SpaceX Falcon 9 first stage Landing Prediction

Marios Smyrnaios

18.12.2023

OUTLINE

- Executive Summary
- Introduction
- Methodology – Data collection and wrangling
- Interactive maps with Folium
- Explorative data analysis & visual analytics
- Explorative data analysis & SQL results
- Predictive analysis results & Model Selection
- Conclusion
- Appendix

EXECUTIVE SUMMARY

- Mission success rate of Falcon 9 for launches after 2019 is above 95%
  - VAFB-SLC launchsite: no rockets launched with heavy payload
  - SSO, HEO, GEO and ES-L1 orbits lanches have 100 % success rate

- Launch success/failure prediction of Falcon 9 has an accuracy of above 87%

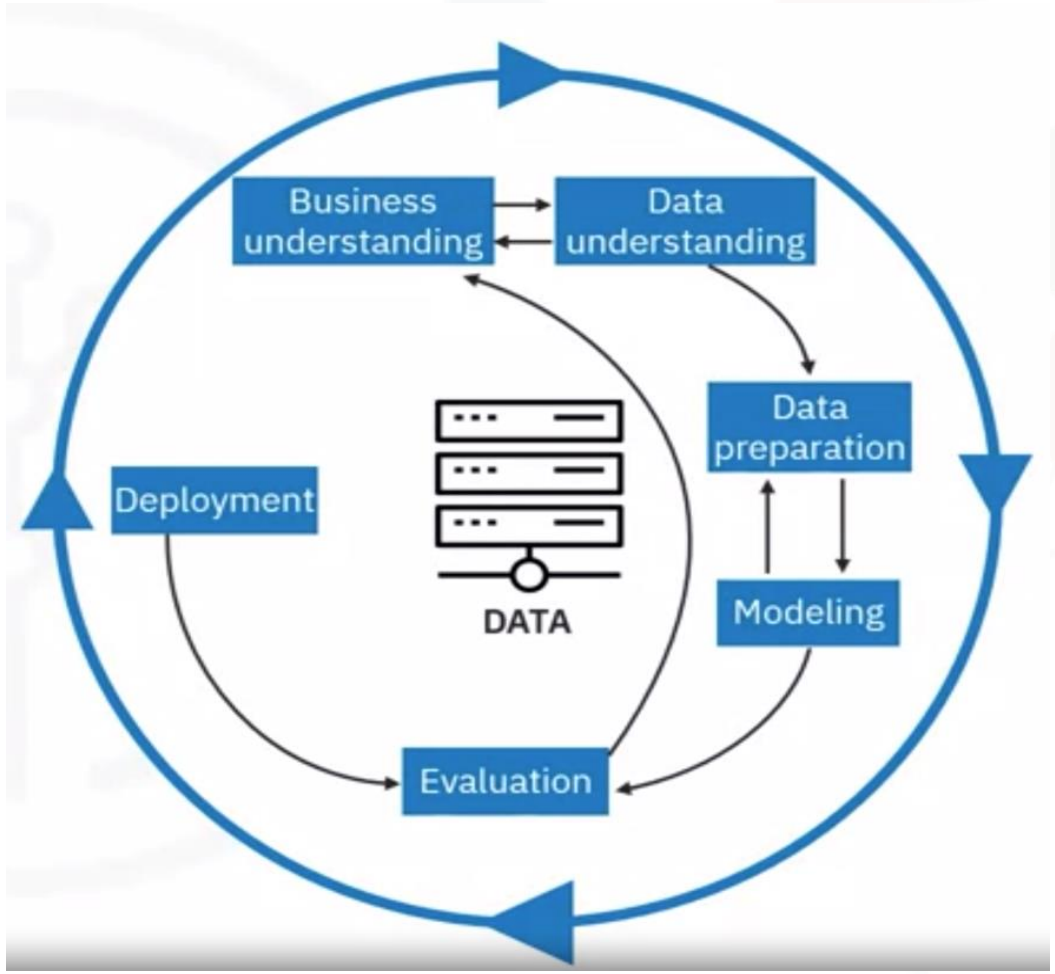
INTRODUCTION

- Falcon 9 rocket launches cost of 62 million dollars
- Other providers cost upward of 165 million dollars
- SpaceX can reuse the first stage & therefore reduce costs

Can we determine if the first stage will land (thus, also determine the cost of a launch)?

