SpaceX Falcon 9 Launch Success Prediction

IBM Data Science Capstone Project

https://github.com/Anandapriya-T/Coursera

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Objective: Predict Falcon 9 first-stage landing success
- Data collected from SpaceX API and Wikipedia
- Performed EDA, SQL analysis, interactive visualizations, and ML modeling
- Best model: Support Vector Machine with RBF kernel
- Outcome supports cost reduction insights for space launches

INTRODUCTION

- SpaceX aims to reduce launch costs via booster reuse
- Predicting landing success is critical for mission planning
- This project applies data science and ML to analyze launch outcomes

Data Collection & Wrangling Methodology

- Data sources: SpaceX REST API and Wikipedia
- Handled missing values and inconsistent formats
- Converted categorical variables using One-Hot Encoding
- Standardized numerical features

EDA & Interactive Visual Analytics Methodology

- Used Pandas and Seaborn for exploratory analysis
- SQL used to aggregate and filter structured data
- Folium for geospatial visualization
- Plotly Dash for interactive dashboards

Predictive Analysis Methodology

- Target variable: Class (Landing success)
- Train-test split: 80% training, 20% testing
- Models: Logistic Regression and SVM
- Hyperparameter tuning with GridSearchCV (cv=10)

EDA with Visualization Results

- Higher success rates observed at KSC LC-39A
- Payload mass shows non-linear relationship with success
- Certain orbits have consistently higher success probability

EDA with SQL Results

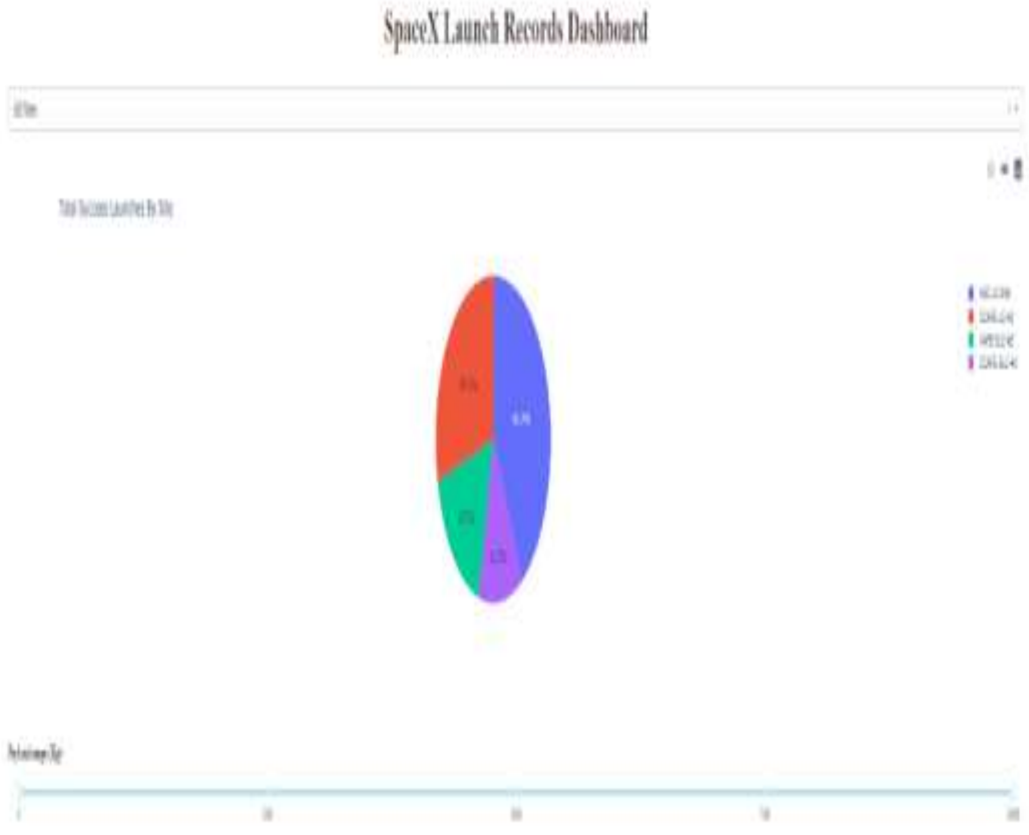
- SQL queries used to compute success rates by site and orbit
- KSC LC-39A recorded highest overall success rate
- Insights validated EDA findings

