

IBM Applied Data Science Capstone Project

Submitted by

Sarsha J S

June 10, 2023.

Table of Contents

- Introduction
- Methodologies
 - Data collection
 - Using REST API
 - Using web scrapping
 - Data wrangling
 - Exploratory Data Analysis
 - Using SQL
 - Using visualization labs
 - Interactive Visual Analytics
 - Using Folium
 - Using Plotly Dash
 - Predictive Analysis
- Additional screenshots
- Results
- Conclusion

Introduction

The commercial space age is here, companies are making space travel affordable for everyone. Virgin Galactic is providing suborbital spaceflights. Rocket Lab is a small satellite provider. Blue Origin manufactures sub-orbital and orbital reusable rockets. Perhaps the most successful is SpaceX. SpaceX's accomplishments include: Sending spacecraft to the International Space Station. Starlink, a satellite internet constellation providing satellite Internet access. Sending manned missions to Space. One reason SpaceX can do this is the rocket launches are relatively inexpensive. 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. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. SpaceX's Falcon 9 launch like regular rockets. Unlike other rocket providers, SpaceX's Falcon 9 can recover the first stage. Sometimes the first stage does not land. Sometimes it will crash. Other times, Space X will sacrifice the first stage due to the mission parameters like payload, orbit, and customer.

In this capstone, as the role of a data scientist working for a new rocket company, Space Y that would like to compete with SpaceX founded by Billionaire industrialist Allon Musk to determine the price of each launch by gathering information about Space X and creating dashboards. Also to determine if SpaceX will reuse the first stage.

Data Collection

- **Using REST API**

REST (Representational State Transfer) API is a popular way to retrieve data from web servers. It allows you to interact with remote servers and request specific data using standard HTTP methods (e.g., GET, POST, PUT, DELETE). This API will give data about launches including information about the rocket used, payload delivered, launch specifications, landing specifications and landing outcomes. Use the URL to target a specific endpoint of the API to get past launch data. Perform a get request using requests library to get the data from API. Response will be in JSON form. Use json_normalize function to convert the JSON and convert it into pandas dataframe.

- **Using web scrapping**

Web scrapping involves extracting data from websites by parsing the underlying HTML or XML structure of web pages. It enables to gather data from websites that do not provide an API or have restricted access to their data. Using the Python BeautifulSoup package to web scrape some HTML tables that contain valuable Falcon 9 launch records. Then parse the data from those tables and convert them into a Pandas dataframe for further visualization and analysis.

Data collection I performed is available on the GitHub link:

https://github.com/SarshaJS18/myrepo/blob/main/Week1_webscrapping.ipynb

Data wrangling

Data wrangling typically includes tasks such as handling missing values, removing duplicates, correcting data inconsistencies, formatting data types, merging or splitting datasets and creating new variables or features.

Data wrangling I performed is available on the GitHub link:

https://github.com/SarshaJS18/myrepo/blob/main/week1_data_wrangling.ipynb

Exploratory Data Analysis

- **Using SQL**

Exploratory Data Analysis (EDA) using SQL involves SQL queries to explore and analyze data within relational database. SQL provides a powerful set of tools for data manipulation and analysis.

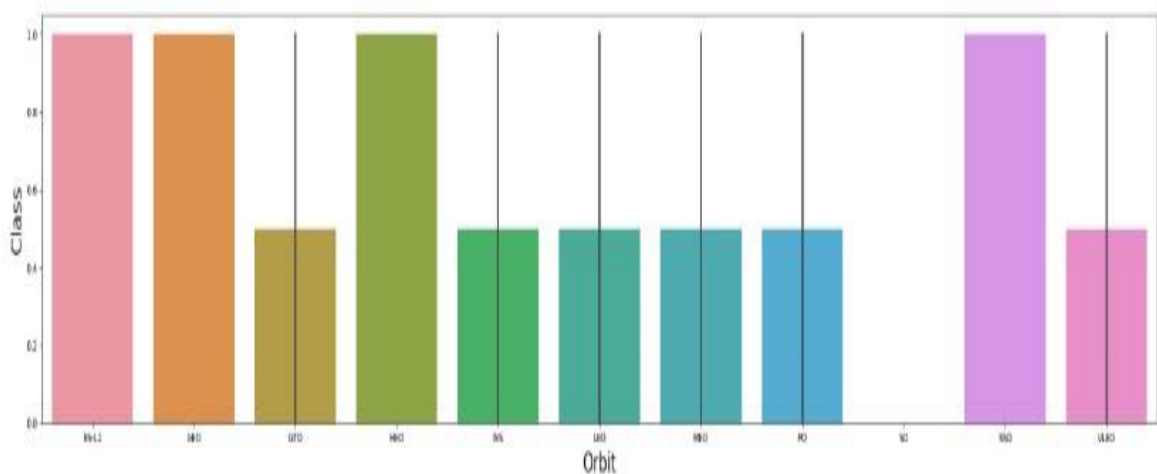
SQL queries I have performed for EDA is available on the GitHub link:

