Winning Space Race with Data Science

Gerald Juwah

2024-02-08

Outline

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Executive Summary

Various methodologies were employed for data analysis in this study:

- Data collection involved web scraping and utilising the SpaceX API.
- Exploratory Data Analysis (EDA) encompassed data wrangling, visualisation, and interactive visual analytics.
- Machine Learning Prediction techniques were applied.

Summary of Key Findings:

- Valuable data was successfully gathered from publicly available sources.
- Through EDA, pivotal features influencing the success of launches were identified.
- Machine Learning Prediction revealed the optimal model for predicting crucial characteristics significantly contributing to successful launches, leveraging the entire dataset.

Introduction

The primary aim of this assessment is to thoroughly examine the feasibility of the emerging company, Space Y, in establishing a competitive stance against its counterpart, Space X. In pursuit of a comprehensive understanding, the analysis seeks to address critical inquiries that will contribute to the strategic positioning of Space Y within the space industry.

Key Objectives:

Cost Estimation: A pivotal aspect of this evaluation involves determining the most effective methodology for estimating the total cost of launches. This will be achieved by leveraging predictive models to discern the likelihood of successful landings for the first stage of rockets. Such insights will be instrumental in formulating precise cost projections.

Optimal Launch Locations: Another key area of exploration is identifying the optimal locations for conducting launches. By scrutinizing various factors and leveraging data-driven approaches, the analysis aims to pinpoint the geographic sites that offer the most favourable conditions for successful and cost-effective space missions.

Through a meticulous exploration of these objectives, this assessment endeavours to provide valuable insights that can inform strategic decisions for Space Y, ultimately enhancing its competitive edge in the dynamic space exploration landscape.

Methodology

Methodology

Data Collection:

- Information from Space X was procured through two distinct channels:
  - Space X API (https://api.spacexdata.com/v4/rockets/)
  - Web Scraping from (https://en.wikipedia.org/wiki/List_of_Falcon/_9/_and_Falcon_Heavy_launches)

Data Processing:

- Conducted data wrangling to ensure cleanliness and coherence.
- Enriched collected data by creating a landing outcome label based on summarized outcome data.

Methodology

Exploratory Data Analysis (EDA):

- Utilized visualization techniques and SQL for comprehensive EDA.
- Performed interactive visual analytics using Folium and Plotly Dash for enhanced insights.

Predictive Analysis:

- Employed classification models for predictive analysis.
- Normalized collected data and divided it into training and test datasets.
- Evaluated the performance of four distinct classification models, employing diverse parameter combinations.
- Assessed the accuracy of each model to identify the most effective predictive approach.

Data Collection (SpaceX API)

API Request and Parse SpaceX data → Filter data for only Falcon 9 launches → Handle Missing Values

Link: [Data Collection (SpaceX API)]

Data Collection (Scraping)

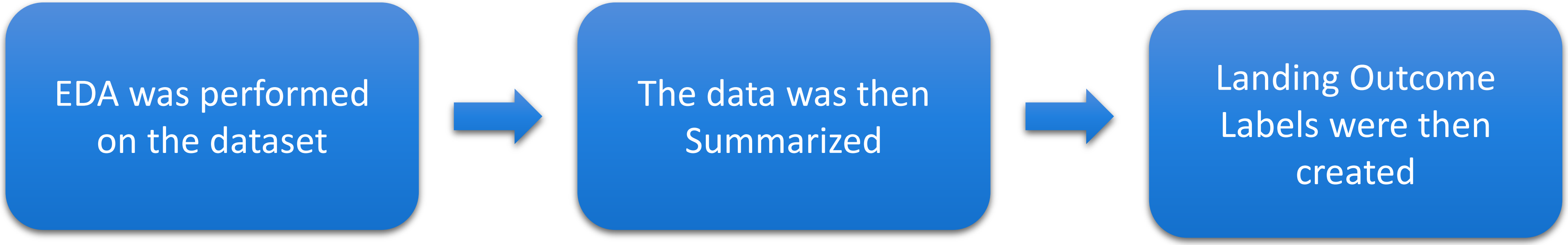
Link: Data Collection (Scraping)

Data Wrangling

EDA was performed on the dataset → The data was then Summarized → Landing Outcome Labels were then created

Link: [Data Wrangling]

