IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

<Narendra>
<08/19/2024>

Outline

Executive Summary

- In this capstone project, we will predict if the SpaceX Falcon 9 first stage will land successfully using several machine learning classification algorithms.
- The main steps in this project include Data collection, wrangling, and formatting
- Exploratory data analysis
• Interactive data visualization • Machine learning prediction
- Our graphs show that some features of the rocket launches have a correlation with the outcome of the launches, i.e., success or failure.
- It is also concluded that decision tree may be the best machine learning algorithm to predict if the Falcon 9 first stage will land successfully

Introduction

- SpaceX's reusable Falcon 9 rocket offers a significant cost advantage over competitors, with a launch cost of $62 million compared to $165 million or more. The key factor in this savings is the reusability of the first stage. Therefore, accurately predicting whether the first stage will land successfully is crucial for determining the cost of a launch.
- Objective
- Our goal is to develop a predictive model that answers the following question: Given a set of features about a Falcon 9 rocket launch, such as payload mass, orbit type, launch site, and others, will the first stage land successfully?
- Key Considerations
- Most unsuccessful landings are planned, with some involving controlled ocean landings.
- Accurate predictions can inform alternate companies bidding against SpaceX for rocket launches.
- Features for Prediction
- Payload mass
- Orbit type
- Launch site
- [Other relevant features]
- By refining our understanding of the factors influencing first stage landing success, we can create a valuable tool for predicting launch costs and informing competitive bidding strategies.

- The overall methodology includes:

1. Data collection, wrangling, and formatting, using:
   - SpaceX API
   - Web scraping

2. Exploratory data analysis (EDA), using:
   - Pandas and NumPy
   - SQL

3. Data visualization, using:
   - Matplotlib and Seaborn
   - Folium
   - Dash

4. Machine learning prediction, using
   - Logistic regression
   - Support vector machine (SVM)
   - Decision tree
   - K-nearest neighbors (KNN)

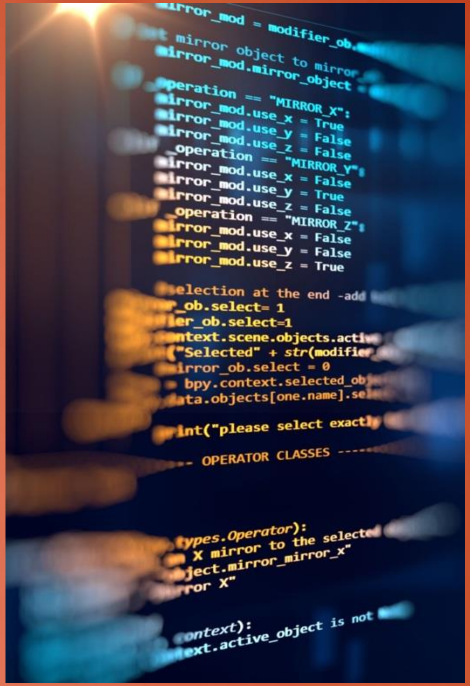
Methodology Data Collection Wrangling and formatting

- Executive Summary:

- • SpaceX API
- • The API used is https://api.spacexdata.com/v4/rockets/.
- • The API provides data about many types of rocket launches done by SpaceX, the data is
- therefore filtered to include only Falcon 9 launches.
- • Every missing value in the data is replaced the mean the column that the missing value
- belongs to.
- • We end up with 90 rows or instances and 17 columns or features. The picture below shows
- the first few rows of the data:

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857

Data Collection



- Web scraping
- • The data is scraped from
- https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&
- oldid=1027686922
- • The website contains only the data about Falcon 9 launches.
- • We end up with 121 rows or instances and 11 columns or features. The picture below shows
- the first few rows of the data:

	Flight No.	Launch site	Payload	Payload mass	Orbit	Customer	Launch outcome	Version Booster	Booster landing	Date	Time
0	1	CCAFS	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success\n	F9 v1.0B0003.1	Failure	4 June 2010	18:45
1	2	CCAFS	Dragon	0	LEO	NASA	Success	F9 v1.0B0004.1	Failure	8 December 2010	15:43
2	3	CCAFS	Dragon	525 kg	LEO	NASA	Success	F9 v1.0B0005.1	No attempt\n	22 May 2012	07:44
3	4	CCAFS	SpaceX CRS-1	4,700 kg	LEO	NASA	Success\n	F9 v1.0B0006.1	No attempt	8 October 2012	00:35
4	5	CCAFS	SpaceX CRS-2	4,877 kg	LEO	NASA	Success\n	F9 v1.0B0007.1	No attempt\n	1 March 2013	15:10

