



IBM Developer
SKILLS NETWORK

Winning Space Race with Data Science

Vivek Kumar Vimal
10 Sept 2023



Outline

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

Executive Summary

- Summary of methodologies
 - Data collection
 - Data wrangling
 - EDA with data visualization
 - EDA with SQL
 - Building an interactive map with Folium
 - Building a Dashboard with Plotly Dash
 - Predictive analysis (Classification)
- Summary of all results
 - EDA results
 - Interactive analytics
 - Predictive analysis

Introduction

- Project background and context
 - 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.
- Problems you want to find answers
 - The project task is to predicting if the first stage of the SpaceX Falcon 9 rocket will land successfully



Section 1

Methodology

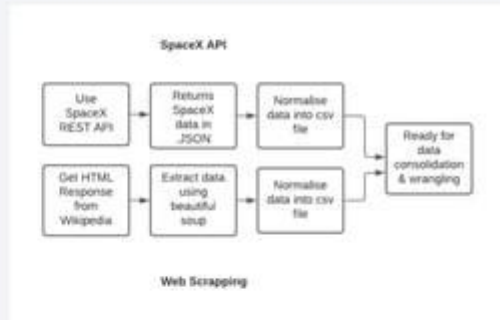
Methodology

Executive Summary

- Data collection methodology:
 - SpaceX Rest API
 - Web Scrapping from Wikipedia
- Perform data wrangling
 - One Hot Encoding data fields for Machine Learning and data cleaning of null values and irrelevant columns
- Perform exploratory data analysis (EDA) using visualization and SQL
- Perform interactive visual analytics using Folium and Plotly Dash
- Perform predictive analysis using classification models
 - LR, KNN, SVM, DT models have been built and evaluated for the best classifier

Data Collection

- The following datasets was collected:
 - SpaceX launch data that is gathered from the SpaceX REST API.
 - This API will give us data about launches, including information about the rocket used, payload delivered, launch specifications, landing specifications, and landing outcome.
 - The SpaceX REST API endpoints, or URL, starts with `api.spacexdata.com/v4/`.
 - Another popular data source for obtaining Falcon 9 Launch data is web scraping Wikipedia using BeautifulSoup.



Data Collection - Scraping

- Web Scrapping from Wikipedia

1. Getting Response from HTML

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

2. Creating BeautifulSoup Object

```
soup = BeautifulSoup(page.text, 'html.parser')
```

3. Finding tables

```
html_tables = soup.find_all('table')
```

4. Getting column names

```
column_names = []  
temp = soup.find_all('th')  
for i in range(len(temp)):  
    key:  
    name = extract_column_from_header(temp[i])  
    if (name is not None and len(name) > 0):  
        column_names.append(name)  
except:  
    pass
```

5. Creation of dictionary

```
launch_dict = dict.fromkeys(column_names)  
  
# Remove an irrelevant column  
del launch_dict['Date and time (UTC)']  
  
launch_dict['Flight No.'] = []  
launch_dict['Launch site'] = []  
launch_dict['Payload'] = []  
launch_dict['Payload mass'] = []  
launch_dict['Orbit'] = []  
launch_dict['Customer'] = []  
launch_dict['Launch outcome'] = []  
launch_dict['Version Number'] = []  
launch_dict['Booster landing'] = []  
launch_dict['Data'] = []  
launch_dict['Time'] = []
```

6. Appending data to keys (refer) to notebook block 12

```
for i in range(len(html_tables)):  
    extracted_row = 0  
    # Extract each table  
    for table in html_tables:  
        # Get only row  
        for row in table.find_all('tr'):  
            # Extract the row of first table
```

7. Converting dictionary to dataframe

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

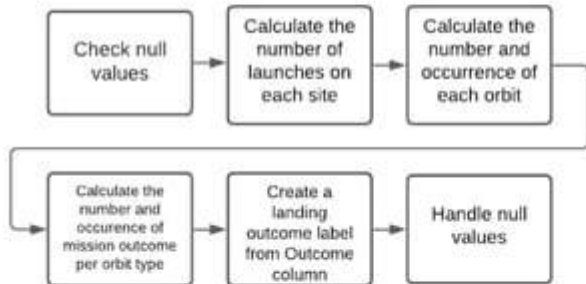
8. Dataframe to .CSV

```
df.to_csv('spacex_web_scraped.csv', index=False)
```