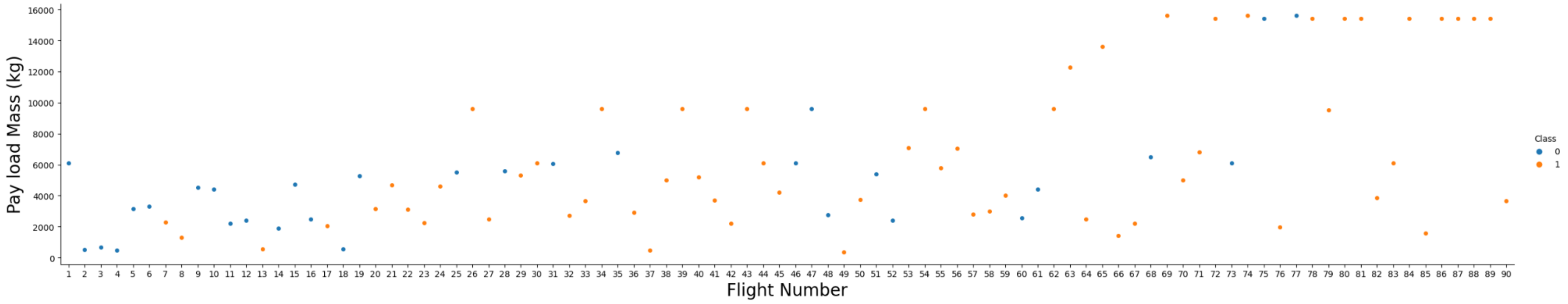
Data Wrangling

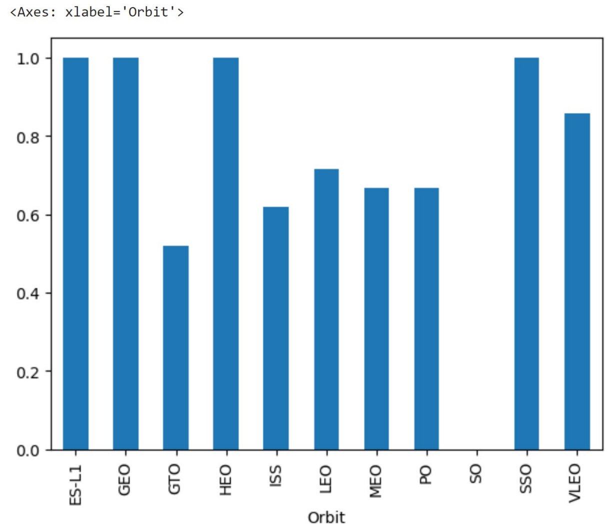


Link: [Data Wrangling]

EDA

For data exploration, scatterplots and bar plots were employed to visualize feature relationships, including Payload Mass vs. Flight Number, Launch Site vs. Flight Number, Launch Site vs. Payload Mass, Orbit vs. Flight Number, and Payload vs. Orbit.





Link: [Exploratory Data Analysis]

EDA and SQL

Performed SQL queries including retrieving the names of unique launch sites, the top 5 launch sites with names starting with 'CCA', the total payload mass carried by NASA-launched boosters (CRS), the average payload mass carried by booster version F9 v1.1, the date of the first successful landing outcome on a ground pad, names of boosters with success on a drone ship and payload mass between 4000 and 6000 kg, total counts of successful and failed mission outcomes, names of booster versions carrying the maximum payload mass, failed landing outcomes on a drone ship with their booster versions and launch site names in the year 2015, and the rank of the count of landing outcomes between June 4, 2010, and March 20, 2017.

Link: Exploratory Data Analysis with SQL

Visualization of Launch Site location

Folium Maps utilize various visual elements:

- Marker's pinpoint locations, such as launch sites.
- Circles highlight specific areas around coordinates, exemplified by NASA Johnson Space Centre.
- Marker clusters group events at each coordinate, representing launches in a launch site.
- Lines visually depict distances between two coordinates.

Link: Launch Site location

Dash App Build

Various graphs and plots were employed for data visualization:

- Launch percentage by site
- Payload range analysis This approach facilitated a swift assessment of the correlation between payloads and launch sites, aiding in the identification of optimal launch locations based on payload considerations.

Link: Dash App Build

Predictive analysis using machine learning

Four classification models were evaluated: logistic regression, support vector machines, decision trees, and k-nearest neighbours.

Data preparation and standardization → Test of each model with combinations of hyperparameter tuning → Comparing results

