IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Outline

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

Executive Summary

- Analysis of data specifically for different launching site shows that the maximum number of launches for Falcon 9 Rockets were from Cape Carnival site and VLEO orbit accounted for maximum deployment.
- Support Vector Machine performed best among other classification algorithms to predict the successful landing of Falcon 9 Rockets.
- KSC LC launch site showed the highest success rate while Orbit of payload deployment and payload mass were found to have good correlation with the success of the mission.

Introduction

- **Project background and context**
- We predicted if the Falcon 9 first stage will land successfully. Space X advertises Falcon 9 rocket on it website, with a cost of 62 million dollars; other provider cost upward of 165 million dollar each, much of the saving is because Space X can reuse the first stage. Therefore, if we can determine it the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to be against Space X for a rocket launch. .
- **Problems you want to find answers**
- What factor affect successful landing of rockets
- The effect of each relationship of rocket variables to outcome
- Which conditions will enable spaceX in a successful landing

Section 1

Methodology

Methodology

Executive Summary

- Data collection methodology:
  - Via SpaceX Rest API
  - Web Scraping from Wikipedia
- Perform data wrangling
  - By dropping irrelevant columns and perform one hot encoding
- Perform exploratory data analysis (EDA)
  - Using Bar graph and Scatter plot to show relationship between variables
- Perform predictive analysis using classification models
  - Used several classification model, tuned each model using gridsearchcv for best parameters and evaluated model accuracy score and confusion matrix

Data Collection

For the purpose of finding the cost of launches, we needed amount of rocket launches data of at least of 5 years. Only falcon 9 Rockets was considered for the purpose of study and related fields like date of launch, location of launch, payload mass, booster version, orbit of deployment, etc. were taken into consideration. Data Collection was done through REST API from SpaceX API website website and web scrapping was carried out from wikipeadia

Data Collection – SpaceX API

```
spacex_url="https://api.spacexdata.com/v4/launches/past"
response = requests.get(spacex_url).json()
```

```
getLaunchSite(data)
getPayloadData(data)
getCoreData(data)
```


```
response = requests.get(static_json_url).json()
data = pd.json_normalize(response)
```

```
data_falcon9.to_csv('dataset_part_1.csv', index=False)
```

Data Collection - Scraping

```
static url = "https://en.wikipedia.org/w/index.php?title=List of Falcon 9 and Falcon Heavy launches&oldid=1027686922"
```

Getting response from HTML

```
response= requests.get(static_url)
```

Creating Beautifulsoup object

```
soup= BeautifulSoup(response.content, 'html.parser')
```

