APPLIED DATA SCIENCE CAPSTONE:

SPACE X FALCON 9 FIRST LANDING PREDICTION

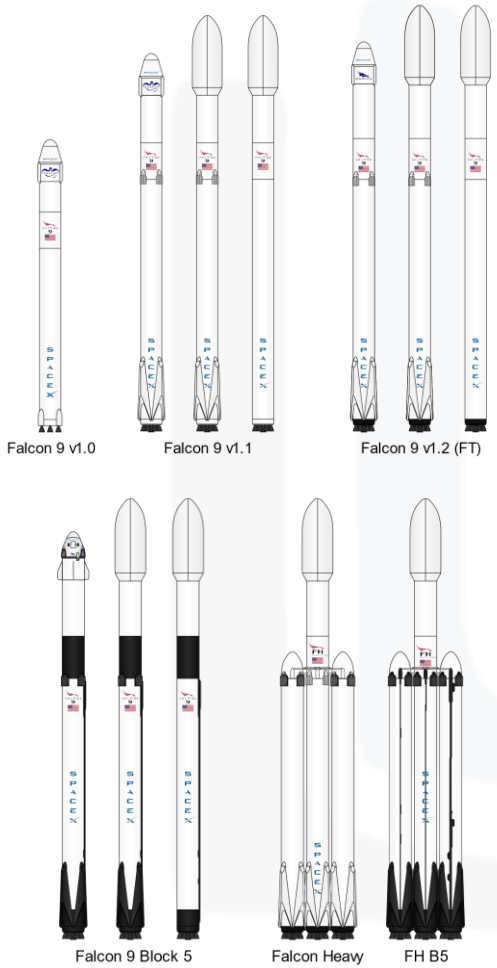
Stanly H.

4 October 2023

IBM Developer

SKILLS NETWORK

OUTLINE



- Executive Summary
- Introduction
- Methodology
- Results
  - Visualization – Charts
  - Dashboard
- Discussion
  - Findings & Implications
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Main Goal : **Predict** if the **Falcon 9** first stage **will land successfully**.

Logistic Regression 0.83	Support Vector Machine 0.83	**Tree 0.94**	K Nearest Neighbour

Selected Strategy : Tree Predictive Model

How do we know Falcon 9 first stage land successfully?

Predictive Model

Predictive modeling is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data. The outcome predicted will be successful or unsuccessful (class 1 or 0) shown in the dataset

INTRODUCTION

In this capstone, we will **predict** if the **Falcon 9** first stage **will land successfully**.

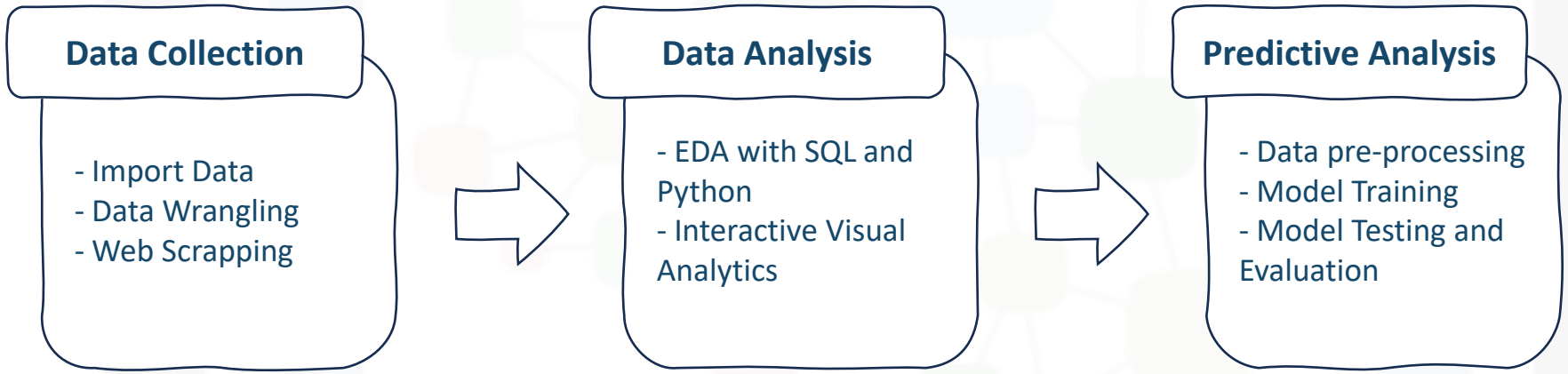
Based on prediction and data availability, we can determine the **cost of a launch**.