Link: Predictive analysis using machine learning

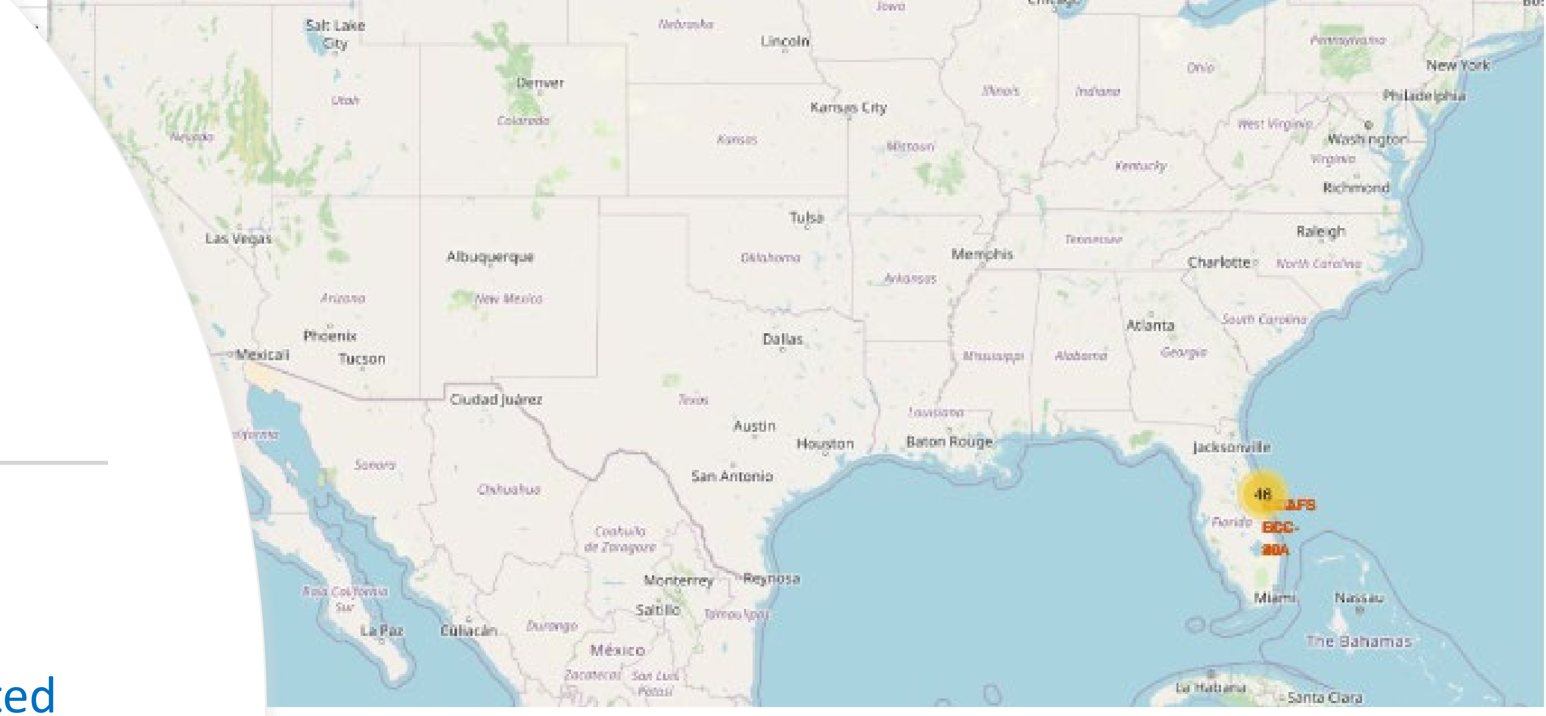
Results

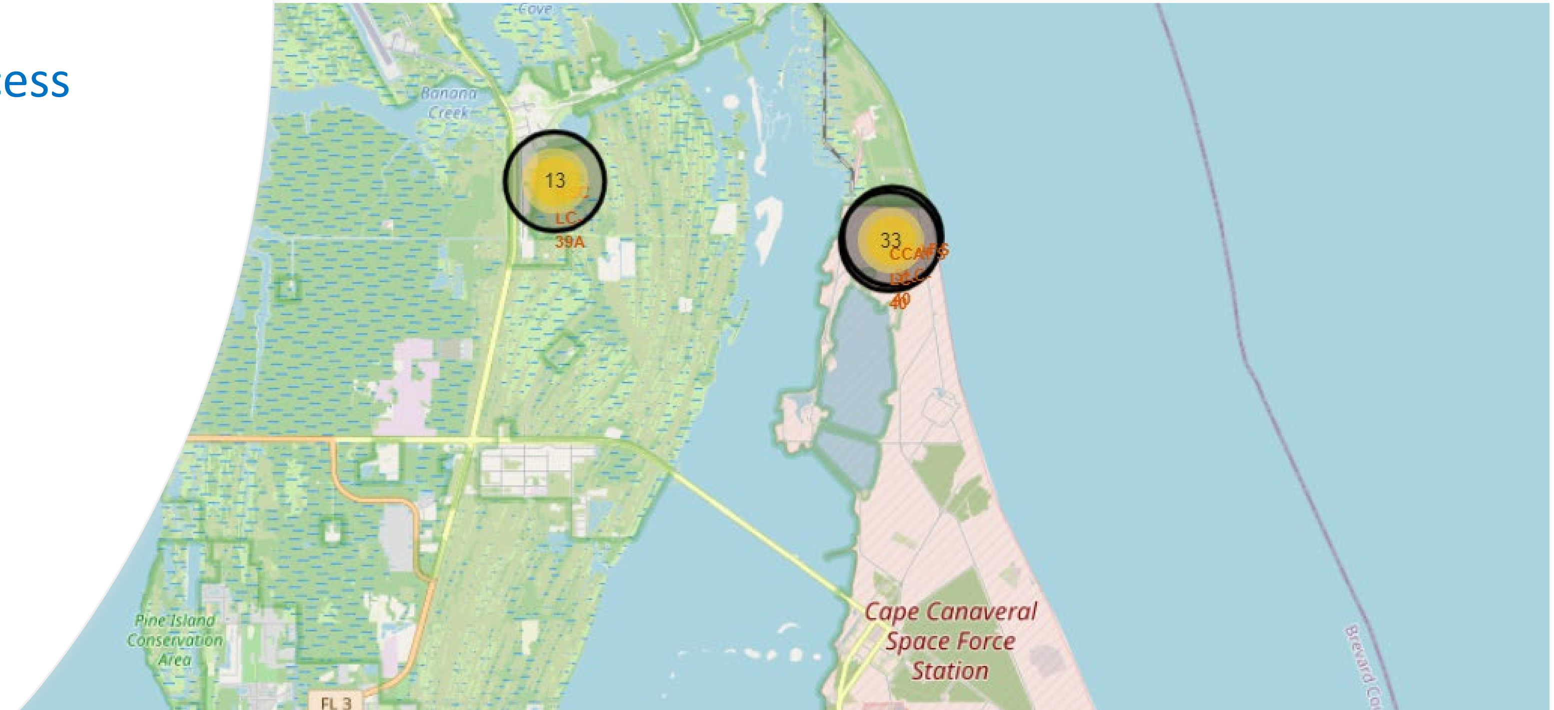
Key findings from exploratory data analysis:

- SpaceX operates 4 distinct launch sites.
- Initial launches were conducted by SpaceX and NASA.
- The average payload of the F9 v1.1 booster is 2,928 kg.
- The first successful landing occurred in 2015, five years after the initial launch.
- Several Falcon 9 booster versions successfully landed on drone ships with payloads exceeding the average.
- Nearly 100% of mission outcomes were successful.
- Two booster versions, F9 v1.1 B1012 and F9 v1.1 B1015, failed to land on drone ships in 2015.
- The success rate of landing outcomes improved over the years.

