



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Ma Jingjing
2021/9/29



CONTENT

01. Summary

02. Introduction

03. Methodology

04. Conclusion

05. Appendix





01

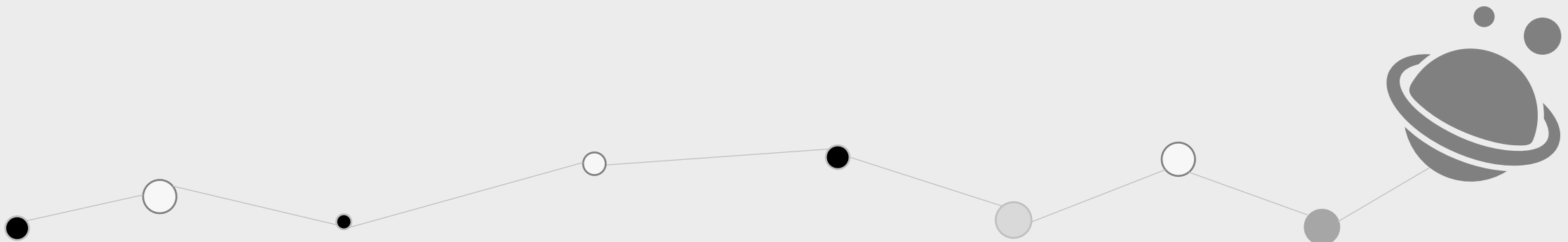
Summary

Summary of methodologies and results

01. Summary

Summary of methodologies and results

- This project is designed to predict the success rate of Falcon 9 launch.
- I used the historical data collected from SpaceX API, performed exploratory data analysis and visual analytics to understand the relationship between variables. Then, I performed feature engineering, and made predictive analysis. I used 4 supervised algorithms (K Nearest Neighbors, SVM, Decision Tree, Logistic Regression) to deal with the problem, finally, found that the decision tree is the best way to predict the success rate for Falcon 9 launch.





02

Introduction

Project background and problems we want to find answers

02. Introduction

Project background and problems we want to find answers

SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore, if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Background

Problems

- Which parameters have the higher impact in success landings for SpaceX, Falcon 9.
- Which is the best way to predict the success rate for rocket launch.

The background features a light grey field with a network of thin grey lines connecting small dots. Some dots are solid black, some are solid grey, and some are hollow white. A large, rounded grey rectangle is centered on the page. The number '03' is prominently displayed in the upper half of this rectangle. Below it, the word 'Methodology' is written in a large, bold, black sans-serif font. A thin horizontal line is positioned below the word. At the bottom of the rectangle, the phrase 'Methodology of every stage' is written in a smaller, grey sans-serif font. Four small, isolated geometric shapes (triangles or quadrilaterals with dots at vertices) are scattered around the central rectangle.

