Predicting Falcon 9 First-Stage Landing Success

Author: Michelle Thys

Date: December 8, 2025

Table of contents

Executive Summary

This project aimed to predict the success of Falcon 9 first-stage landings, a key factor in SpaceX's cost advantage through rocket reusability. Using data collected from the SpaceX API and enriched with Wikipedia sources, we prepared a clean dataset and performed exploratory analysis to identify patterns in landing success across payload ranges, orbit types, and launch sites. Interactive visualizations and dashboards provided insights into geographic and operational trends.

Four machine learning models—Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN) —were evaluated using cross-validation, test accuracy, and ROC–AUC. While most models achieved similar accuracy, SVM delivered the highest ROC–AUC score (0.9583), indicating superior ability to distinguish between successful and failed landings. This makes SVM the most reliable choice for practical prediction.

Despite promising results, limitations remain, including a small dataset, class imbalance, and lack of temporal features. Future work should focus on expanding the dataset and incorporating time-based variables to improve predictive power. These enhancements will enable stakeholders to better estimate costs and make informed decisions in the competitive space launch market.

Introduction

SpaceX has transformed space travel by introducing the reuse of the Falcon 9 rocket's first stage, significantly lowering launch costs. The ability to predict whether a landing will be successful is valuable, as it enables competitors to better estimate costs and make informed decisions when bidding for contracts. This research set out to answer two main questions: Can we accurately predict the success of a Falcon 9 first-stage landing, and is it possible to develop a predictive model that supports cost estimation and competitive strategies?

Methodology

The methodology for this project involved several key steps, starting with comprehensive data collection from two primary sources: the SpaceX API and Wikipedia. The SpaceX API provided structured JSON data on launches, booster versions, payload mass, and landing outcomes. This data was retrieved using Python's requests library and converted into a Pandas DataFrame for analysis.
To enrich the dataset and fill gaps, additional information was scraped from Wikipedia using HTTP requests and HTML parsing techniques. The scraped data included launch dates, sites, booster details, payload mass, orbit type, and landing results.

Once both datasets were obtained, they were merged into a single DataFrame. Data wrangling steps included inspecting for consistency, handling missing values—particularly payload mass—by imputing averages and creating a target variable to indicate landing success (1 for success, 0 for failure).
Exploratory Data Analysis (EDA) was then conducted using a combination of visualizations and SQL queries. Scatter plots, bar charts, and line graphs were generated to uncover patterns in success rates across orbit types, payload ranges, and launch sites, as well as trends over time. SQL queries were used to compute statistics, such as total payload by booster type and success rates by site.

Interactive analytics were implemented using Folium maps to visualize the geographic distribution of launch sites and their proximity to infrastructure like highways and coastlines. Plotly Dash was employed to build a dashboard featuring filters for launch sites and payload ranges, pie charts for success rates, and scatter plots for payload-performance relationships.

Finally, predictive modeling was performed using four classification algorithms: Logistic Regression, Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors (KNN). The dataset was standardized, split into training and test sets, and hyperparameter tuning was conducted via grid search. Model performance was evaluated using accuracy scores, ROC–AUC score and confusion matrices.

Results & Discussion

The analysis revealed several important insights. First, success rates varied significantly across orbit types and launch sites. For example, KSC LC-39A emerged as the most reliable site with a success rate of 76.9%. Orbit type also played a role, with certain orbits consistently associated with higher landing success.
Over time, the overall success rate improved markedly, as shown by the yearly trend analysis (figure 1), reflecting technological advancements and operational refinements.