Results

- Key findings from exploratory data analysis:
    - Launch sites are strategically located in safe areas, often near the sea, ensuring safety measures and access to logistic infrastructure.
    - The majority of launches occur at launch sites located on the East Coast.

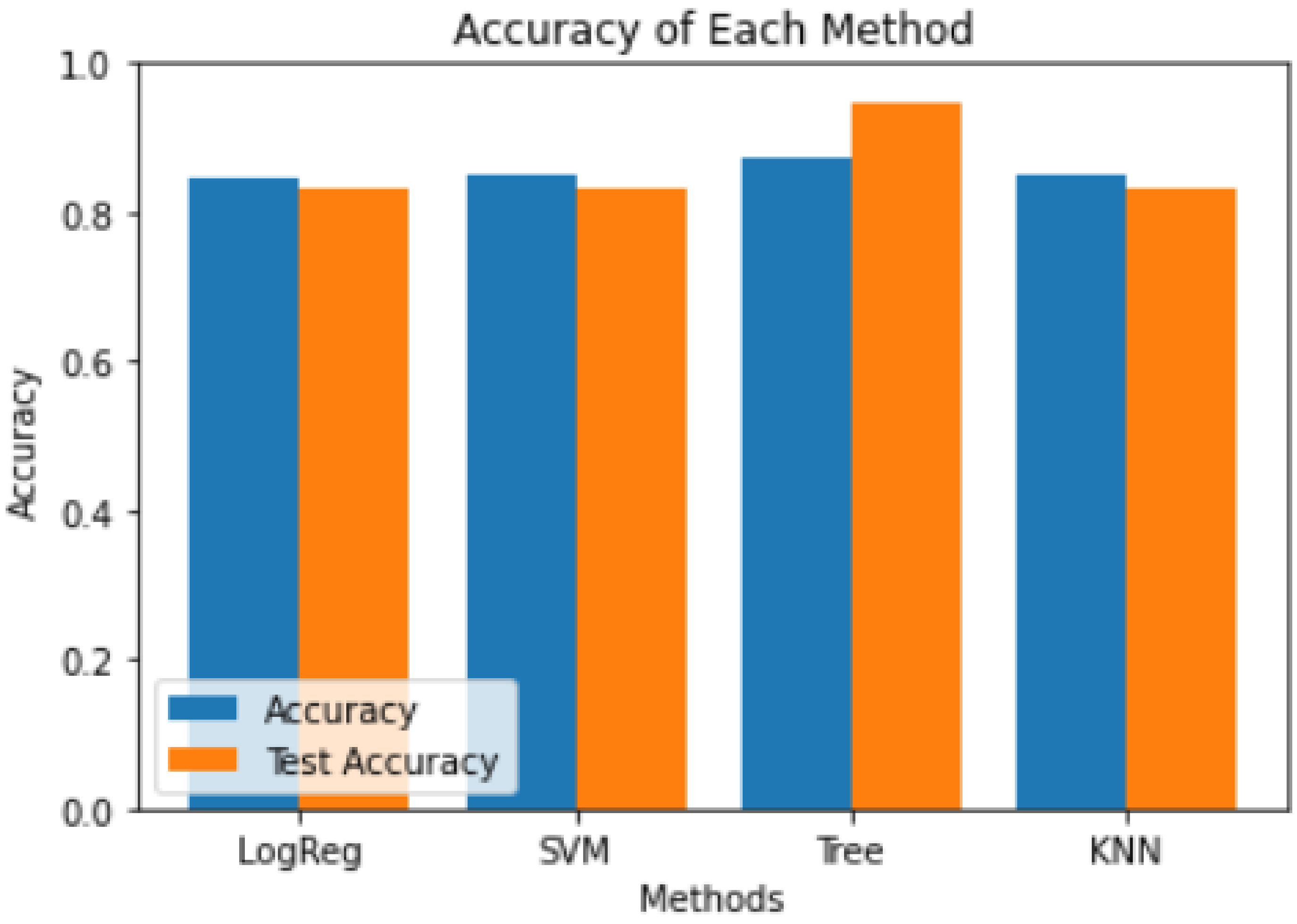




Results

Predictive Analysis findings:

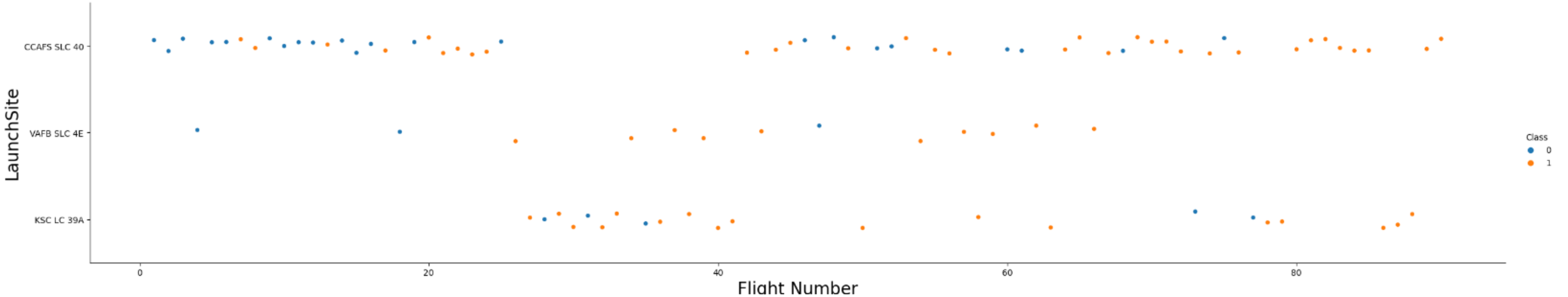
- Decision Tree Classifier emerged as the most effective model for predicting successful landings, boasting an accuracy exceeding 87% and achieving a test data accuracy surpassing 94%.