03

Methodology

Methodology of every stage

03. Methodology

Methodology of every stage



Data processing

- Collect the launch data from SpaceX API
- Clean requested data



Exploratory data analysis

- Finding relationships in the data by means of scatterplots, bar charts

Interactive visual analytics

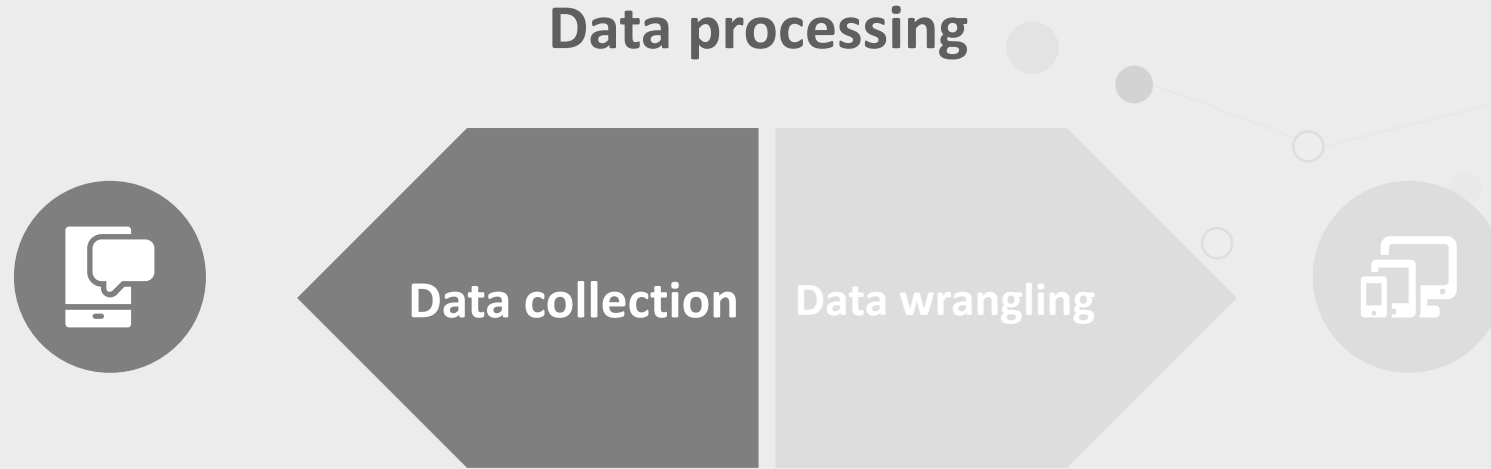
- Build maps with Folium and dashboard with Plotly Dash

Predictive analysis

- Defining suitable prediction model
- Setting and running the models
- Choosing the most accurate prediction model

03. Methodology

Methodology of every stage



- Request and parse the SpaceX launch data using the **GET** request
- Read raw data and send it to a data frame
- Filter the data frame to only include Falcon 9 launches
- Find missing value and replace them with mean

- Remove missing values
- Eliminating irrelevant columns
- Identify successful and unsuccessful launches by **One Hot Encoding**
- Get information for each column and place it into the dictionary, then convert it into a data frame

03. Methodology

Methodology of every stage

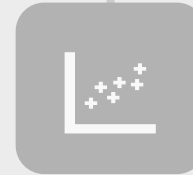


Exploratory data analysis

(EDA)

Perform Exploratory Data Analysis and Feature Engineering using Pandas and Matplotlib

