IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Raihan Fadhilah Baihaqi
24 Oct 2024

Outline

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

Executive Summary

In this capstone project, the goal is to predict whether the first stage of SpaceX's Falcon 9 will successfully land using various machine learning classification algorithms. The project involves several key steps:

- Data collection, wrangling, and formatting
- Exploratory data analysis
- Interactive data visualization
- Machine learning prediction

Our visualizations suggest that certain launch features are correlated with the success or failure of the rocket landing. Based on the results, we concluded that the decision tree algorithm may be the most effective in predicting Falcon 9 landing success.

Introduction

In this capstone project, the objective is to predict whether the Falcon 9 first stage will successfully land. SpaceX promotes its Falcon 9 rocket launches at a cost of $62 million, compared to other providers that charge over $165 million. A significant portion of these savings comes from SpaceX's ability to reuse the first stage. Therefore, predicting whether the first stage will land successfully can help estimate launch costs, which could be useful for companies looking to compete with SpaceX for rocket launch contracts.

Most unsuccessful landings are intentional, as SpaceX sometimes opts for controlled ocean landings. The core question we aim to address is: given various features of a Falcon 9 rocket launch, such as payload mass, orbit type, launch site, etc., can we predict if the first stage will land successfully?

Section 1

Methodology

Methodology

The overall methodology includes:

1. Data collection, wrangling, and formatting, using:
- SpaceX API
- Web scraping
2. Exploratory data analysis (EDA), using:
- Pandas and NumPy
- SQL
3. Data visualization, using:
- Matplotlib and Seaborn
- Folium
- Dash
4. Machine learning prediction, using:
- Logistic regression
- Support vector machine (SVM)
- Decision tree
- K-nearest neighbors (KNN)

Data Collection

- SpaceX API
- Web scraping

Data Collection – SpaceX API

- Present your data collection with SpaceX REST calls using key phrases and flowcharts

- Add the GitHub URL of the completed SpaceX API calls notebook (must include completed code cell and outcome cell), as an external reference and peer-review purpose

Place your flowchart of SpaceX API calls here

Data Collection - Scraping

- Present your web scraping process using key phrases and flowcharts

- Add the GitHub URL of the completed web scraping notebook, as an external reference and peer-review purpose

Place your flowchart of web scraping here

Data Wrangling

- Describe how data were processed
- You need to present your data wrangling process using key phrases and flowcharts
- Add the GitHub URL of your completed data wrangling related notebooks, as an external reference and peer-review purpose

EDA with Data Visualization

Pandas and NumPy

Functions from the Pandas and NumPy libraries are used to derive basic information about the data collected, which includes:

- The number of launches on each launch site
- The number of occurrence of each orbit
- The number and occurrence of each mission outcome

Matplotlib and Seaborn

Functions from the Matplotlib and Seaborn libraries are used to visualize the data through scatterplots, bar charts, and line charts.

The plots and charts are used to understand more about the relationships between several features, such as:

- The relationship between flight number and launch site
- The relationship between payload mass and launch site
- The relationship between success rate and orbit type

EDA with SQL

SQL

The data is queried using SQL to answer several questions about the data such as:

- The names of the unique launch sites in the space mission
- The total payload mass carried by boosters launched by NASA (CRS)
- The average payload mass carried by booster version F9 v1.1

Build an Interactive Map with Folium

Folium

Functions from the Folium libraries are used to visualize the data through interactive maps.

The Folium library is used to:

- Mark all launch sites on a map
- Mark the succeeded launches and failed launches for each site on the map
- Mark the distances between a launch site to its proximities such as the nearest city, railway, or highway

Build a Dashboard with Plotly Dash

Dash

Functions from Dash are used to generate an interactive site where we can toggle the input using a dropdown menu and a range slider.