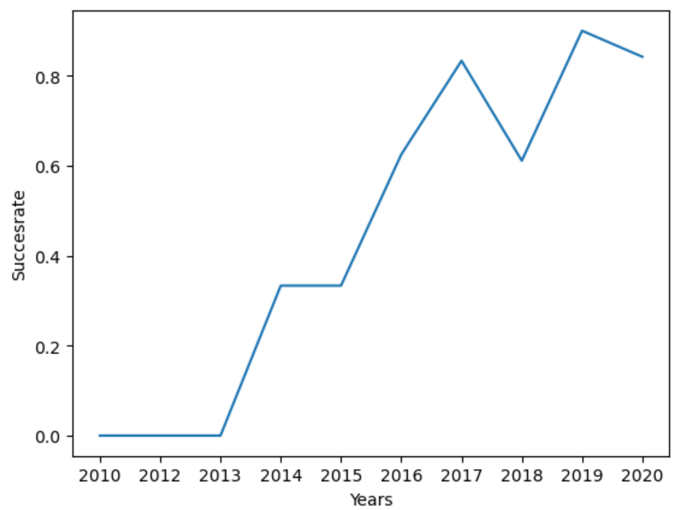

Figure 1: Launch success yearly trend

Predictive modeling confirmed that landing success can be forecasted with reasonable accuracy. Among the models tested (Logistic Regression, KNN, Decision Tree, and SVM) the SVM demonstrated the most robust performance.
While most models achieved similar test accuracy (approximately 83%), SVM stood out with the highest ROC–AUC score (0.9583), indicating superior ability to distinguish between successful and failed landings across all thresholds. This metric is particularly important given the class imbalance in the dataset, where successful landings dominate.
The confusion matrix for SVM (figure 2) confirmed consistent classification with fewer misclassifications compared to Decision Tree, which showed weaker generalization and a lower ROC–AUC score. These results suggest that SVM not only performs well at a fixed threshold but also ranks predictions effectively across varying decision boundaries.

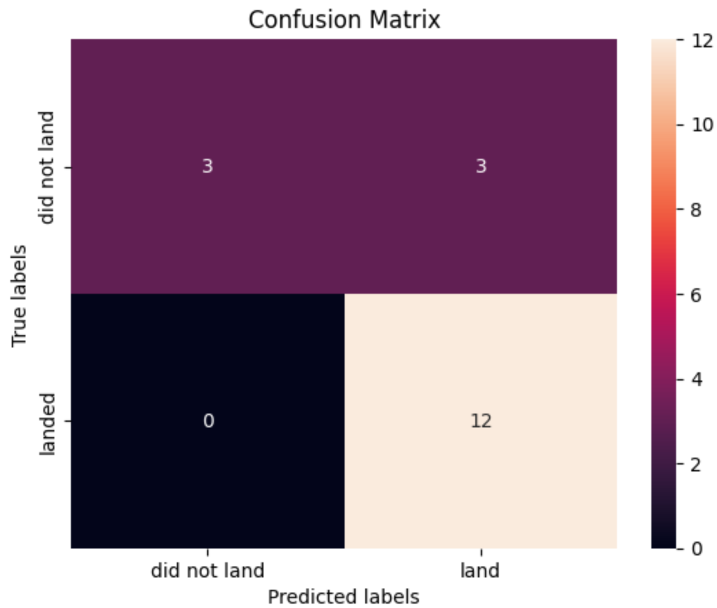

Figure 2: Confusion matrix of the SVM model

Overall, the results highlight the importance of launch site selection, orbit type, and payload considerations in predicting landing outcomes. They also underscore the need for larger datasets and more balanced classes to improve model robustness.

Conclusion

This project demonstrates that predicting Falcon 9 first-stage landing success is achievable with reasonable accuracy using machine learning models. Among the tested algorithms, the Support Vector Machine (SVM) emerged as the most reliable choice, achieving the highest ROC–AUC score and strong test accuracy. Its ability to distinguish between successful and failed landings across all thresholds makes it particularly suitable for practical applications.

Despite these promising results, the analysis revealed limitations, including a relatively small dataset, class imbalance favoring successful landings, and the lack of temporal features to capture technological improvements over time. Future work should focus on expanding the dataset, incorporating time-based variables, and addressing skewness to enhance predictive power. By refining these models and improving data quality, stakeholders can better estimate costs and make informed decisions in the competitive space launch market.

References

The links used to retrieve data from online sources:

- SpaceX API
    - https://api.spacexdata.com/v4/rockets/
    - https://api.spacexdata.com/v4/launchpads/
    - https://api.spacexdata.com/v4/payloads/
    - https://api.spacexdata.com/v4/cores/
    - https://api.spacexdata.com/v4/launches/past
- Wikipedia
    - https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922

GitHub Repository: https://github.com/michellethys2/IBM-datascience-capstone-project