Extracting all columns

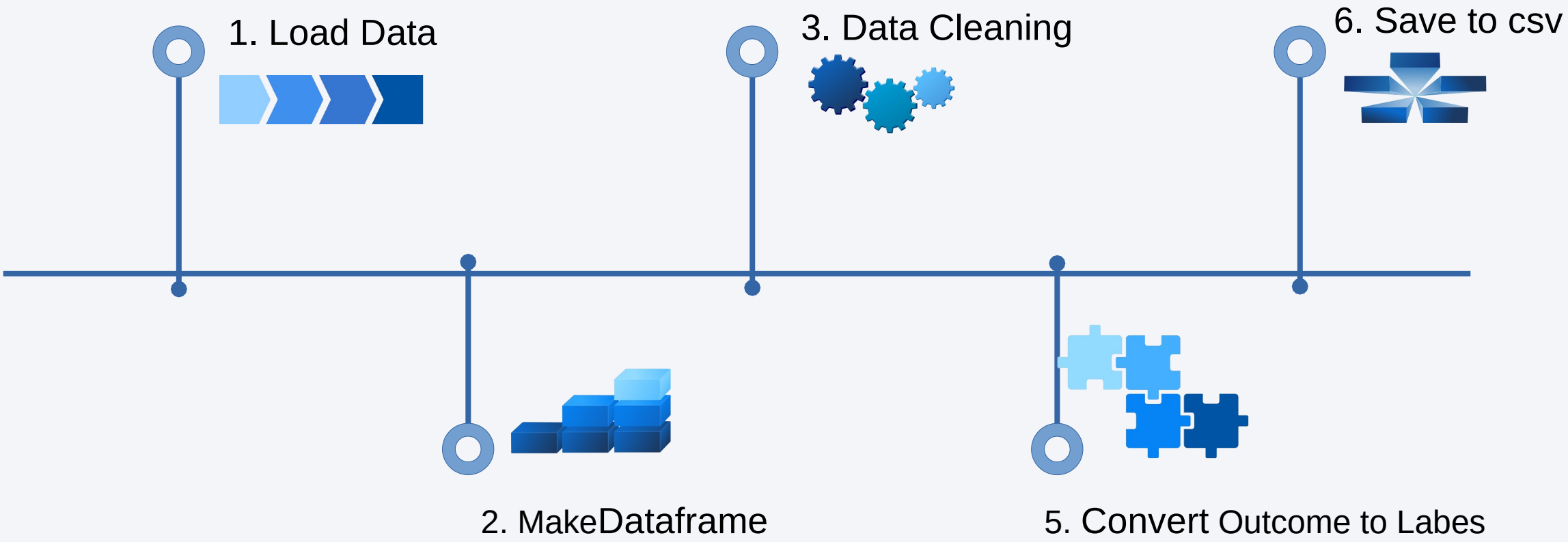
```
html_tables=soup.find_all('table')
first_launch_table.find_all('th')
for row in first_launch_table.find_all('th'):
    name =extract_column_from_header(row)
    if name != None and len(name) > 0:
        column_names.append(name)
```

Creating pandas dataframe

```
df=pd.DataFrame(launch_dict)
```

Data Wrangling

The majoy cleaning here was converting the outcome to labels where 1 and 0 mean sucessfull and unsucssfull landing



[Notebook]

EDA with Data Visualization

Visualisation was done Seaborn Scatter plot, Bar plot Line plot and Plotly Scatter map;

Scatter plot was used to show the relationship between fight number and launch sight, payload and launch sight, flight number and orbit type, and pay load and orbit type

Bar chart was used to show the comparisons between mean success rate for each orbit. Oribit is a categorical data hence the choice of Bar chart.

Line plot was used to show the the trend in successfull landing from 2010 to 2022

Predictive Analysis (Classification)

