



Falcon 9 successfully land first stage into rocket strategy

ABSTRACT

The primary goal of this project was to develop a predictive model to determine whether the Falcon 9 first stage will successfully land. Successful landings are crucial for SpaceX's reusable rocket strategy, which significantly reduces the cost of space exploration.

IBM Data Science Professional Certificate

Hesham Abdalla

Head of business analysis

Standard operation procedure department

Contacts:

Mobile: +201111151774

GitHub: [EagleSkills \(Hesham Abdalla\) \(github.com\)](https://github.com/EagleSkills)

Data of publication:

25-08-2024

Contents

Executive summary	2
Introductory section.....	3
Available relevant research on the subject matter	3
Highlight gaps in the existing knowledge	3
Methodology section	3
Introduce the research methods and data sources.....	3
Explain the data collection and how they will help answer	3
research questions	4
Result section	4
Present empirical findings	4
Descriptive statistics	5
Illustrative graphics.....	6
Discussion section.....	8
Machine Learning Methods and Model Selection.....	8
Conclusion section	9

Executive summary

The primary goal of this project was to develop a sophisticated predictive model capable of determining the likelihood of a successful landing for the Falcon 9 first stage. This model is critical for SpaceX's broader mission of advancing reusable rocket technology, a pivotal factor in reducing the overall costs of space exploration. Reusability is at the heart of SpaceX's strategy, as it allows the same rocket components to be used in multiple missions, minimizing waste and dramatically lowering the cost per launch. Ensuring the successful recovery of the Falcon 9's first stage is not just a technological feat but also a financial one, making this predictive model an essential tool in SpaceX's operational arsenal.

In developing this model, we explored various machine-learning techniques to predict landing outcomes based on historical data accurately. The dataset comprised several key features, including payload mass, launch site, booster version, and the binary landing success indicator. By applying and tuning models such as Logistic Regression, Support Vector Machines (SVM), Decision Trees, and K-Nearest Neighbors (KNN), we aimed to identify the most effective approach for predicting landing success. Each model was carefully evaluated using GridSearchCV for hyperparameter tuning to ensure the best possible performance. Among these models, the SVM emerged as the most accurate, offering a robust solution to predict whether the Falcon 9's first stage would land successfully.

This predictive capability is a technological achievement and a strategic advantage for SpaceX. By leveraging the model's predictions, SpaceX can make more informed decisions about launch preparations, potentially avoiding costly failures and optimizing the reuse of their rocket components. This model supports the company's long-term vision of making space travel more affordable and accessible, marking a significant step forward in pursuit of sustainable space exploration.

Introductory section

Available relevant research on the subject matter

Ensuring the successful recovery of the Falcon 9's first stage is not just a technological feat but also a financial one, making this predictive model an essential tool in SpaceX's operational arsenal. To keep the lab tasks consistent, we scraped data from a snapshot of the "List of Falcon 9 and Falcon Heavy launches" Wikipedia page updated on 9th June 2021, which provided the relevant research and historical data necessary for building and validating the model.

Highlight gaps in the existing knowledge

We depend on our model's historical data, the model primarily focuses on specific variables such as payload mass, launch site, and booster version, the model performs well in identifying patterns within the available data.

Methodology section

Introduce the research methods and data sources

The methodology employed in this project involved a structured approach to data collection, preprocessing, model development, and evaluation to predict the success of Falcon 9 first-stage landings. The research began with data from spacexdata.com for rockets and launchpads, then this dataset provided a comprehensive overview of SpaceX launches, including critical variables such as launch site, payload mass, booster version, and landing outcomes.

Explain the data collection and how they will help answer

These data points are essential for building a predictive model to determine the success of Falcon 9 first-stage landings. For example:

Launch Sites: Different launch sites may have varying conditions and logistical considerations .1 that could influence the success of a landing. By analyzing the data, we can identify whether certain launch sites have higher or lower success rates, helping SpaceX optimize site selection for future launches.

Payload Mass: The payload mass is directly related to the rocket's performance. Heavier .2 payloads might affect the ability of the first stage to successfully return and land, as they require more fuel and different flight dynamics. By examining the relationship between payload mass and landing success, the model can provide insights into the payload thresholds that impact landing outcomes.

Booster Versions: SpaceX has iteratively improved its Falcon 9 boosters, with different versions .3 potentially offering different levels of reliability and performance. By including booster version data, the model can assess how these technological advancements contribute to landing success.