Data Collection – SpaceX API

- *The data is later processed so that there are no missing entries and categorical features are encoded using one-hot encoding.*
- *An extra column called 'Class' is also added to the data frame. The column 'Class' contains 0 if a given launch is failed and 1 if it is successful.*
- *In the end, we end up with 90 rows or instances and 83 columns or features.*

EDA with Data Visualization

- • Pandas and NumPy
- • Functions from the Pandas and NumPy libraries are used to derive basic information about
- the data collected, which includes:
- • The number of launches on each launch site
- • The number of occurrence of each orbit
- • The number and occurrence of each mission outcome
- • SQL
- • The data is queried using SQL to answer several questions about the data such as:
- • The names of the unique launch sites in the space mission
- • The total payload mass carried by boosters launched by NASA (CRS)
- • The average payload mass carried by booster version F9 v1.1

Build an Interactive Map with Folium

- • Folium
- • Functions from the Folium libraries are used to visualize the data through interactive maps.
- • The Folium library is used to:
- • Mark all launch sites on a map
- • Mark the succeeded launches and failed launches for each site on the map
- • Mark the distances between a launch site to its proximities such as the nearest city, railway, or highway

Build a Dashboard with Plotly Dash

- • Dash
- • Functions from Dash are used to generate an interactive site where we can toggle the input
- using a dropdown menu and a range slider.
- • Using a pie chart and a scatterplot, the interactive site shows:
- • The total success launches from each launch site
- • The correlation between payload mass and mission outcome (success or failure) for each launch site

Machine Learning Prediction

- Functions from the Scikit-learn library are used to create our machine learning
- models.
- • The machine learning prediction phase include the following steps:
- • Standardizing the data
- • Splitting the data into training and test data
- • Creating machine learning models, which include:
- • Logistic regression
- • Support vector machine (SVM)
- • Decision tree
- • K nearest neighbors (KNN)
- • Fit the models on the training set
- • Find the best combination of hyperparameters for each model
- • Evaluate the models based on their accuracy scores and confusion matrix