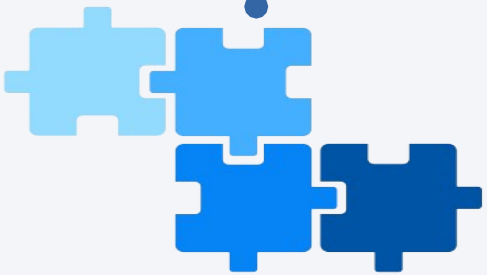
1. Load and standardize data

2. Split data into train and test data

3. Fit model and tune hyperparameter

4. Evaluate Model

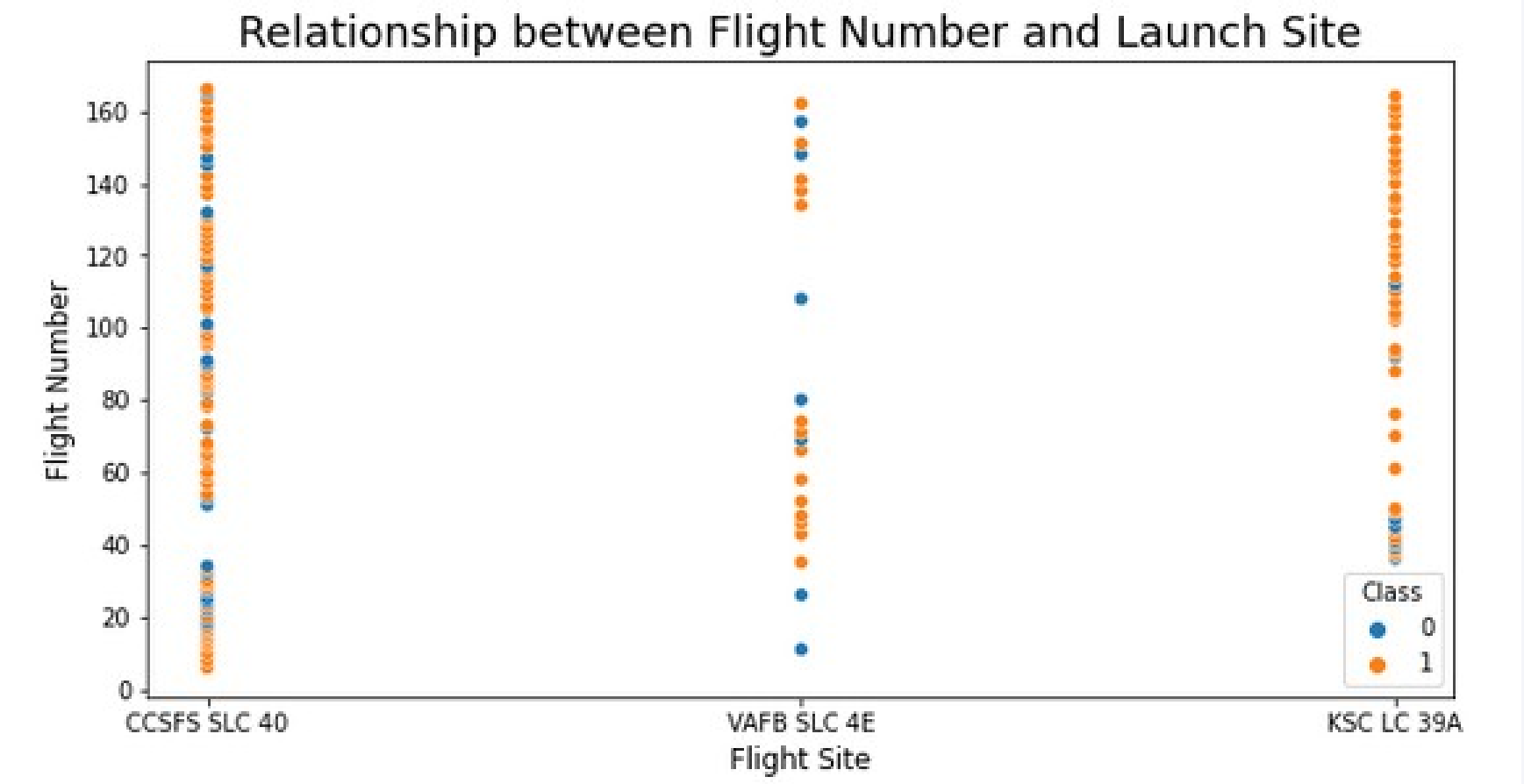