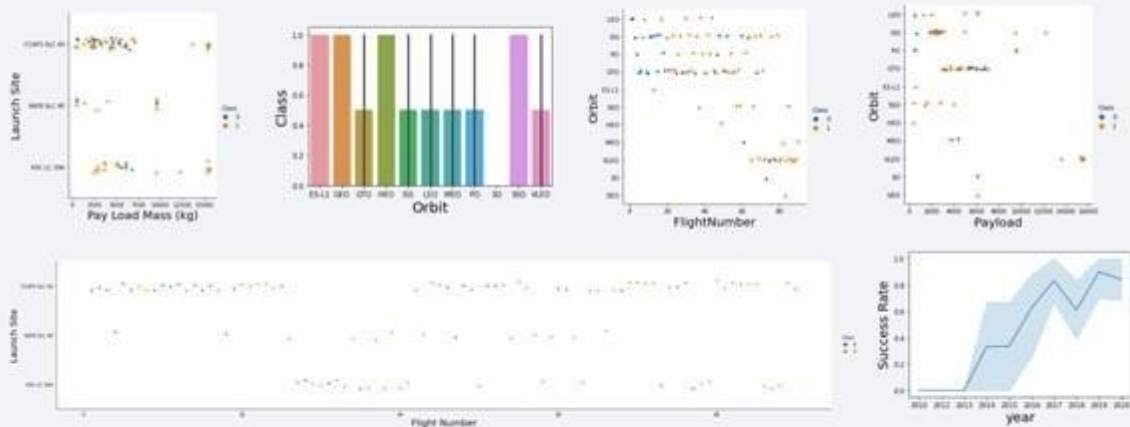
<https://github.com/initiative1972/data-science/blob/master/10-IBM%20DS%20Capstone%20project-lab2-web%20scrapping.ipynb>

Data Wrangling

EDA analysis



EDA with Data Visualisation



EDA with SQL

- SQL queries performed include:
 - Displaying the names of the unique launch sites in the space mission
 - Displaying 5 records where launch sites begin with the string 'KSC'
 - Displaying the total payload mass carried by boosters launched by NASA (CRS)
 - Displaying average payload mass carried by booster version F9 v1.1
 - Listing the date where the successful landing outcome in drone ship was achieved.
 - Listing the names of the boosters which have success in ground pad and have payload mass greater than 4000 but less than 6000
 - Listing the total number of successful and failure mission outcomes
 - Listing the names of the booster_versions which have carried the maximum payload mass.
 - Listing the records which will display the month names, successful landing_outcomes in ground pad ,booster versions, launch_site for the months in year 2017
 - Ranking the count of successful landing_outcomes between the date 2010 06 04 and 2017 03 20 in descending order.

Build an Interactive Map with Folium

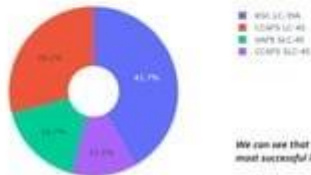


Map markers have been added to the map with aim to finding an optimal location for building a launch site

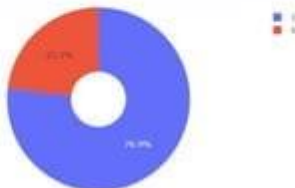
<https://github.com/initiative1972/data-science/blob/master/10-IBM%20DS%20Capstone-lab5-Folium.ipynb>

Build a Dashboard with Plotly Dash

Total Success Launches By all sites

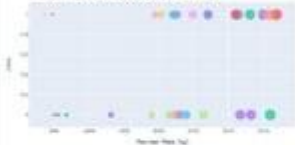


We can see that KSC LC-39A had the most successful launches from all the sites

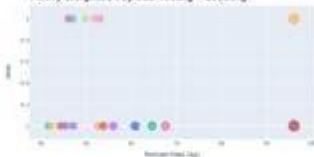


KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Low Weighted Payload 0kg - 4000kg



