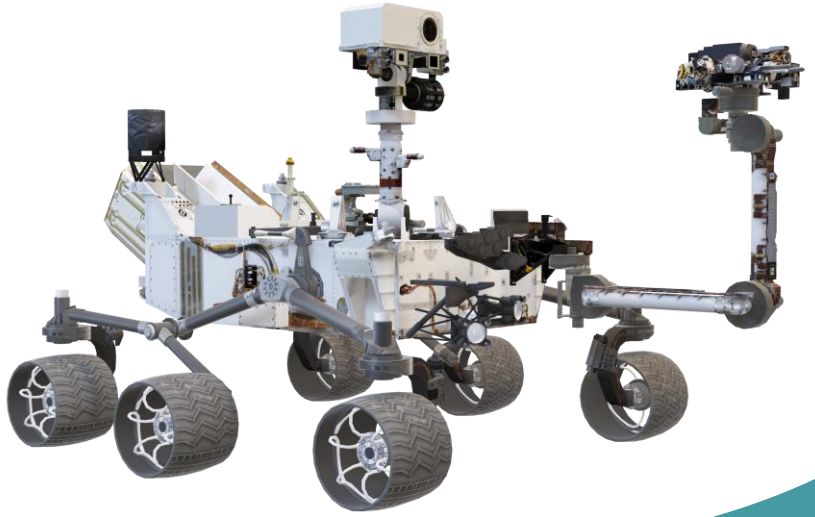
EDA And Interactive Visual Analytics

EDA WITH VISUALIZATION

- Relationship between flight number and launch site
- Relationship between payload and launch site
- Relationship between success rate of each orbit type
- Relationship between flight number and orbit type
- Relationship between payload and orbit type
- Launch success yearly trend

EDA WITH SQL

- Understand the SpaceX Dataset
- Load the dataset into the corresponding table in a Db2 database
- Execute SQL queries to answer assignment questions



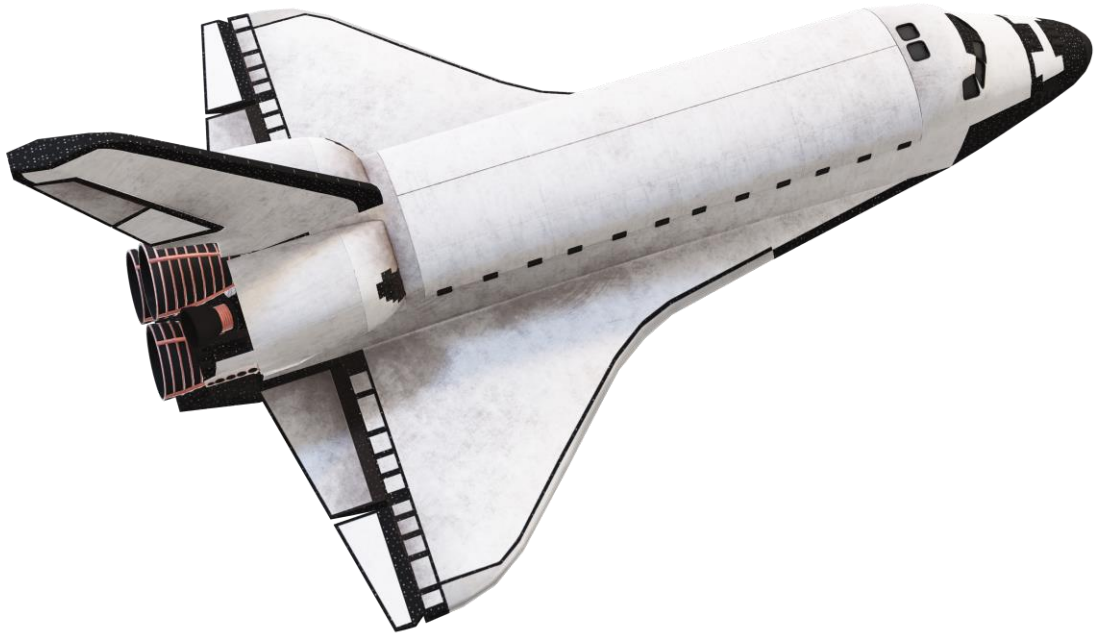
Predictive Analysis

OBJECTIVES

- Perform exploratory data analysis and determine training labels

1. Create a column for the class
2. Standardize the data
3. Split into training data and test data

- Find best hyperparameter for SVM, classification trees and logistic regression
- Find the method performs best using test data