https://github.com/SarshaJS18/myrepo/blob/main/Week_2_lab1_EDA_withSQL.ipynb

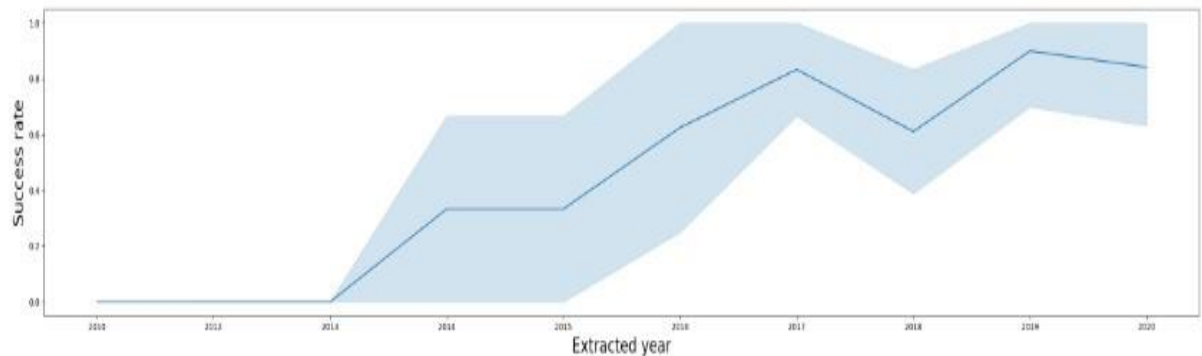
- **Using Visualization labs**

Visualization plays a crucial role in Exploratory Data Analysis (EDA) as it enables to visually understand and interpret the data. It helps to identify patterns, trends, outliers and relationships within the dataset.

Some of the screenshots are as follows:



The above screenshot shows the relationship between Orbit and class.



The above screenshot shows the yearly trend success rate. From this we can see that success rate is steadily increasing from the year 2013.

EDA with visualization lab I performed is available on the GitHub link: https://github.com/SarshaJS18/myrepo/blob/main/Week2_lab2_EDA_with_Visualization.ipynb

