IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Alaxander Hamilton
02/05/2022

Outline

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

Executive Summary

- Space X Falcon 9 First Stage Landing data were studied with machine learning classification techniques to predict the landing. Logistic regression, Support Vector Machine, Decision Tree and K-Nearest Neighbors algorithms were used and produced accuracy metrics.
- All four classification algorithms are comparable in performance in this dataset. However, decision tree algorithm may perform slightly better.

Introduction

We want to perform exploratory Data Analysis and determine Training Labels on the data and

•create a column for the class
•Standardize the data
•Split into training data and test data
-Find best Hyperparameter for SVM, Classification Trees and Logistic Regression
•Find the method performs best using test data

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - This dataset was downloaded from its repository.

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/api/dataset_part_2.csv

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/api/dataset_part_3.csv

- Perform data wrangling
  - The dataset was preprocessed and prepared for further analysis with pandas and sklearn libraries.
- Perform predictive analysis using classification models
  - Predictive analysis was performed using sklearn. The data set was split into two :training and test to train and validate the classifier.

Data Collection

- This dataset is from SpaceX and it is publicly available. For this project, the dataset was facilitated by IBM and available at:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/api/dataset_part_2.csv

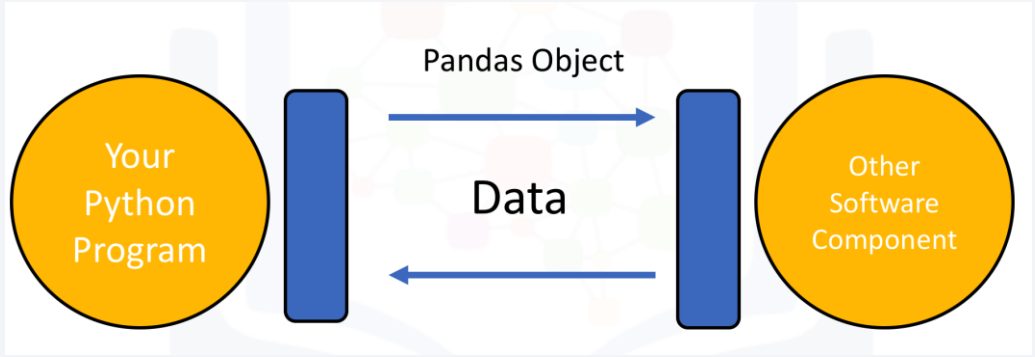
https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBMDeveloperSkillsNetwork-DS0701EN-SkillsNetwork/api/dataset_part_3.csv

