IBM DATA SCIENCE CAPSTONE PROJECT

SPACEX FALCON 9 FIRST STAGE LANDING PREDICTION

OUTLINE

- Executive Summary slide.
- Introduction .
- Methodologies
- Exploratory Data Analysis & Visualization
- Models
- Results
- Conclusion

Executive Summary slide

Methodology:

- Data Collection
- Data Wrangling
- EDA

Results:

- Classification Models
- Comparison
- Conclusion

Introduction

In this capstone project, our goal is to predict whether the Falcon 9 first stage will successfully land. SpaceX offers Falcon 9 rocket launches at a cost of 62 million dollars, significantly lower than other providers, which can charge upwards of 165 million dollars per launch. The key to SpaceX's cost efficiency is the reusability of the Falcon 9's first stage. Accurately predicting the landing outcome can provide valuable insights into the potential cost savings for a launch. This information is particularly useful for companies considering competing with SpaceX for rocket launch contracts. This project will provide an overview of the problem, as well as the tools and methodologies needed to develop a predictive model for Falcon 9 landings.

Methodology

1. Data Collection

- Performed GET requests to the SpaceX REST API
- Conducted web scraping to gather additional data

2. Data Wrangling

- Handle missing values

outcome label to indicate:

- 0 for unsuccessful
- 1 for successful

3. Exploratory Data Analysis (EDA)

- Executed SQL queries to manipulate and evaluate the SpaceX dataset
- Employed Pandas and Matplotlib for data visualization to identify relationships and patterns between variables

4. Interactive Visual Analytics

- Conducted geospatial analysis using Folium
- Developed an interactive dashboard with Plotly Dash

5. Data Modeling and Evaluation

- Leveraged Scikit-Learn for:
    - Data standardization
    - Splitting data into training and test sets using train_test_split
    - Training various classification models
    - Hyperparameter tuning with GridSearchCV
- Plotted confusion matrices for each classification model
- Evaluated the accuracy of each classification model

EDA with SQL

1 - Unique Launch Sites

- **Query:** SELECT DISTINCT LAUNCH_SITE FROM SPACEXTBL;
- **Insight:** Identified four unique launch sites used by SpaceX.

2. Launch Sites Starting with 'CCA'

- **Query:** SELECT LAUNCH_SITE FROM SPACEXTBL WHERE LAUNCH_SITE LIKE 'CCA%' LIMIT 5;
- **Insight:** Consistently showed CCAFS LC-40 for the initial records.

3. Total Payload Mass by NASA (CRS)

- **Query:** SELECT SUM(PAYLOAD_MASS__KG_) AS TOTAL_PAYLOAD_MASS FROM SPACEXTBL WHERE CUSTOMER = 'NASA (CRS)';
- **Insight:** NASA (CRS) missions have a total payload mass of 45,596 kg.

4. Average Payload Mass for Booster Version F9 v1.1

•**Query:** SELECT AVG(PAYLOAD_MASS__KG_) AS AVERAGE_PAYLOAD_MASS FROM SPACEXTBL WHERE BOOSTER_VERSION = 'F9 v1.1';

•**Insight:** The average payload mass for F9 v1.1 is 2,928.4 kg.

5. Date of First Successful Ground Pad Landing

•**Query:** SELECT MIN(DATE) AS FIRST_SUCCESSFUL_GROUND_LANDING FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (ground pad)';

•**Insight:** Identified the earliest successful landing date (pending correct column name).

6. Boosters with Successful Drone Ship Landings and Specific Payload Mass

•**Query:** SELECT BOOSTER_VERSION FROM SPACEXTBL WHERE "Landing_Outcome" = 'Success (drone ship)' AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;

•**Insight:** Listed boosters meeting the criteria (pending correct column name).