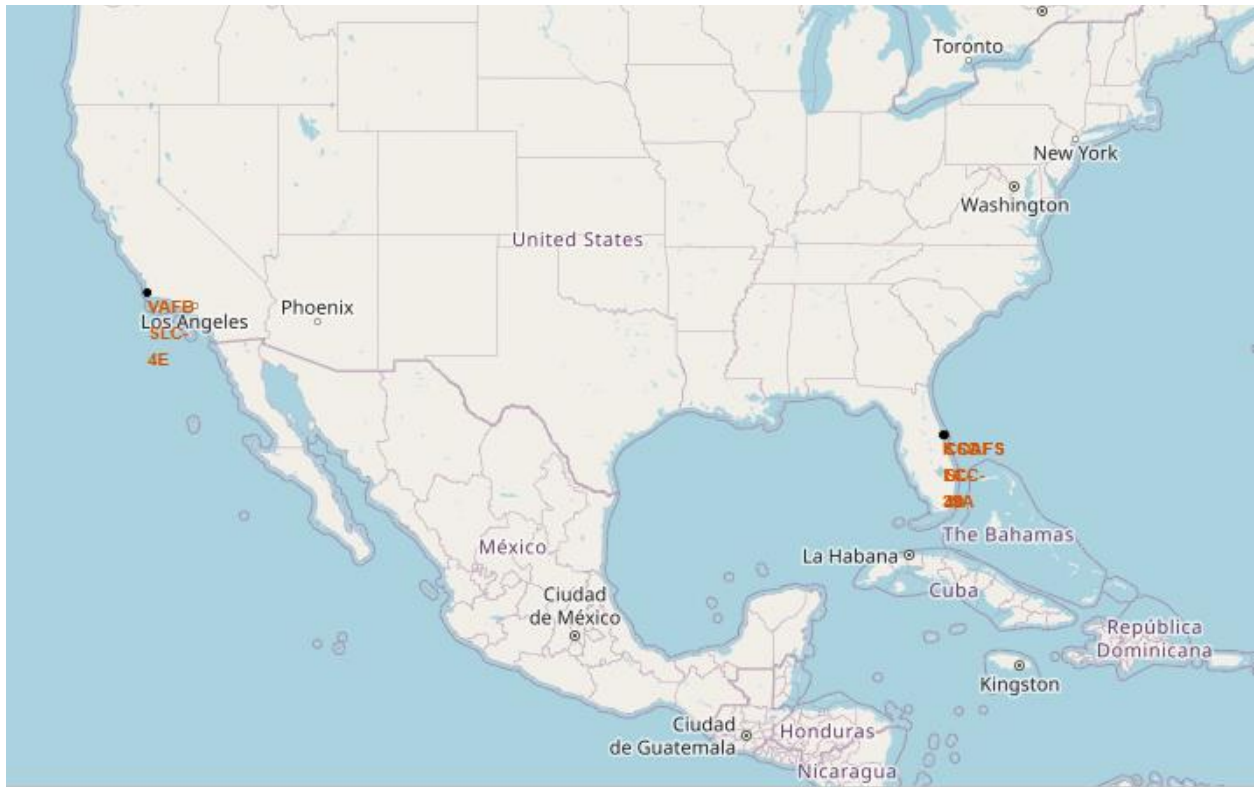
Interactive Visual Analytics

Interactive Visual Analytics enables users to explore and manipulate data in an interactive and real-time way. Common interactions include zoom-in, zoom-out, pan, filter, search and link. With Interactive Visual Analytics, users could find visual patterns faster and more effectively.

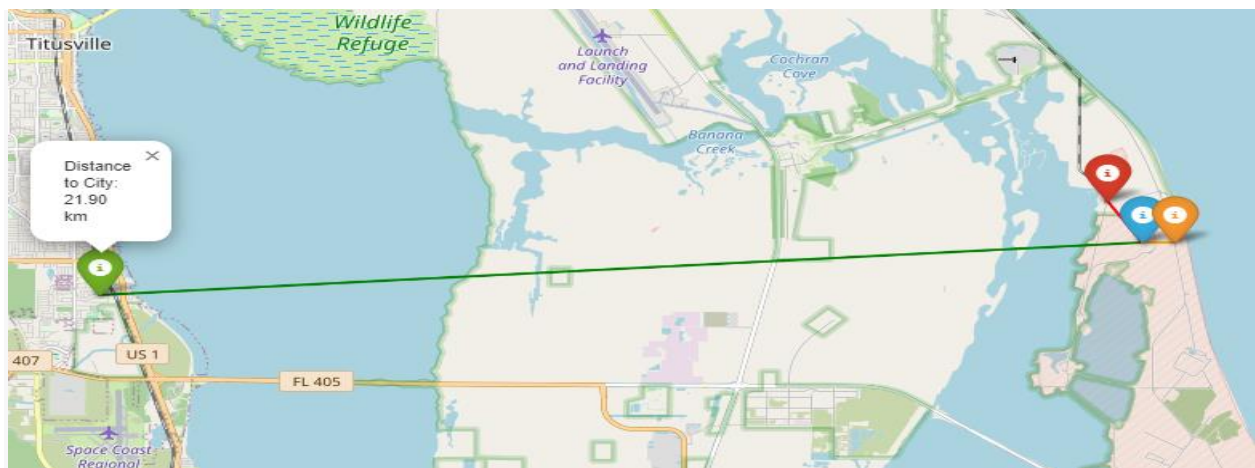
- **Using Folium**

Folium is a python library that allows creating interactive maps and visualizing geospatial data. It leverages the leaflet.js library to generate interactive and customizable maps.