Using a pie chart and a scatterplot, the interactive site shows:

- The total success launches from each launch site
- The correlation between payload mass and mission outcome (success or failure) for each launch site

Predictive Analysis (Classification)

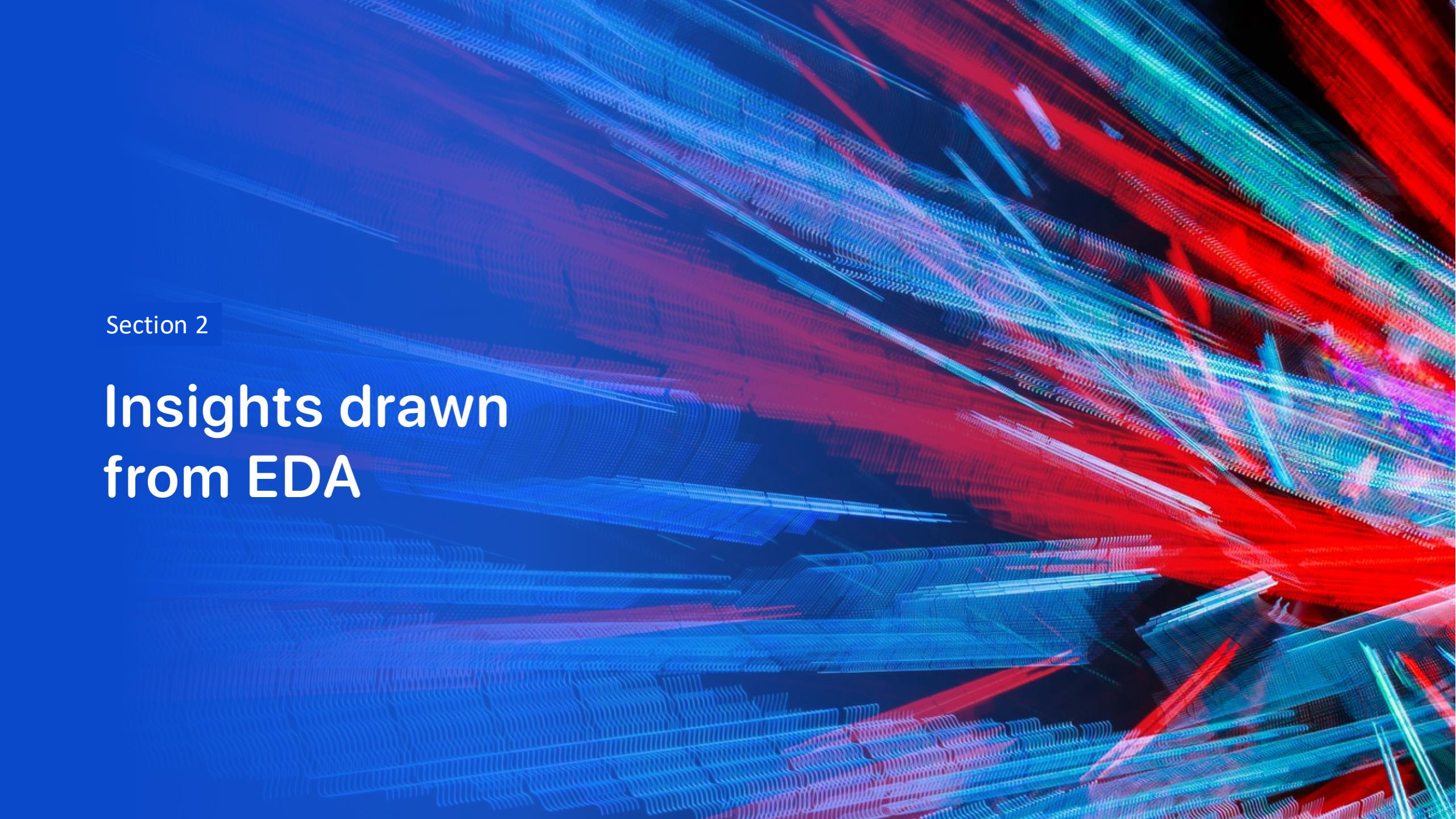
Functions from the Scikit-learn library are used to create our machine learning models.

Results

The results are split into 5 sections:

- SQL (EDA with SQL)
- Matplotlib and Seaborn (EDA with Visualization)
- Folium
- Dash
- Predictive Analysis

In all of the graphs that follow, class 0 represents a failed launch outcome while class 1 represents a successful launch outcome.


