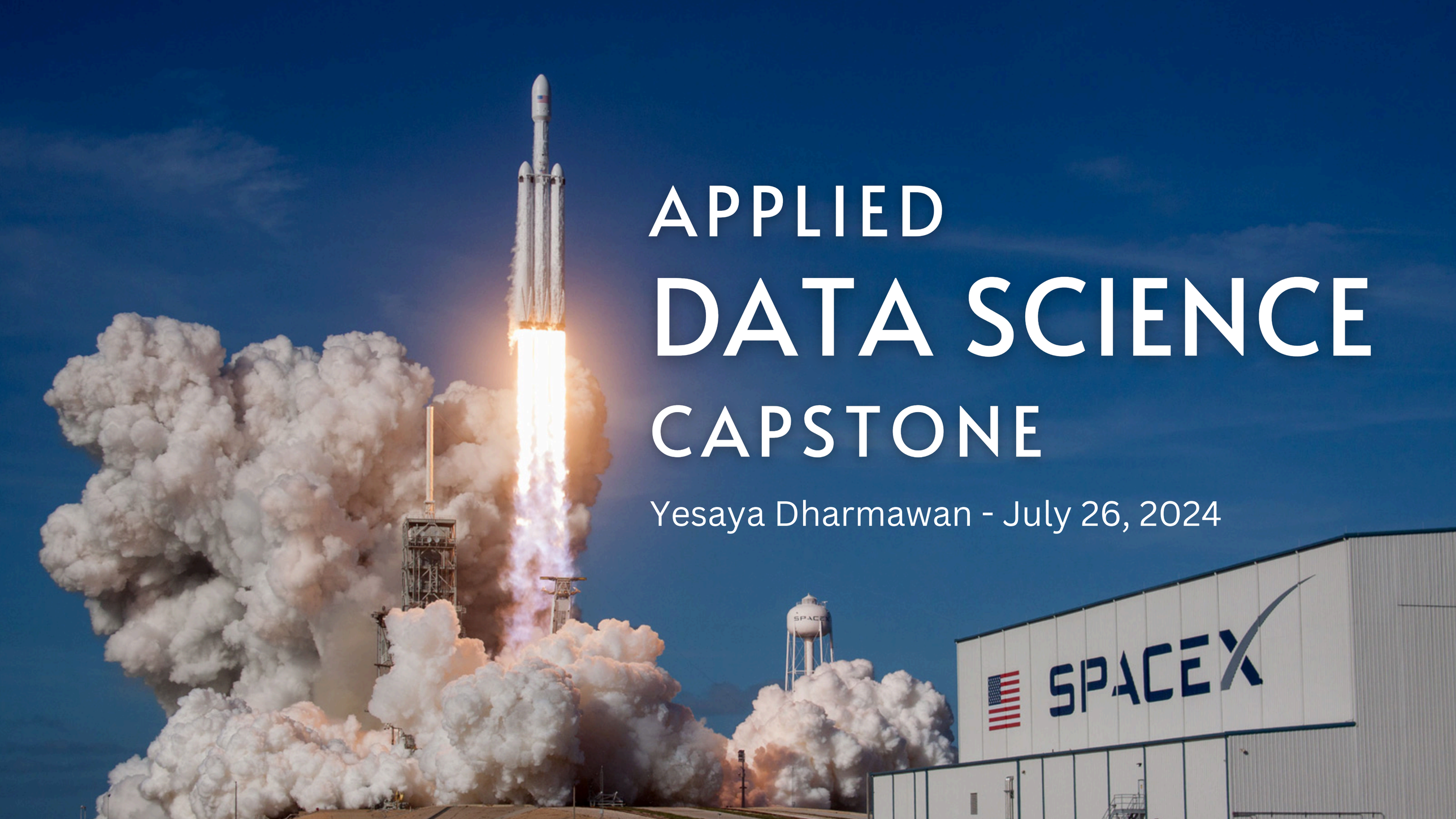


APPLIED DATA SCIENCE CAPSTONE

Yesaya Dharmawan - July 26, 2024

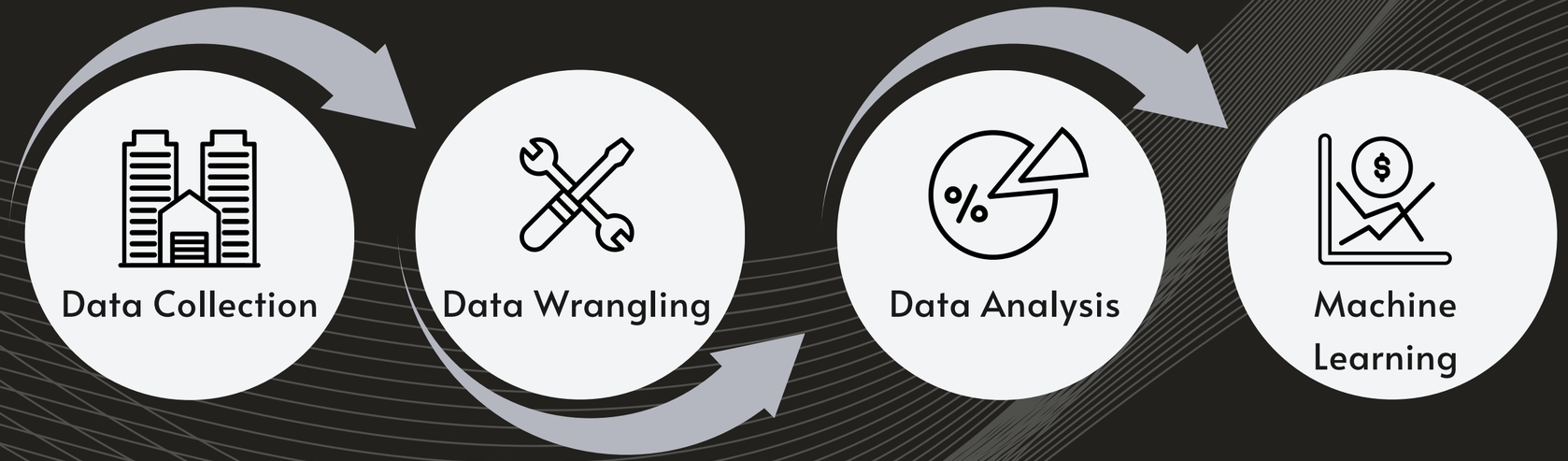
Overview

- Executive Summary
- Introduction
- Methodology
- Conclusion

Executive Summary

In this capstone project, we want to predict Space X Falcon 9 First Stage Landing result using several machine learning method.

Before we get into the machine learning, we must understand the data with following steps:



We also found out that some features have correlation to the outcome of the launches and we also found that the launch site is design for specific weightload

Decision Tree is the best algorithm to predict the landing result

Introduction

Space X advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because Space X can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against space X for a rocket launch.