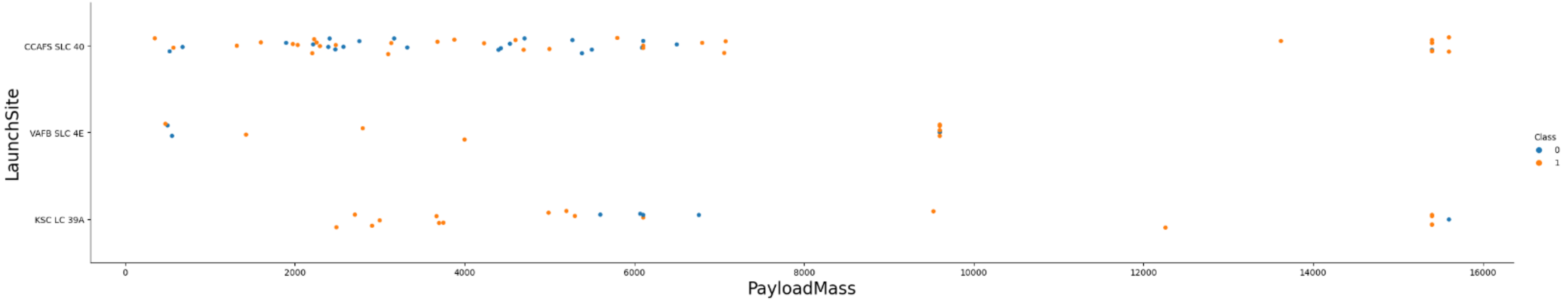


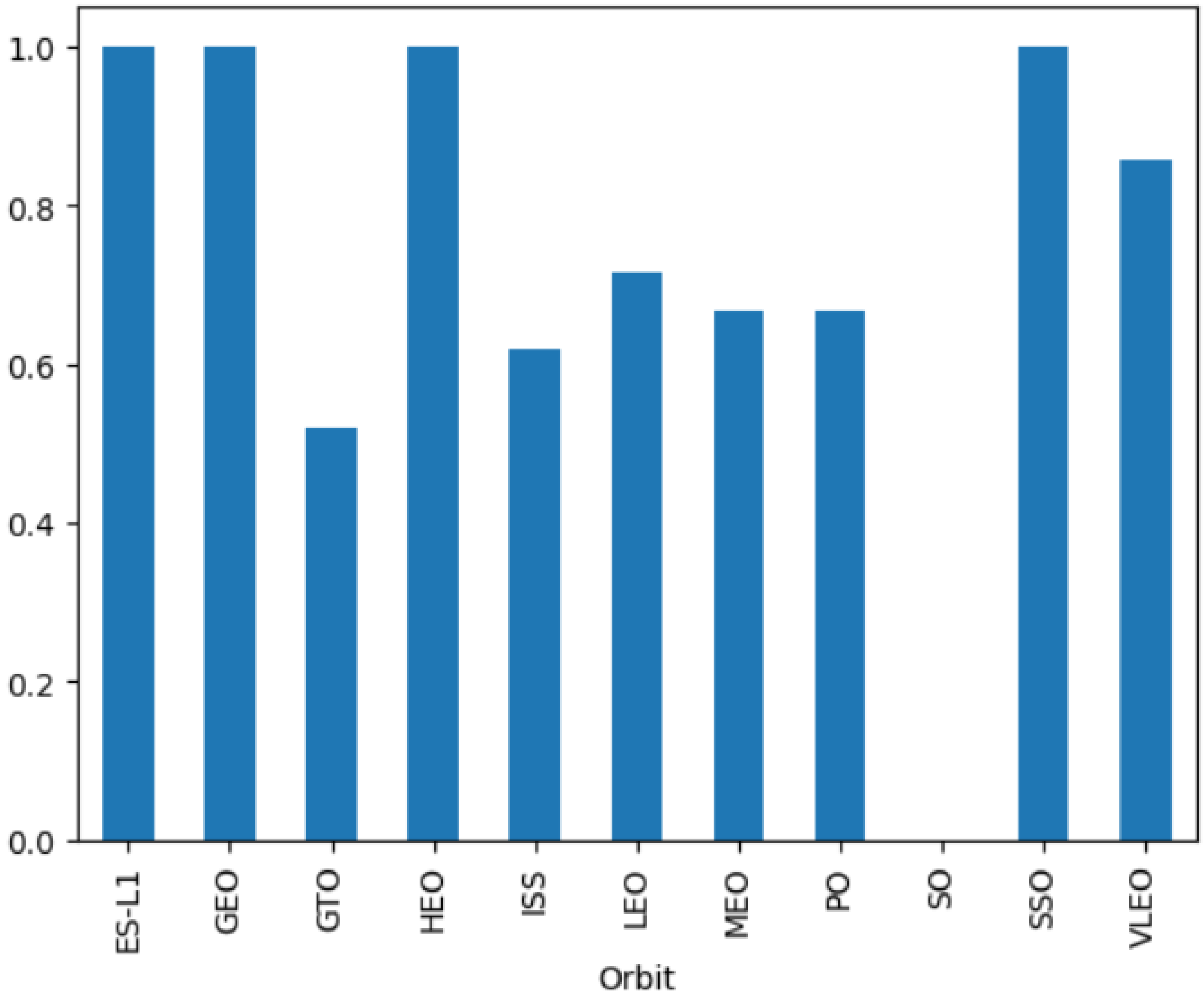
Insight from EDA

Insights from EDA

- Key insights from the plot:
  - CCAFS SLC 40 emerges as the top-performing launch site presently, boasting a high success rate for recent launches.
  - Following closely are VAFB SLC 4E in second place and KSC LC 39A in third place.
  - Over time, there is a noticeable improvement in the overall success rate of launches.
  - Payloads exceeding 9,000 kg exhibit an excellent success rate.
  - Payloads surpassing 12,000 kg appear feasible primarily at CCAFS SLC 40 and KSC LC 39A launch sites.