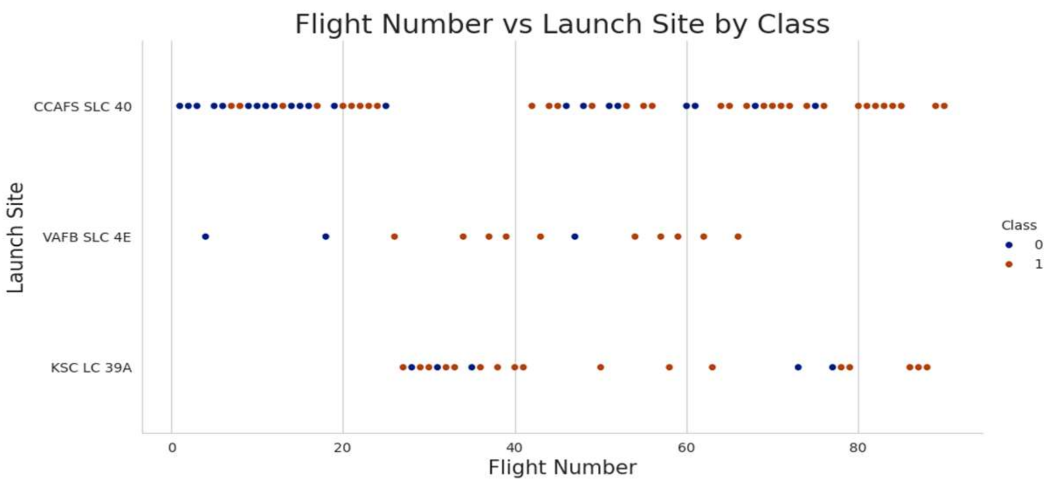
7. Total Number of Mission Outcomes

•**Query:** SELECT MISSION_OUTCOME, COUNT(*) AS TOTAL_NUMBER FROM SPACEXTBL GROUP BY MISSION_OUTCOME;

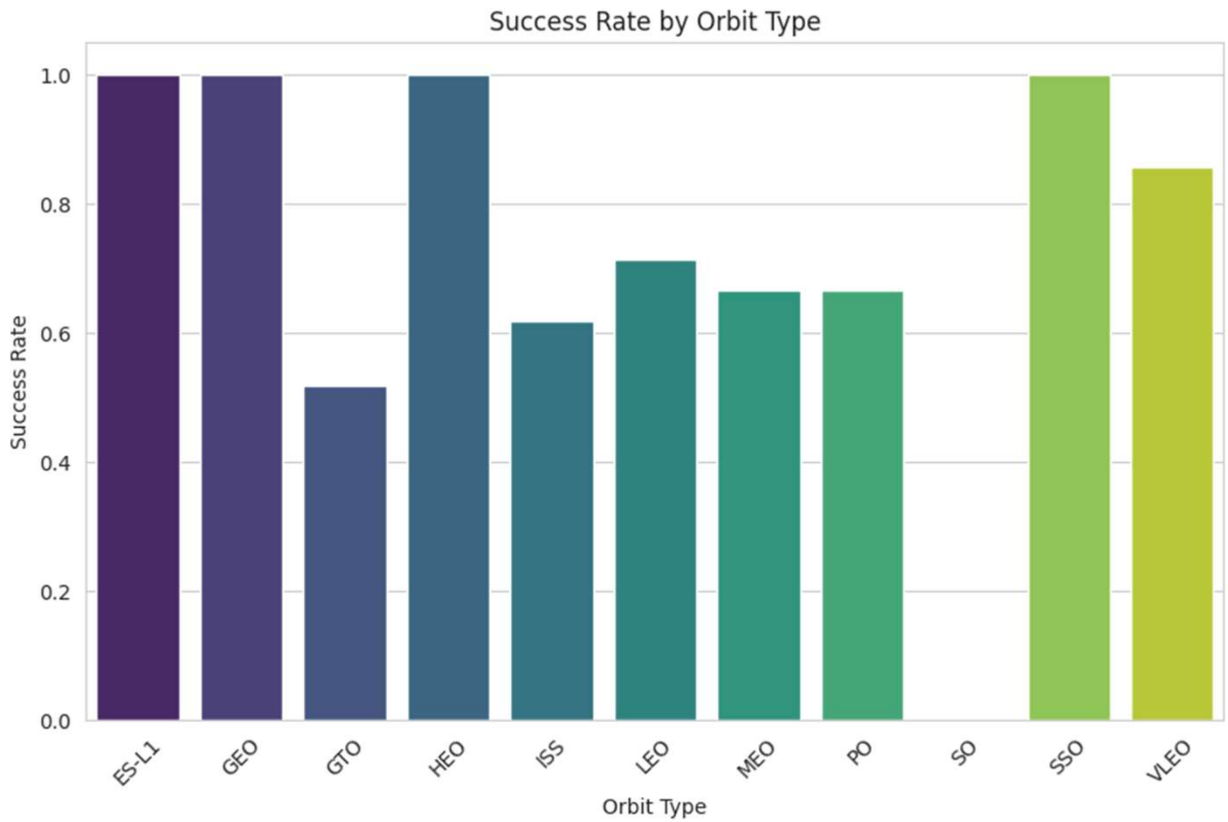
•**Insight:** Majority of missions are successful (98 out of 100).