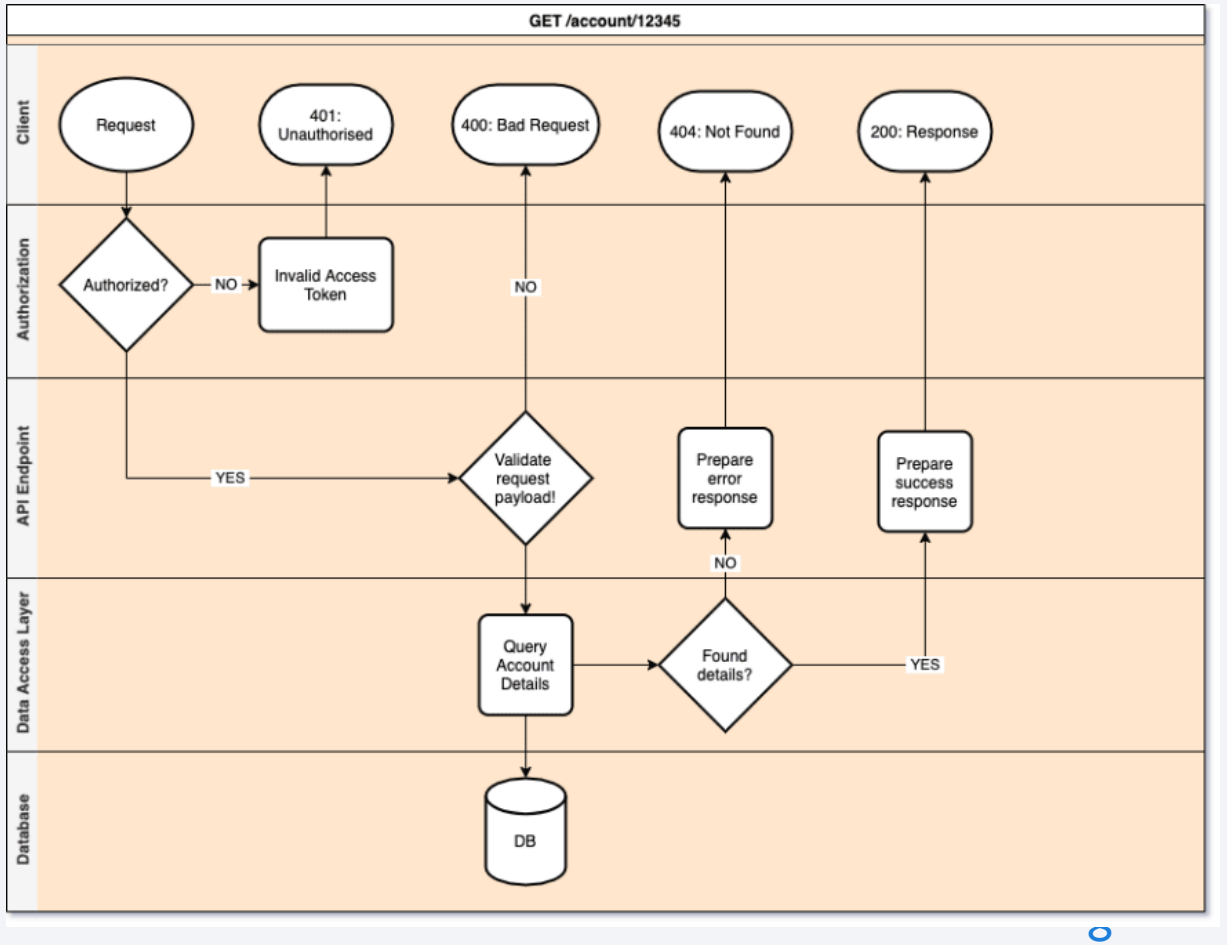
```
spacex_url="https://api.spacexdata.com/v4/launches/past"


response = requests.get(spacex_url)
```