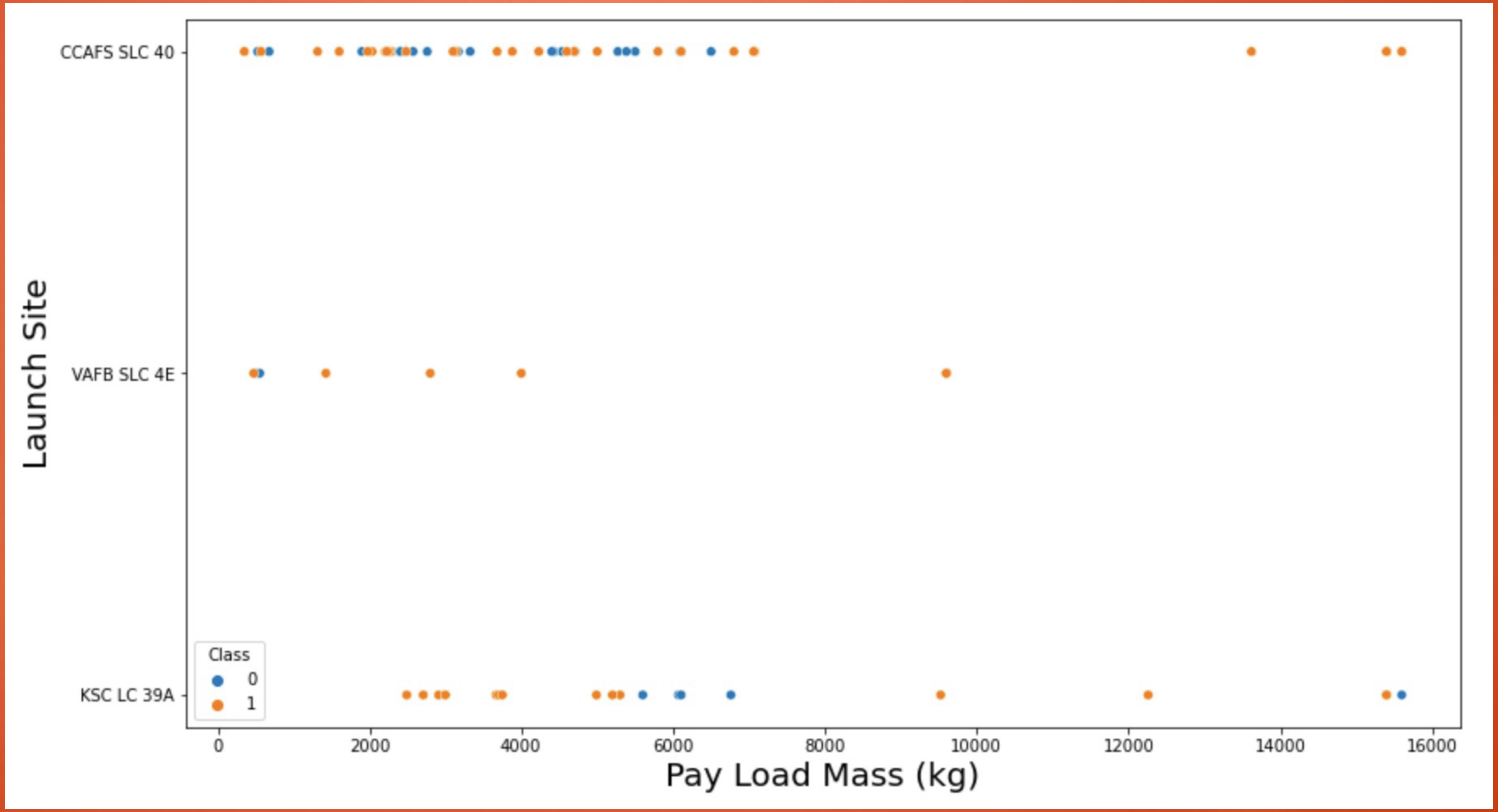
Results

- • The results are split into 5 sections:
- • SQL (EDA with SQL)
- • Matplotlib and Seaborn (EDA with Visualization)
- • Folium
- • Dash
- • Predictive Analysis
- • In all of the graphs that follow, class 0 represents a failed launch outcome while
- class 1 represents a successful launch outcome.