Results

EDA With Visualization

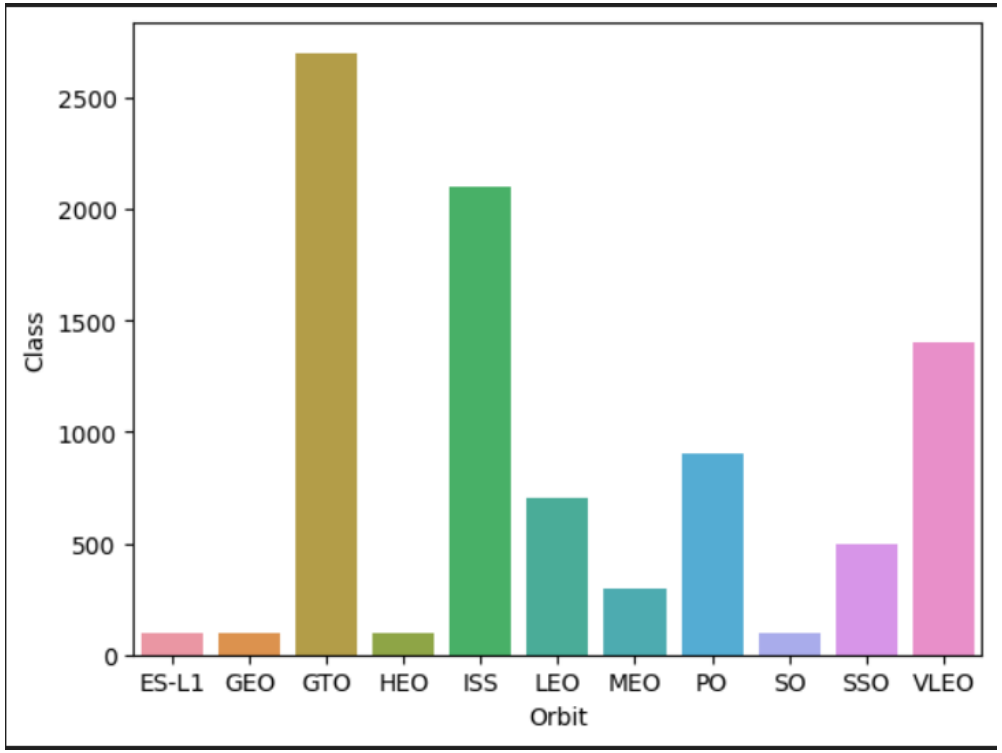
EDA With SQL Results

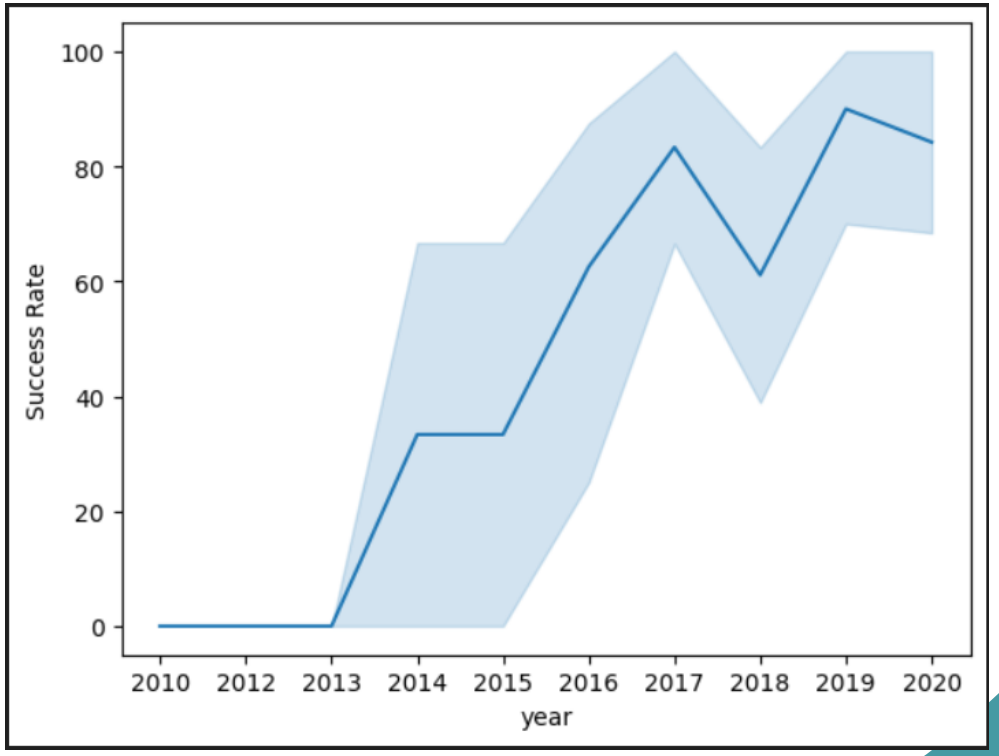
Interactive Map With Folium And Plotly Dash

