SpaceX Falcon 9 first stage Landing Prediction Report

VINCENT

22 APRIL 2024

Executive Summary

- This report analyse SpaceX Company's launch data and establish a predictive model. This model will give a investment reference if we want to bid against SpaceX for a rocket launch
- Firstly we will collection data from web and APIs, then we clean data, after that we will do a EDA and find relative variables for the model with visualization tools. Finally we reprocess the data and create four different model to test which one is optimized.
- The success predict rate of final model is about 83% in test set

Introduction

In this report, we will predict if the Falcon 9 first stage will land successfully. SpaceX 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 SpaceX 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 we wants to bid against SpaceX for a rocket launch.

Data collection and wrangling methodology

1. Collection data. Using request API and webscraping Launches Records from Wikipedia, extracting required data from html table
2. Dealing with Missing Values. Replace numerical data with mean and categorical data with median, delete rows with too many data missed
3. Filter Falcon 9 related data
4. Create a landing outcome label from column

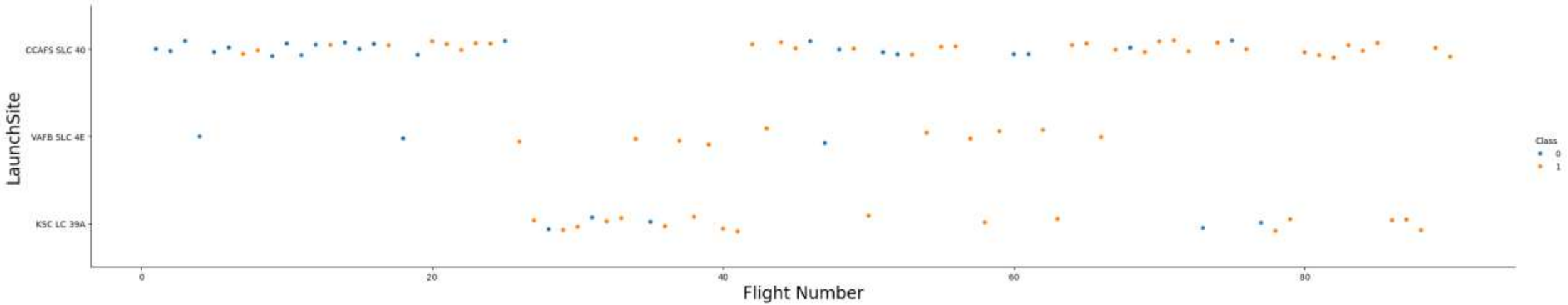
EDA and interactive visual analytics methodology

- SQL and Pandas for some descriptive statistics analysis
- Matplotlib and Seaborn for static visualization
- Folium for map visualization
- Plotly and Dash for interactive visualization

