IBM Data Science Capstone Project

Ahmet Mahmut Gökkaya

2025

Contents

1 Executive Summary

This report documents the IBM Data Science Capstone Project, which focuses on analyzing historical Falcon 9 rocket launches by SpaceX to predict the likelihood of successful booster landings. The project involves data collection, wrangling, exploratory data analysis (EDA), predictive modeling, and visualization using various data science tools and techniques.

1.1 Project Components

- **Data Collection:** API extraction and web scraping.
- **Data Wrangling:** Cleaning, transformation, and feature engineering.
- **EDA:** Visualization and SQL-based analysis.
- **Predictive Analysis:** Machine learning models for classification.
- **Interactive Maps:** Folium-based geographic visualizations.
- **Interactive Dashboard:** Built using Plotly Dash.

2 Introduction

2.1 Background

SpaceX is a private aerospace company that has revolutionized the satellite launch industry by drastically reducing costs—up to 70% lower than competitors—through reusable booster technology. Successful booster landings are crucial for cost efficiency.

2.2 Problem Statement

The objective of this project is to predict whether a Falcon 9 booster will successfully land based on various factors such as:

- Launch site location.
- Payload orbit and mass.
- Landing pad location.
- Version of the booster.

3 Methodology

3.1 Data Collection

REST API: Extracted JSON data from SpaceX API, converted into a Pandas DataFrame using normalization.
Web Scraping: Scraped Falcon 9 launch data from Wikipedia.

1. API Request → JSON Normalization → DataFrame
2. Web Scraping → Data Extraction → CSV Storage

3.2 Data Wrangling

Data cleaning and transformation included:

- Handling missing values.
- Encoding categorical variables.
- Feature engineering for predictive modeling.

3.3 Exploratory Data Analysis (EDA)

EDA techniques were applied to identify trends:

- Flight number vs. launch site.
- Payload mass vs. launch site.
- Success rate by orbit type.
- Success rate trends over time.

4 Results

4.1 Key Insights from EDA

- Most launches occurred at CCAFS-SLC-40.
- Higher payloads were commonly launched to VLEO orbit.
- Success rates increased significantly over time.
- Orbit types like GEO and HEO had higher success rates.

4.2 Machine Learning Models

Several classification models were trained to predict booster landing success:

- **K-Nearest Neighbors (KNN)**
- **Decision Tree**
- **Logistic Regression**
- **Support Vector Machine (SVM)**

The best-performing model was **KNN**, achieving:

- Accuracy: **77%**
- Best Score: **87%**

5 Interactive Visualizations

5.1 Folium Maps

- Interactive maps showing launch sites.
- Visual representation of successful vs. failed landings.

5.2 Plotly Dash Dashboard

- Launch site success rate visualization.
- Payload success rate by range slider.
- Filtering launches by booster version.

6 Conclusion

This capstone project successfully analyzed historical SpaceX Falcon 9 launch data to develop predictive models for booster landing success. Key takeaways:

- Launch site, payload mass, and orbit type significantly impact booster success.
- The best machine learning model (KNN) achieved a high accuracy of 77%.
- Interactive visualizations provided meaningful insights for analysis.

6.1 Future Work

- Incorporate real-time SpaceX API data updates.
- Implement deep learning techniques for improved prediction accuracy.
- Extend analysis to Falcon Heavy launches.

A Appendix

A.1 Project Links

- **Data Collection:** GitHub Repository
- **Data Wrangling:** GitHub Repository
- **EDA:** GitHub Repository
- **Predictive Analysis:** GitHub Repository
- **Interactive Dashboard:** GitHub Repository

A.2 References

- SpaceX API: https://api.spacexdata.com
- Wikipedia Falcon 9 Data: https://en.wikipedia.org/wiki/List_of_Falcon_9_and_Falcon_Heavy_launches
- Scikit-learn Documentation: https://scikit-learn.org