Predictive Analysis

Result

EDA WITH VISUALIZATION





Result

EDA WITH SQL RESULTS

```
[17]
%%sql
select DISTINCT Launch_Site from SPACEXTBL;
```

```
* sqlite:///my_data1.db
Done.
```

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40
None

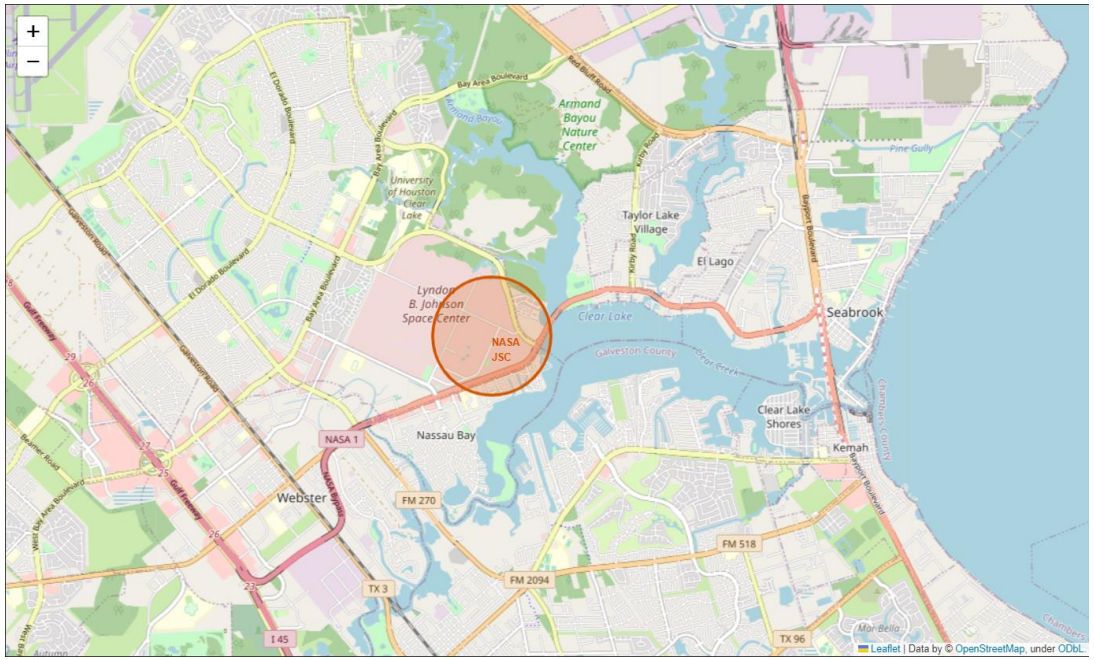
```
[63]
%%sql
select Booster_Version from SPACEXTBL where Landing_Outcome =
'Success (drone ship)' AND PAYLOAD_MASS__KG_ Between 4000 and 6000;
```