This information can be used if an alternate company wants to bid against SpaceX for a rocket launch.

Therefore, the main question in this project will be

- **How do we know Falcon 9 first stage land successfully?**
- **How accurate was the prediction model?**
- **Which Model was used?**

METHODOLOGY



METHODOLOGY

Data Collection

- Import Data
- Web Scrapping
- Data Wrangling

Import Data

Obtaining Data from available data files
In this project we use available .json file (https://api.spacexdata.com/v4/launches/past)

Web Scrapping

Means of obtaining more data from a website, usually displayed in unstructured form.
Extracting several tables in the html sites or remade them into structured for use.

Data Wrangling

The processing raw data into usable form. Involving normalization, restructuration, renaming, selection or replacing of available / unavailable data

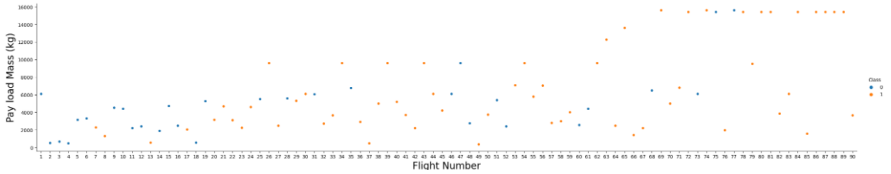
METHODOLOGY

Data Analysis

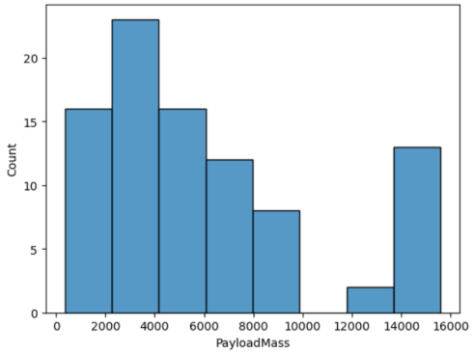
- EDA with SQL and Python
- Interactive Visual Analytics

EDA

Known as Exploratory Data Analysis
to analyze and investigate data sets and summarize their main characteristics, often employing data visualization methods.



