SpaceX Falcon 9 First Stage Landing Prediction

IBM Data Science Capstone Project

Executive Summary

- Objective: Predict whether the Falcon 9 first stage will successfully land.
- Motivation: Successful landings significantly reduce launch costs.
- Approach: Data collection, EDA, visualization, and machine learning classification.
- Outcome: Built predictive models with strong classification performance.

Introduction

- SpaceX aims to reuse rocket boosters to lower space launch costs.
- Landing success depends on mission parameters and launch conditions.
- This project analyzes historical launch data to predict landing outcomes.

Data Collection

- SpaceX REST API used to collect launch and rocket data.
- Web scraping performed on Wikipedia for additional mission details.
- Multiple datasets merged into a unified analysis table.

Data Wrangling

- Handled missing values and inconsistent formats.
- Converted categorical variables to numerical representations.
- Defined target variable: successful landing (1) or failure (0).

EDA Methodology

- Statistical summaries to understand distributions.
- Visualization using Matplotlib and Seaborn.
- Interactive analysis using Plotly and Folium.

EDA – Visualization Results

- Launch success rate increased over time.
- Certain launch sites show higher landing success.
- Payload mass influences landing probability.

EDA with SQL – Methodology

- Loaded cleaned dataset into SQLite database.
- Used SQL queries for aggregation and filtering.
- Analyzed success rates by site, orbit, and booster version.

EDA with SQL – Results

- KSC LC-39A showed the highest landing success rate.
- LEO missions had higher success compared to other orbits.
- Newer booster versions improved landing outcomes.

Interactive Map – Folium

- Mapped launch sites using geographic coordinates.
- Visualized success and failure markers.
- Distance analysis from coastlines and infrastructure.

Plotly Dash Dashboard

- Interactive dashboard for payload vs. success outcome.
- Dropdowns to filter by launch site.
- Scatter plots and pie charts for dynamic exploration.

Predictive Analysis – Methodology

- Framed problem as a binary classification task.
- Split data into training and testing sets.
- Evaluated Logistic Regression, SVM, KNN, and Decision Trees.

Predictive Analysis – Results

- Decision Tree and SVM achieved highest accuracy.
- Cross-validation used to avoid overfitting.
- Model effectively predicts landing success.

Conclusion

- Landing success can be predicted using historical launch data.
- Data-driven insights highlight key success factors.
- Future work: include weather and real-time telemetry data.