IBM Data Science Capstone

By Tatevik Khachatryan

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

- Methodology Summary
  - Data Collection
  - Data Wrangling
  - EDA
  - EDA with SQL
  - Predictive Analysis
  - Interactive Map with Folium
  - Interactive Dashboard Using Plotly
- Results Summary
  - EDA Results
  - Predictive Analysis Results
-

Introduction

Background: SpaceX a rocket company launches satellites at low price like 70% less than their competitor since they lan their satellites for reusing them to launch .

Problem: We use the previous data of launches of Falcon 9 rocket to predict the probability of the booster landing back to the pad influenced/correlated with the space launch site,the payload orbit, mass,landing pad location and the version of the booster.

Methodology

- Data Collection - Web Scraping and API
- Data Wrangling
- Data Cleaning
- EDA with Visualization and EDA with SQL
- Interactive Map with Folium & Plotly Dash
- Predictive Analysis (ML modeling)

Data is collected by web scraping, using BeautifulSoup. The source was the Falcon 9 wikipedia URL.

As they were several tables, data was collected by iterating through tables and saved to ‘spacex_web_scraped.csv’

In [6]:

```
extracted_row = 0

# count number of iterations
for table_number,table in enumerate(soup.find_all('table',"wikitable plainrowheaders collapsible")):
    for rows in table.find_all("tr"):  # table row
        if rows.th:
            if rows.th.string:
                flight_number = rows.th.string.strip() # remove 'enters' between numbers
                flag = flight_number.isdigit() # this gives True or False
            else:
                flag=False
            row = rows.find_all('td') # table element

            if flag: # if it is a number save cells in a dictionary
                extracted_row+=1
                launch_dict['Flight No.'].append(extracted_row)

                datetimelist = [data_time.strip() for data_time in list(row[0].strings)][0:2] # extract full datetime
```

EDA with Visualization

As a result of EDA we've learned following insights:

- Flight number & Launch Sites-Visualizing the launch from every site.
- Payload & Launch Sites-Payload launch from sites
- Success rate & Orbit type-Success rate compared to the orbit type
- Flight number & Orbit Type - Type of orbit for each launch
- Payload & Orbit type - Payload and the orbit .
- Trend of success rate - Trend of the success rate over the years.

EDA with SQL

Performed SQL queries:

- Displaying the names of the unique launch sites in the space mission
- Displaying 5 records where launch sites begin with the string ‘CCA'
- Displaying the total payload mass carried by boosters launched by NASA
- Displaying average payload mass carried by booster version F9 v1.1
- Listing the total number of successful and failure mission outcomes
- Listing the names of the booster versions which have carried the maximum payload mass
- Ranking the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20 in descending order

Build an interactive map with Folium

Markers of all Launch Sites: - Added Marker with Circle, Popup Label and Text Label of NASA Johnson Space Center using its latitude and longitude coordinates as a start location. - Added Markers with Circle, Popup Label and Text Label of all Launch Sites using their latitude and longitude coordinates to show their geographical locations and proximity to Equator and coasts.

Coloured Markers of the launch outcomes for each Launch Site: - Added coloured Markers of success (Green) and failed (Red) launches using Marker Cluster to identify which launch sites have relatively high success rates.

Distances between a Launch Site to its proximities: - Added coloured Lines to show distances between the Launch Site KSC LC-39A (as an example) and its proximities like Railway, Highway, Coastline and Closest City

Build a Dashboard with Plotly Dash

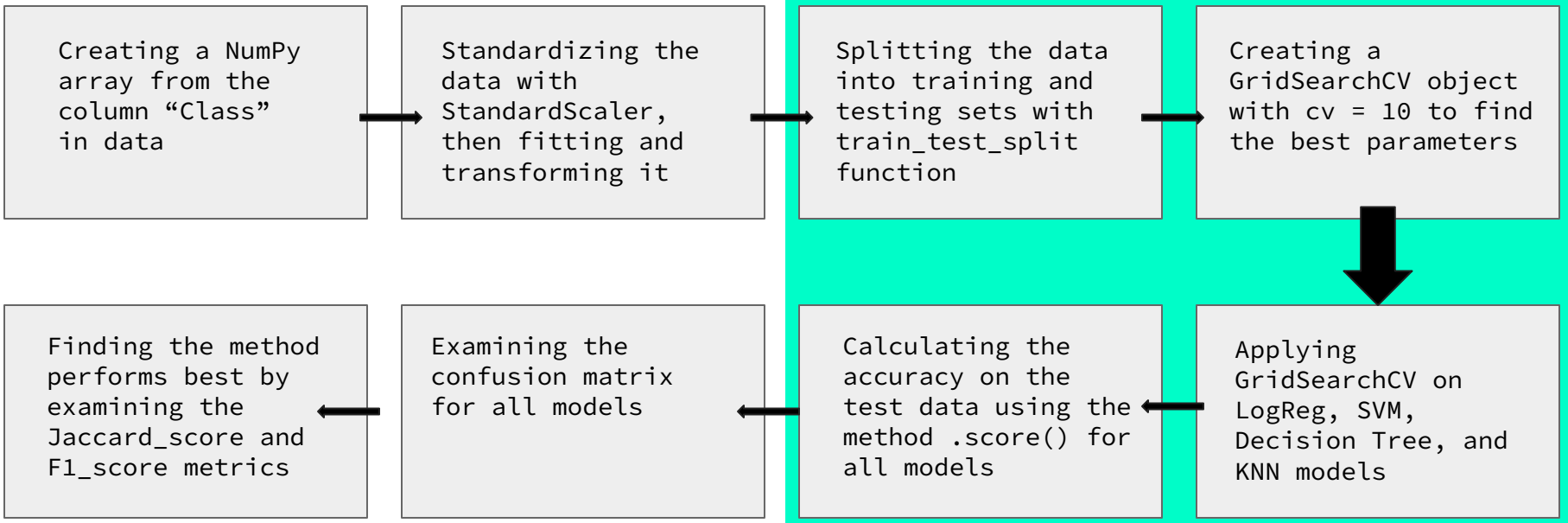
Launch Sites Dropdown List: - Added a dropdown list to enable Launch Site selection.

Pie Chart showing Success Launches (All Sites/Certain Site): - Added a pie chart to show the total successful launches count for all sites and the Success vs. Failed counts for the site, if a specific Launch Site was selected.

Slider of Payload Mass Range: - Added a slider to select Payload range.

Scatter Chart of Payload Mass vs. Success Rate for the different Booster Versions: - Added a scatter chart to show the correlation between Payload and Launch Success.

Predictive analysis (Classification)



Conclusion

- Decision Tree Model is the best algorithm for this dataset.
- Launches with a low payload mass show better results than launches with a larger payload mass.
- Most of launch sites are in proximity to the Equator line and all the sites are in very close proximity to the coast.
- The success rate of launches increases over the years.
- KSC LC-39A has the highest success rate of the launches from all the sites.
- Orbits ES-L1, GEO, HEO and SSO have 100% success rate.