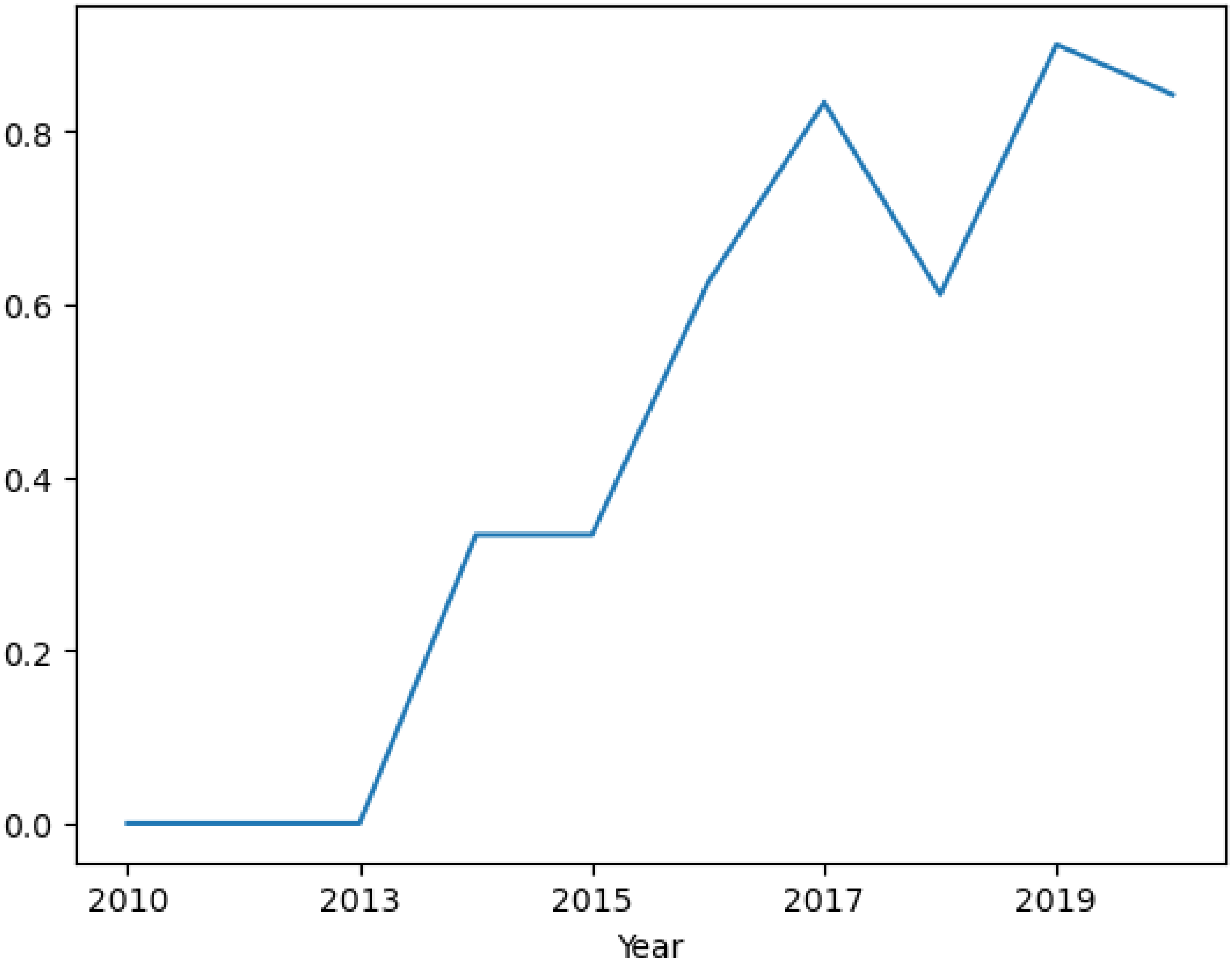






Success Rate vs Orbit Type

- Key insights on success rates by orbit:
  - Highest success rates are observed for orbits:
    - ES-L1
    - GEO
    - HEO
    - SSO
  - Additionally, notable success rates are seen for:
    - VLEO (over 80%)
    - LFO (over 70%)



Launch Success Yearly Trend

Key insights on success rate trends:

- Success rate saw a consistent increase from 2013 to 2020.
- The initial three years may have represented a period of adjustments and technological advancements.

All Launch site Names

- Key information:
  - There are four distinct launch sites.