Data Collection – SpaceX API

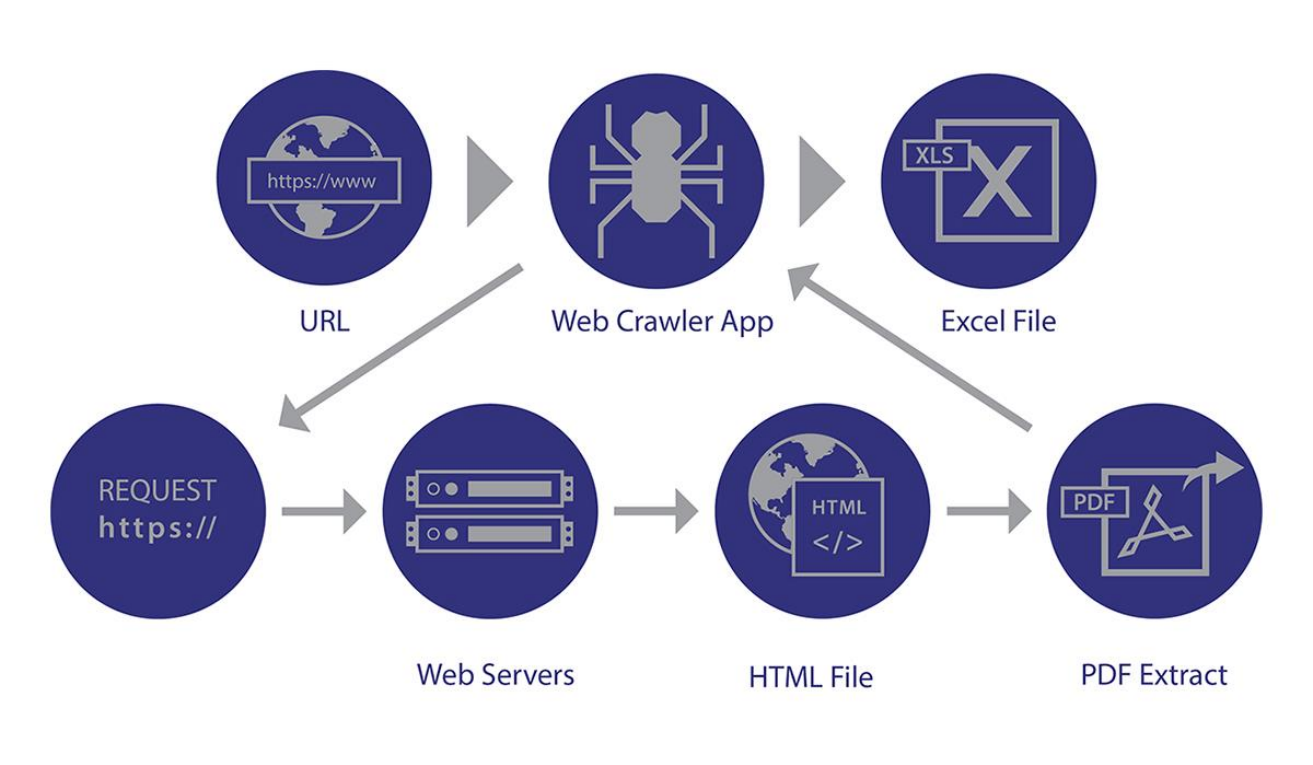
https://github.com/r-spacex/SpaceX-API





Data Collection - Scraping

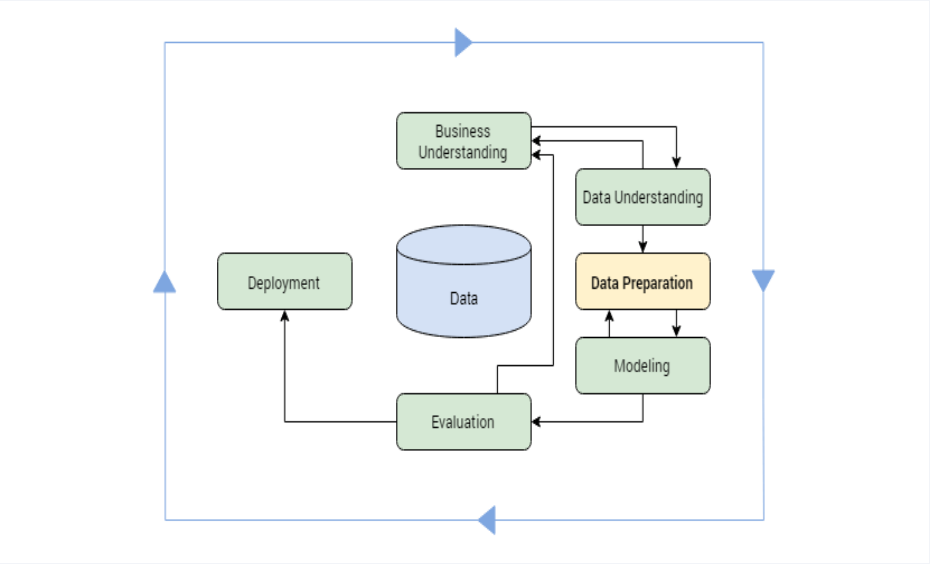
- https://github.com/dennislamcv1/IBMDAProject/blob/main/Web-Scraping-Review-Lab.ipynb



Data Wrangling

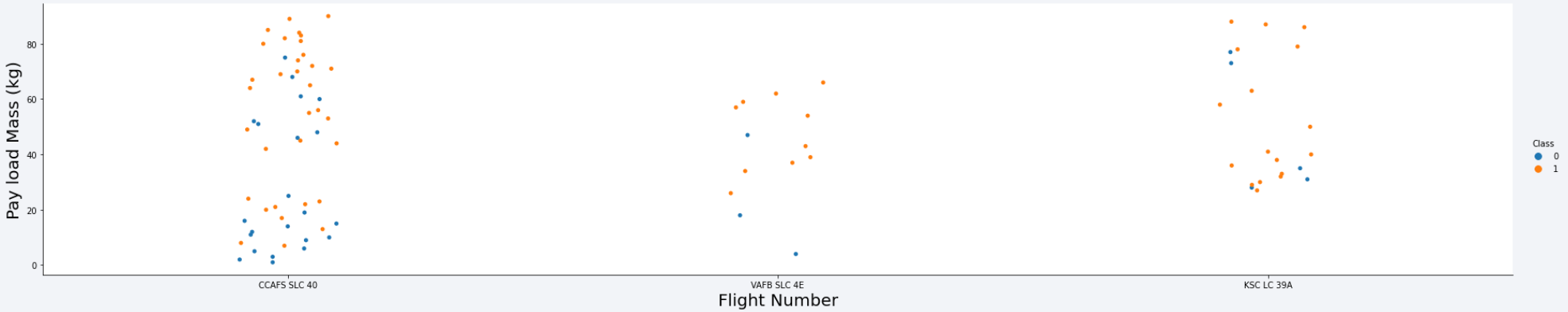
- Preprocessing — the initial state that occurs right after the acquiring of data.