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

Data Science METHODOLOGY



- Data collection & data wrangling
- Explorative data analysis
- Modeling
- Model Evaluation
- Conclusion

Data collection & Data Wrangling (1)

- Request to the SpaceX API
- Keep only Falcon 9 Launches

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

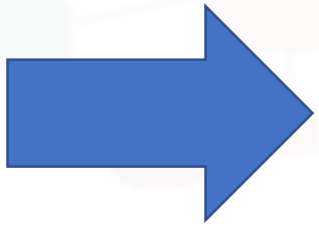
90 rows × 17 columns

Data collection & Data Wrangling (2)

- Replace “PayloadMass” missing values with the “mean” value

```
data_falcon9.isnull().sum()
```

```
FlightNumber        0
Date                0
BoosterVersion      0
PayloadMass         5
Orbit               0
LaunchSite          0
Outcome             0
Flights             0
GridFins            0
Reused              0
Legs                0
LandingPad         26
Block               0
ReusedCount         0
Serial              0
Longitude           0
Latitude            0
dtype: int64
```



```
data_falcon9.isnull().sum()
```

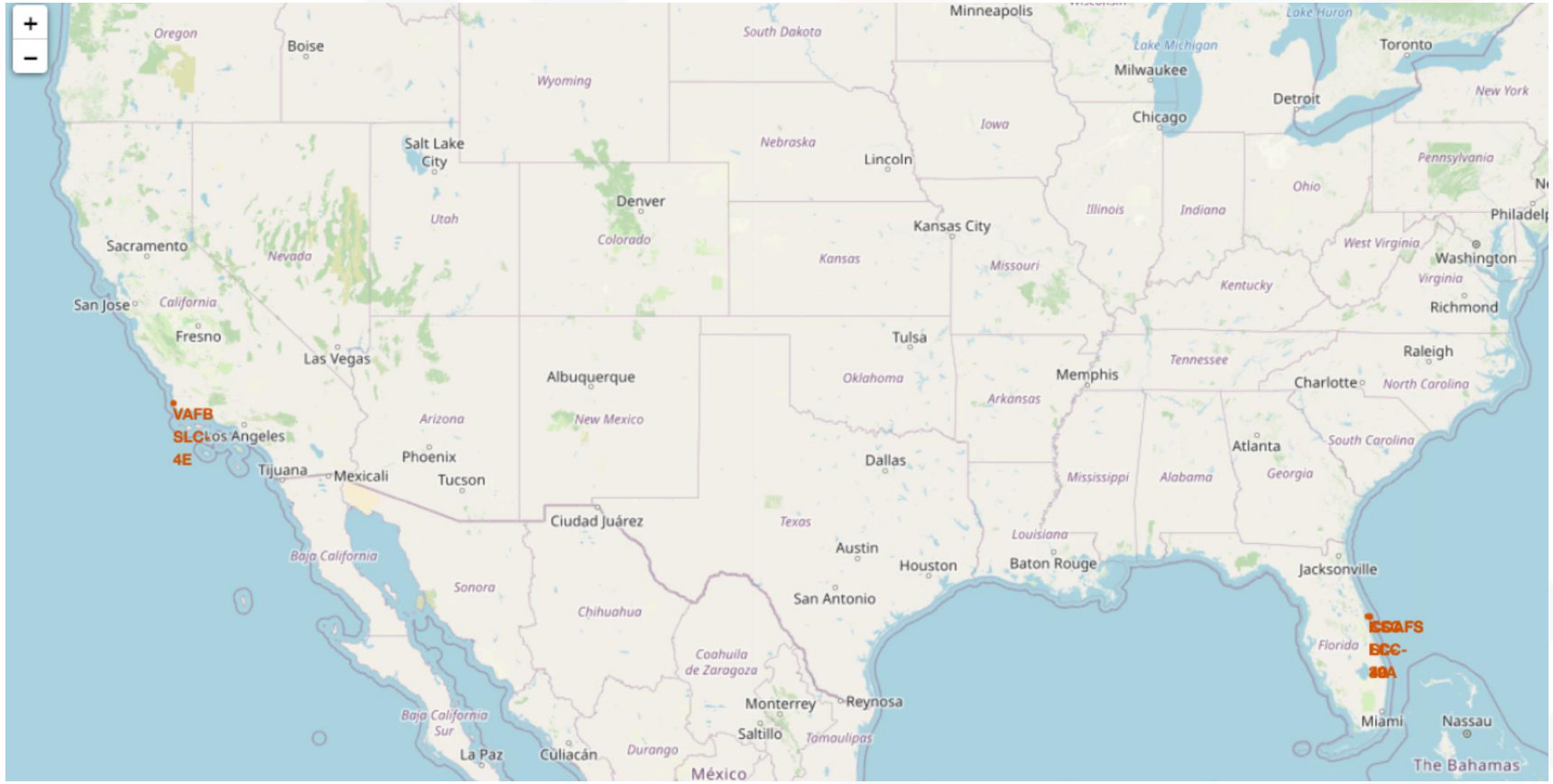
```
FlightNumber        0
Date                0
BoosterVersion      0
PayloadMass         0
Orbit               0
LaunchSite          0
Outcome             0
Flights             0
GridFins            0
Reused              0
Legs                0
LandingPad         26
Block               0
ReusedCount         0
Serial              0
Longitude           0
Latitude            0
dtype: int64
```