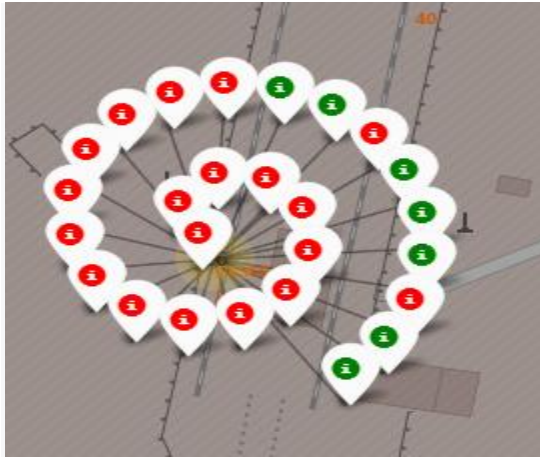
Some of the screenshots of Interactive Visual Analytics using Folium are as follows:



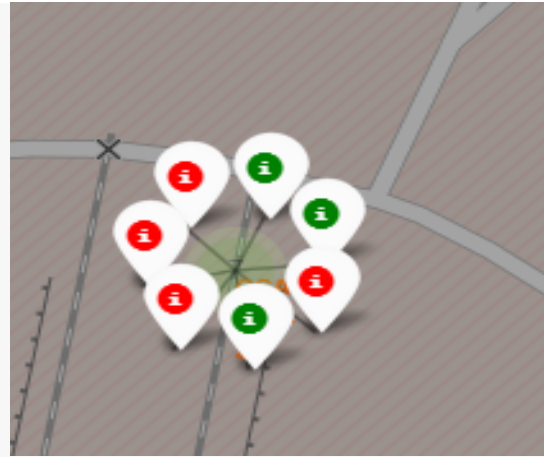
In the above screenshots names of the launch sites are labeled in red color.



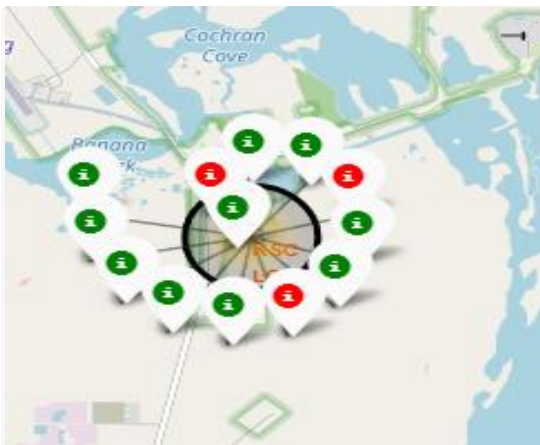
The above screenshot shows the distance from launch site which is blue icon to the city in green icon, railroad in red icon and highway in orange icon.



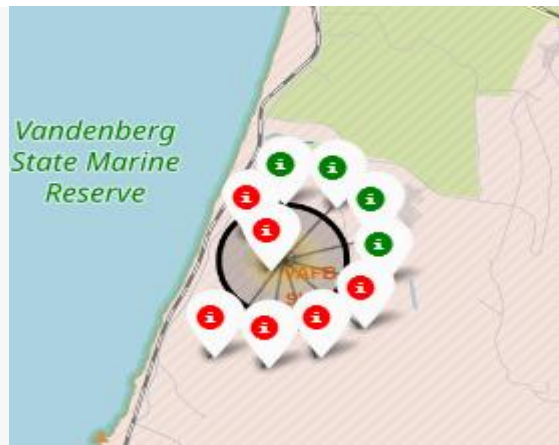
CCAF SLC-40



CCAFSSLC-40



KSC LC-39A



VAFB SLC-4E

The above screen shots show the success and failure of launch sites where green icon is success and red icon is failure. From this we can conclude that KSC LC-39A has more success rates.