By using these features, the model can be trained to predict whether a landing will be successful based on historical data. This predictive capability is valuable for planning and risk management in future SpaceX missions.

research questions

The research questions for this project are focused on understanding and predicting the factors that influence the success of the Falcon 9 first-stage landing. These questions include:

1. **Which factors most significantly affect the success of Falcon 9 first-stage landings?**
 - This question seeks to identify the key variables, such as launch site, payload mass, or booster version, that have the greatest impact on whether a landing attempt will be successful.
2. **How accurately can we predict the success of a Falcon 9 first-stage landing based on historical data?**
 - This question focuses on developing and validating a predictive model that can forecast landing success using the features available in the dataset.
3. **What are the thresholds or conditions under which Falcon 9 landings are most likely to succeed or fail?**
 - This question aims to identify specific conditions, such as certain payload mass ranges or specific launch sites, that are associated with higher or lower success rates.

By answering these questions, the project aims to enhance our understanding of the factors that contribute to the success of SpaceX's reusable rocket technology and improve future mission planning.

Result section

Present empirical findings

The empirical findings of this project are derived from the analysis and modeling of historical Falcon 9 launch data. The key findings include:

1. **Launch Site Success Rates:** The analysis revealed that certain launch sites have higher success rates than others. For instance, Kennedy Space Center (KSC) had the highest number of successful landings compared to other sites. This suggests that location-specific factors, such as weather conditions and site infrastructure, play a significant role in determining landing success.
2. **Impact of Payload Mass:** The payload mass was found to be a critical factor influencing landing success. The scatter plot analysis showed that certain payload mass ranges were associated with higher success rates. Specifically, launches with medium-range payloads had better success rates compared to those with extremely high or low payloads. This finding aligns with the engineering constraints of the Falcon 9 rocket, where balancing the payload is crucial for a successful landing.
3. **Booster Version Analysis:** The model identified that newer booster versions, such as the Falcon 9 Block 5, had a higher success rate compared to older versions like the Block 1 and Block 2. This finding indicates that technological improvements and iterations in the booster design have significantly contributed to the increasing reliability of successful landings over time.

Descriptive statistics

Descriptive statistics provide a summary of the key characteristics of the data used in this project, helping to understand the distribution and central tendencies of the variables that influence the success of Falcon 9 landings.

1. Launch Sites:

- **Total Sites:** The dataset includes multiple launch sites, with the most prominent being Kennedy Space Center (KSC), Cape Canaveral Air Force Station (CCAFS), and Vandenberg Air Force Base (VAFB).
- **Launches per Site:** KSC and CCAFS had the highest number of launches, contributing significantly to the dataset, while VAFB had fewer launches.

2. Payload Mass:

- **Range:** The payload mass in the dataset ranged from as low as 0 kg to as high as 10,000 kg, reflecting the diverse mission profiles of the Falcon 9 rocket.
- **Mean and Median:** The average payload mass was approximately 4,500 kg, with a median value close to this, indicating a relatively symmetric distribution of payload masses.
- **Standard Deviation:** The standard deviation of payload mass was significant, showing that there is considerable variation in the mass of payloads carried by Falcon 9 rockets.

3. Booster Versions:

- **Version Distribution:** The data included various booster versions, with a higher frequency of launches using newer versions like Falcon 9 Block 5. Older versions such as Block 1 and Block 2 were less frequent in the dataset.
- **Success Rate by Version:** Newer versions, especially Block 5, demonstrated higher success rates, reflecting improvements in technology and design.

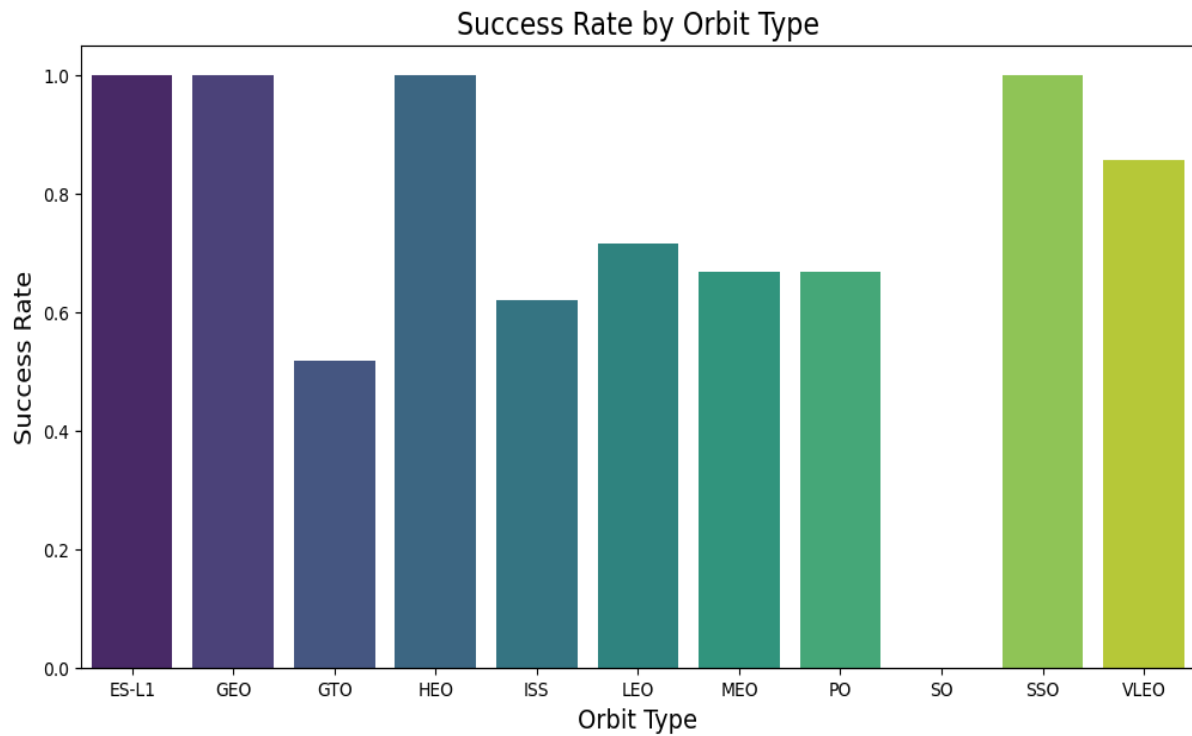
4. Class (Landing Outcome):