5. Choose the best model



Section 2

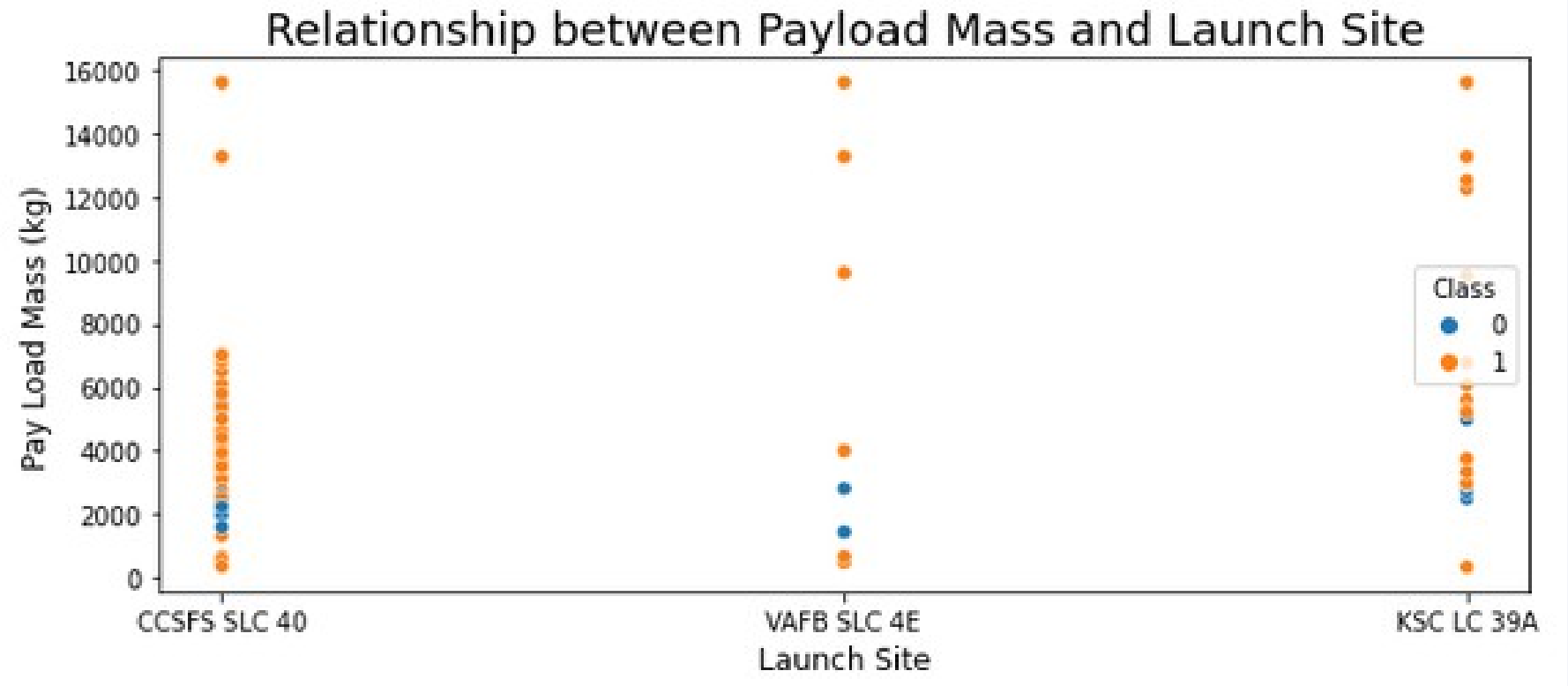
Insights drawn from EDA

Flight Number vs. Launch Site



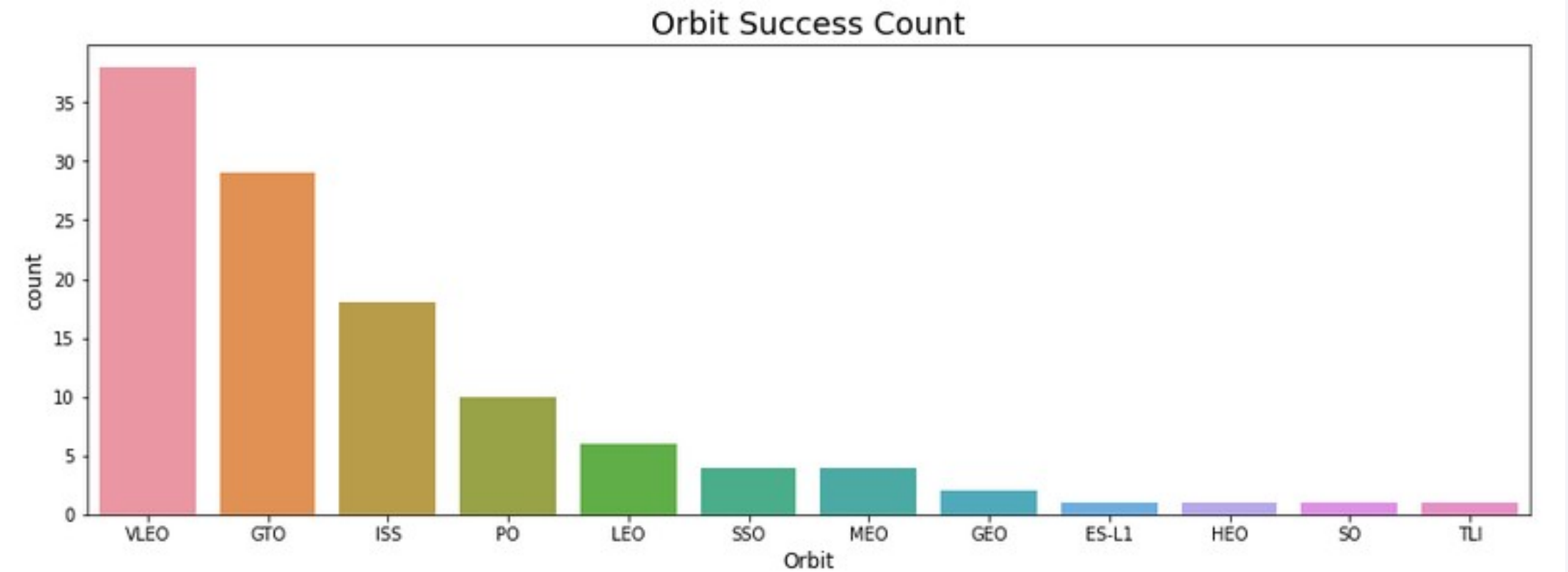