Success Rate: 66.67%



Scatter chart

Visualize the relationships between variables



Bar chart

Find out the relationship between success rate and orbit type

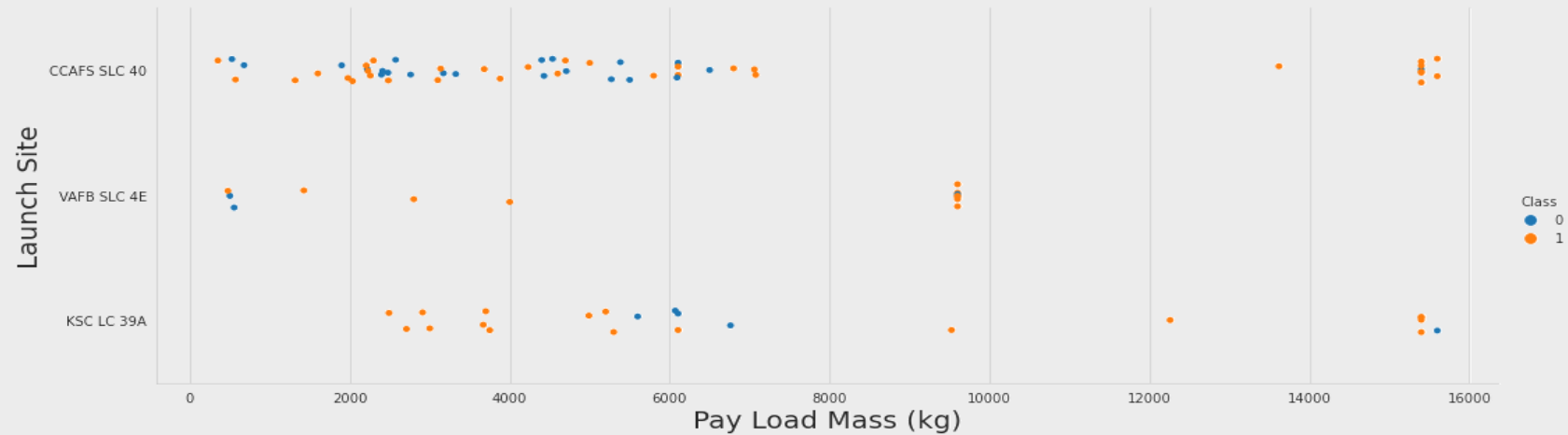


Line chart

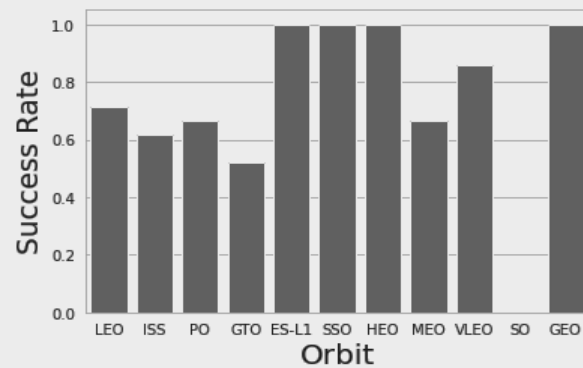
Visualize the launch success yearly trend

03. Methodology

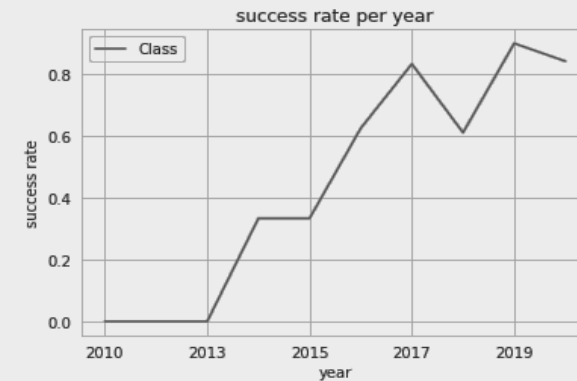
Methodology of every stage



- Different launch sites have different success rates. CCAFS LC-40: 60 %, KSC LC-39A and VAFB SLC 4E: 77%.
- In general, CCAFS SLC 40 with payload larger than 12000kg and KSC LC 39A with payload lower than 5000 have the best performance.