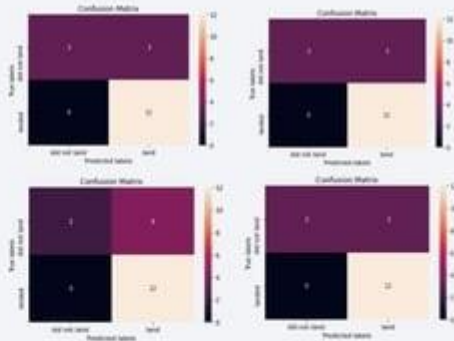
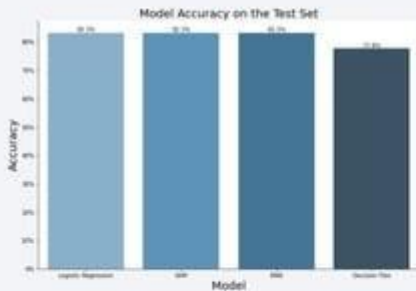
Heavy Weighted Payload 4000kg - 20000kg



We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

Predictive Analysis (Classification)

- The SVM, KNN, and Logistic Regression model achieved the highest accuracy at 83.3%, while the SVM performs the best in terms of Area Under the Curve at 0.958.



Results

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

The background of the slide is an abstract composition of numerous diagonal streaks and brushstrokes in vibrant blue and red colors, set against a solid dark blue background. The streaks vary in thickness and intensity, creating a sense of dynamic movement and energy.

Section 2

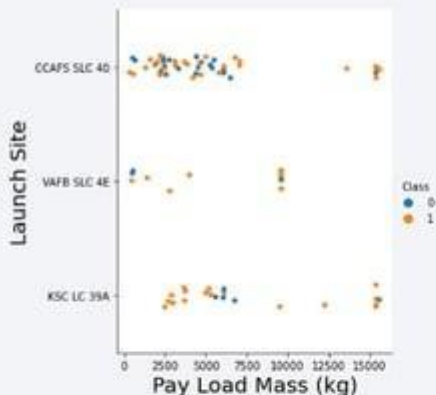
Insights drawn from EDA

Flight Number vs. Launch Site



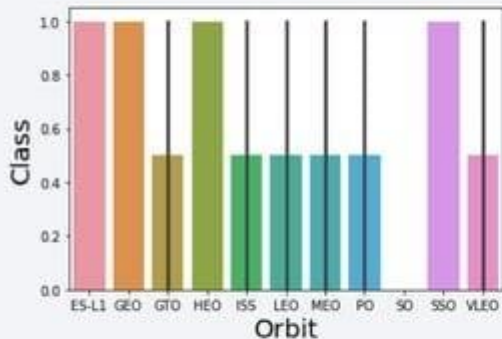
- Launches from the site of CCAFS SLC 40 are significantly higher than launches from other sites.

Payload vs. Launch Site



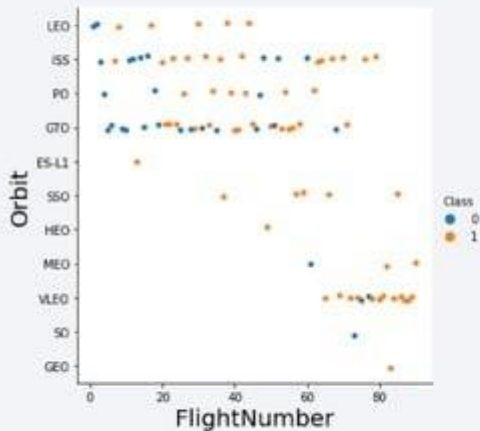
- The majority of IPay Loads with lower Mass have been launched from CCAFS SLC 40.

