Applied Data Science Capstone

By Anderson Quiroz Gonzales

August 9, 2024

Executive Summary

Project Objective: "The objective of this project was to develop a predictive model capable of determining whether the first stage of SpaceX's Falcon 9 rocket will land successfully, leveraging historical data and advanced data science techniques."

Key Results:

"The Logistic Regression and KNN models achieved perfect accuracy of 100%."
"The SVM's sigmoid kernel achieved an accuracy of 96.67%."

Significance: "These results can help SpaceX and other competitors optimize costs and improve accuracy in launch planning."

Introduction

Context: "SpaceX has revolutionized the aerospace industry by reusing the first stages of its rockets, significantly reducing launch costs."

Purpose: "This project aims to predict the success of Falcon 9 first stage landings using a data science-driven approach."

Overall Methodology: "Exploratory data analysis (EDA), predictive modeling, and interactive visualization techniques were used to develop and evaluate several models."

Data Collection and Wrangling Methodology

- Data Sources: "The data used was sourced from historical SpaceX launch records, including variables such as date, launch site, and landing outcome."
- Processing: "Data cleaning and transformation procedures, such as removing null values and normalizing variables, were carried out to ensure analysis quality."
- Tools Used: "Python (pandas, numpy) was the main tool for data manipulation."

Exploratory Data Analysis (EDA) and Interactive Visualization Methodology

- EDA: "Exploratory analyses were performed to understand the distribution of the data and the relationships between key variables."

- Visualizations: "We used Plotly and Folium to create interactive charts and maps, allowing for deeper and more dynamic data exploration."

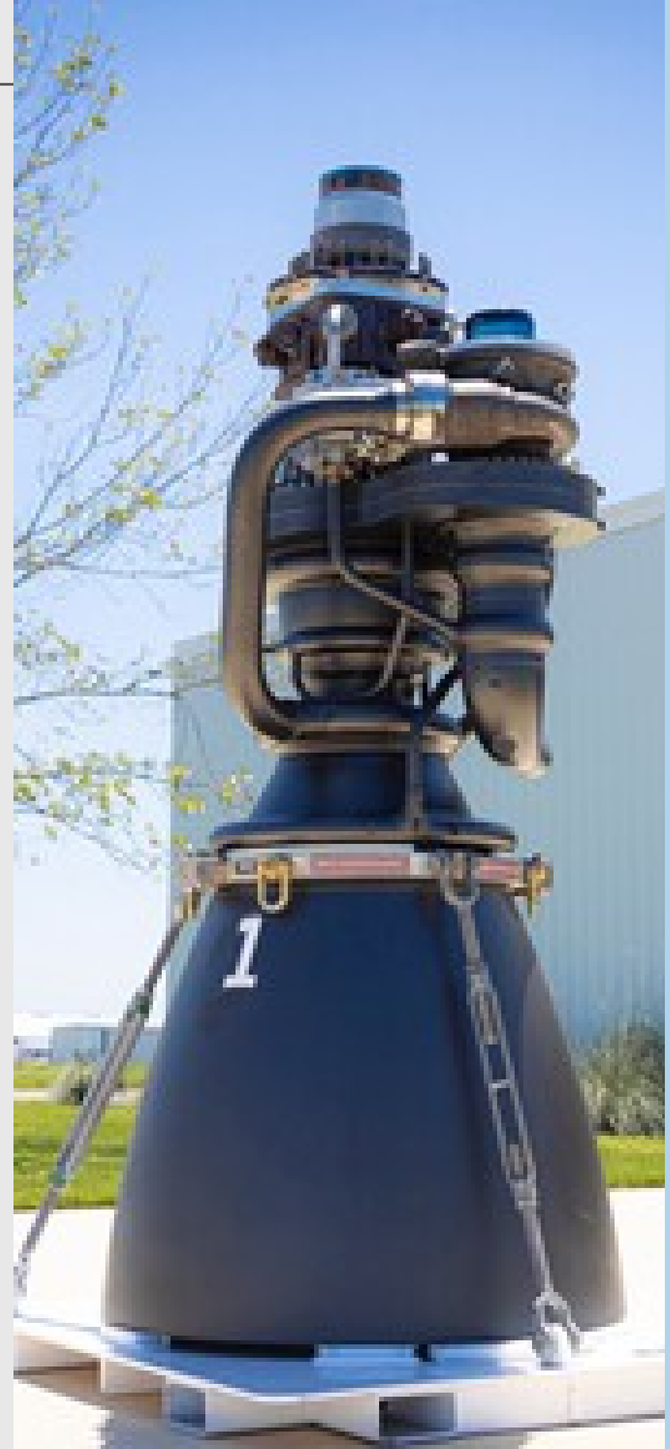
BORCELLE BUSINESS

Predictive Analysis Methodology

- Models Used: "Four classification models were tested: Logistic Regression, SVM, Decision Tree, and KNN."
- Optimization: "GridSearchCV techniques were used to find the best hyperparameters for each model."
- Evaluation Criteria: "Accuracy was the main metric used to evaluate the performance of the models."

EDA Results with Visualizations

- Correlation Plot: "The correlation matrix showed that certain variables, such as 'rocket mass' and 'wind speed,' had a significant relationship with landing success."
- Histograms: "Distributions of key variables such as 'landing distance' showed patterns that were useful for predictive modeling."



EDA Results with SQL

- SQL Queries: "We executed SQL queries to extract specific data on launch outcomes and the factors that most influenced landing success."
- Key Results: "We identified that launches conducted in certain seasons had a higher success rate."

Interactive Map with Folium

- Launch Map: "An interactive map was created to visualize the geographic locations of successful and failed launches and landings."
- Additional Layers: "The map includes layers showing additional details, such as weather conditions during each launch."

Dashboard Results with Plotly Dash

- Dashboard Screenshots: "The interactive dashboard allows for exploration of SpaceX launch and landing data across different dimensions."
- Features: "It includes filters to analyze data by year, rocket type, and landing outcomes."

Predictive Analysis Results

Model Comparison:

- "Logistic Regression: 100% Accuracy"
- "SVM (sigmoid kernel): 96.67% Accuracy"
- "Decision Tree: 93.33% Accuracy"
- "KNN: 100% Accuracy"

Conclusion: "Logistic Regression and KNN proved to be the most reliable models for predicting Falcon 9 landing success."

Conclusion

- Summary: "This project successfully developed a highly accurate predictive model for Falcon 9 landings, which can be of great utility to the aerospace industry."
- Impact: "The results obtained can improve planning and reduce costs in future SpaceX missions."
- Future Research: "We propose further exploration of the impact of weather conditions and payload on landing outcomes."

Creativity and Innovation

- Creative Elements: "We incorporated interactive visualizations and a dynamic dashboard to facilitate data analysis."
- Innovation: "The use of interactive maps and detailed analysis allowed for the discovery of non-obvious patterns in the data."

BORCELLE BUSINESS

Questions and Acknowledgements

- Questions: "We are open to any questions you may have about this project."

- Acknowledgements: "Thanks to everyone who contributed to this project and to my fellow classmates for their support."

References

Sources List: "Includes all data sources, tools, and references used throughout the project."