Launch Site
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

  - These sites are derived by extracting unique occurrences of "launch_site" values from the dataset.

Launch Site Names Begin with ‘CCA’

- 5 records where launch sites begin with `CCA`:

Date	Time UTC	Booster Version	Launch Site	Payload	Payload Mass kg	Orbit	Customer	Mission Outcome	Landing Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	07:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attemp

Total Payload Mass

The total payload carried by NASA boosters:

- Calculated by summing the payloads associated with codes containing 'CRS', which corresponds to NASA.
- The total payload carried by NASA boosters is 111,268 kg. Calculated by summing the payloads associated with codes containing 'CRS', which corresponds to NASA

Average Payload Mass by F9 v 1.1

- The average payload mass carried by booster version F9 v1.1 is 2,928 kg. This value was obtained by filtering the data for the F9 v1.1 booster version and calculating the average payload mass

First Successful Ground Landing Date

- The first successful landing outcome on the ground pad occurred on December 22, 2015. This information was obtained by filtering the data for successful landing outcomes on the ground pad and retrieving the minimum date value.

Successful Drone ship landing with payload between 4000 and 6000

- Four distinct booster versions have successfully landed on a drone ship and had a payload mass greater than 4000 kg but less than 6000 kg

Booster Version
F9 FT B1021.2
F9 FT B1031.2
F9 FT B1022
F9 FT B1026

Total Number of Successful and Failure Mission Outcomes

Mission Outcome	Occurrences
Success	99
Success (payload status unclear)	1
Failure (in flight)	1

- The summary above shows the count of successful and failed mission outcomes, obtained by grouping the mission outcomes and counting the records for each group.

Booster Carried Maximum Payload

Booster Version (...)
F9 B5 B1048.4
F9 B5 B1048.5
F9 B5 B1049.4
F9 B5 B1049.5
F9 B5 B1049.7
F9 B5 B1051.3

Booster Version
F9 B5 B1051.4
F9 B5 B1051.6
F9 B5 B1056.4
F9 B5 B1058.3
F9 B5 B1060.2
F9 B5 B1060.3

- The listed boosters represent those that have carried the maximum payload mass recorded in the dataset.

2015 Launch Records

Booster Version	Launch Site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

- The list above shows the only two occurrences of failed landing outcomes on drone ships, along with their respective booster versions and launch site names, all from the year 2015

Rank Landing Outcomes Between 2010-06-04 and 2017-03-20

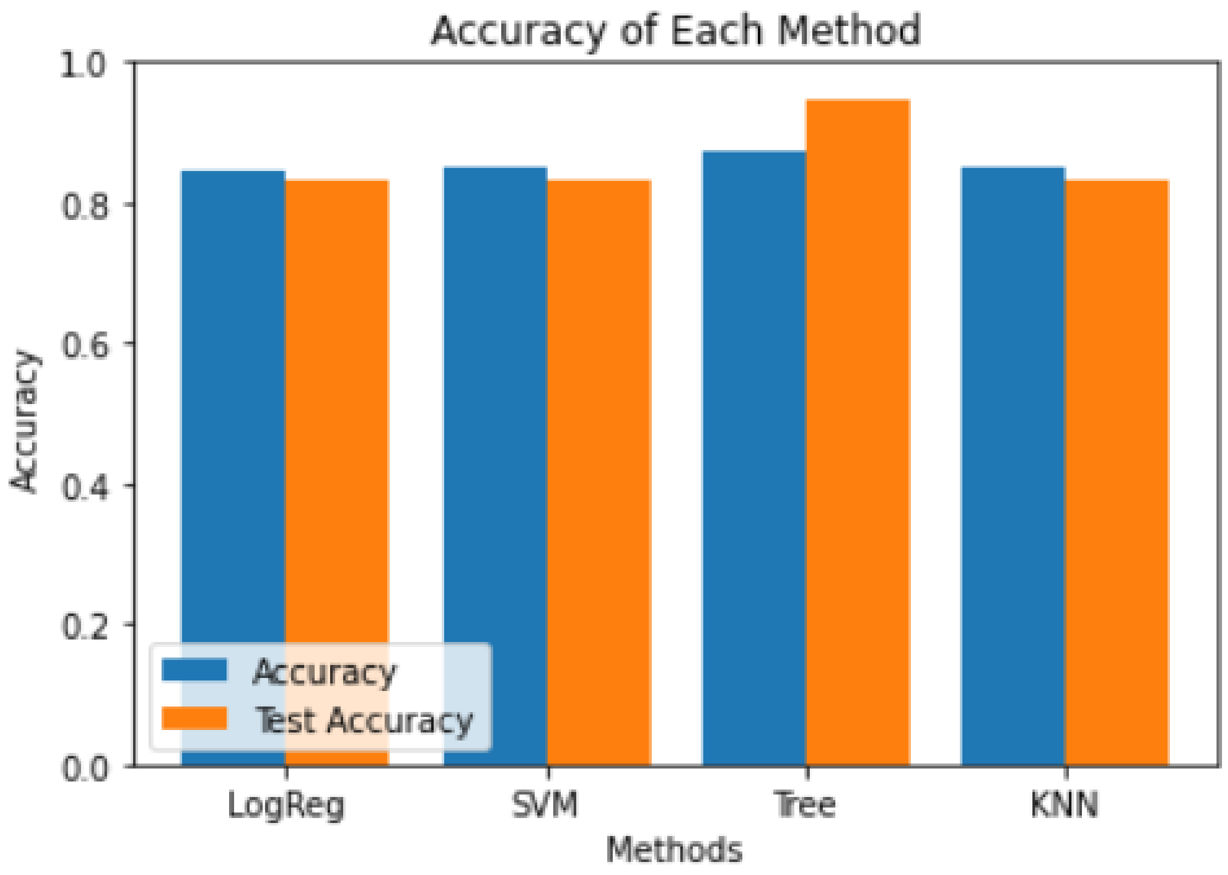
Landing Outcome	Occurrences
No attempt	10
Failure (drone ship)	5
Success (drone ship)	5
Controlled (ocean)	3
Success (ground pad)	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