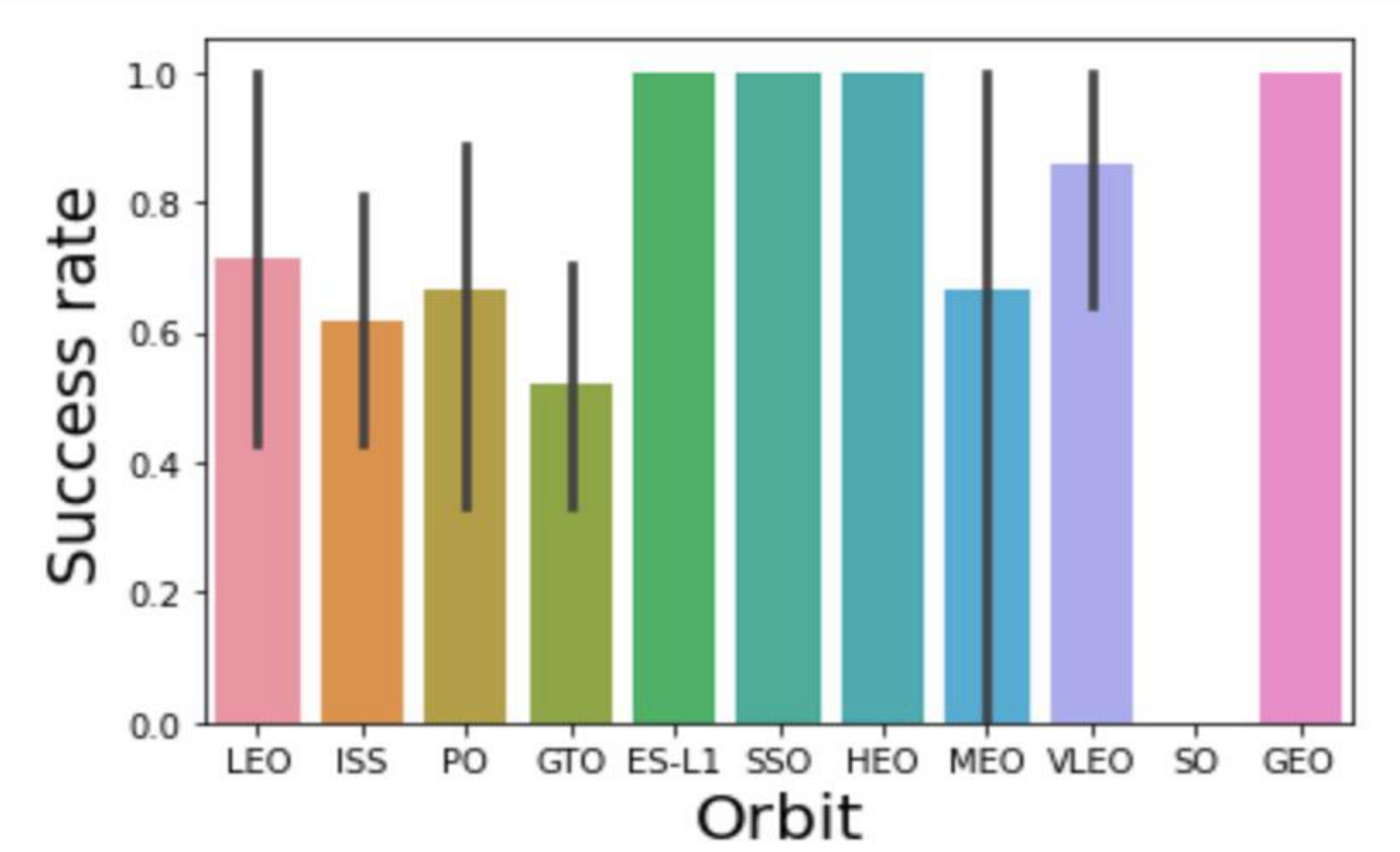
Section 2

Insights drawn from EDA

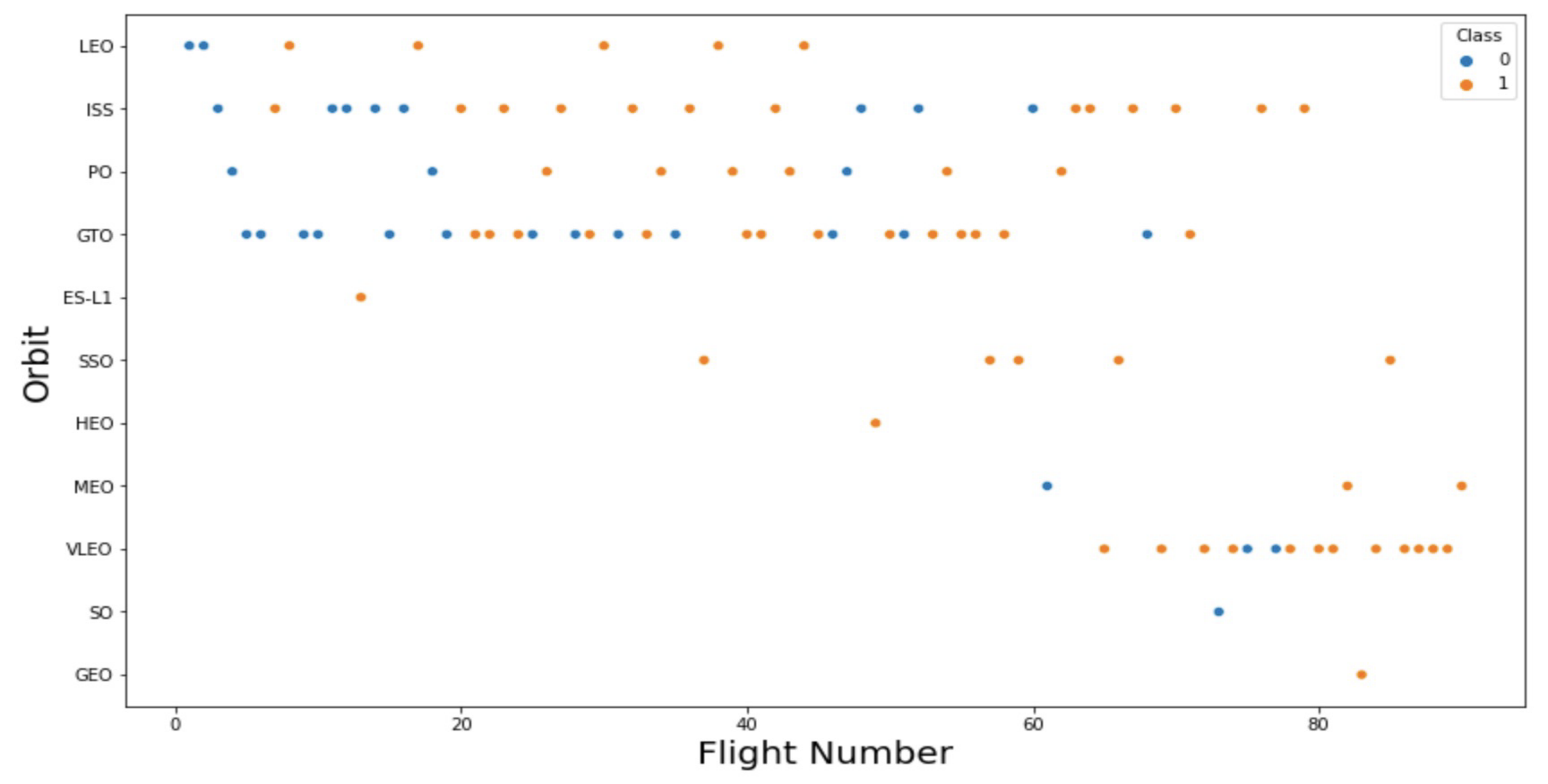
Flight Number vs. Launch Site



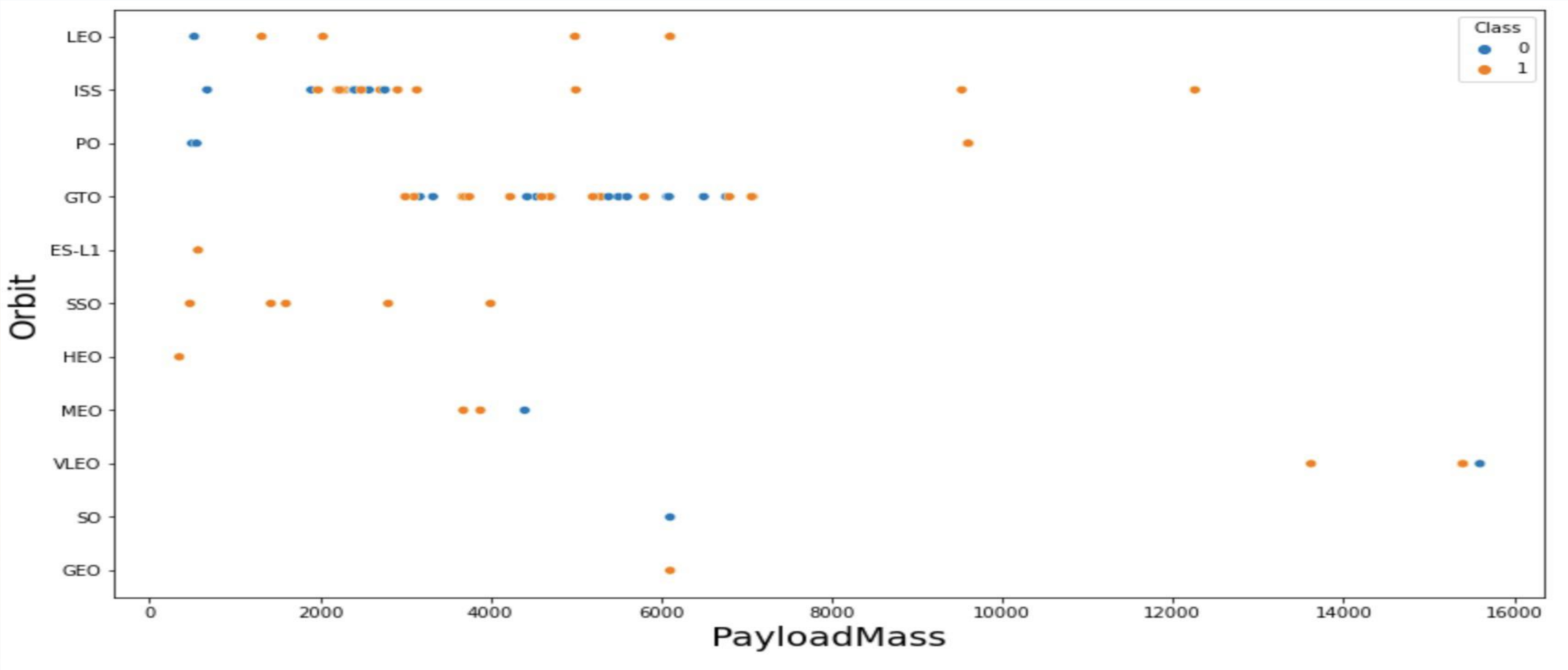
Success Rate vs. Orbit Type



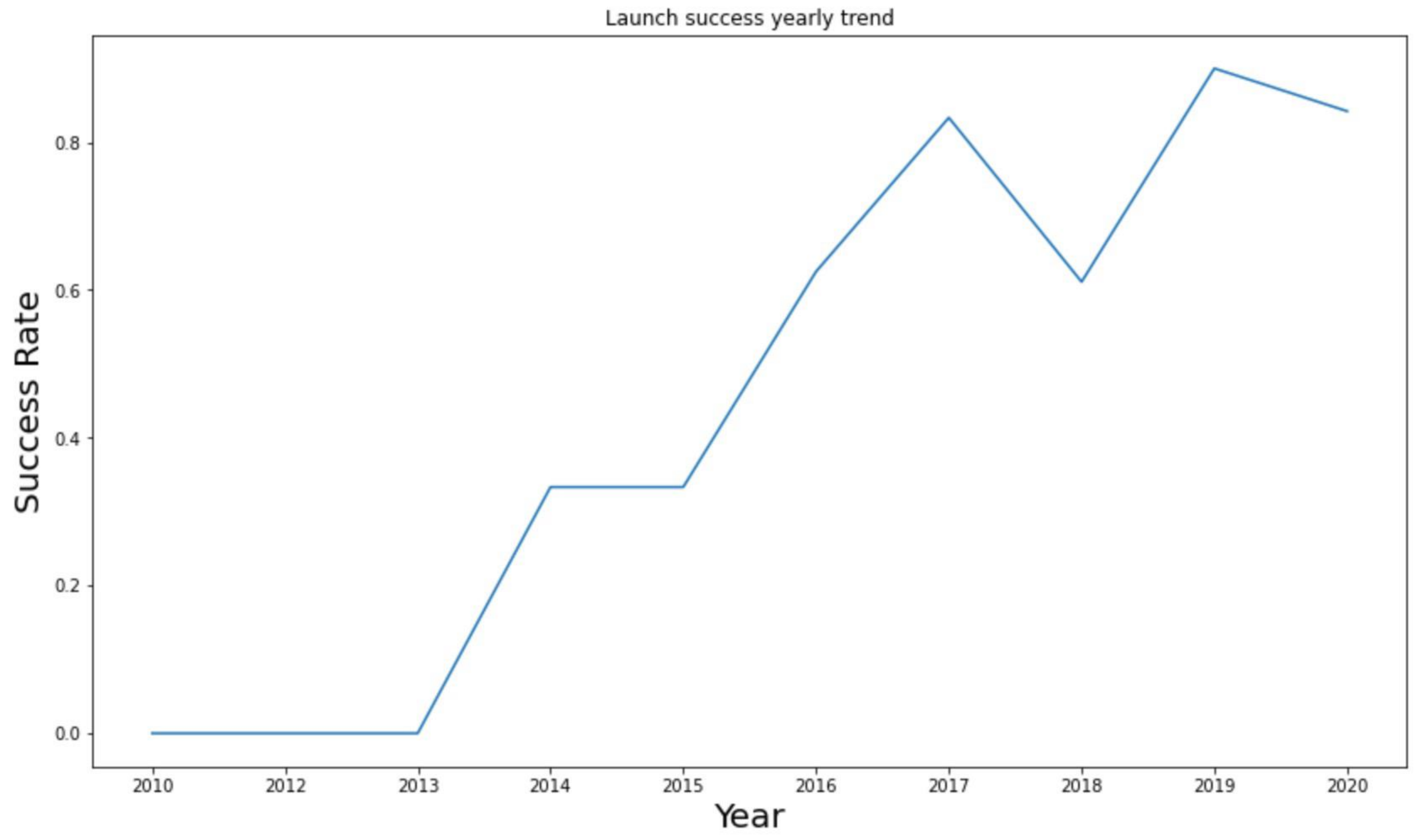