Visualization with SeaBorn

1-Visualize the relationship between Flight Number and Launch Site

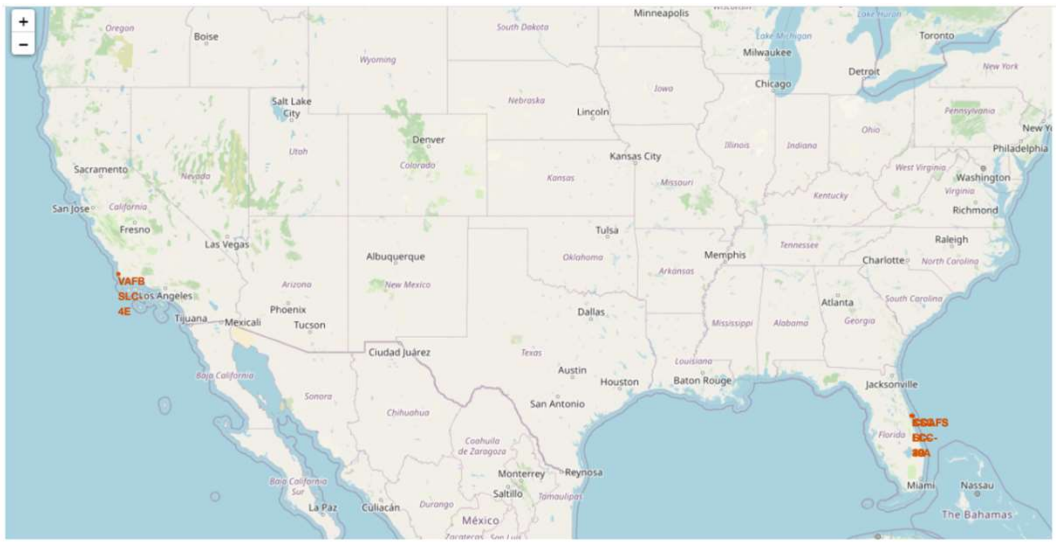


2-Visualize the relationship between success rate of each orbit type

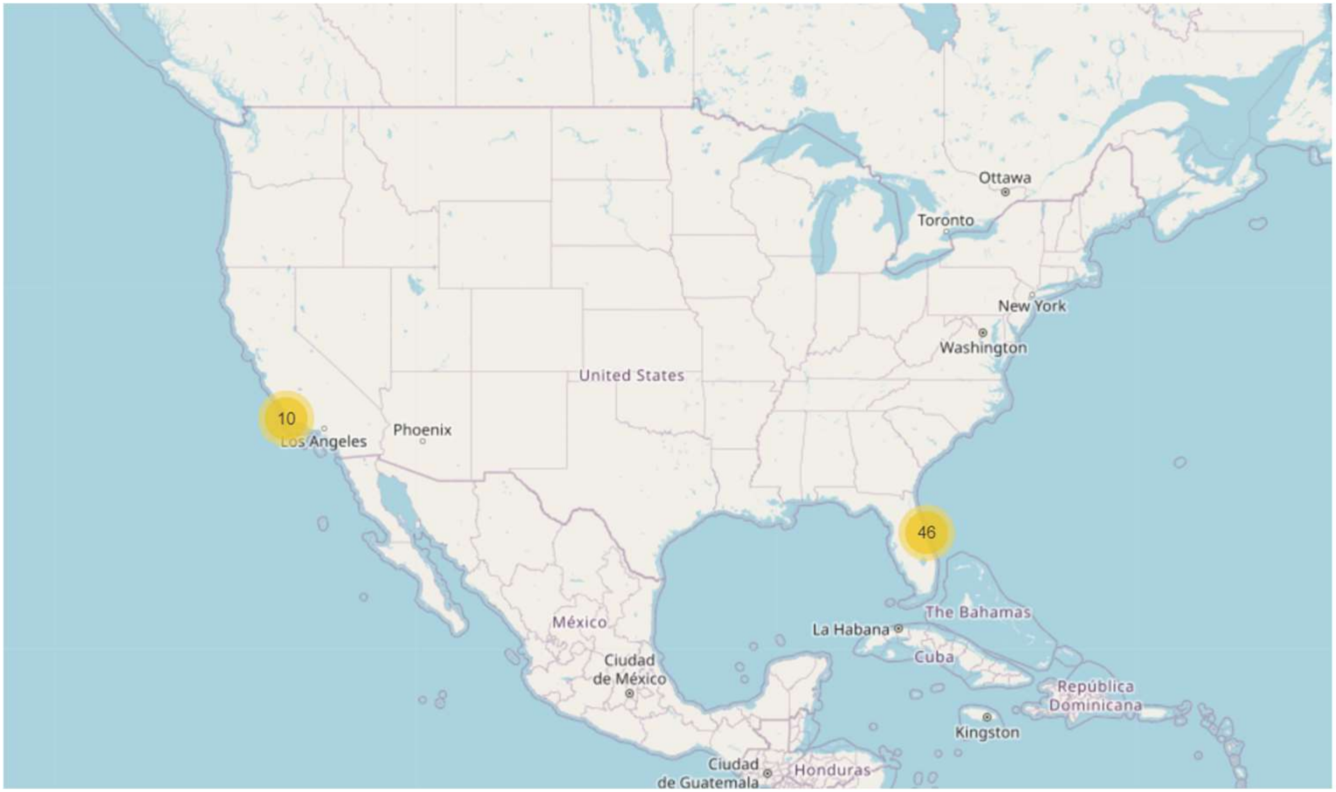


