Predictive Analysis of SpaceX Launch Outcomes

Applied Data Science Capstone Project

Presented by: Tuomas V

Date: 16th July 2024

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Conclusion

Executive Summary

- **Summary of methodoligies:** Data collection from API and webscraping, Exploratory analysis with SQL and visualization, Machine learning prediction
- Objective was to analyze and predict the outcomes of SpaceX Falcon 9 rocket launches.
- Key factors influencing launch success were identified, and predictive models were developed.
- Successful prediction of launch outcomes with insights for future improvements were found.

Introduction

- **Background:** SpaceX aims to make space travel more cost-effective and reliable.
- **Problem Statement:** Predicting the success of Falcon 9 launches based on historical data.
- **Goals:** Improve prediction accuracy and identify factors correlating with launch success.

Methodology

- **Data Collection:** SpaceX API and web scraping.
- **Data Wrangling:** Cleaning and preprocessing using Python libraries.
- **Exploratory Data Analysis (EDA):** Visualizing data trends and distributions.
- **Machine Learning Models:** Logistic Regression, Decision Trees, Support Vector Machines (SVM).
- **Tools:** Jupyter Notebooks, Pandas, Matplotlib, Seaborn, Scikit-Learn, SQL, Github.

Methodology: Data Collection

- Sources used were SpaceX API and web scraping Wikipedia.
- Tools used in data collection were Jupyter Notebooks and Python libraries like Requests and BeautifulSoup.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/jupyter-labs-spacex-data-collection-api.ipynb and https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/jupyter-labs-webscraping.ipynb

Methodology: Data Wrangling

- Data wrangling is cleaning and preprocessing data.
- Pandas library was used for data manipulation.
- Challenges included handling missing values and normalizing data.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/labs-jupyter-spacex-Data%20wrangling.ipynb

Methodology: Exploratory Data Analysis (EDA)

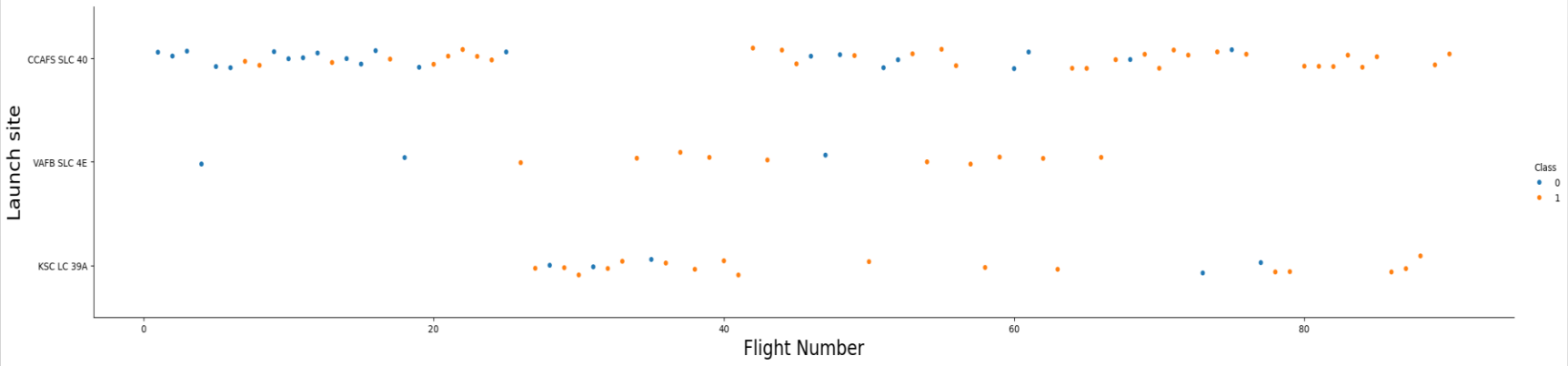
- EDA is visualizing data distributions and relationships and using them to get initial insight into the dataset.
- Tools used for EDA included Matplotlib, Seaborn and SQL.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb and https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

Methodology: Machine Learning Models

- Machine Learning Models used were Logistic Regression, Decision Trees, Support Vector Machines (SVM) and K-Nearest Neighbour (KNN) from the Scikit-Learn library.
- Evaluation Metrics for finding the most suitable model were Accuracy, Precision, Recall, F1 Score.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Results: EDA - Visualizations

- Many visualizations were made of the data in order to get an understanding of it.
- For example, from this graph we can see Falcon 9 launches by launch site. The orange dots indicate successful launches and blue ones unsuccessful ones.



Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/edadataviz.ipynb

Results: EDA - SQL

- Using SQL in EDA allowed extracting a lot of categorical and numerical data from the data set.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/jupyter-labs-eda-sql-coursera_sqllite.ipynb

Landing_Outcome	Outcome_Count
No attempt	10
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Controlled (ocean)	3
Uncontrolled (ocean)	2
Failure (parachute)	2
Precluded (drone ship)	1

Results: Folium mapping

- Folium maps were used to add circles and markers to launch sites and calculating distances between points of interest.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/lab_jupyter_launch_site_location.ipynb

Results: Dash

- An interactive app with visualizations was made using Plotly Dash
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/spacex_dash_app.py

Result: Predictive Modelling

- **Model Performance:** Evaluated using accuracy, precision, recall, and F1 score.
- Best Model could not be reliably determined as all models tested gave the same accuracy result, likely due to the relatively small sample size.
- Significant factors correlating to success include launch site, payload mass, and booster version.
- Code: https://github.com/hmllss/Applied-Data-Science-Capstone/blob/main/SpaceX_Machine%20Learning%20Prediction_Part_5.ipynb

Conclusion

- Success rate of Falcon 9 launches has increased over time.
- KSC LC-39A is the most successful launch site.
- Heavier payloads have a higher success rate though it may be due to payload being increased alongside other development in reliability.
- Full Github reposotory: https://github.com/hmllss/Applied-Data-Science-Capstone

Discussion

- Launch site and payload mass appear to be critical factors for predicting success.
- **Challenges:** Dealing with the sample size being relatively small and handling missing data and model overfitting.
- In the future repeating the steps with a larger sample size would be useful and integrating additional data sources and enhancing model complexity.