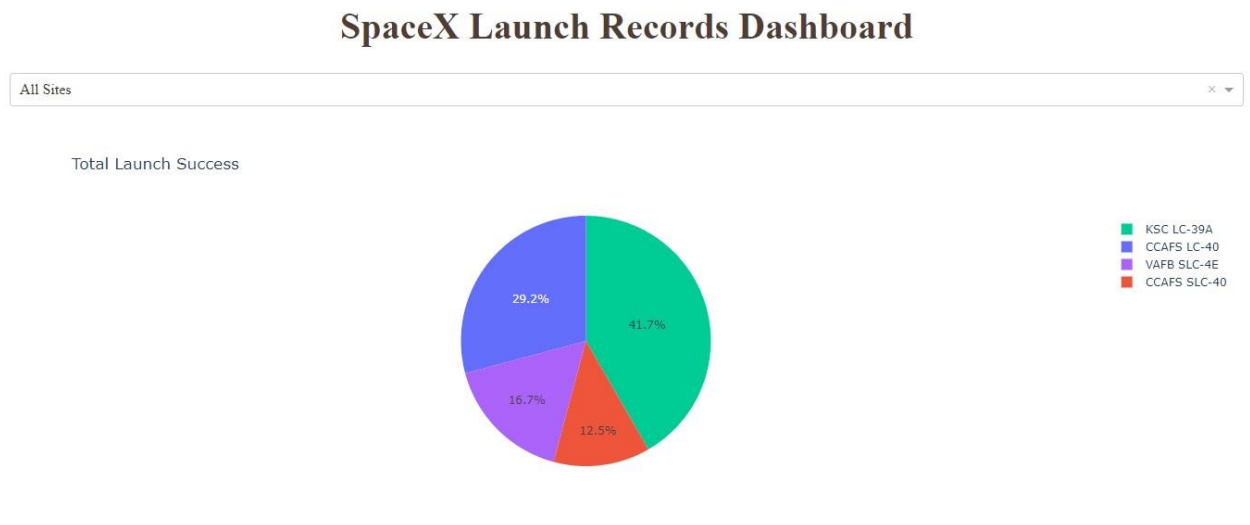
Interactive Data Analytics using Folium I performed is available on the GitHub link:

https://github.com/SarshaJS18/myrepo/blob/main/Week3_lab1_Interactive_visual_analytic_with_folium.ipynb

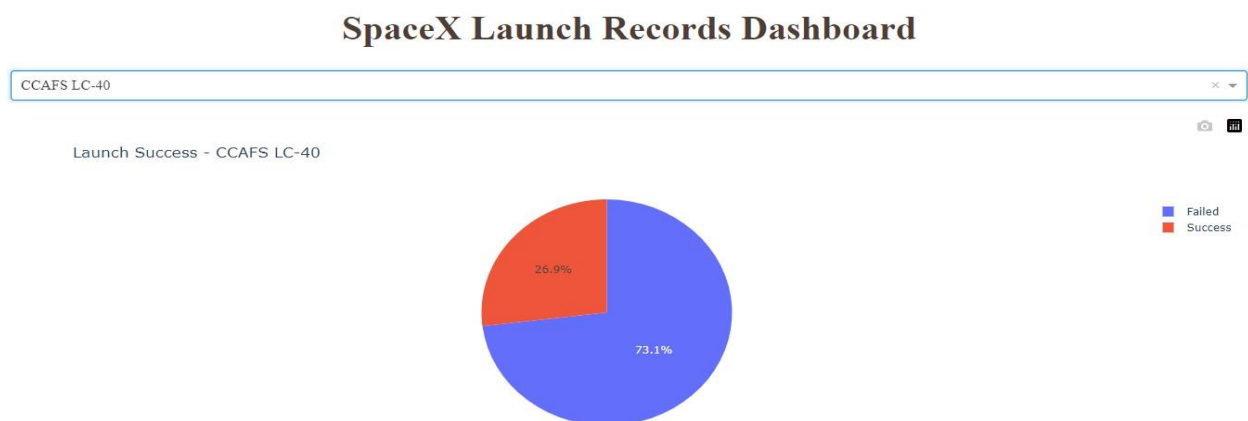
- **Using Plotly Dash**

Plotly Dash is a python framework for building interactive web-based dashboards and applications. It combines the power of Plotly for interactive visualizations with Flask for web development.