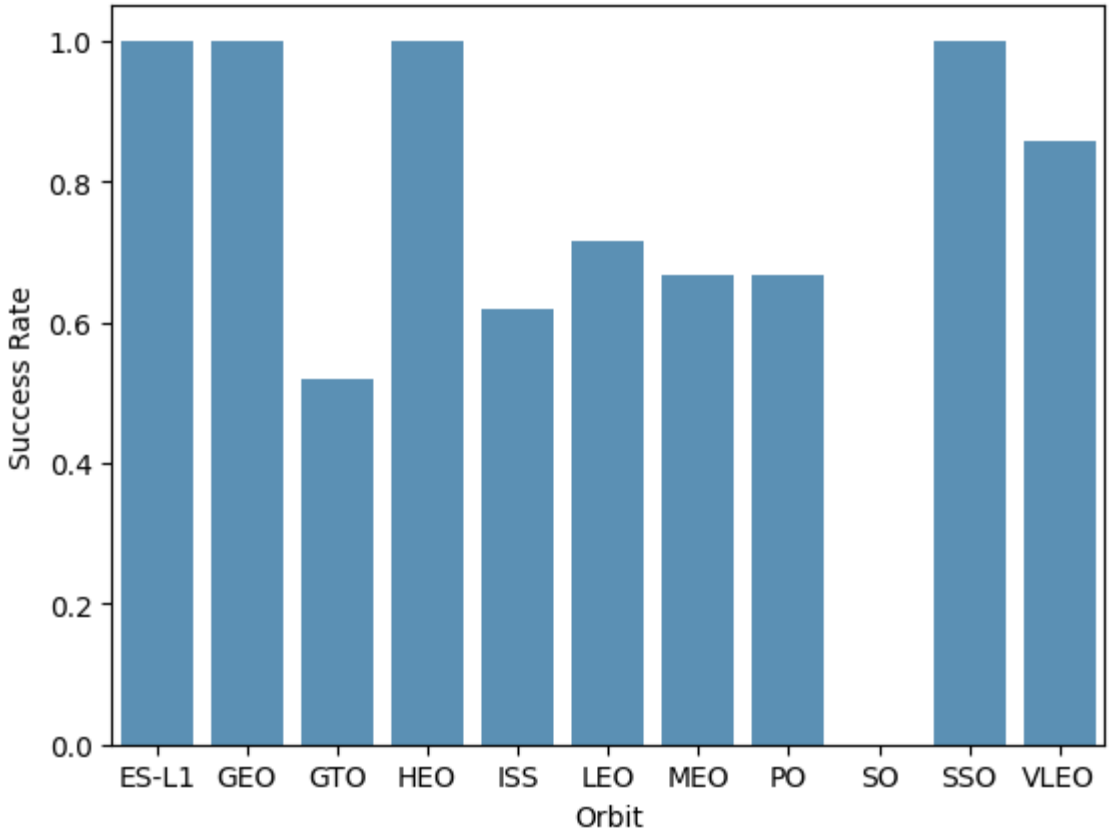
Predictive analysis methodology

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

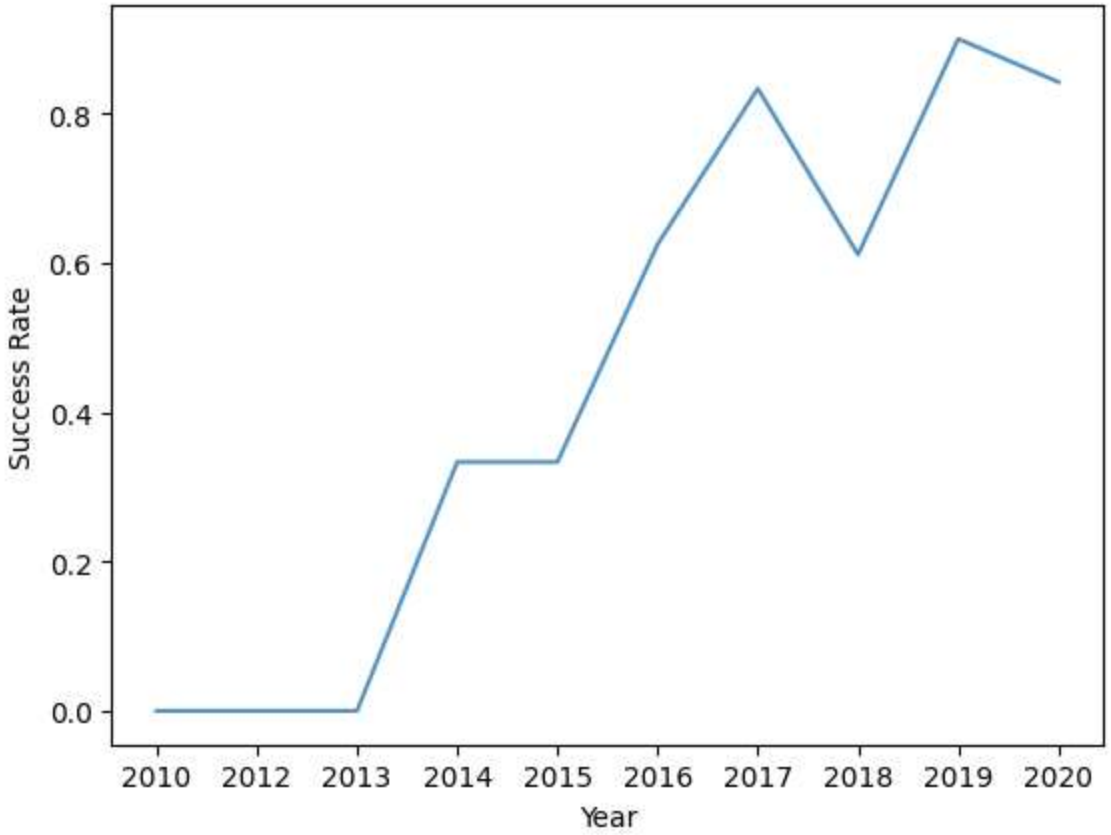
EDA with visualization results



In early age, launch mostly occurred in CCAFS SLC 40 with low launch success rate. Success rate increased by time.



ES-L1, GEO, HEO, SSO and VLEO have success rate higher than 80%



The success rate since 2013 kept increasing till 2017 (stable in 2014) and after 2015 it started increasing

EDA with SQL results

1. There are four launch site: CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40
2. Average payload mass carried by booster version F9 v1.1 is about 2534.7 kg
3. the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000 are F9 FT B1026, F9 FT B1022, F9 FT B1021.2, F9 FT B1031.2