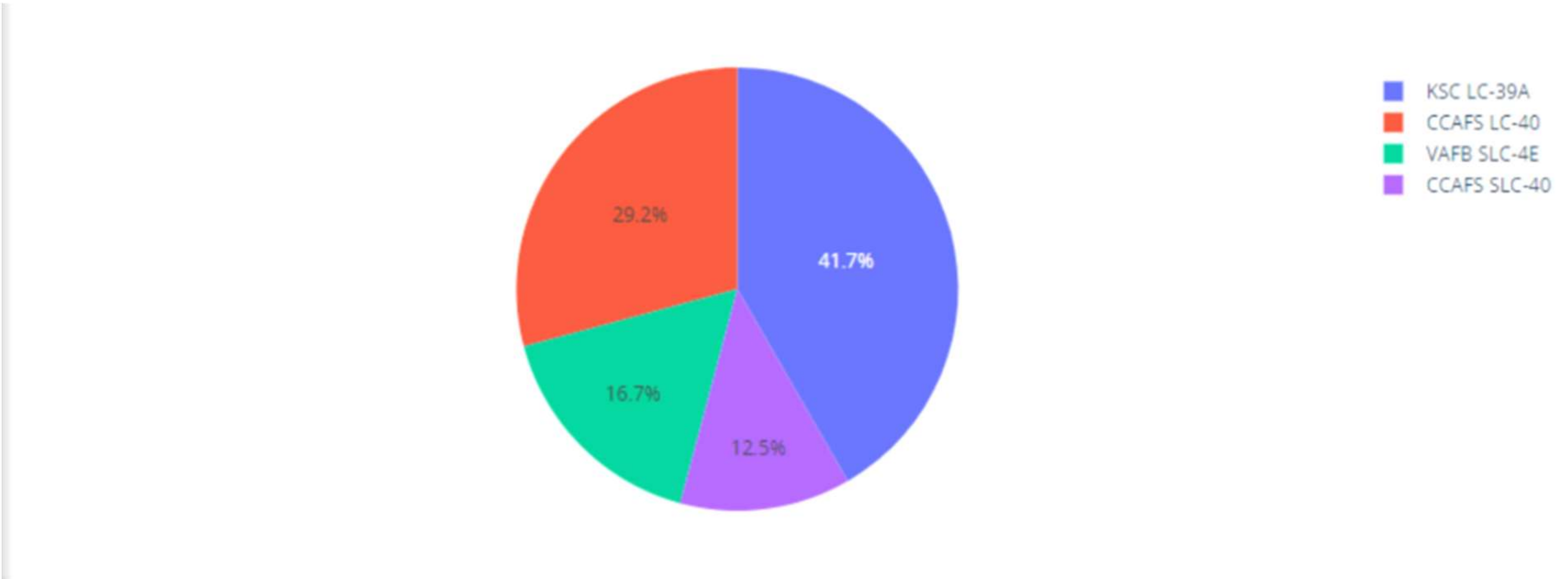
Visualization with Folium

1- Mark all launch sites on a map



2-Mark the success/failed launches for each site on the map





Visualization with Plotly

Models

In our predictive analysis, we utilized various classification