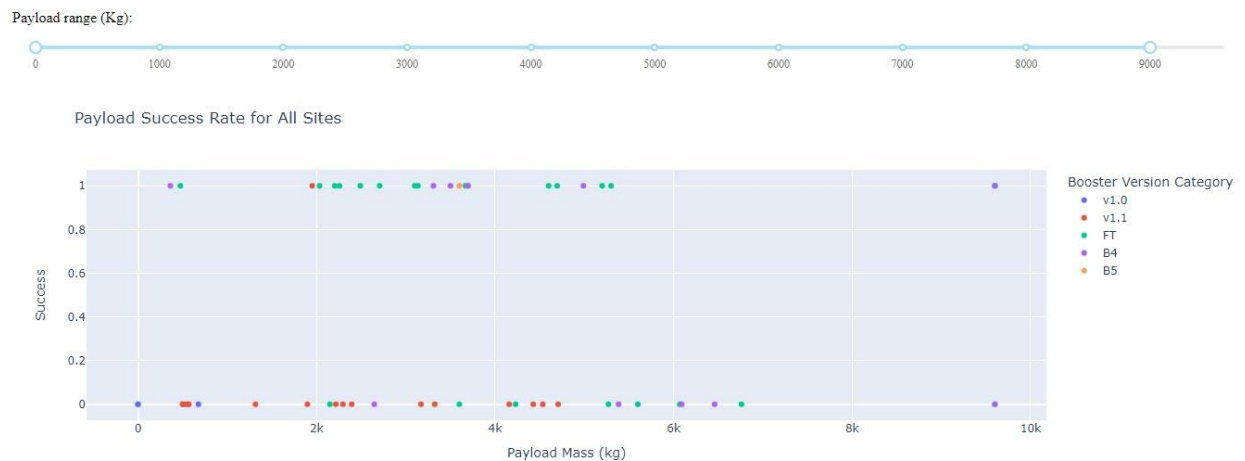
Some of the screenshots of the dashboard are as follows:



The above screenshot shows the launch success for all sites. From this pie chart we can conclude KSC LC-39A has high launch success.



The above screenshot shows the success and failure rate of CCAFS LC-40 where the success is red portion and failure is blue portion.



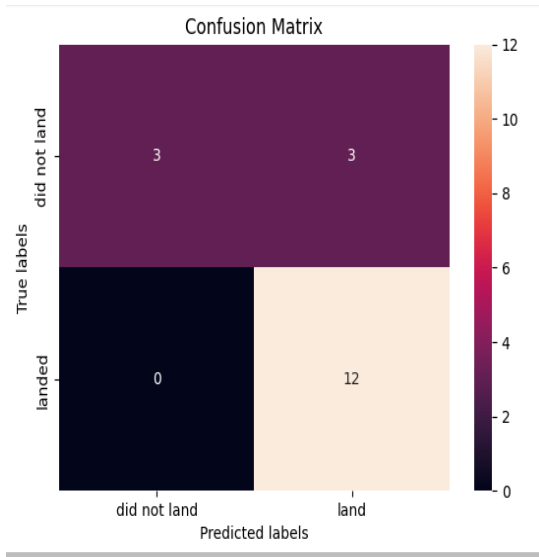
The above screenshot shows the payload success rate for all sites.

Dashboard I created using Plotly Dash is available on the GitHub link:
https://github.com/SarshaJS18/myrepo/blob/main/spacex_dash_app.py