Either to know relationships or distribution of data



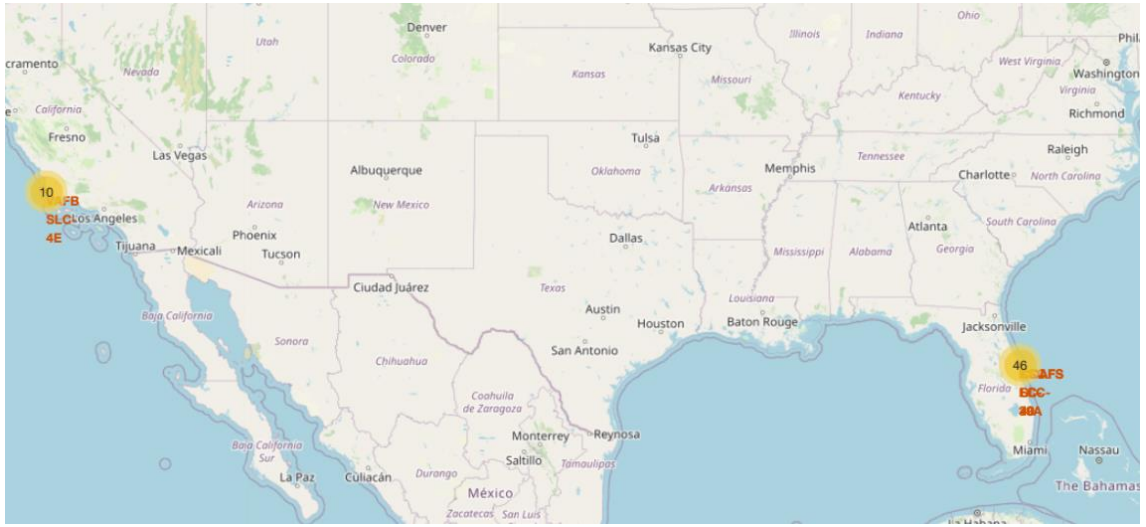
METHODOLOGY

Data Analysis

- EDA with SQL and Python
- Interactive Visual Analytics

Interactive Visual

was used w interactivity to opt, pan, zoom into the targeted location



METHODOLOGY

Predictive Analysis

- Data pre-processing
- Model Training
- Model Testing and Evaluation

Model Building

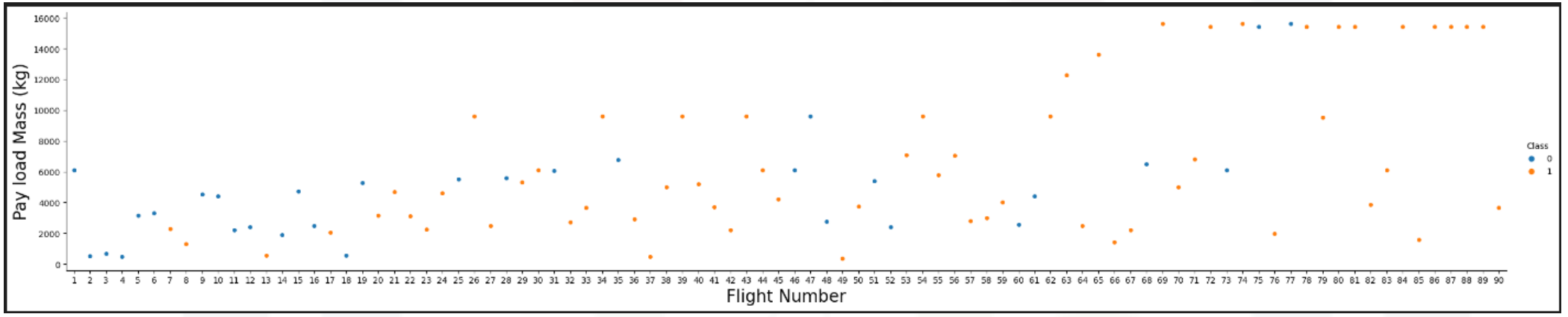
Data Pre-processing
refers to the cleaning, transforming, and integrating of data in order to improve the quality of the data and to make it more suitable for the specific data mining task.

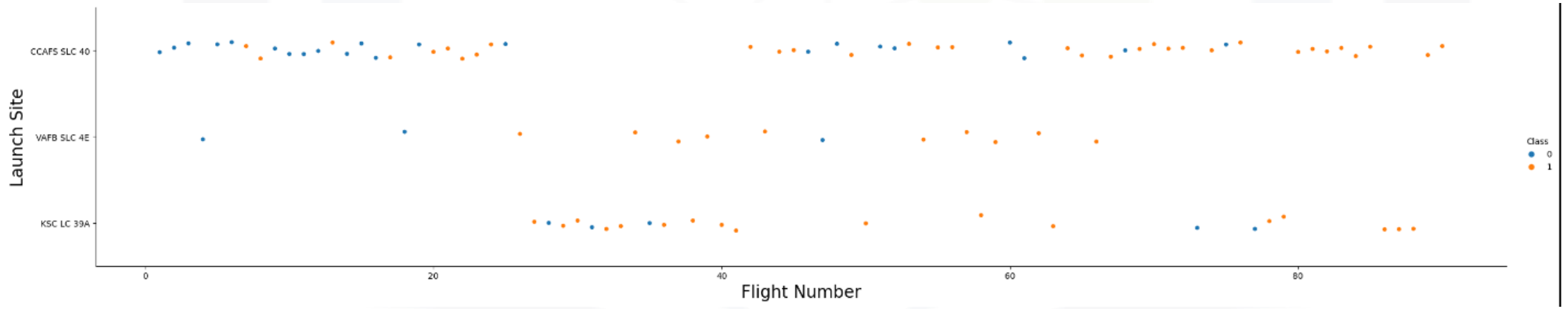
Model Training
Model training is the phase in the data science development lifecycle where practitioners try to fit the best combination of weights and bias to a machine learning algorithm to minimize a loss function over the prediction range.