algorithms to determine the likelihood of a successful landing for the Falcon 9 first stage. The classification models included:

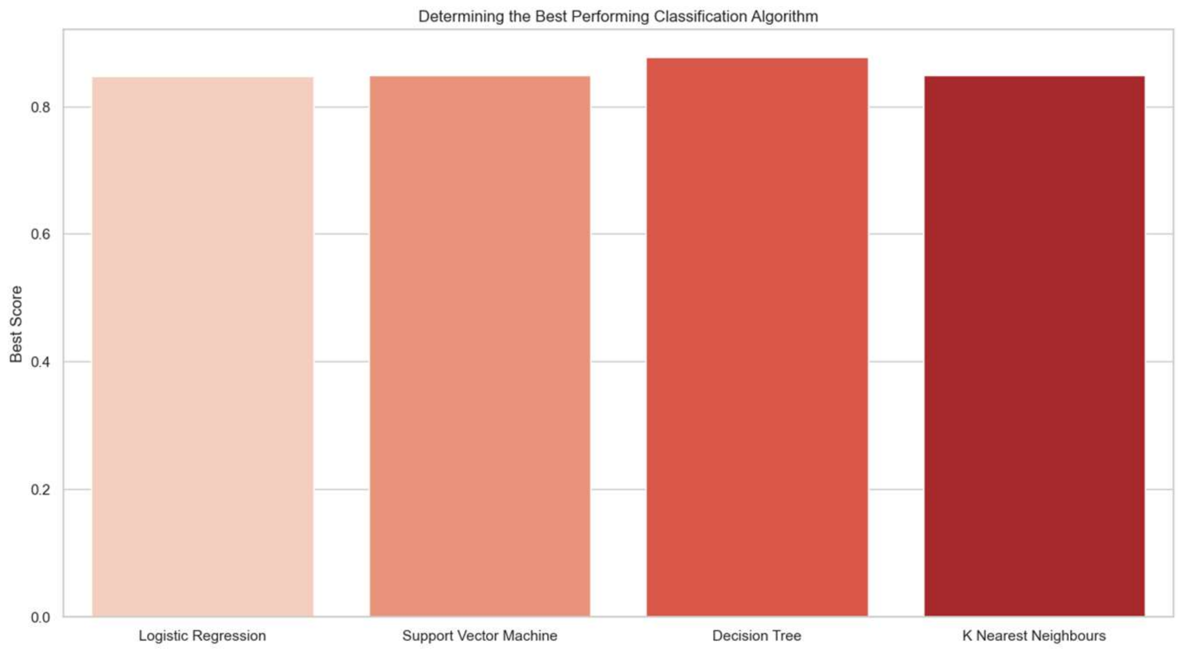
1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree
4. K Nearest Neighbors (KNN)