Success Rate vs. Orbit Type



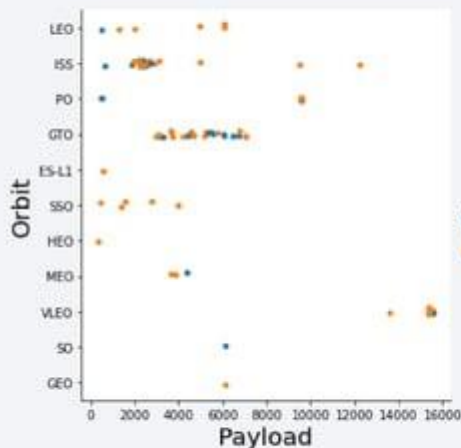
- The orbit types of ES-L1, GEO, HEO, SSO are among the highest success rate.

Flight Number vs. Orbit Type



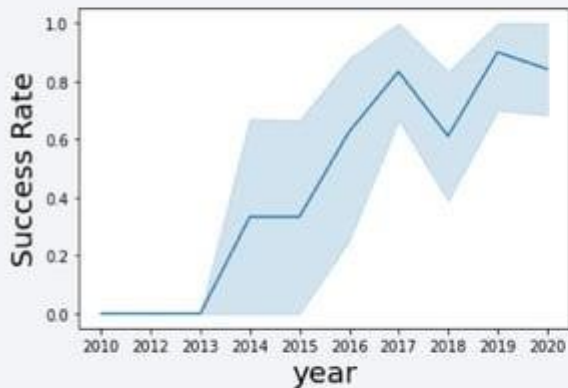
- A trend can be observed of shifting to VLEO launches in recent years.

Payload vs. Orbit Type



- There are strong correlation between ISS and Payload at the range around 2000, as well as between GTO and the range of 4000-8000.

Launch Success Yearly Trend



- Launch success rate has increased significantly since 2013 and has stabilised since 2019, potentially due to advance in technology and lessons learned.

All Launch Site Names

- %sql select distinct(LAUNCH_SITE) from SPACEXTBL

launch_site

CCAFS LC-40

CCAFS SLC-40

KSC LC-39A

VAFB SLC-4E

Launch Site Names Begin with 'CCA'

- %sql select * from SPACEXTBL where LAUNCH_SITE like 'CCA%' limit 5

DATE	time_utc	booster_version	launch_site	payload	payload_mass_kg	orbit	customer	mission_outcome	landing_outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	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
2012-10-08	00:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

Total Payload Mass

- %sql select sum(PAYLOAD_MASS__KG_) from SPACEXTBL where CUSTOMER = 'NASA (CRS)'

45596

Average Payload Mass by F9 v1.1

- %sql select avg(PAYLOAD_MASS__KG_) from SPACEXTBL where BOOSTER_VERSION = 'F9 v1.1'

2928.400000

First Successful Ground Landing Date

- %sql select min(DATE) from SPACEXTBL where Landing__Outcome = 'Success (ground pad)'

2015-12-22

Successful Drone Ship Landing with Payload between 4000 and 6000

- %sql select BOOSTER_VERSION from SPACEXTBL where Landing__Outcome = 'Success (drone ship)' and PAYLOAD_MASS__KG_ > 4000 and PAYLOAD_MASS__KG_ < 6000

booster_version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

Total Number of Successful and Failure Mission Outcomes

- %sql select count(MISSION_OUTCOME) from SPACEXTBL where MISSION_OUTCOME = 'Success' or MISSION_OUTCOME = 'Failure (in flight)'

100

Boosters Carried Maximum Payload

- %sql select BOOSTER_VERSION from SPACEXTBL where PAYLOAD_MASS_KG_ = (select max(PAYLOAD_MASS_KG_) from SPACEXTBL)

booster_version

F9 B5 B1048.4

F9 B5 B1049.4

F9 B5 B1051.3

F9 B5 B1056.4

F9 B5 B1048.5

F9 B5 B1051.4

F9 B5 B1049.5

F9 B5 B1060.2

F9 B5 B1058.3

F9 B5 B1051.6

F9 B5 B1060.3

F9 B5 B1049.7

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

- %sql select * from SPACEXTBL where Landing__Outcome like 'Success%' and (DATE between '2010-06-04' and '2017-03-20') order by date desc