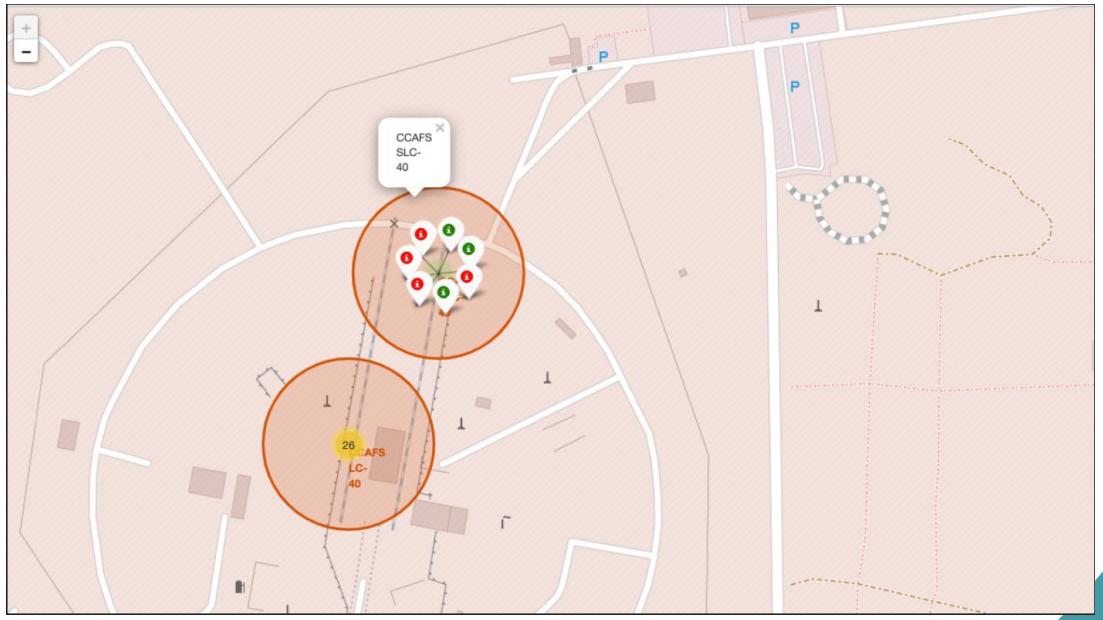
```
* sqlite:///my_data1.db
Done.
```

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Result

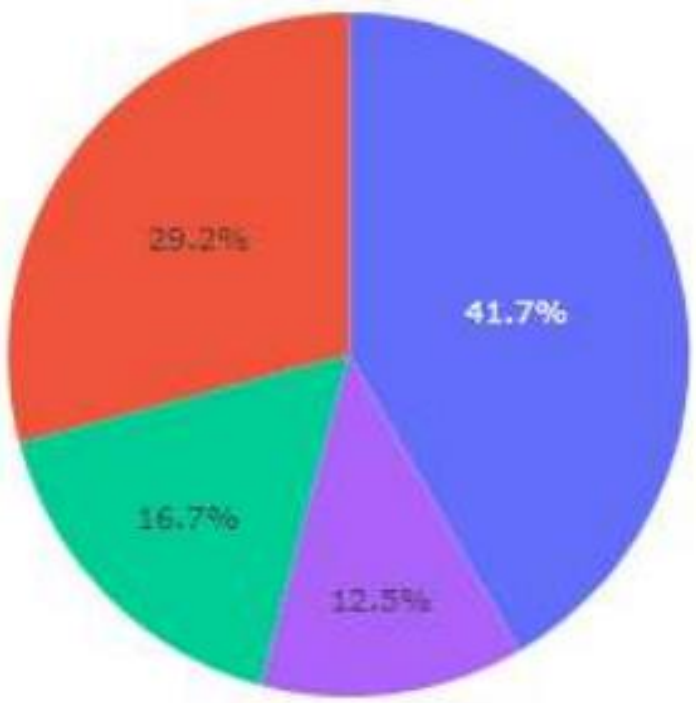
INTERACTIVE MAP WITH FOLIUM

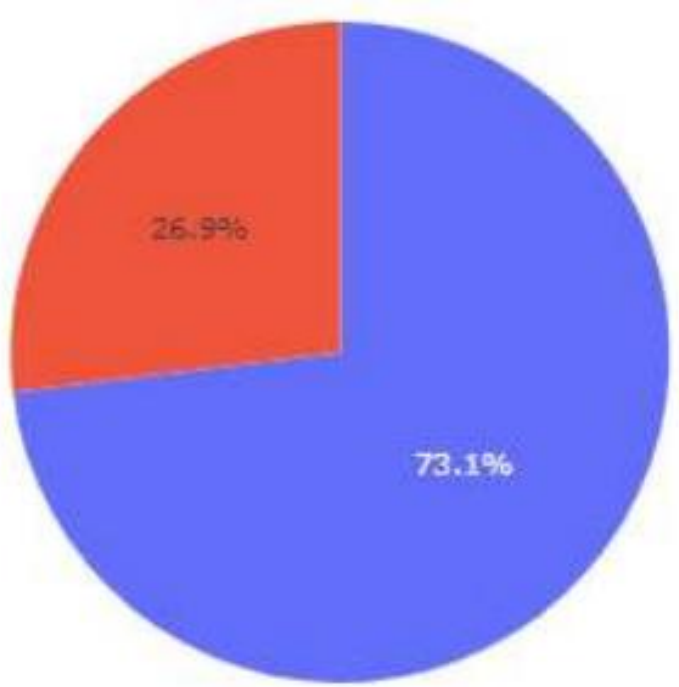




Result

INTERACTIVE MAP WITH
PLOTLY DASH





Result

PREDICTIVE ANALYSIS

GridSearchCV

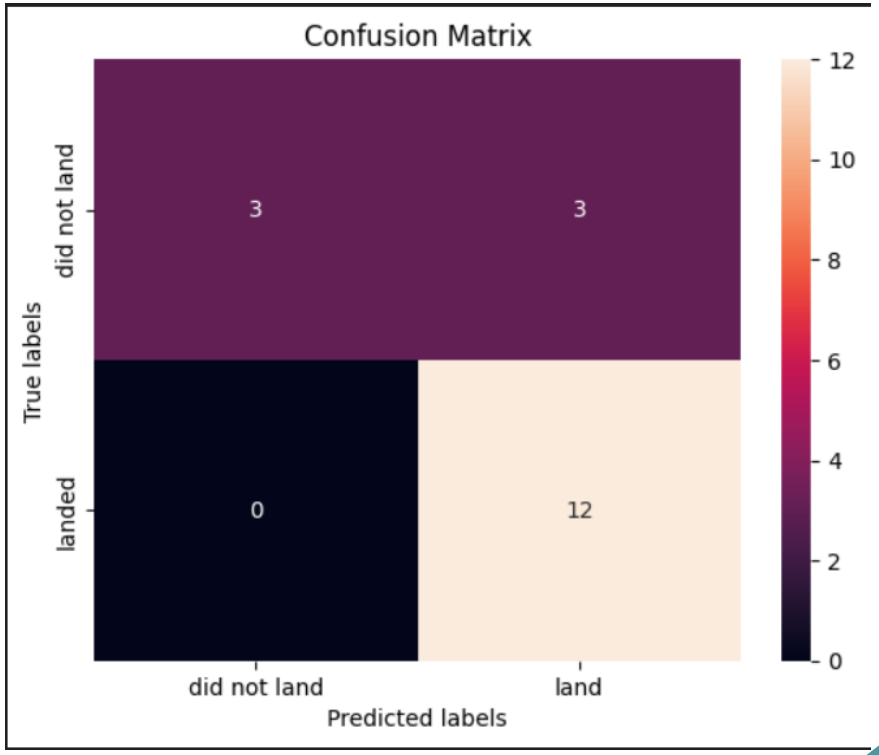
```
GridSearchCV(cv=10, estimator=LogisticRegression(),
             param_grid={'C': [0.01, 0.1, 1], 'penalty': ['l2'],
                         'solver': ['lbfgs']})
```

estimator: LogisticRegression

```
LogisticRegression()
```

LogisticRegression

```
LogisticRegression()
```



Discussion

Success generally increase over the time since 2013 and in recent it is around 80%.

Launch sites are close to coast and relatively far from cities so that the launch failures can land in sea and doesn't cause any damage to cities.

All models have nearly same accuracy about 83.33%.

Conclusion

We have used public SpaceX data, formatted it to get some insights.