Section 2
Insights drawn from EDA

Flight Number vs. Launch Site

- Show a scatter plot of Flight Number vs. Launch Site

- Show the screenshot of the scatter plot with explanations

Payload vs. Launch Site

- Show a scatter plot of Payload vs. Launch Site

- Show the screenshot of the scatter plot with explanations

Success Rate vs. Orbit Type

- Show a bar chart for the success rate of each orbit type

- Show the screenshot of the scatter plot with explanations

Flight Number vs. Orbit Type

- Show a scatter point of Flight number vs. Orbit type

- Show the screenshot of the scatter plot with explanations

Payload vs. Orbit Type

- Show a scatter point of payload vs. orbit type

- Show the screenshot of the scatter plot with explanations

Launch Success Yearly Trend

- Show a line chart of yearly average success rate

- Show the screenshot of the scatter plot with explanations

All Launch Site Names

- Find the names of the unique launch sites
- Present your query result with a short explanation here

Launch Site Names Begin with 'CCA'

- Find 5 records where launch sites begin with `CCA`
- Present your query result with a short explanation here

Total Payload Mass

- Calculate the total payload carried by boosters from NASA
- Present your query result with a short explanation here

Average Payload Mass by F9 v1.1

- Calculate the average payload mass carried by booster version F9 v1.1
- Present your query result with a short explanation here

First Successful Ground Landing Date

- Find the dates of the first successful landing outcome on ground pad
- Present your query result with a short explanation here

Successful Drone Ship Landing with Payload between 4000 and 6000

- List the names of boosters which have successfully landed on drone ship and had payload mass greater than 4000 but less than 6000

- Present your query result with a short explanation here

Total Number of Successful and Failure Mission Outcomes

- Calculate the total number of successful and failure mission outcomes
- Present your query result with a short explanation here

Boosters Carried Maximum Payload

- List the names of the booster which have carried the maximum payload mass
- Present your query result with a short explanation here