Model Testing and Evaluation
the process where the performance of a fully trained model is evaluated on a testing set.

RESULTS

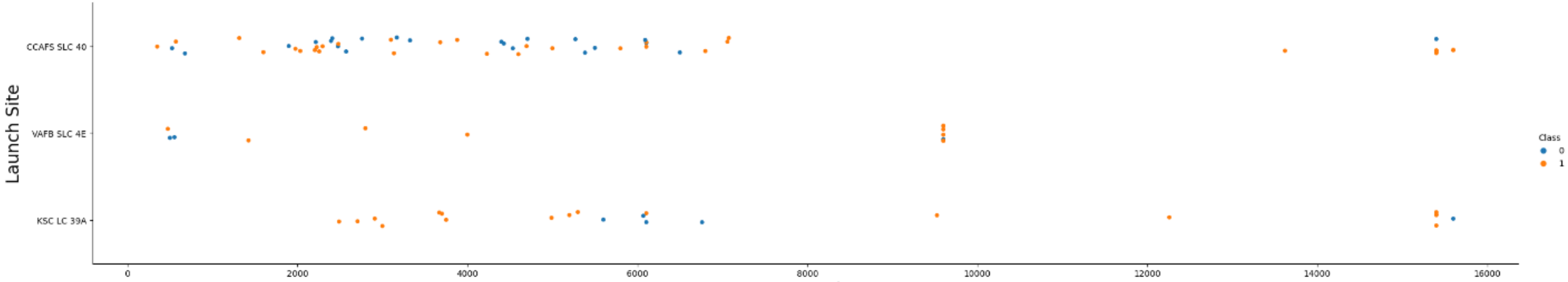
EDA with visualization result

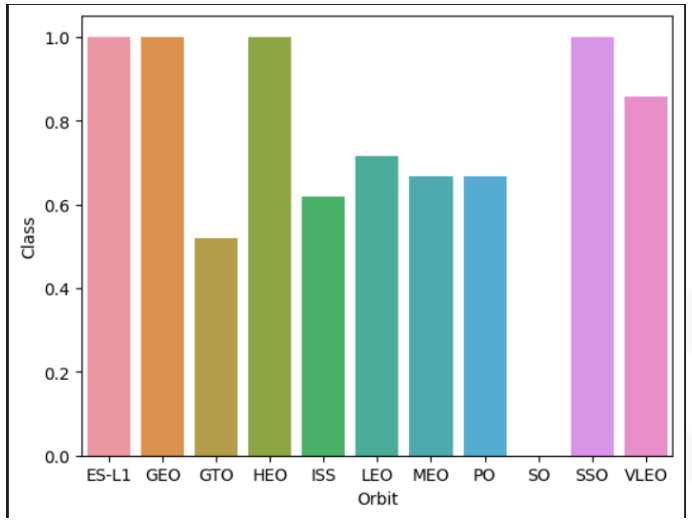


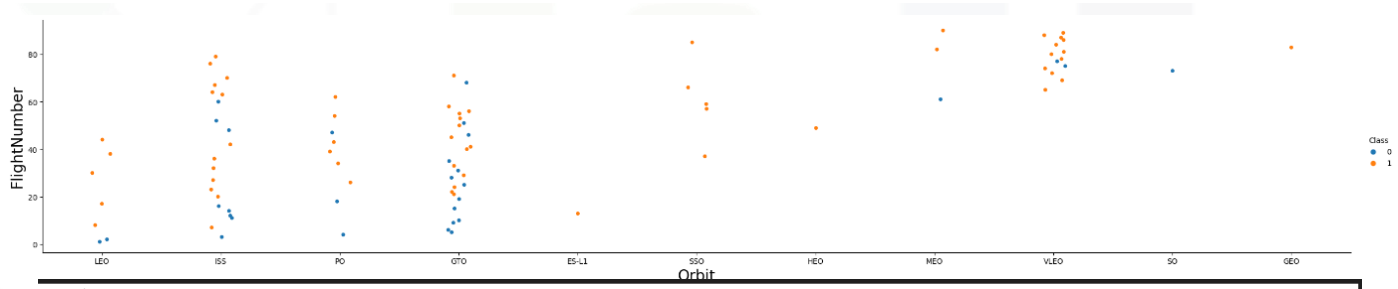


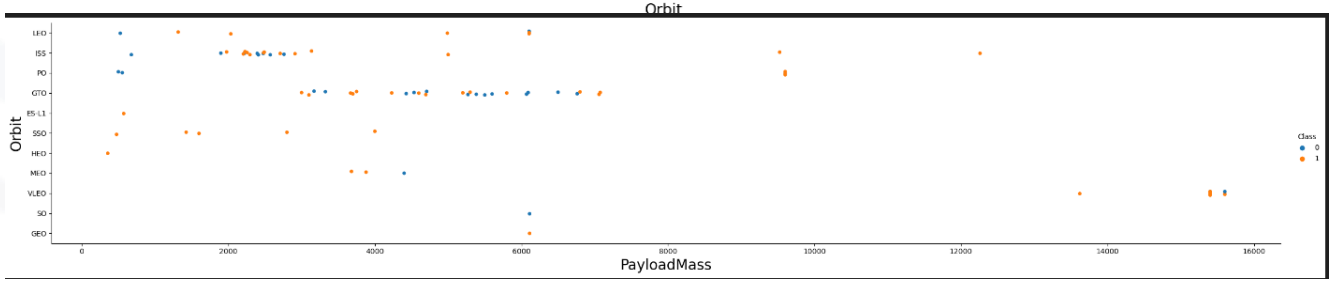
RESULTS

EDA with visualization result









RESULTS

EDA with SQL

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

min(PAYLOAD_MASS__KG_)
0

Customer	Total Mass
NASA (COTS) NRO	107010

Booster_Version	Total Mass
F9 v1.1 B1003	2534.6666666666665

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-04-06	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-08-12	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