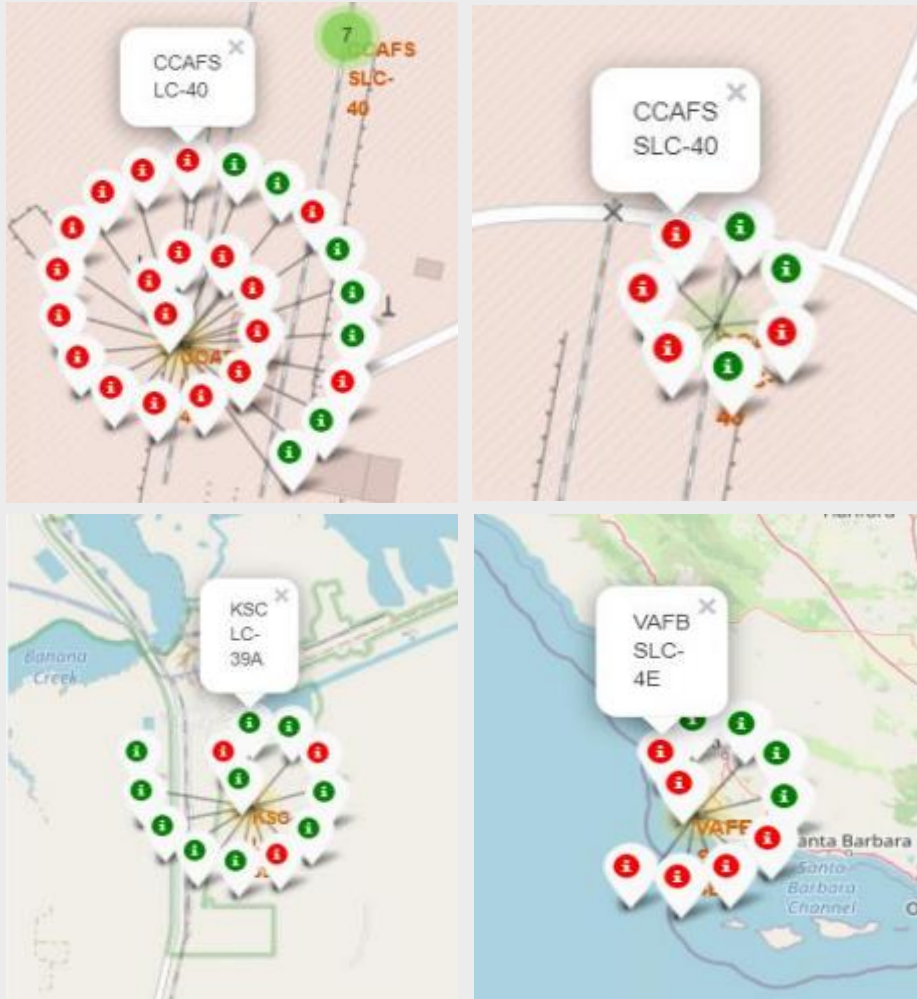
- 4 types of orbit has the best success rate : ES-L1/SSO/HEO/GEO



- The success rate since 2013 kept increasing till 2020

03. Methodology

Methodology of every stage



Interactive visual analytics

- Failure/Success labels were placed on each of the launching sites.
- KSC LC 39A in FLORIDA Shows very good results.
- The rest of sites does deliver good result, but the majority of launches have been failures.

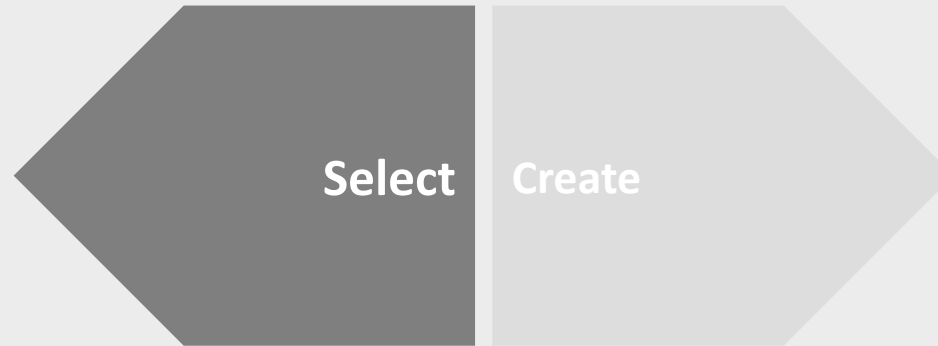
● success

● failure

03. Methodology

Methodology of every stage

Features Engineering



- Select the features that will be used in success prediction in the future module: 'FlightNumber', 'PayloadMass', 'Orbit', 'LaunchSite', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad', 'Block', 'ReusedCount', 'Serial'

- Create dummy variables to categorical columns. Use the function `get dummies` and features data frame to apply `OneHotEncoder` to the column: `Orbits`, `LaunchSite`, `LandingPad`, and `Serial`

03. Methodology

Methodology of every stage

Predictive analysis

- create a machine learning pipeline to predict if the first stage will land given the data from the preceding labs.