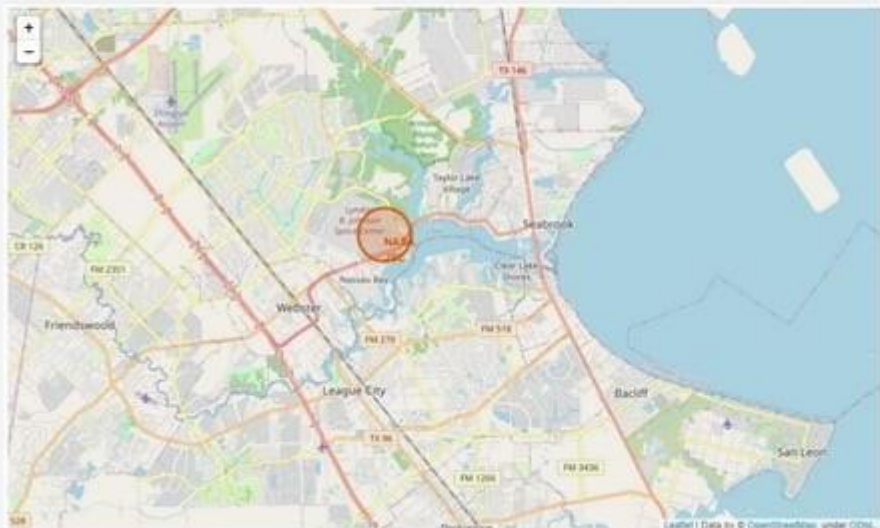
2016-05-27	21:39:00	F9 FT B1023.1	CCAFS LC-40	Thalcom 8	3100	GTO	Thalcom	Success	Success (drone ship)
2016-05-06	05:21:00	F9 FT B1022	CCAFS LC-40	JCSAT-14	4696	GTO	SKY Perfect iSAT Group	Success	Success (drone ship)
2016-04-08	20:43:00	F9 FT B1021.1	CCAFS LC-40	SpaceX CRS-8	3138	LEO (ISS)	NASA (CRS)	Success	Success (drone ship)
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)

The background of the slide is a high-quality photograph of Earth taken from space. The top half of the image shows the dark, starry void of space. Below it, the curved horizon of the Earth is visible, with a thin layer of white clouds. The lower half of the image shows the Earth's surface at night, with numerous bright yellow and orange lights from cities and towns scattered across the dark landmasses. The overall color palette is dominated by deep blues and blacks, with the warm glow of the city lights providing a stark contrast.