2015 Launch Records

- List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

- Present your query result with a short explanation here

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

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

- Present your query result with a short explanation here

Section 3

Launch Sites Proximities Analysis

<Folium Map Screenshot 1>

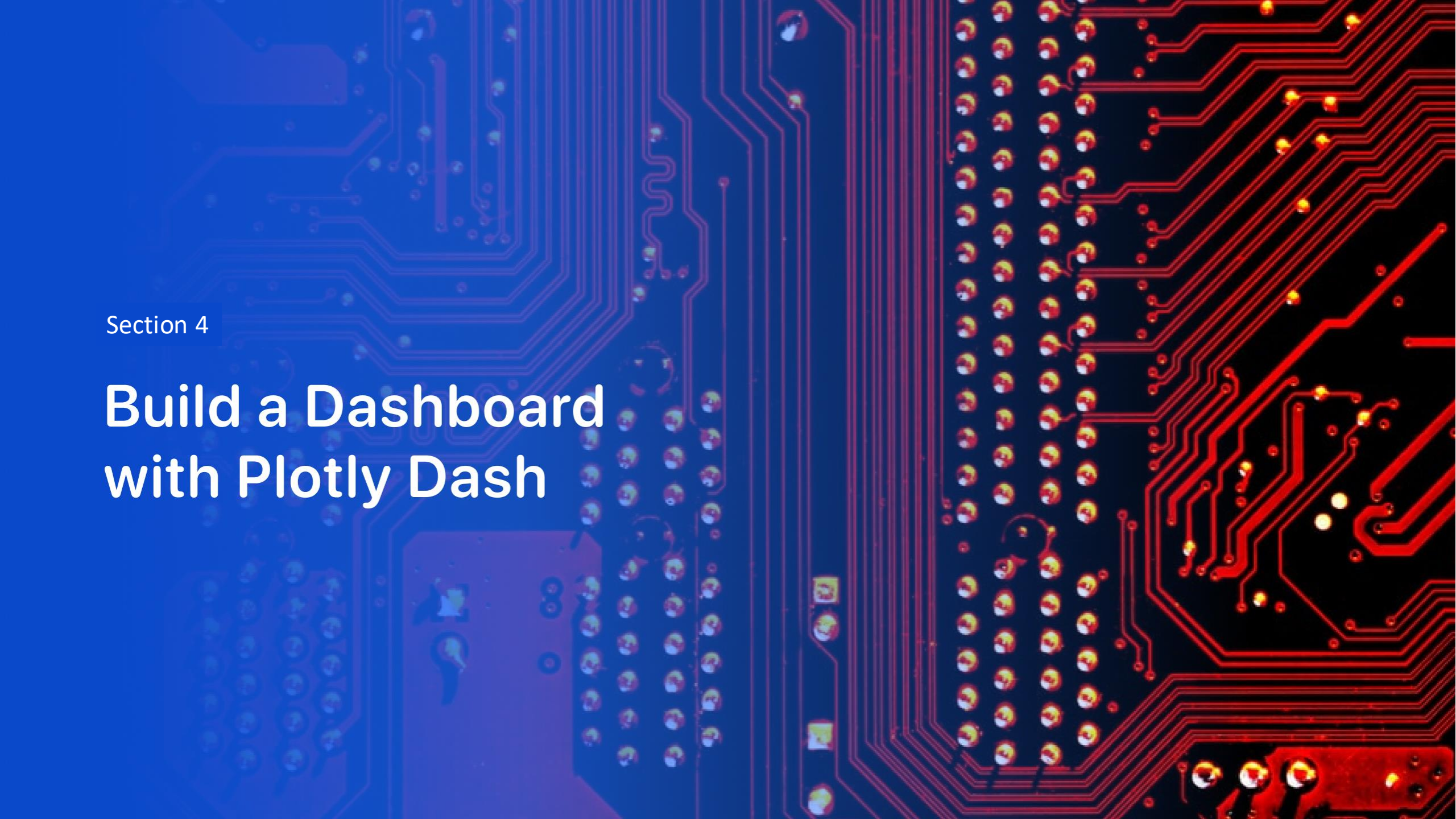
- Replace <Folium map screenshot 1> title with an appropriate title
- Explore the generated folium map and make a proper screenshot to include all launch sites' location markers on a global map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 2>

- Replace <Folium map screenshot 2> title with an appropriate title
- Explore the folium map and make a proper screenshot to show the color-labeled launch outcomes on the map
- Explain the important elements and findings on the screenshot

<Folium Map Screenshot 3>

- Replace <Folium map screenshot 3> title with an appropriate title
- Explore the generated folium map and show the screenshot of a selected launch site to its proximities such as railway, highway, coastline, with distance calculated and displayed
- Explain the important elements and findings on the screenshot


Section 4
Build a Dashboard with Plotly Dash

<Dashboard Screenshot 1>

- Replace <Dashboard screenshot 1> title with an appropriate title
- Show the screenshot of launch success count for all sites, in a piechart
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 2>

- Replace <Dashboard screenshot 2> title with an appropriate title
- Show the screenshot of the piechart for the launch site with highest launch success ratio
- Explain the important elements and findings on the screenshot

<Dashboard Screenshot 3>

- Replace <Dashboard screenshot 3> title with an appropriate title
- Show screenshots of Payload vs. Launch Outcome scatter plot for all sites, with different payload selected in the range slider
- Explain the important elements and findings on the screenshot, such as which payload range or booster version have the largest success rate, etc.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

- Visualize the built model accuracy for all built classification models, in a bar chart

- Find which model has the highest classification accuracy

Confusion Matrix

- Show the confusion matrix of the best performing model with an explanation

Conclusions

In this project, the goal is to predict whether the first stage of a Falcon 9 launch will successfully land, which can help determine the launch cost. Various features of a Falcon 9 launch, such as payload mass and orbit type, may influence the outcome of the mission. Several machine learning algorithms are used to analyze patterns in past Falcon 9 launch data and build predictive models. Among the four algorithms tested, the decision tree model demonstrated the best performance in predicting the outcome of a Falcon 9 launch.

Appendix

- Include any relevant assets like Python code snippets, SQL queries, charts, Notebook outputs, or data sets that you may have created during this project

Thank you!