Determine
training labels

Create a column
for the class

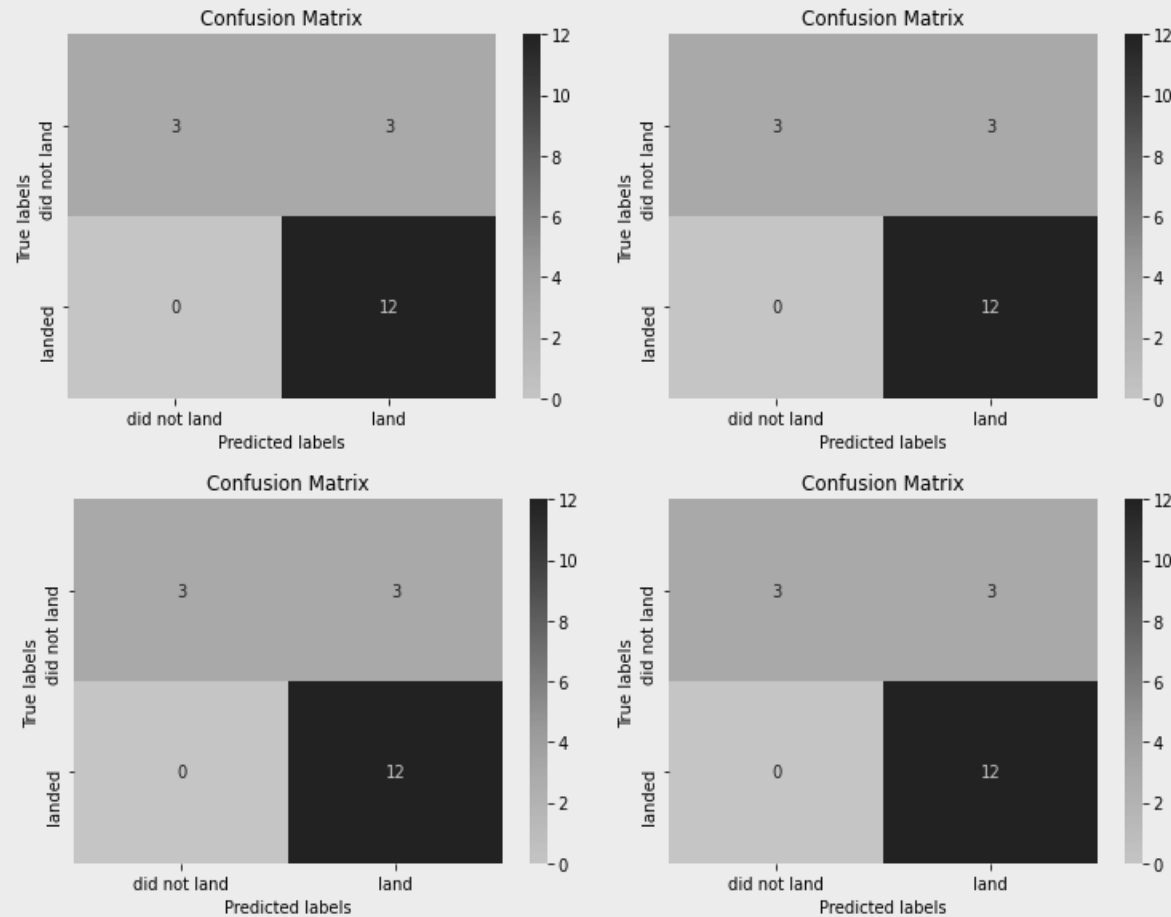
Standardize
the data

Split into training
data and test
data

Find the method
which has the best
accuracy

03. Methodology

Methodology of every stage



Predictive analysis

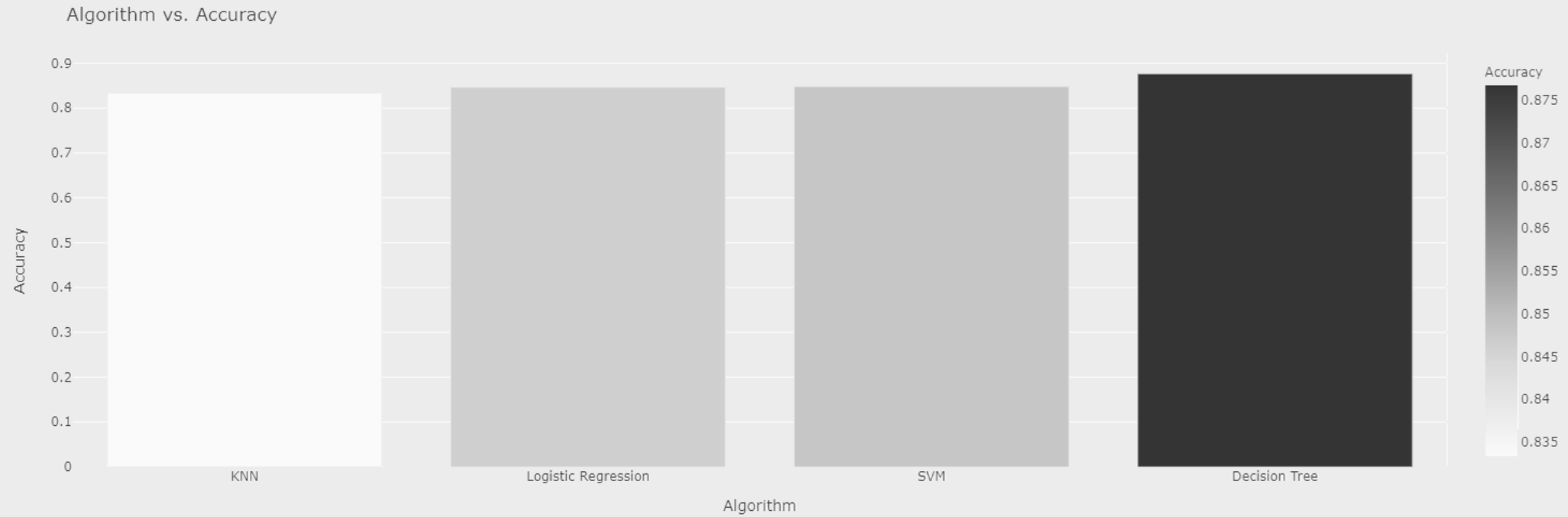
Confusion matrix for all models is the same

- Overall Accuracy=ACC= 0.8333
- precision=PPV=0.5
- Sensitivity=Recall=TPR=0.5
- specificity=TNR=1

The models are good enough
for predicting success rate

03. Methodology

Methodology of every stage



- Decision tree classifier has the highest classification accuracy

The background features a light grey field with several abstract geometric elements. A large, rounded grey rectangle is centered, containing the main text. Scattered around this rectangle are various line patterns: some consist of three points forming a triangle, others are more complex networks of interconnected points and lines. The points are in different shades of grey and black, and the lines are thin and light grey.

04

Conclusion

Conclusion of the project

04. Conclusion

Conclusion of the project



Orbits ES-L1/SSO/HEO/GEO has the best success rate

KSC LC 39A with Payloads lower than 5000 kg has best success rates

Decision Three Classifier has the best accuracy(0.88)

The background features a light gray field with a network of thin gray lines connecting small black and white dots. A large, rounded gray rectangle is centered on the page. Inside this rectangle, the number '05' is prominently displayed in a large, black, sans-serif font. Below it, the word 'Appendix' is written in a smaller, black, sans-serif font. A thin horizontal line is positioned between 'Appendix' and the text 'Relevant assets', which is written in a gray, sans-serif font. The overall aesthetic is minimalist and modern, with a focus on geometric shapes and a clean layout.

05

Appendix

Relevant assets

05. Appendix

Conclusion of the project

Notebooks about this project:

- [GitHub Link](#)

Project source:

- [IBM Data Science](#)

The background is a light gray field filled with a complex network of thin, light gray lines connecting small dots. The dots are in three colors: black, white, and a medium gray. Some dots are isolated, while others are part of small clusters or larger, more interconnected groups. The overall effect is a subtle, modern geometric pattern.

Thanks