Created visual charts, graphs and dashboards to get better understanding of the data which has increased our knowledge relate to landing of rockets.

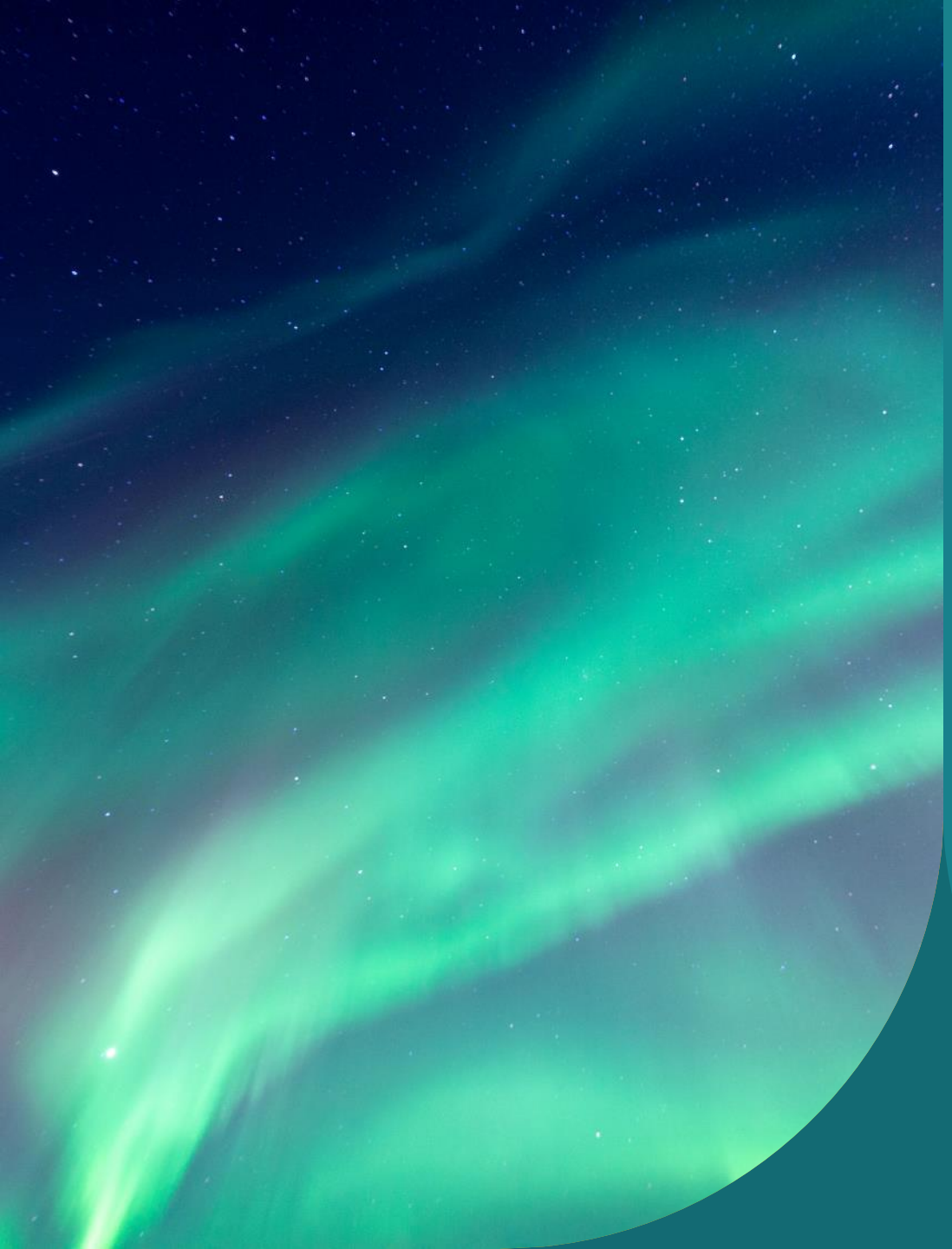
We have build a machine learning model with accuracy of 83.33%.

Allon Mask of SpaceY and use this findings to determine whether the launch will have a successful stage 1 landing.

Appendix

GitHub repository url:

https://github.com/SumitKaware/Applied-Data-Science-Capstone.git



Thank You

Sumit Kaware