KSC seems to have the hihgest success rate . The more amount of flights at a launch site the greater the success rate at a launch site.

Payload vs. Launch Site



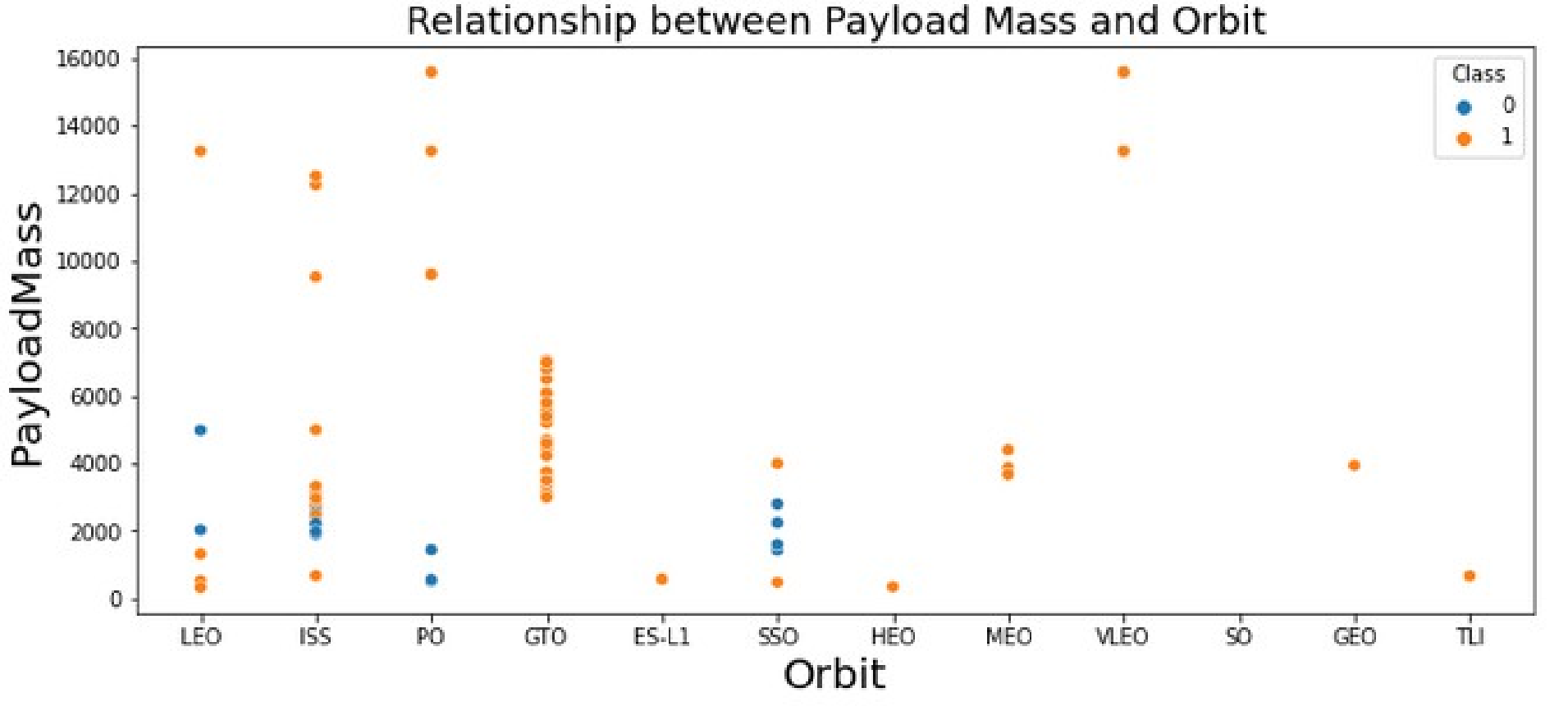
There is not quite a clear pattern to be found using this visualization to make a decision if the Launch Site is dependant on Pay Load Mass for a success launch. payloads less than 8000 tends to be lauched from ccsfs slc 40

