Falcon 9 First Stage Landing Prediction Using Machine Learning

IBM Data Science Capstone

Presented by: Sai Sanjana Mummaneni

Date: July 2025

Executive Summary

- Predicting first-stage landing success using ML
- Methods: Data wrangling, EDA, modeling
- Tools: Python, Pandas, Sklearn, Dash, Folium
- Best accuracy: 83.33% (LogReg/Decision Tree)

Introduction

- Problem: Reduce cost via booster reusability
- Target: Predict success/failure of landing
- Data: SpaceX launch records from IBM
- Flow: Cleaning → EDA → Modeling → Dash

Data Collection and Wrangling

- Source: IBM CSVs
- Merged 3 datasets
- Encoded Orbit, LaunchSite, LandingPad, Serial
- Final Features: Numeric, Binary, Encoded

EDA & Visual Analytics

- Tools: matplotlib, seaborn
- Class Distribution, Payload vs Class, Site/Orbit vs Outcome
- Heatmaps to reveal correlation
- Goal: Identify important features

SQL-Based EDA

- Ran queries on IBM DB2
- Launch counts by site, Min/Max Payload
- 41 landings on drone ships (True ASDS)
- Visualized queries with charts

Folium Interactive Map

- Used folium, MarkerCluster, CircleMarker
- Mapped CCAFS, KSC, VAFB launch sites
- Success/failure via color-coded markers
- Zoom, popups for interactivity

Plotly Dash Dashboard

- Filters: Launch Site, Year
- Visuals: Pie, Bar, Line charts
- Interactive & responsive
- Helps non-tech users explore trends

Modeling Methodology

- Label: Class (0=Fail, 1=Success)
- Models: LogReg, SVM, Tree, KNN
- Used GridSearchCV with 10-fold CV
- Scaled data, train-test split (80/20)

Classification Results

- LogReg & Tree: 83.33%
- SVM: 77.77%, KNN: 66.66%
- Confusion Matrix → LogReg & Tree: TP=12, FP=3
- KNN had lowest recall

Conclusion

- Best models: LogReg & Tree
- Important features: Payload, Orbit, LandingPad, Reused
- GridSearch & scaling improved performance
- Dashboard/map: usable insights

Appendix

- Extra: Feature importance, SQL charts, Confusion matrices

- Tools: pandas, seaborn, sklearn, dash