* Standardizing data into an understandable format. Sort the by types of events and time stamps and converting categorical variables into 1 or 0.

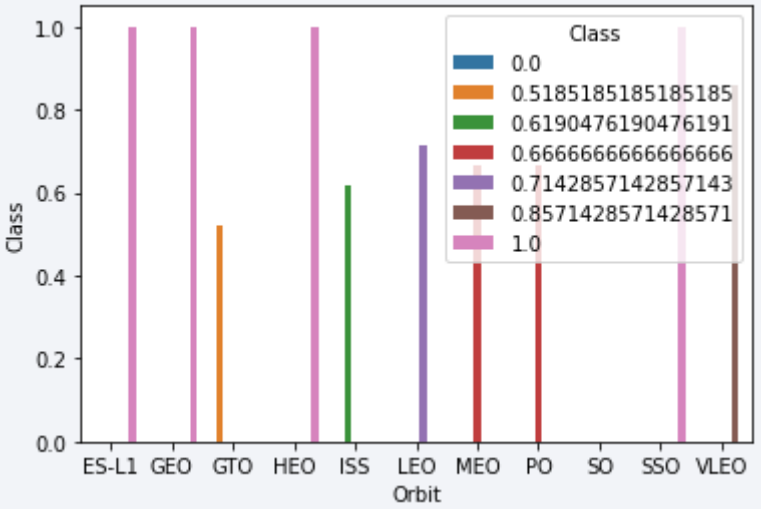



```
Y=data['Class'].to_numpy()
Y
array([0, 0, 0, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 0, 0, 1, 1, 1,
       1, 1, 0, 1, 1, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 0, 0, 0, 1, 1, 0, 0, 1, 1, 1, 1, 1, 1, 1, 0, 0, 1, 1, 1, 1, 1,
       1, 0, 1, 1, 1, 1, 0, 1, 0, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,
       1, 1])
```

EDA with Data Visualization



EDA and data visualization were done using pandas and seaborn.



EDA with SQL

- EDA with SQL was fascinating.
- We could do queries with select statement. It’s a powerful command. There are two categories of command. They are DDL and DML. Select command belongs to database management language whereas CREATE belongs to DDL.