Section 4

Launch Sites Proximities Analysis

All launch sites marked on a map



Success/failed launches marked on the map



Distances between a launch site to its proximities



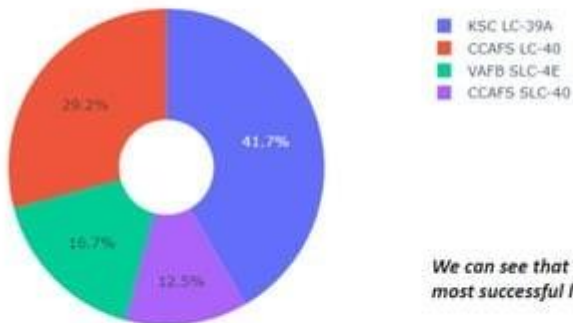


Section 5

Build a Dashboard with Plotly Dash

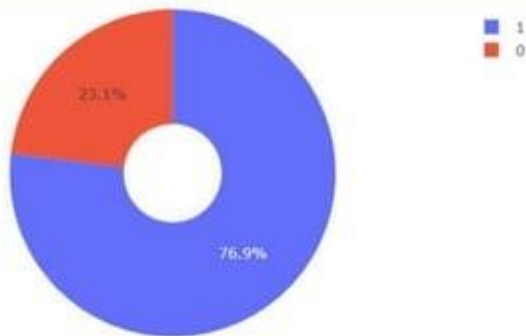
Total success launches by all sites

Total Success Launches By all sites



We can see that KSC LC-39A had the most successful launches from all the sites

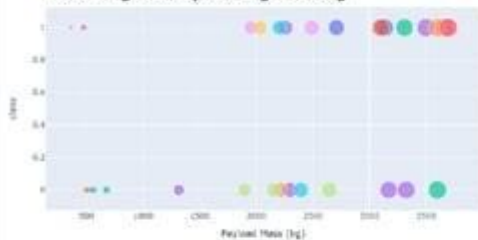
Success rate by site



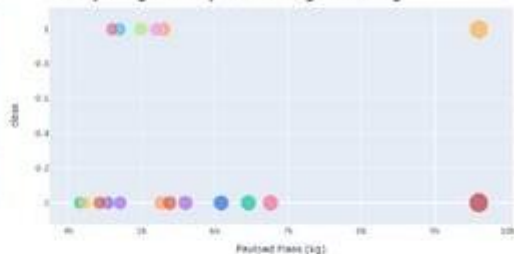
KSC LC-39A achieved a 76.9% success rate while getting a 23.1% failure rate

Payload vs launch outcome

Low Weighted Payload 0kg – 4000kg



Heavy Weighted Payload 4000kg – 10000kg



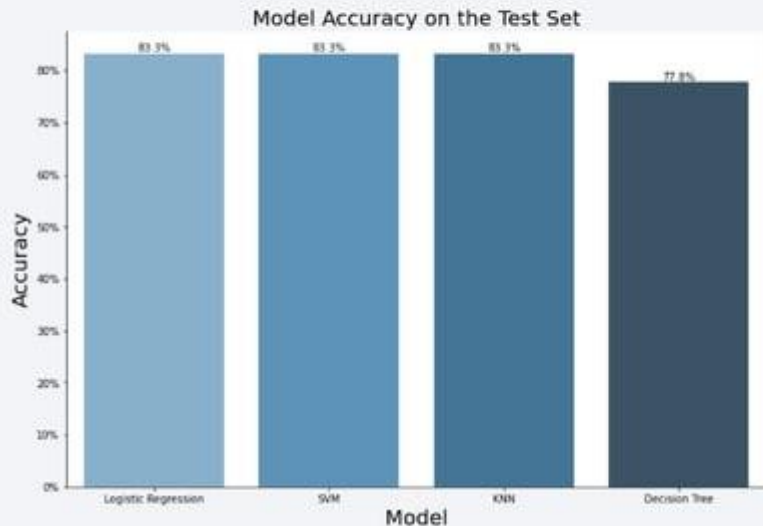
We can see the success rates for low weighted payloads is higher than the heavy weighted payloads

The background of the slide is an abstract composition of flowing, curved lines. On the left, there is a solid blue area. This transitions into a series of concentric, curved bands of varying shades of blue and white that sweep across the frame from the bottom left towards the top right. On the far right, a portion of a white, curved structure is visible, resembling the interior of a modern tunnel or a futuristic architectural element, with some faint lights visible in the distance.

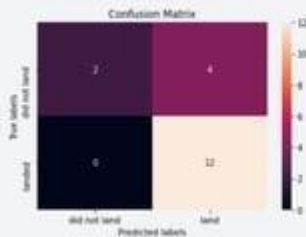
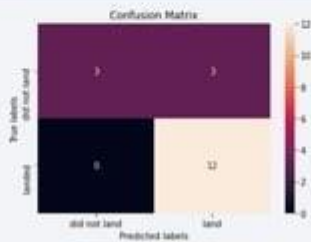
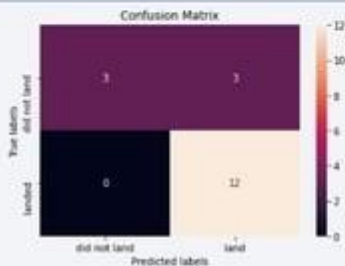
Section 6

Predictive Analysis (Classification)

Classification Accuracy



Confusion Matrix



Conclusions

- The SVM, KNN, and Logistic Regression models are the best in terms of prediction accuracy for this dataset.
- Low weighted payloads perform better than the heavier payloads.
- The success rates for SpaceX launches is directly proportional time in years they will eventually perfect the launches.
- KSC LC 39A had the most successful launches from all the sites.
- Orbit GEO,HEO,SSO,ES L1 has the best Success Rate.

Thank you!