Falcon 9 Landing Outcomes

Landing outcomes for **2016** and **2017**

Date	Landing_Outcome
2017-03-16	No attempt
2017-02-19	Success (ground pad)
2017-01-14	Success (drone ship)
2016-08-14	Success (drone ship)
2016-07-18	Success (ground pad)
2016-06-15	Failure (drone ship)
2016-05-27	Success (drone ship)
2016-05-06	Success (drone ship)
2016-04-08	Success (drone ship)
2016-03-04	Failure (drone ship)
2016-01-17	Failure (drone ship)

Launch Site vs Launches per Site



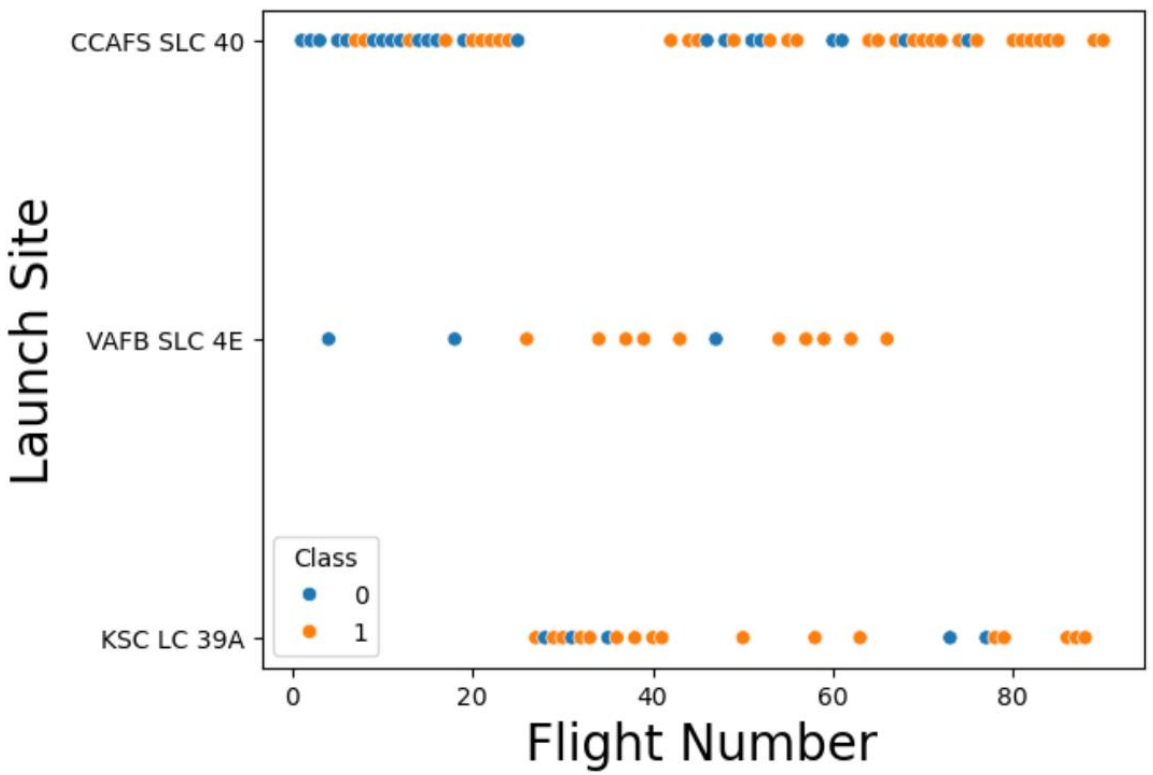
Launch Site	Lat	Long
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745

