DS Capstone Project

Table of Contents

Executive Summary

- Summary of methodologies:

  - Apply craping data knowledge from API and webpages using BeautifulSoup

  - Use pandas, SQL, Matplotlib, Seaborn, Plotly Dash and other libraries for analyzing and visualizing data

  - Apply Sklearn library of Python for different ML models

- Summary of all results

  - Dataset collected

  - EDA results

  - Interactive Visual Analytics and Dashboards

  - ML models for prediction

Introduction

The commercial space industry is experiencing a revolution, with companies like SpaceX, Virgin Galactic, Blue Origin, and Rocket Lab making space travel more accessible and cost-effective. SpaceX, led by Allon Musk, has been particularly successful, with a track record of sending spacecraft to the International Space Station, launching the innovative Starlink satellite internet constellation, and conducting manned space missions. A significant factor in SpaceX's cost-efficiency is its ability to recover and reuse the first stage of its Falcon 9 rockets, a feat that many other providers cannot replicate. To compete with SpaceX, a new player in the industry, Space Y, aims to determine competitive pricing for rocket launches. This endeavor involves extensive data collection, the creation of informative dashboards, and the development of a machine learning model to predict the successful reusability of SpaceX's first stage based on publicly available data. These efforts will enable Space Y to make informed business decisions and take on the space launch market with confidence. In this capstone, we will predict if the Falcon 9 first stage will land successfully using data from its website.

Methodology

- Data collection from Webpages and API
- Perform data wrangling
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform geographical data & interactive visual analytics using Folium and Plotly Dash
- Predict results from different ML models
- Tuning model's parameters for best results

Data collection

- 2 sources:
    - SpaceX API (advertised on its webpage)
    - Wikipedia
- Knowledge required:
    - RESTful API
    - BeautifulSoup
    - Request library
    - HTML

Data Wrangling

- Data was filtered using the BoosterVersion column to only keep the Falcon 9 launches
- Missing data values in PayloadMass were replaced using mean value of column.

EDA

Perform data Analysis and Feature Engineering using Pandas and Matplotlib.

- Use scatter chart to visualize the relationship between Flight Number, Payload, Launch Site and Orbit type

- Use bar chart to visualize the success rate of each orbit type

- Use line chart to visualize the launch success yearly trend

Build an Interactive Map with Folium

- Created folium map to mark all launch sites, as well as map objects such as markers, circles, and lines to indicate the success or failure of launches for each launch site.

Build a Dashboard with Plotly Dash

- Adding a Launch Site Drop-down Input Component
- Adding a callback function to render success pie chart based on selected site dropdown
- Adding a Range Slider to Payload Selection

ML Models

- In this project, we apply 4 ML models: LR, SVM, Dicision Tree, KNN for prediction
- Apply GridSearchCV library for tuning parameters of each model

Result

	Model	Accuracy
0	Logistics Regression	0.833333
1	SVM	0.833333
2	Tree	0.833333
3	KNN	0.833333

Results

- Insights from EDA
- Interactive dashboard
- Model choice