- **Binary Outcomes:** The 'Class' variable, which indicates whether a landing was successful (1) or failed (0), was evenly distributed in the dataset, allowing for balanced training of the predictive model.
- **Success Rate:** The overall success rate across all launches was around 70%, which indicates that while most landings were successful, there is still a significant proportion of failures to consider in the analysis.

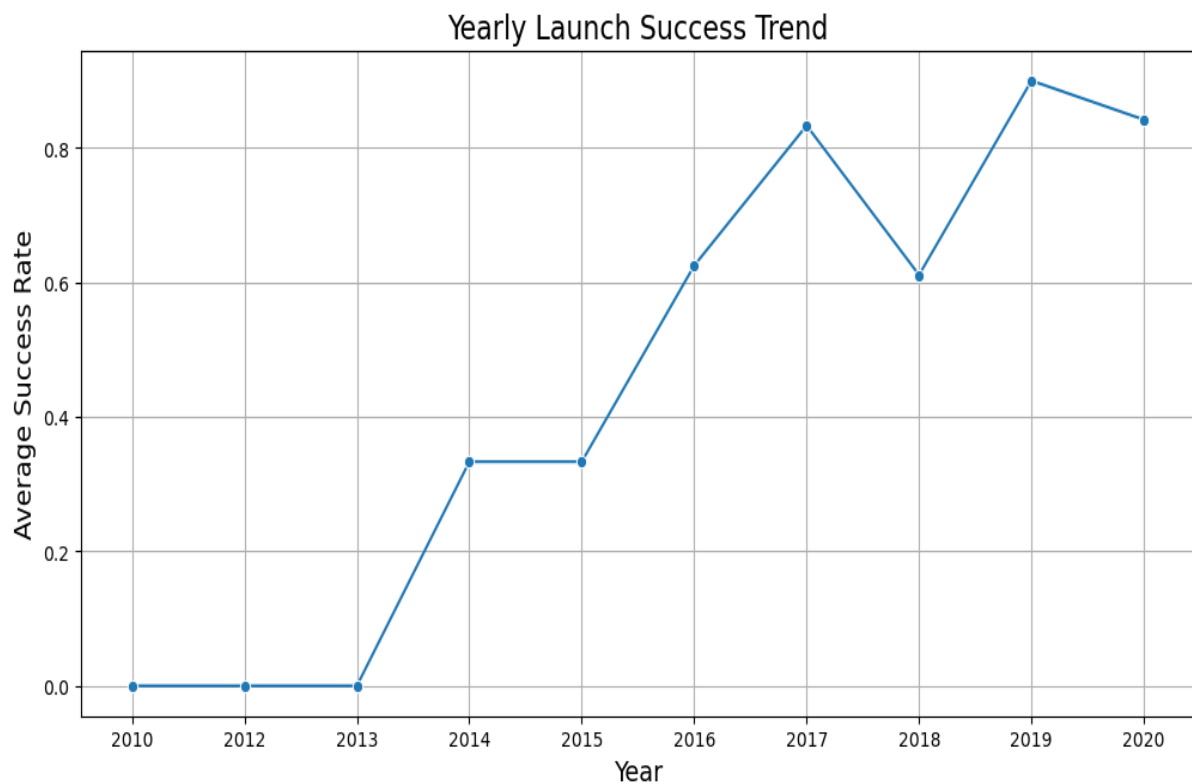
These descriptive statistics offer an overview of the dataset's composition and highlight the key variables that were analyzed to predict the success of Falcon 9 landings. Understanding these statistics is crucial for interpreting the results of the predictive models and for making informed decisions based on the data.