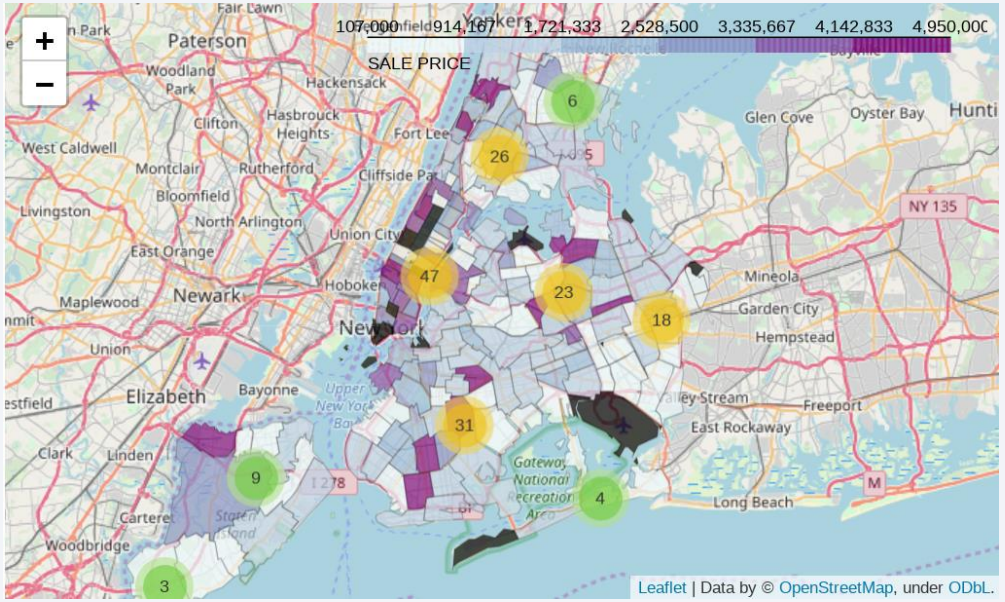
Build an Interactive Map with Folium

- Folium is a versatile package.
- We can do varieties of maps including terrains and toner.

```
# create map of Toronto using latitude and longitude values
map_toronto = folium.Map(location=[latitude, longitude], zoom_start=12)

# add markers to map
for lat, lng, borough, neighborhood in zip(toronto_data['Latitude'], toronto_data['Longitude'],
                                           toronto_data['Borough'], toronto_data['Neighborhood']):
    label = '{}, {}'.format(neighborhood, borough)
    label = folium.Popup(label, parse_html=True)
    folium.CircleMarker(
        [lat, lng],
        radius=3,
        popup=label,
        color='blue',
        fill=True,
        fill_color='#3186cc',
        fill_opacity=0.7,
        parse_html=False).add_to(map_toronto)

map_toronto
```



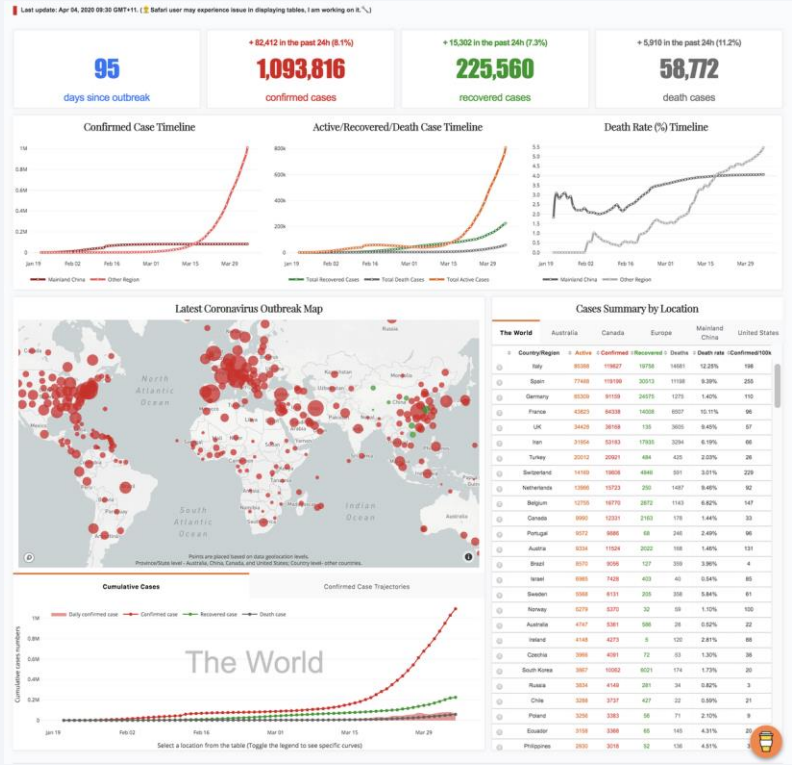
Build a Dashboard with Plotly Dash

- Plotly dash can produce interactive as well as real time charts and plots.
- Very useful in financial markets and in any such application.

```
import dash
import dash_html_components as html
import dash_core_components as dcc
import plotly.graph_objects as go
import plotly.express as px
```

```
app = dash.Dash()   #initialising dash app
df = px.data.stocks() #reading stock price dataset
```