2012-08-10	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-01-03	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2015-12-22	01:29:00	F9 FT B1019	CCAFS LC-40	OG2 Mission 2 11 Orbcomm-OG2 satellites	2034	LEO	Orbcomm	Success	Success (ground pad)

Booster_Version	Landing_Outcome	PAYLOAD_MASS__KG_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1026	Success (drone ship)	4600
F9 FT B1021.2	Success (drone ship)	5300
F9 FT B1031.2	Success (drone ship)	5200

Date	Landing_Outcome
2015-10-01	Failure (drone ship)
2015-04-14	Failure (drone ship)

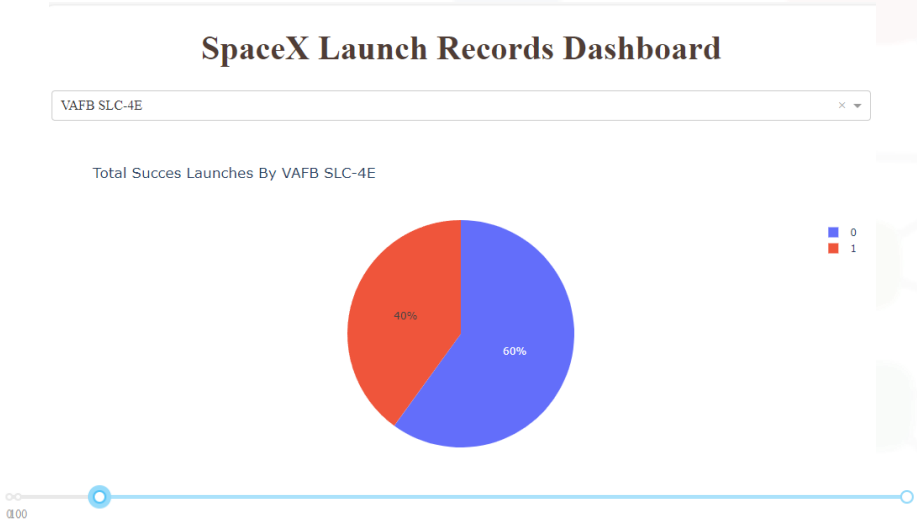
Landing_Outcome	Counted Outcome
No attempt	10
Success (ground pad)	5
Success (drone ship)	5
Failure (drone ship)	5
Controlled (ocean)	3
Uncontrolled (ocean)	2
Precluded (drone ship)	1
Failure (parachute)	1