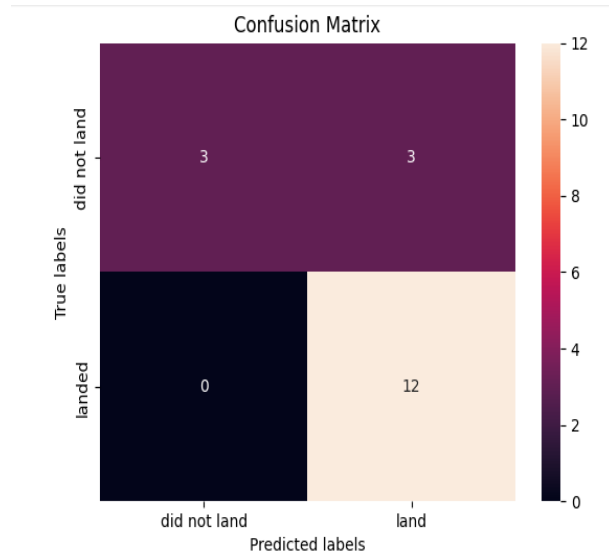
Predictive Analysis

Machine learning model is built to predict if the first stage of the Falcon 9 will land successfully or not. The data is standardized and split into training data and testing data. Then GridSearch is performed to find the hyperparameters that allows a given algorithm to perform best. Logistic Regression, Support Vector Machines, Decision Tree classifier and KNN algorithms were tested to find the best method based on accuracy. Result is shown in confusion matrix.

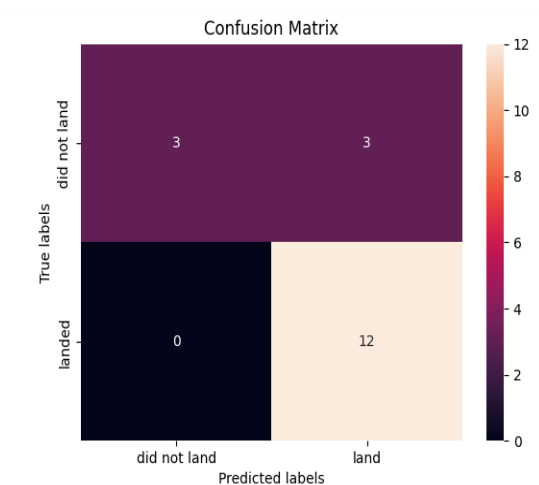
Some of the screenshots of the confusion matrix are as follows:



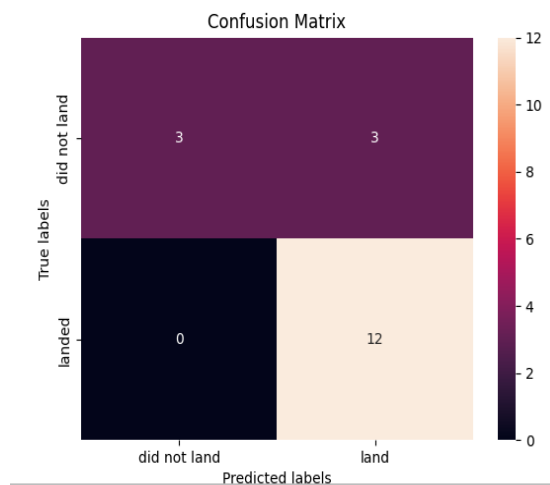
Logistic Regression



Support Vector Machine



Decision tree



K-Nearest Neighbors

From accuracy of the algorithms we can conclude Decision Tree classifier is the best method.

Predictive Analysis I performed is available on the Github link:

https://github.com/SarshaJS18/myrepo/blob/main/Week4_lab1_Machine_learning_predictive.ipynb

Additional screenshots

MONTH	Landing_Outcome	Booster_Version	Launch_Site
OCTOBER	Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40
APRIL	Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40

The above screenshot shows 2015 launch records.

done.

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

The above screenshot shows the names of all launch sights.

Booster_Version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

The above screenshot shows the booster version that carried maximum payload.

Result

From the Analysis it is proven that KSC LC-39A has the high launch rate when compared to others.

SpaceX first successful ground landing happened in August 1, 2018.

Decision Tree classifier is the best method for predicting with accuracy of 0.8888.

Conclusion

By analyzing the data existed rocket companies like SpaceX can predict the cost of launches, success and failure of each launch sites.

Visualization can be used to make the clients understand the situation easily.

PowerPoint presentation I prepared is available on the GitHub link:

https://github.com/SarshaJS18/myrepo/blob/main/Week4_lab1_Machine_learning_predictive.ipynb