Methodology

Data collection, wrangling, and formatting

- SpaceX API
- Web Scraping

Exploratory data analysis (EDA)

- Pandas and NumPy
- SQL

Data visualization

- Matplotlib and Seaborn
- Folium
- Dash

Machine learning prediction

- Logistic Regression
- Support Vector Machine
- Decision Tree
- K-Nearest Neighbors

DATA COLLECTION, WRANGLING, AND FORMATTING

- The API used is https://api.spacexdata.com/v4/rockets/
- This API provides data about several rocket launches by SpaceX, so we must filter the data to Falcon 9 launches only
- Every missing value is replaced by mean of the current column
- The final shape of the data is 90 rows and 17 columns

DATA COLLECTION, WRANGLING, AND FORMATTING

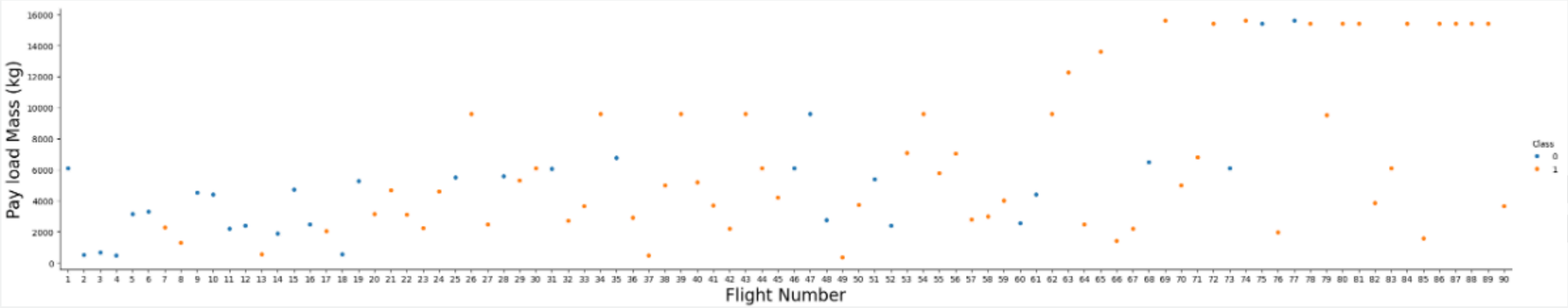
- The data scraped from https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- The shape of the data is 121 rows and 11 columns