Success Rate vs. Orbit Type



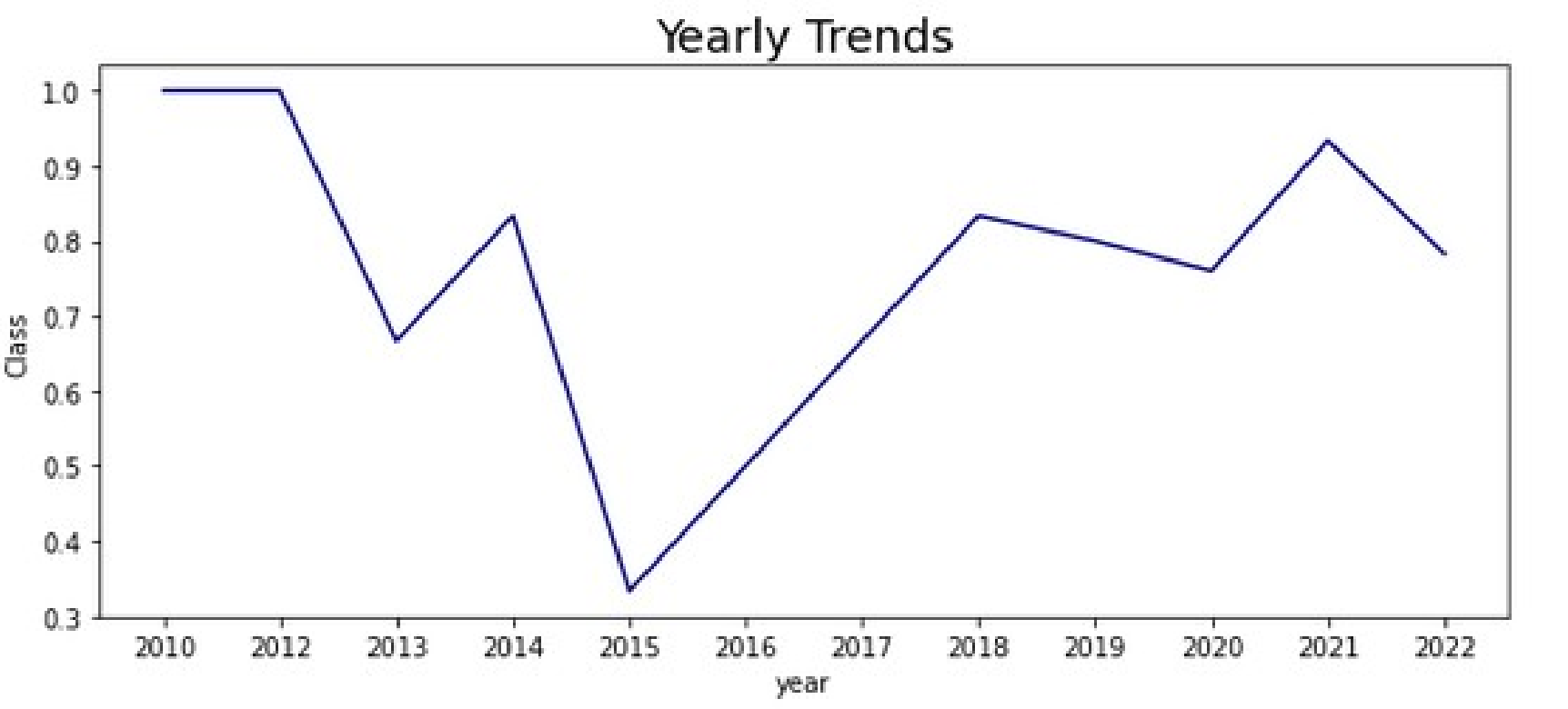
`VLEO` > `ISS` > `GTO` > `LEO` > `PO` > `SSO` > `MEO` > `GEO` > `ES-L1` > `HEO` > `TLI` success rate