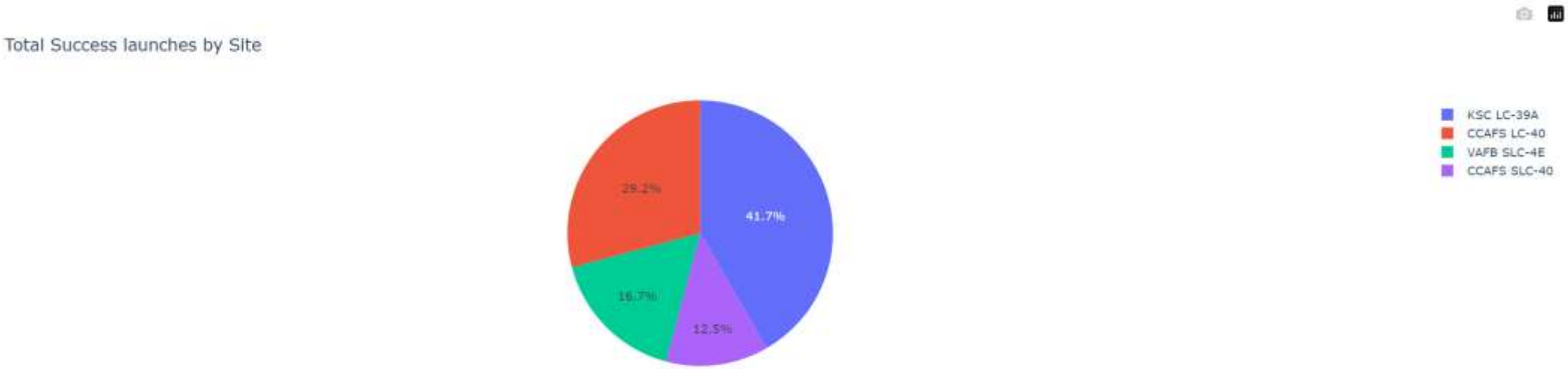
Interactive map with Folium-related

Most of launches is in Florida

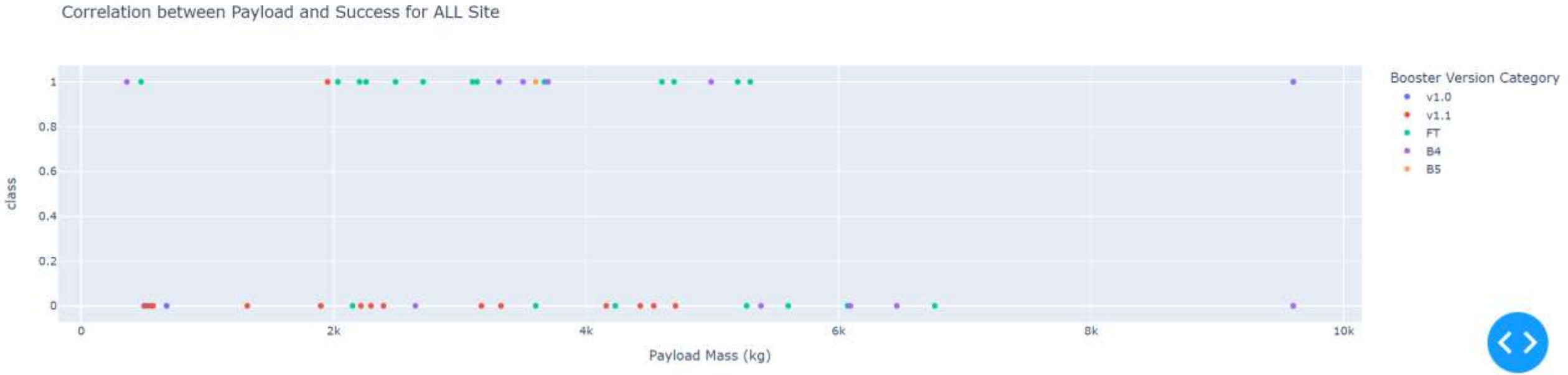
In ccafs lc-40, most launches are failed

About 900m distance between launch site and coastline

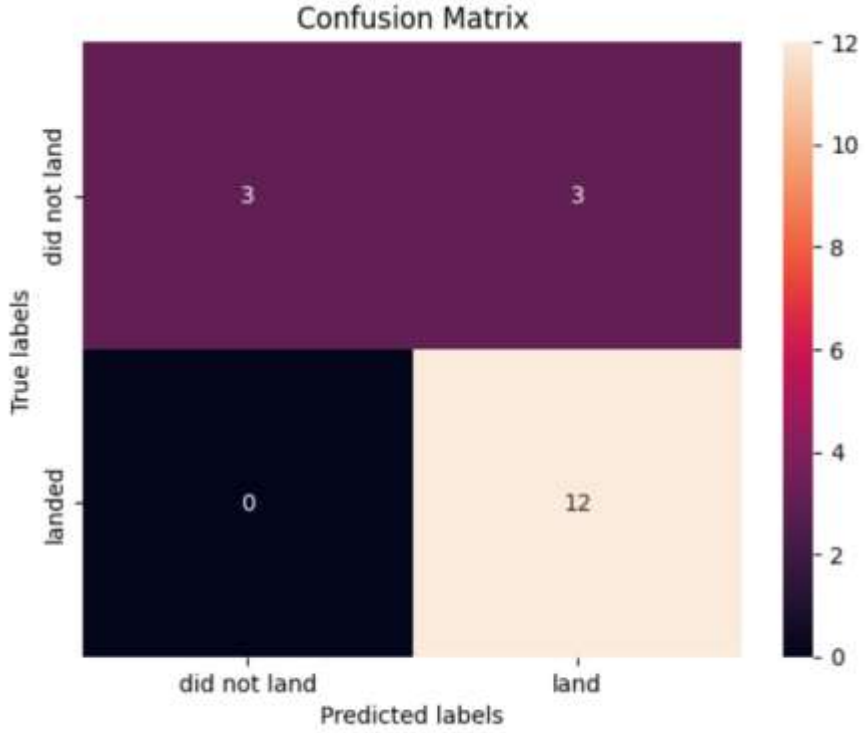
Plotly Dash-related



Payload range (Kg):



Predictive analysis result



Logistic regression, svm and knn have the same accuracy, same confusion matrix in test set. The three method are all suitable in the case.

All about 83.3% accuracy in test site

Conclusion

- We use the attributes: BoosterVersion, PayloadMass,Orbit, LaunchSite, Outcome, Flights, GridFins, Reused, Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude to predict launch success rate.
- The model is about 83%accuracy
- Launch site should close to equator and close to coastline
- ES-L1, GEO, HEO, SSO and VLEO are recommended due to >80% success rate
- We should expect higher than 80% success rate in the future