DATA COLLECTION, WRANGLING, AND FORMATTING

- The categorical features in the data are processed with one-hot encoding
- Added an extra column called “Class” that contains 0 if the launch is failed and 1 if the launch is successful
- The final data shape is 90 rows and 83 columns

EXPLORATORY DATA ANALYSIS (EDA)

- Using several plot to find correlation and relationship between features, for example:



EXPLORATORY DATA ANALYSIS (EDA)

- Using SQL queries to also get insight about launch site, total and average payload of specific boosters, first successful landing, and etc.

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

Average_Payload
2928.4

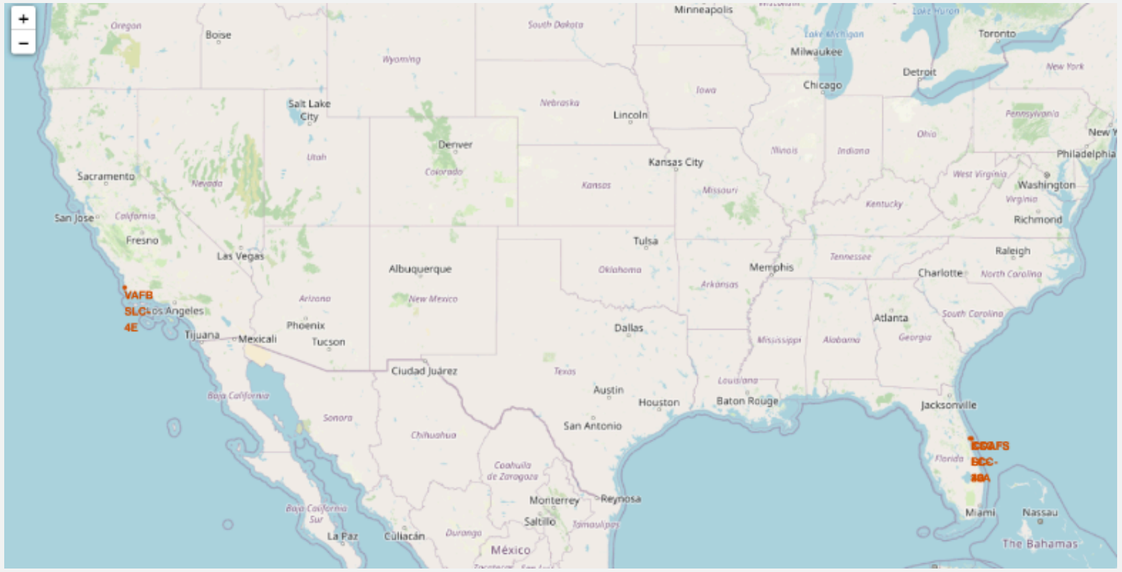
Total_Payload
45596

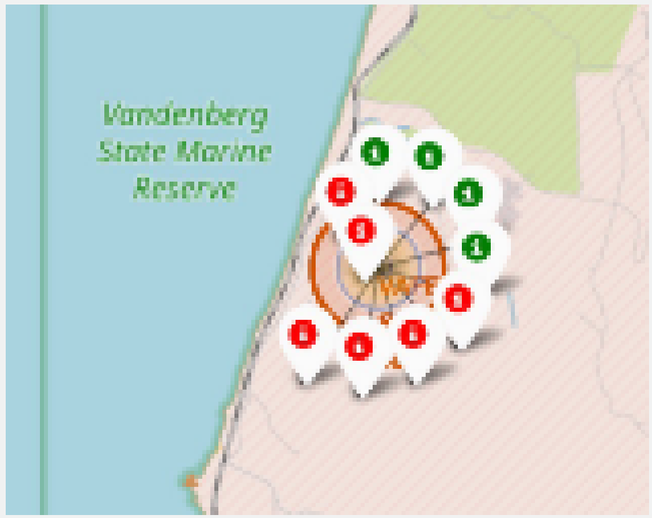
Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

first_landing
2015-12-22

EXPLORATORY DATA ANALYSIS (EDA)