Interactive Map Results (Folium)

- Mapped launch sites with success/failure markers
- Calculated distances to coastline, railways, and roads
- Geographical proximity influences recovery feasibility

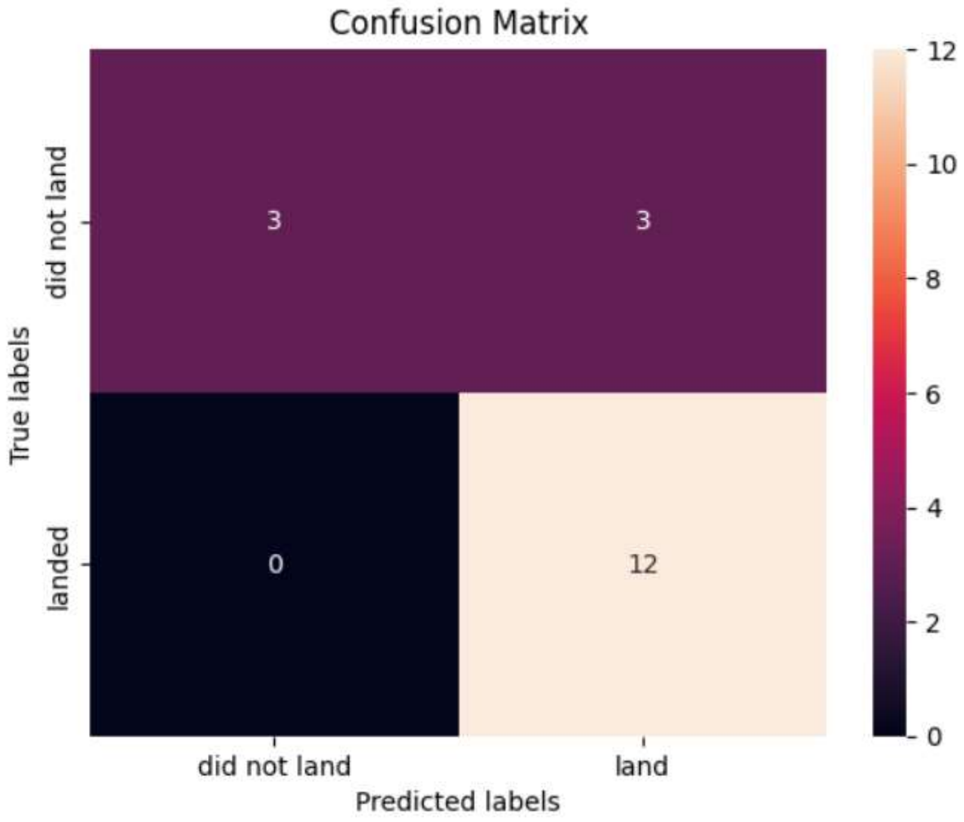
Plotly Dash Dashboard Results

- Interactive dashboard with dropdown filters
- Visual comparison of launch success by site and payload
- Improved stakeholder exploration of data



Predictive Analysis Results

- Logistic Regression accuracy evaluated
- SVM with RBF kernel achieved best validation performance
- Model generalizes well on unseen test data



Conclusion

- Machine learning effectively predicts launch success
- Geography and payload significantly affect outcomes
- SVM (RBF) is the best-performing model
- Future work: include weather and temporal features

Creativity & Innovative Insights

- Integrated multiple analytics techniques into one pipeline
- Linked geographic constraints to engineering outcomes
- Dashboard enables decision-making beyond static analysis