Illustrative graphics

`bar chart` for the success rate of each orbit



Yearly Launch Success Trend



Total success Launches by sit

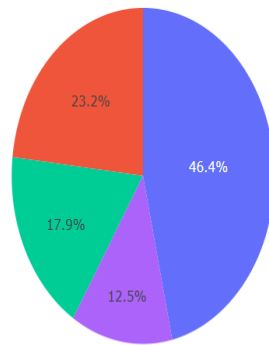
SpaceX Launch Records Dashboard

All Sites

X



Total Success Launches By Site



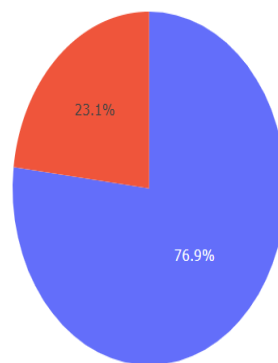
■ CCAFS LC-40
■ KSC LC-39A
■ VAFB SLC-4E
■ CCAFS SLC-40

Total Success Launches has the most successful land

KSC LC-39A

X

Total Success Launches for site KSC LC-39A



■ 1
■ 0

Discussion section

Machine Learning Methods and Model Selection

In this project, several machine learning methods were employed to predict the success of Falcon 9 first-stage landings, including logistic regression, support vector machines (SVM), decision trees, and k-nearest neighbors (KNN). Each of these methods was evaluated for its ability to accurately classify landing. To determine the best model for predicting Falcon 9 first stage landing success, we compared several machine learning methods based on their accuracy and other characteristics. Here is a detailed comparison of the models:

1. Logistic Regression

Accuracy:

- Logistic Regression achieved an accuracy of 0.8333. This performance indicates that the model is a strong baseline for binary classification, but it may not capture complex patterns in the data as effectively as some other methods.

2. Support Vector Machine (SVM)

Accuracy:

- SVM also achieved an accuracy of 0.8333. This suggests that SVM performs comparably to Logistic Regression in this case, handling complex decision boundaries effectively with appropriate kernels.

3. Decision Tree

Accuracy:

- The Decision Tree model performed the best with an accuracy of 0.8889. This high accuracy indicates that the Decision Tree is effective at capturing the patterns in the data and making accurate predictions.

4. K-Nearest Neighbors (KNN)

Accuracy:

- KNN achieved an accuracy of 0.8333, similar to Logistic Regression and SVM. This performance shows that KNN is competitive but may not be the best choice in this context.

Summary of Performance Comparison

1. **Logistic Regression:** Achieved an accuracy of 0.8333. Good for linear relationships but may not handle complex patterns as well.
2. **SVM:** Also achieved an accuracy of 0.8333. Effective for complex, high-dimensional data but requires careful tuning and can be computationally demanding.
3. **Decision Tree:** The best performer with an accuracy of 0.8889. Offers high accuracy and interpretability but may overfit if not properly regularized.
4. **KNN:** Achieved an accuracy of 0.8333. Simple and intuitive but can be computationally expensive and sensitive to feature scaling.

Best Model

Based on the accuracy results, the **Decision Tree** is the best-performing model with an accuracy of 0.8889. It provides high accuracy and clear interpretability, making it the most effective choice for this prediction task. While other models like SVM and Logistic Regression also performed well, the Decision Tree's superior accuracy and ease of interpretation make it the preferred option for predicting Falcon 9 first-stage landing success.

Conclusion section

This project aimed to develop a predictive model to determine the success of Falcon 9 first-stage landings, a critical factor in SpaceX's cost-effective space exploration strategy. By leveraging machine learning techniques, the project analyzed various features such as launch sites, payload mass, and booster versions to predict landing success. The analysis included data preprocessing, feature engineering, and the application of several classification models including logistic regression, support vector machines, decision trees, and k-nearest neighbors.

Overall, this project provided valuable insights into the factors contributing to Falcon 9 landing success and demonstrated the practical application of machine learning techniques in aerospace engineering. The results contribute to SpaceX's ongoing efforts to enhance rocket reusability and reduce the costs of space missions.

The marketing value:

approach for the Falcon 9 first-stage landing prediction model focuses on demonstrating its value to key stakeholders, leveraging content and digital marketing, building strategic partnerships, and continuously improving based on feedback. By effectively communicating the model's benefits and applications, the aim is to position it as a valuable tool for enhancing mission planning and achieving greater success in space exploration.