- This data view highlights the ranking of all landing outcomes recorded between June 4, 2010, and March 20, 2017. It's essential to consider instances labelled as "No attempt" in the analysis.

Predictive Analysis

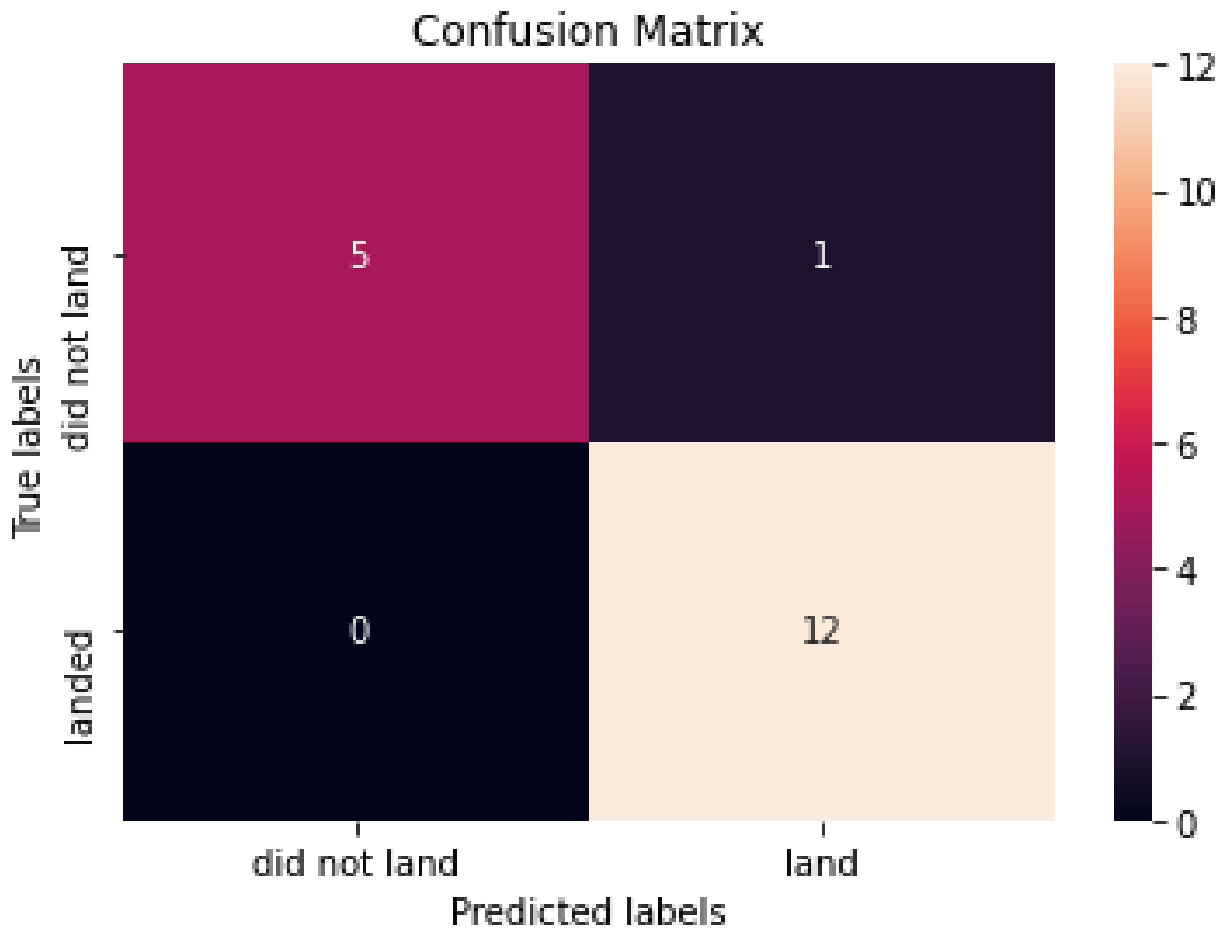
Predictive Analysis

- Classification Accuracy:
  - Four classification models were tested, and their accuracies are plotted beside;
  - The model with the highest classification accuracy is the Decision Tree Classifier, which has accuracies over 87%.



Predictive Analysis

- Confusion Matrix of Decision Tree Classifier:
    - The confusion matrix of the Decision Tree Classifier demonstrates its accuracy by revealing significant numbers of true positives and true negatives in comparison to the false values



Conclusions

- Multiple data sources were scrutinized, refining conclusions iteratively.
- KSC LC-39A emerged as the optimal launch site.
- Payloads exceeding 7,000kg pose a lower risk.
- While overall mission outcomes tend to succeed, successful landing rates demonstrate improvement over time, reflecting advancements in processes and rocket technology.
- Utilizing the Decision Tree Classifier can effectively predict successful landings, potentially boosting profitability

Thanks