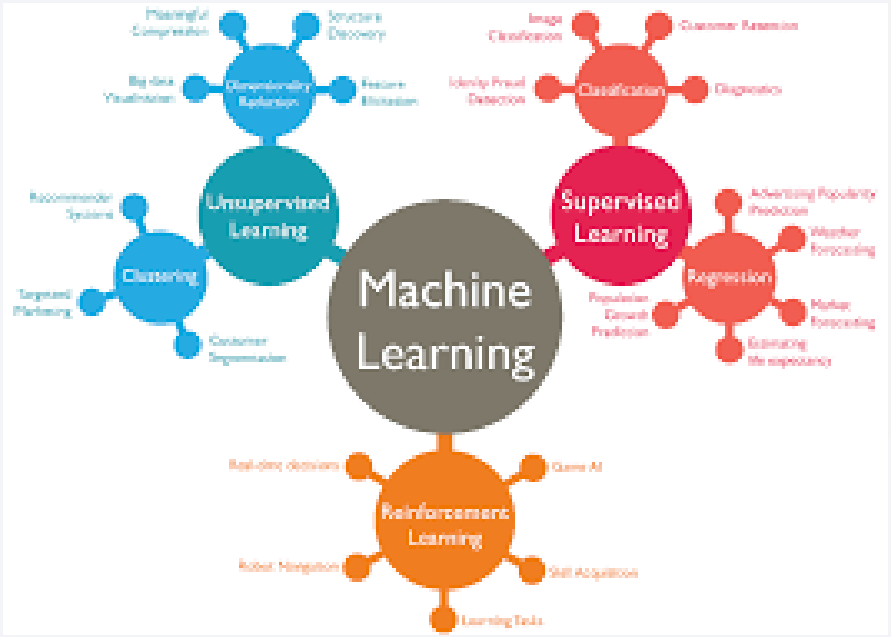
https://github.com/plotly/dash-sample-apps

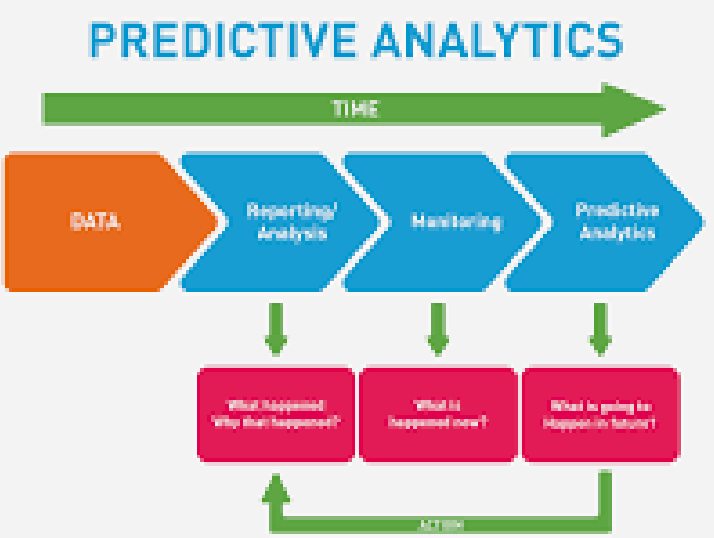


Predictive Analysis (Classification)

- Predictive analysis is essential component in data analysis.
- It can be used to predict outcome or classify new data.