Flight Number vs. Orbit Type



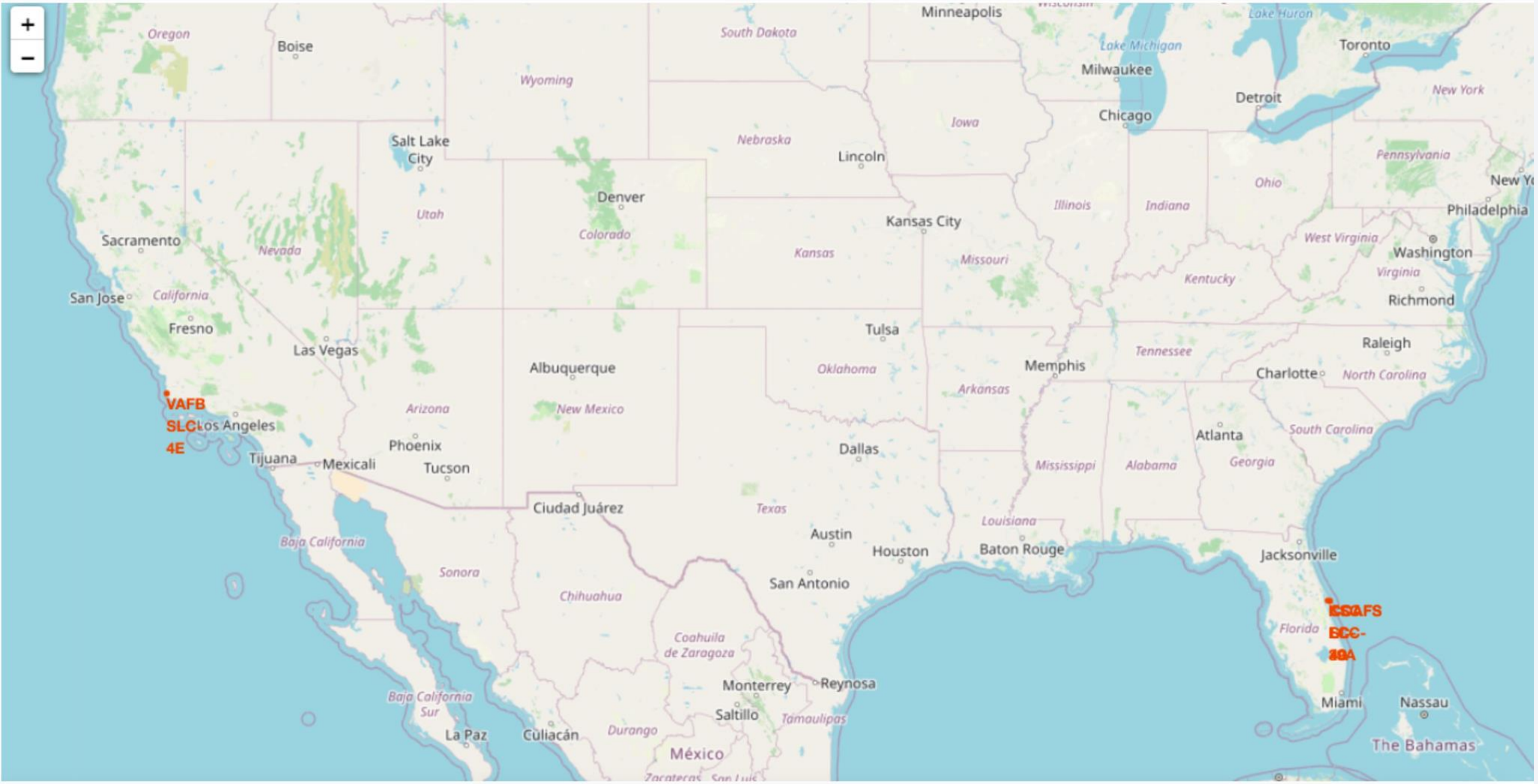
Payload vs. Orbit Type



Launch Success Yearly Trend



All Launch Site Names



Conclusions

- • In this project, we try to predict if the first stage of a given Falcon 9 launch will
- land in order to determine the cost of a launch.
- • Each feature of a Falcon 9 launch, such as its payload mass or orbit type, may
- affect the mission outcome in a certain way.
- • Several machine learning algorithms are employed to learn the patterns of past
- Falcon 9 launch data to produce predictive models that can be used to predict the
- outcome of a Falcon 9 launch.
- • The predictive model produced by decision tree algorithm performed the best
- among the 4 machine learning algorithms employed.

Thank you!