These models were evaluated for their accuracy and ability to predict the landing outcome of the Falcon 9 first stage. The models were fine-tuned using GridSearchCV to find the best hyperparameters, and their performance was visualized through confusion matrices to identify strengths and weaknesses in distinguishing between successful and unsuccessful landings.

Model Evaluation Summary

Algorithm	Accuracy Score	Accuracy Score (Hyperparameter)
Logistic Regression	0.833333	0.846429
Support Vector Machine	0.833333	0.848214
Decision Tree	0.944444	0.876786
K Nearest Neighbors	0.833333	0.848214

Visualization of Results



Conclusion

In conclusion, our predictive analysis aimed to determine the likelihood of a successful landing for the Falcon 9 first stage using various classification algorithms. After evaluating the performance of Logistic Regression, Support Vector Machine, Decision Tree, and K Nearest Neighbors models, we obtained the following results:

- Logistic Regression and Support Vector Machine models achieved comparable accuracy scores, both around 83.33%. However, after fine-tuning hyperparameters, the accuracy slightly improved to approximately 84.64% and 84.82%, respectively.
- The Decision Tree model outperformed other algorithms with an accuracy score of 94.44%. Even after fine-tuning hyperparameters, the accuracy increased to approximately 87.68%.
- K Nearest Neighbors exhibited similar performance to Logistic Regression and Support Vector Machine models, with an initial accuracy score of 83.33%. After hyperparameter tuning, the accuracy improved slightly to around 84.82%.

Overall, the Decision Tree algorithm demonstrated the highest accuracy in predicting the landing outcome of the Falcon 9 first stage. These results indicate the effectiveness of classification algorithms in analyzing complex data and providing insights into mission success for space exploration endeavors.

Future Work

While our predictive analysis has provided valuable insights into the likelihood of successful landings for the Falcon 9 first stage, there are several avenues for future exploration and improvement:

1. **Feature Engineering**: Explore additional features or engineered variables that could enhance the predictive power of the models. This could include factors such as weather conditions, payload specifications, or historical launch data.
2. **Ensemble Methods**: Investigate ensemble learning techniques, such as Random Forest or Gradient Boosting, to combine the strengths of multiple models and improve overall predictive performance.
3. **Fine-tuning Hyperparameters**: Further fine-tune hyperparameters of the models to optimize their performance and achieve higher accuracy scores.

4.Deep Learning: Explore the application of deep learning algorithms, such as neural networks, to capture complex patterns and relationships in the data that may not be captured by traditional machine learning models.
5.Real-time Prediction: Develop a real-time prediction system that can continuously monitor and predict the outcome of upcoming Falcon 9 launches based on the latest data and information.
6.Integration with Operational Systems: Integrate the predictive models into operational systems used by SpaceX or other space agencies to support decision-making processes related to launch planning and mission success.
By pursuing these avenues for future work, we can further enhance the accuracy and reliability of predictive models for assessing the success of Falcon 9 first stage landings and contribute to the continued advancement of space exploration efforts.