Payload vs. Orbit Type



There seems to be no relationship between flight number when in GTO orbit,
All GTO launches were successful.

Launch Success Yearly Trend



From 2015 they has been steady increase in the mean successfull lauches although they is a deep in 2017 but picked backup in 2020


Section 3
Launch Sites
Proximities Analysis

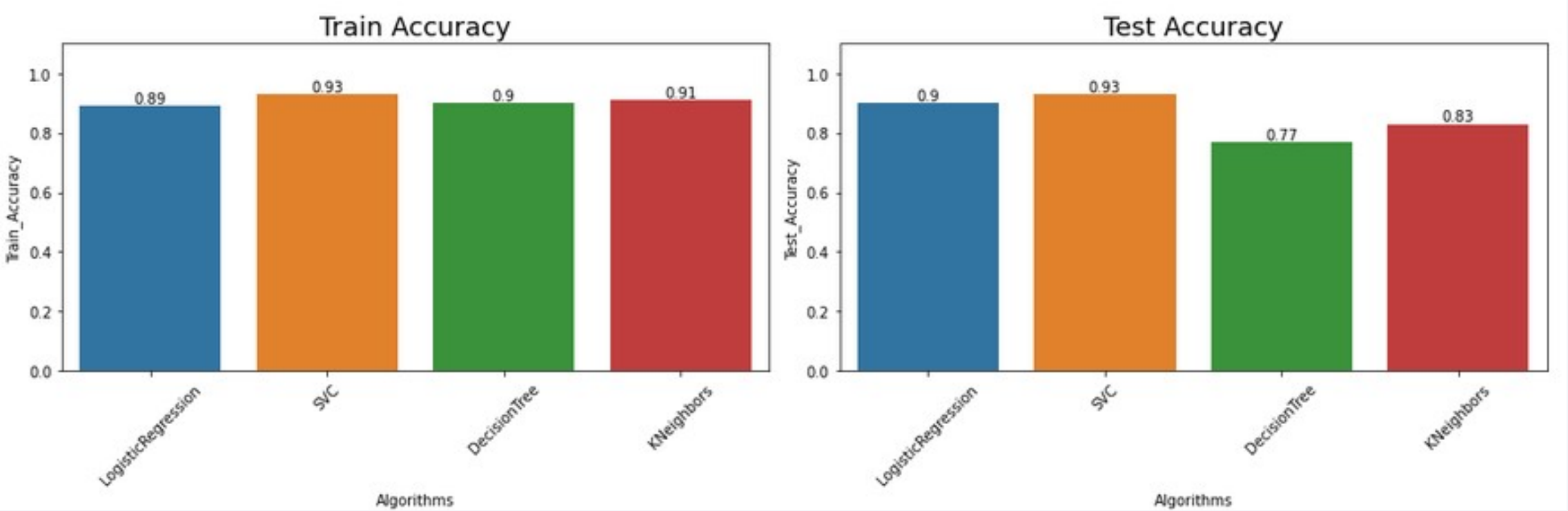
Launch Sites

Launch sites are in very close proximity to the coastline

Section 5

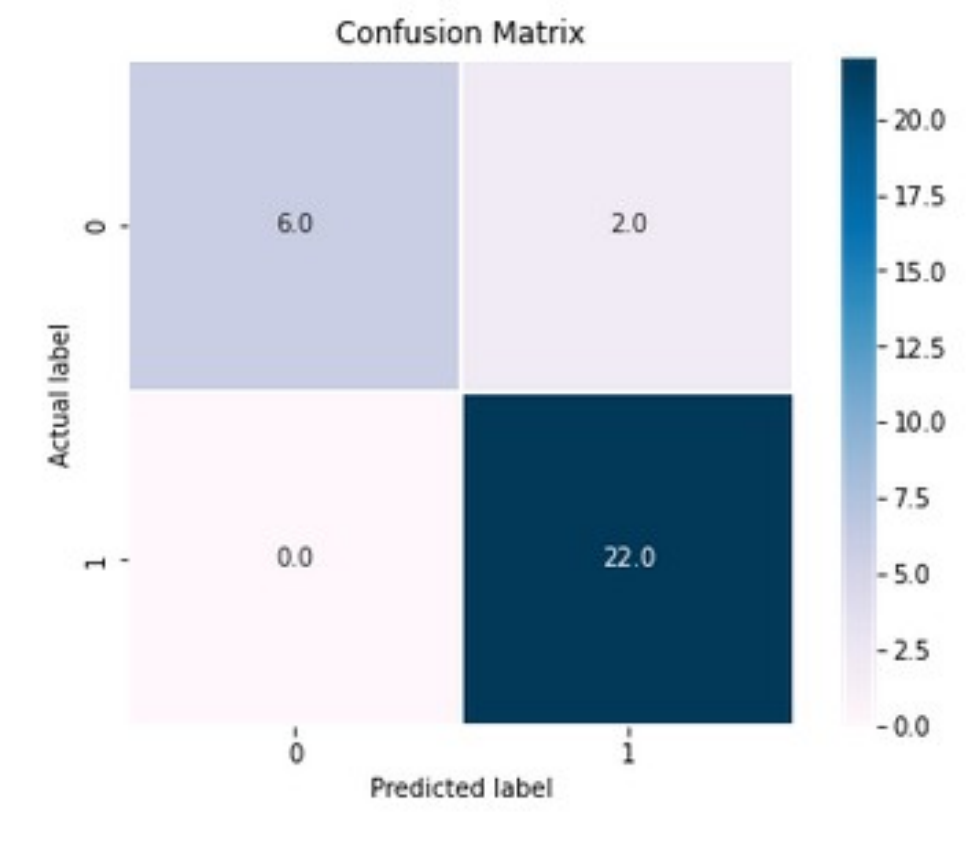
Predictive Analysis (Classification)

Classification Accuracy



From the above Bar chart Support Vector has the best accuracy on the test data. also from it confusion matrix it predicted successful landings 100%correctly and it correctly predictedn unsuccessful landing is 75%.

Confusion Matrix



Two false positive
Zero false negative

Conclusions

1. Falcon 9 rockets launched from KSC are more successful than other launch sites and payload mass between 2k-4k will have higher chances of successful mission
2. Landing success depends on many features like the physical working attributes of a rocket, orbit of payload
3. The success rate for SpaceX is directly proportional to time in years. With more time, they will eventually perfect the launches
4. The best classifier is Support Vector Machine with an accuracy of 88.87%

...

Appendix

Wikipeadia
Sklearn
Plotly

Thank you!