SpaceX Falcon 9 Launch Analysis & Prediction

A Data Science Capstone Project

Executive Summary

This project explores SpaceX Falcon 9 launch data to understand patterns in rocket launches, perform exploratory data analysis (EDA), build interactive dashboards, and develop classification models to predict the success of rocket landings. The findings contribute towards reducing launch costs and support decision-making in the aerospace domain.

Introduction

SpaceX has revolutionized space travel by creating reusable rockets. Predicting rocket landing success can significantly lower costs. This project analyzes Falcon 9 data with the goal of predicting the outcome of landings and visualizing insights through various tools like Folium, Plotly Dash, and machine learning models.

Data Collection & Wrangling

• Data collected from SpaceX API and Wikipedia • Cleaned and structured into pandas DataFrames • Missing values handled (PayloadMass replaced with mean values) • One-hot encoding applied to categorical variables (Orbit, LaunchSite, LandingPad, Serial)

Exploratory Data Analysis (EDA) & Visualization Methodology

• Data distribution visualized using seaborn and matplotlib • Success rates analyzed by Orbit type, Launch Sites, and Payload ranges • Interactive geospatial analysis using Folium maps • Interactive dashboards built with Plotly Dash

SQL Analysis Results

SQL queries used to analyze launch data: • Top launch sites by frequency • Average payload mass by orbit • Success rate comparisons across landing pads

Interactive Map with Folium

• Mapped launch sites with markers and popups • Circle markers represent launch success probability • Cluster markers used for overlapping launch coordinates

Plotly Dash Dashboard Results

• Interactive payload vs. success rate scatter plots • Dropdown filters for launch sites and orbits • Visualization of success trends over flight numbers

Predictive Analysis Methodology

• Machine Learning models applied: Logistic Regression, Decision Trees, KNN, and SVM • Hyperparameter tuning performed using GridSearchCV • Best models selected based on validation accuracy

Predictive Analysis Results

• Decision Tree achieved ~93% accuracy on test set • SVM with RBF kernel provided strong results • Logistic Regression provided baseline comparison • Predictions help estimate landing success probability for new launches

Conclusion

The project successfully demonstrated the process of collecting, wrangling, analyzing, and modeling SpaceX launch data. Predictive models showed high accuracy in determining landing outcomes, and interactive dashboards/maps improved interpretability. These insights contribute to reducing space launch costs and support innovation in aerospace analytics.