IBM Data Science Professional Certificate – Capstone Project

Executive Summary

- This presentation summarizes the full data science pipeline applied to analyze and predict Falcon 9 launch success. It includes data collection, wrangling, EDA, visualization, modeling, and dashboarding.

Introduction

- The objective is to determine key factors influencing the success of Falcon 9 launches and develop a predictive model. Data is sourced from SpaceX and includes mission details, launch site, payload mass, and orbit.

Data Collection & Wrangling

- • Data collected using SpaceX API and Wikipedia tables.
- • Duplicates removed, missing values handled.
- • Columns renamed and formatted for consistency.
- • Feature engineering performed to extract relevant data (e.g. binary outcome labels).

EDA & Interactive Visual Analytics - Methodology

- • Univariate and bivariate analysis with Seaborn & Matplotlib.
- • Visual correlation matrix to identify feature relationships.
- • Interactive charts using Plotly to explore trends in launch success.

Predictive Analysis - Methodology

- • Classification task with Logistic Regression, SVM, Random Forest.
- • Data split: 80% training / 20% test.
- • GridSearchCV used for hyperparameter tuning.
- • Evaluation with Accuracy, Precision, Recall, F1-score.

EDA with Visualization - Results

- • Payload mass vs. success analyzed.
- • Orbit type significantly impacts outcome.
- • Pie charts and box plots used to visualize categorical and continuous features.

EDA with SQL - Results (1/2)

- • SQL used via IBM Db2 magic commands in Jupyter.
- • Count of launches per site.
- • Average payload mass per orbit.

EDA with SQL - Results (2/2)

- • Join queries performed across datasets.
- • Grouped metrics by launch site and mission outcome.
- • Queries support hypotheses drawn in EDA.

Folium Map - Interactive Visualization

- • Folium map created to show launch sites.
- • Markers used with success rate popups.
- • Map highlights geographical patterns in success.

Plotly Dash - Interactive Dashboard

- • Dash app built to interactively explore data.
- • Users can select site, payload range.
- • Dynamic charts update based on filters.

Predictive Analysis - Results

- • Best performing model: Random Forest.
- • Achieved Accuracy: ~92%
- • Confusion matrix and classification report evaluated.
- • Importance of features like orbit, payload shown.

Conclusion

- • Data-driven insights show orbit and payload mass influence launch success.
- • Predictive modeling achieved high accuracy.
- • Tools like Folium and Dash enhance understanding.
- • Project demonstrates end-to-end data science pipeline.

Creativity & Innovation

- • Extended the template with animated visualizations.
- • Integrated external data (payload types) for deeper insight.
- • Created a custom dashboard layout in Dash.