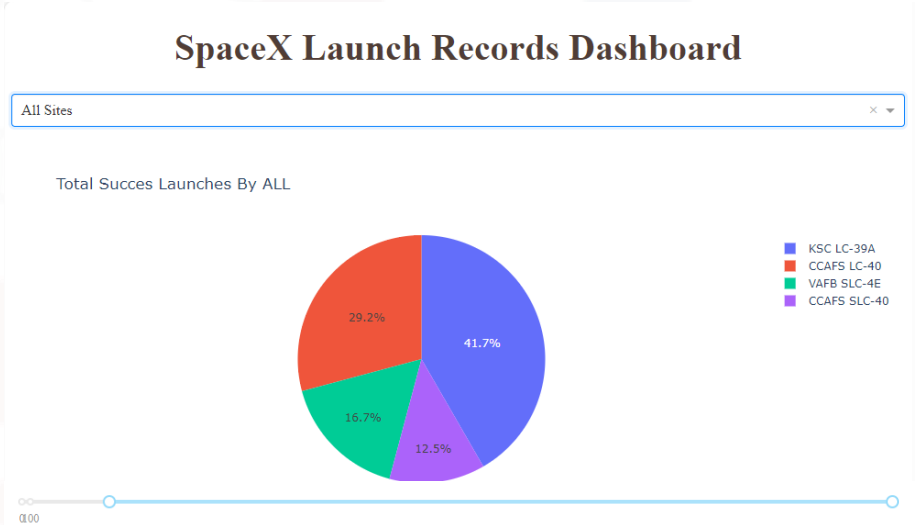
RESULTS

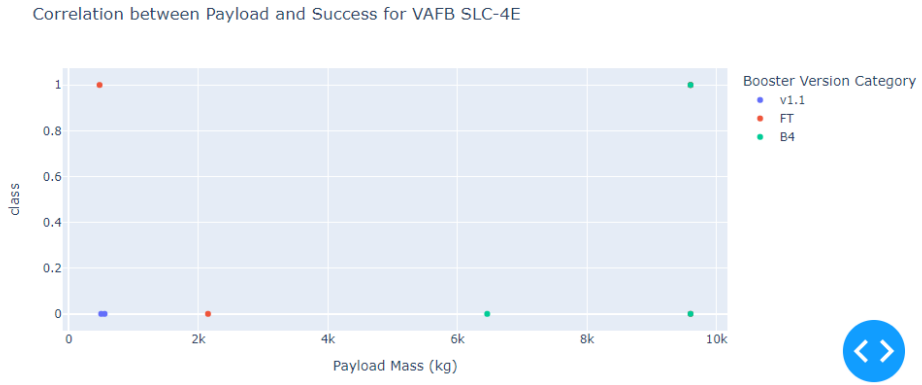
Interactive with Folium

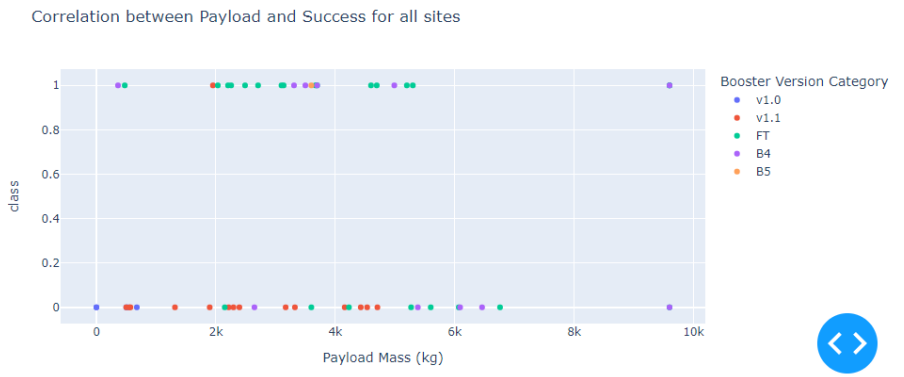
RESULTS

Interactive with Plotly Dash







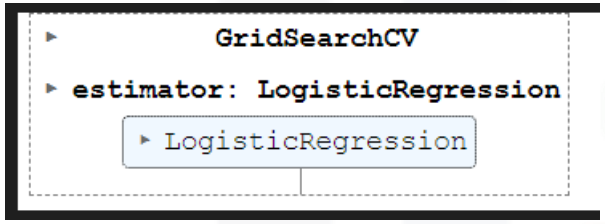


IBM Developer

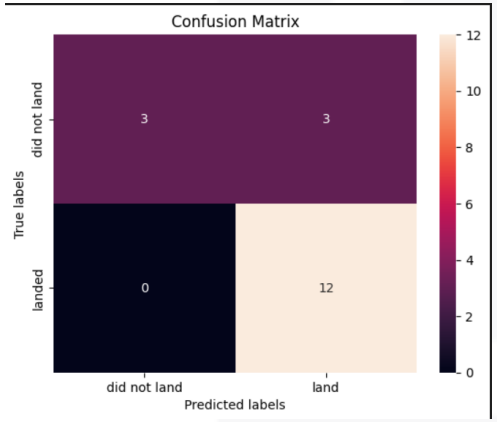
SKILLS NETWORK

RESULTS

Predictive Analysis – Logistic Regression

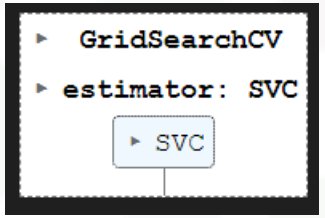



```
