IBM DATA SCIENCE CAPSTONE PROJECT

Space X Falcon 9 Landing Analysis

Daniel Chiebuka Ihenacho

Outline

- Executive Summary
- Introduction
- Methodology
- Results and Discussions
- Conclusion

EXECUTIVE SUMMARY

- Summary of Methodologies:
- This project follows these steps:
- •Data Collection
- •Data Wrangling
- •Exploratory Data Analysis
- •Interactive Visual Analytics
- •Predictive Analysis (Classification)

- Summary of Results:
- This project produced the following outputs and visualizations:
- 1.Exploratory Data Analysis (EDA) results
- 2.Geospatial analytics
- 3.Interactive dashboard
- 4.Predictive analysis of classification models

INTRODUCTION

- SpaceX launches Falcon 9 rockets at a cost of around $62m. This is considerably cheaper than other providers (which usually cost upwards of $165m), and much of the savings are because SpaceX can land, and then re-use the first stage of the rocket.
- •If we can make predictions on whether the first stage will land, we can determine the cost of a launch, and use this information to assess whether or not an alternate company should bid and SpaceX for a rocket launch.
- •This project will ultimately predict if the Space X Falcon 9 first stage will land successfully.

METHODOLOGY SUMMARY

1. Data Collection

- Making GET requests to the SpaceX REST API
- Web Scraping

2.Data Wrangling

- Using the .fillna()method to remove NaN values
- Using the .value_counts()method to determine the following:
- Number of launches on each site
- Number of and occurrence of each orbit
- Number and occurrence of mission outcome per orbit type
- Creating a landing outcome label that shows the following:
  - 0 when the booster did not land successfully
  - 1 when the booster did land successfully

3.Exploratory Data Analysis

- Using SQL queries to manipulate and evaluate the SpaceX dataset
- Using Pandas and Matplotlib to visualize relationships between variables, and determine patterns

4.Interactive Visual Analytics

- Geospatial analytics using Folium
- Creating an interactive dashboard using Plotly Dash

5.Data Modelling and Evaluation

- Using Scikit-Learn to:
  - Pre-process (standardize) the data
  - Split the data into training and testing data using train_test_split
  - Train different classification models
  - Find hyperparameters using GridSearchCV
- Plotting confusion matrices for each classification model
- Assessing the accuracy of each classification model

DATA COLLECTION –SPACE X REST API

Using the SpaceX API to retrieve data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.

- Make a GET response to the SpaceX REST API
- Convert the response to a .json file then to a Pandas DataFrame
- Use custom logic to clean the data
- Define lists for data to be stored in
- Call custom functionsto retrieve data and fill the lists
- Use these lists as values in a dictionary and construct the dataset
- Create a Pandas DataFrame from the constructed dictionary dataset
- Filter the DataFrame to only include Falcon 9 launches
- Reset the FlightNumber column
- Replace missing values of PayloadMass with the mean PayloadMass
- value

EXPLORATORY DATA ANALYSIS (EDA) –SQL

- To gather some information about the dataset, some SQL queries were performed.
- The SQL queries performed on the data set were used to:
- 1.Display the names of the unique launch sites in the space mission
- 2.Display 5 records where launch sites begin with the string ‘CCA’
- 3.Display the total payload mass carried by boosters launched by NASA (CRS)
- 4.Display the average payload mass carried by booster version F9 v1.1
- 5.List the date when the first successful landing outcome on a ground pad was achieved
- 6.List the names of the boosters which had success on a drone ship and a payload mass between 4000 and 6000 kg
- 7.List the total number of successful and failed mission outcomes
- 8.List the names of the booster versions which have carried the maximum payload mass
- 9.List the failed landing outcomes on drone ships, their booster versions, and launch site names for 2015
- 10.Rank 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

GEOSPATIAL ANALYSIS –FOLIUM

- The followingstepsweretakentovisualizethelaunchdataonan interactive map:
- 1.Mark alllaunchsitesonamap
- •Initialise the map using a Folium Mapobject
- •Add a folium.Circleand folium.Markerforeachlaunchsiteonthelaunchmap
- 2.Mark the success/failed launches for each site on a map
- •As many launches have the same coordinates, it makes sense to cluster them together.
- •Beforeclustering them, assign a marker colour of successful (class = 1) as green, and failed (class = 0) as red.
- •To putthelaunchesintoclusters, for each launch, add afolium.Markerto the MarkerCluster()object.
- •Create an icon as a text label, assigning the icon_coloras themarker_colourdetermined previously.
- 3.Calculatethedistancesbetween alaunchsitetoitsproximities
- •To explore the proximities of launch sites, calculations of distances between points can be made using the Latand Longvalues.
- •After marking a point using the Latand Longvalues, create a folium.Markerobject to show the distance.
- •Todisplaythedistanceline between two points, draw a folium.PolyLineand add this to the map.

INTERACTIVE DASHBOARD –PLOTLY DASH

- The following plots were added to a Plotly Dash dashboard to have an interactive visualisation of the data:
- 1.Pie chart (px.pie()) showing the total successful launches per site
- •Thismakesitcleartoseewhichsitesaremostsuccessful
- •The chart could also be filtered (using a dcc.Dropdown()object) to see the success/failureratioforanindividualsite
- 2.Scattergraph(px.scatter()) to showthecorrelationbetweenoutcome (success or not) and payload mass (kg)
- •Thiscouldbefiltered(using a RangeSlider()object) by rangesofpayloadmasses
- •It could also be filtered by boosterversion

PREDICTIVE ANALYSIS -CLASSIFICATION

- This involved the following
  - Model Development
  - Model Evaluation
  - Finding the best classification model

CONCLUSIONS

- •As the number of flights increases, the rate of success at a launch site increases, with most early flights being unsuccessful. I.e. with more experience, the success rate increases.
- •Between 2010 and 2013, all landings were unsuccessful (as the success rate is 0).
- •After 2013, the success rate generally increased, despite small dips in 2018 and 2020.
- •After 2016, there was always a greater than 50% chance of success.
- •Orbit types ES-L1, GEO, HEO, and SSO, have the highest (100%) success rate.
  - •The 100% success rate of GEO, HEO, and ES-L1 orbits can be explained by only having 1 flight into the respective orbits.
  - •The 100% success rate in SSO is more impressive, with 5 successful flights.
  - •The orbit types PO, ISS, and LEO, have more success with heavy payloads:
  - •VLEO (Very Low Earth Orbit) launches are associated with heavier payloads, which makes intuitive sense.
- •The launch site KSC LC-39 Ahad the most successful launches, with 41.7% of the total successful launches, and also the highest rate of successful launches, with a 76.9% success rate.
- •The success for massive payloads (over 4000kg) is lower than that for low payloads.
- •The best performing classification model is the Decision Tree model, with an accuracy of 94.44%.