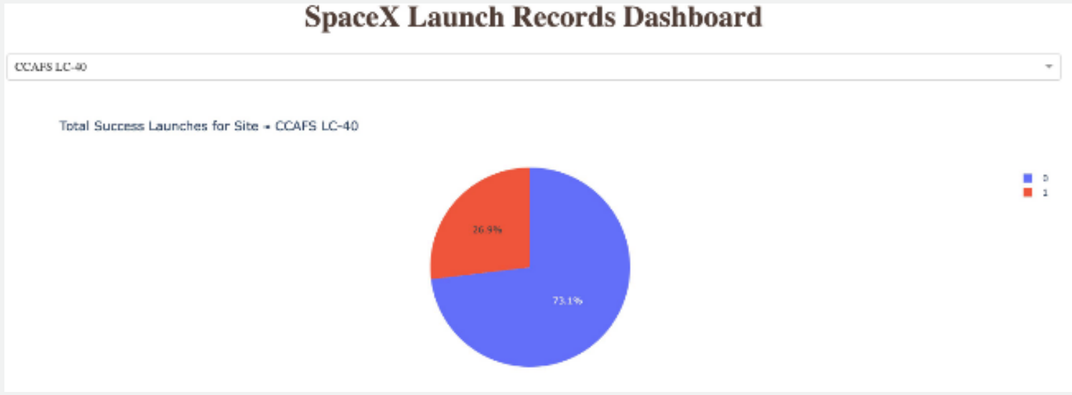
- Using Folium to map and get insight about the launch site location





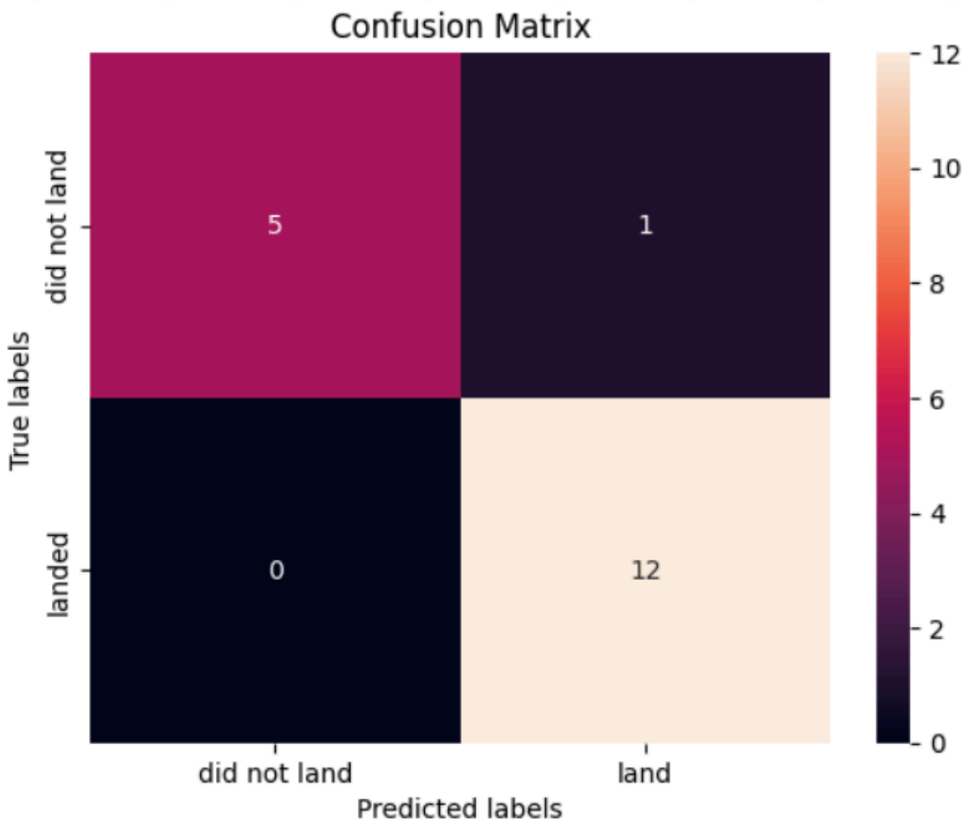
EXPLORATORY DATA ANALYSIS (EDA)

- Using dash to visualize interactive graph in specific launch site and payload range



MACHINE LEARNING

The best machine learning result is Decision Tree that gives the highest accuracy in the train and test set.



SO CAN BE CONCLUDED THAT THE BEST ALGORITHM TO PREDICT THE LAUNCHES RESULT IS DECISION TREE