tuned hpyerparameters :(best parameters)  {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
accuracy : 0.8464285714285713
```

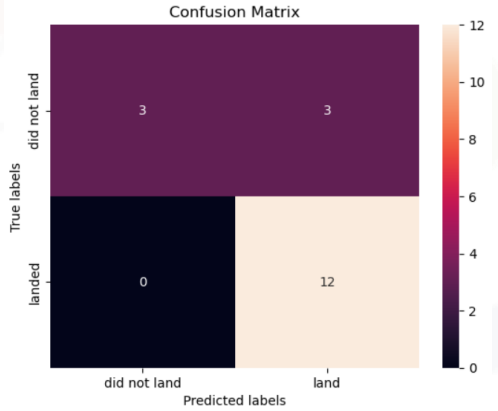



```
Test Score for Logistic Regression 0.8333333333333334
```

Predictive Analysis – SVM



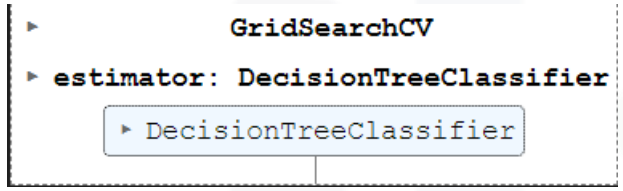

```
tuned hpyerparameters :(best parameters)  {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
```