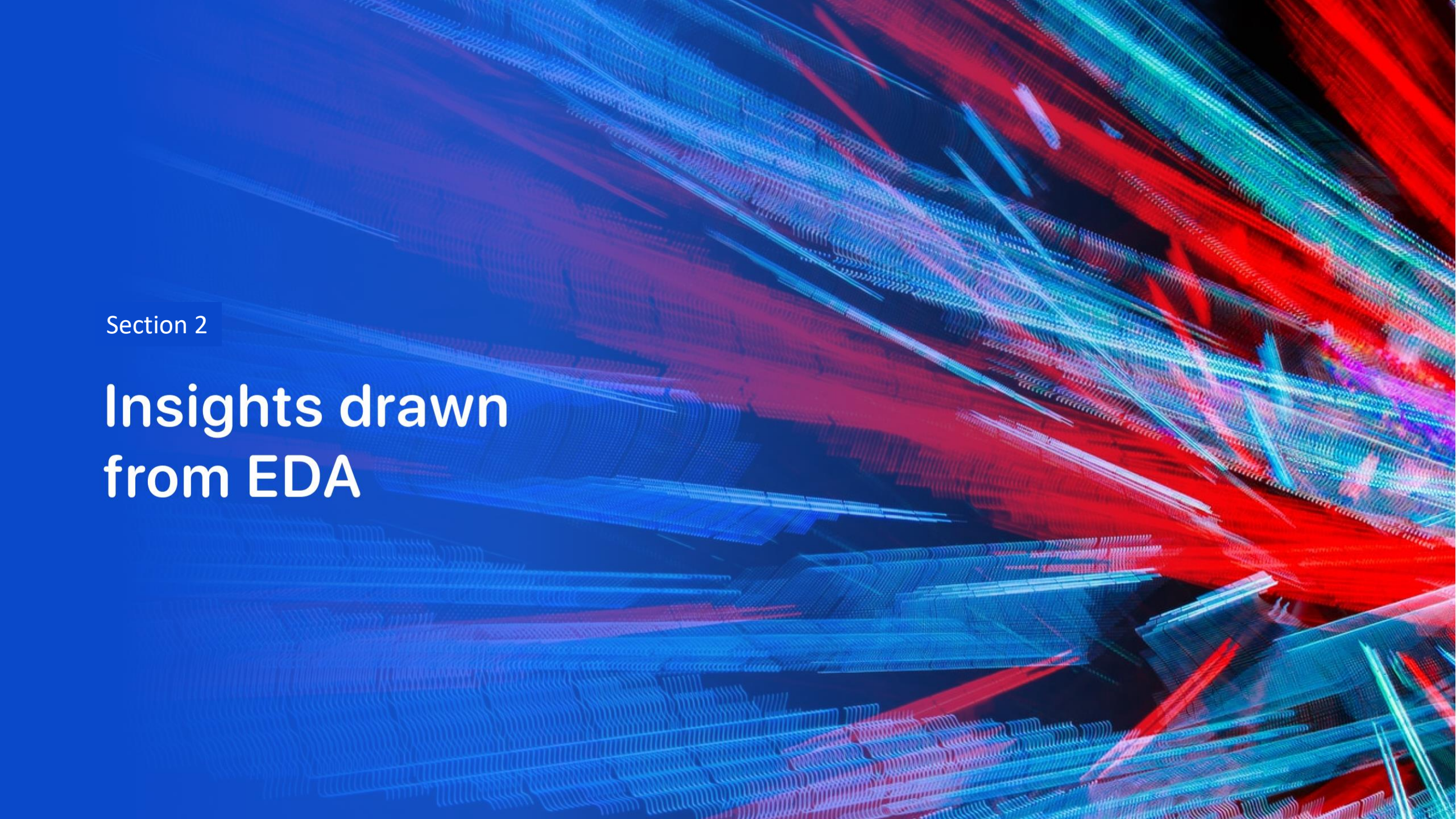
Predictive analysis involves machine learning methods.



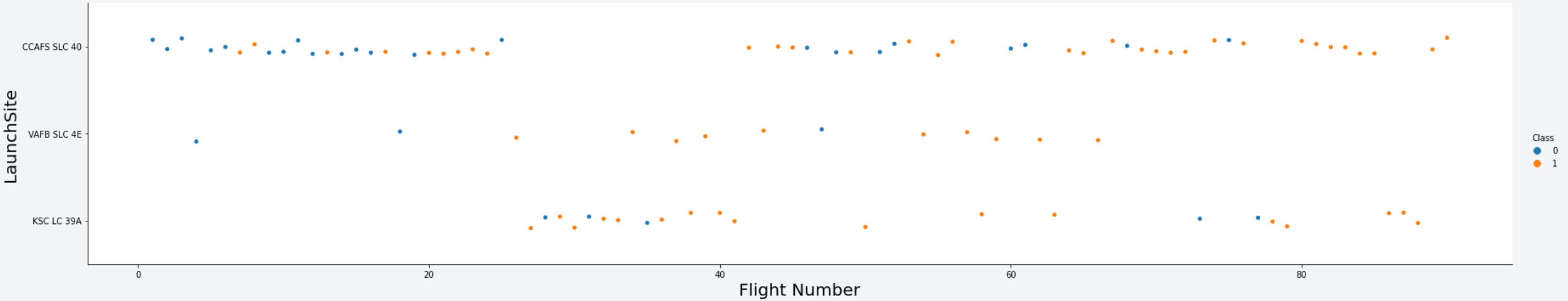


Results

All four algorithms perform almost equal well. However, decision tree algorithm show slightly better accuracy.