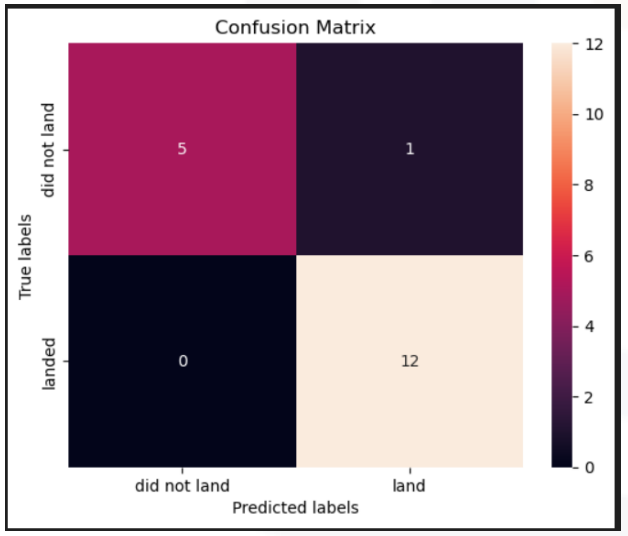

```
Test Score for Logistic Regression 0.8333333333333334
```

RESULTS

Predictive Analysis – Tree

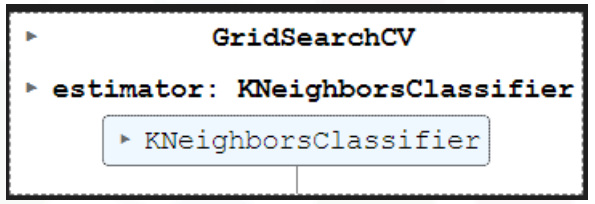



```
tuned hpyerparameters :(best parameters)  {'criterion': 'gini', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 5,
accuracy : 0.9035714285714285
```

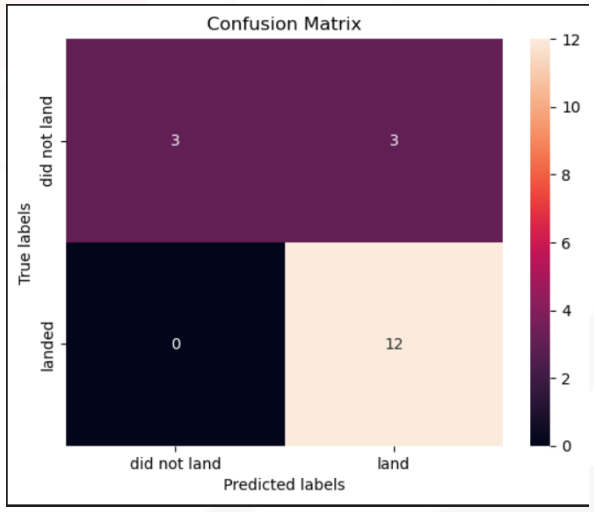



```
Test Score for Logistic Regression 0.9444
```

Predictive Analysis – K Nearest Neighbour




```
tuned hpyerparameters :(best parameters)  {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
accuracy : 0.8482142857142858
```




```
Test Score for Logistic Regression 0.83333
```

CONCLUSION

How do we know Falcon 9 first stage land successfully?

Predictive Model
Predictive modeling is a mathematical process used to predict future events or outcomes by analyzing patterns in a given set of input data. The outcome predicted will be successful or unsuccessful (class 1 or 0) shown in the dataset

How accurate was the prediction model? Which Model was used?

Model Accuracy
The model was built based on a separated dataset which is training dataset and tested on test dataset. Which will allow unbiased accuracy.
In this project, the best model was the tree model with the **accuracy of 0.94**

TO DO

The main grading criteria will be:

- Uploaded the URL of your GitHub repository including all the completed notebooks and Python files (1 pt)
- Uploaded your completed presentation in PDF format (1 pt)
- Completed the required Executive Summary slide (1 pt)
- Completed the required Introduction slide (1 pt)
- Completed the required data collection and data wrangling methodology related slides (1 pt)
- Completed the required EDA and interactive visual analytics methodology related slides (3 pts)
- Completed the required predictive analysis methodology related slides (1 pt)
- Completed the required EDA with visualization results slides (6 pts)
- Completed the required EDA with SQL results slides (10 pts)
- Completed the required interactive map with Folium results slides (3 pts)
- Completed the required Plotly Dash dashboard results slides (3 pts)
- Completed the required predictive analysis (classification) results slides (6 pts)
- Completed the required Conclusion slide (1 pts)
- Applied your creativity to improve the presentation beyond the template (1 pts)
- Displayed any innovative insights (1 pts)