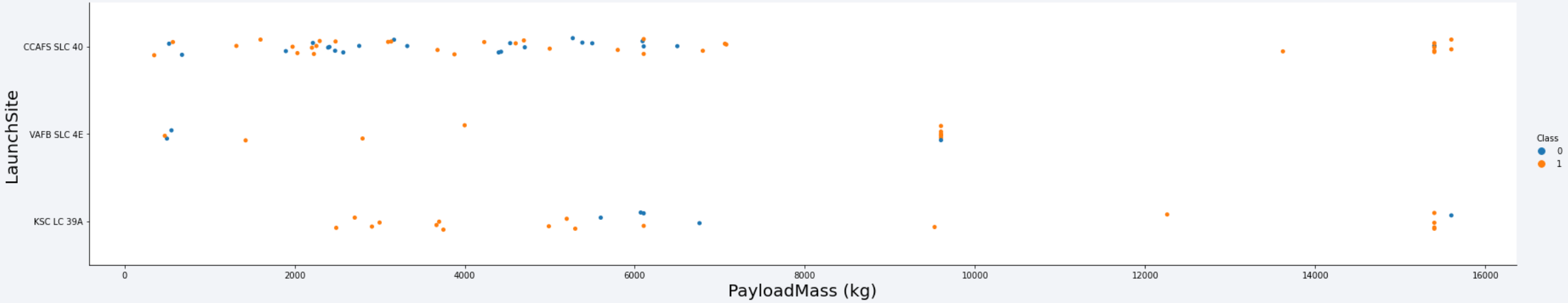

Section 2
Insights drawn from EDA

Flight Number vs. Launch Site



Launch site CCAFS SLC 40 also has low flight numbers.

Payload vs. Launch Site



Launch site CCAFS SLC 40 has good distribution of pay load mass(kg). WAFB SLC 4E does not have extreme payloads.

Section 5

Predictive Analysis (Classification)

Classification Accuracy

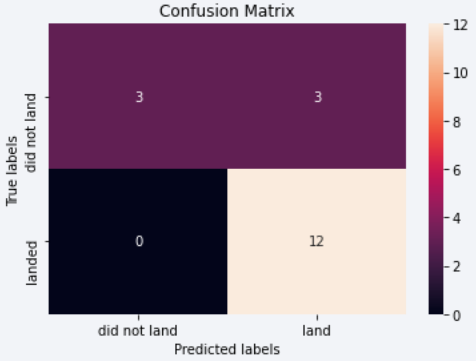
- Classification Accuracy was done using score method.

Decision tree has an accuracy about 0.94 whereas all other models as about 0.83

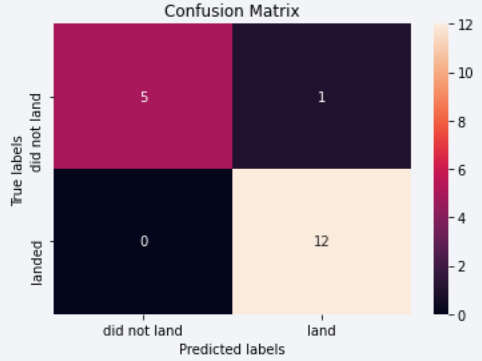
Confusion Matrix

- Decision Tree confusion matrix has the highest accuracy, and therefore decision tree model is the best performing model for this dataset.

Logistic Regression



Support Vector Machine



Decision Tree

KNN

Conclusions

- Machine learning can successfully be used to predict the success of a future flight. Models should be fine tuned with various parameters.

Appendix

- https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/bdb33af8-ee66-49dc-b1e6-7aa3c140ae26/view?access_token=62f8714ddbb9380c4c162fa4671b90b50220dbc49507878e5761f2578a6f0af3
- https://www.spacex.com/launches/
- https://github.com/IBM
- https://www.coursera.org/

Thank you!