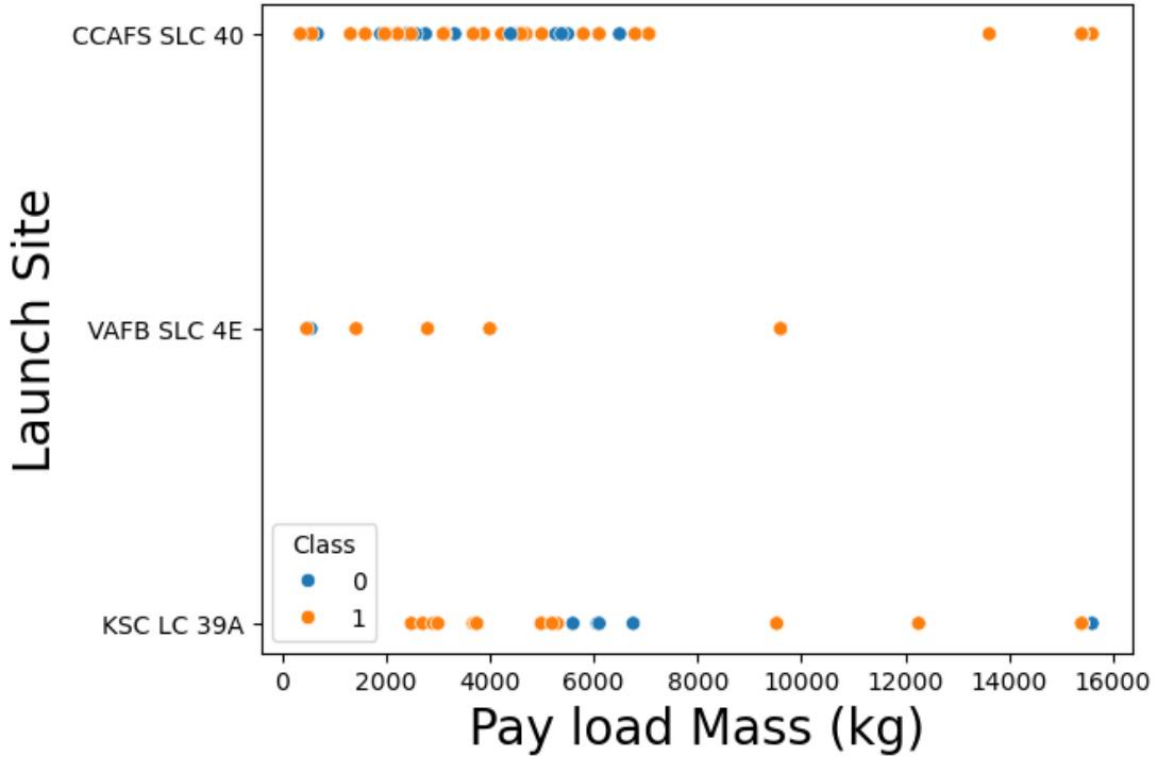
Launches per Site

CCAFS SLC 40	55
KSC LC 39A	22
VAFB SLC 4E	13

Launch Site vs Flight Number and Payload Mass

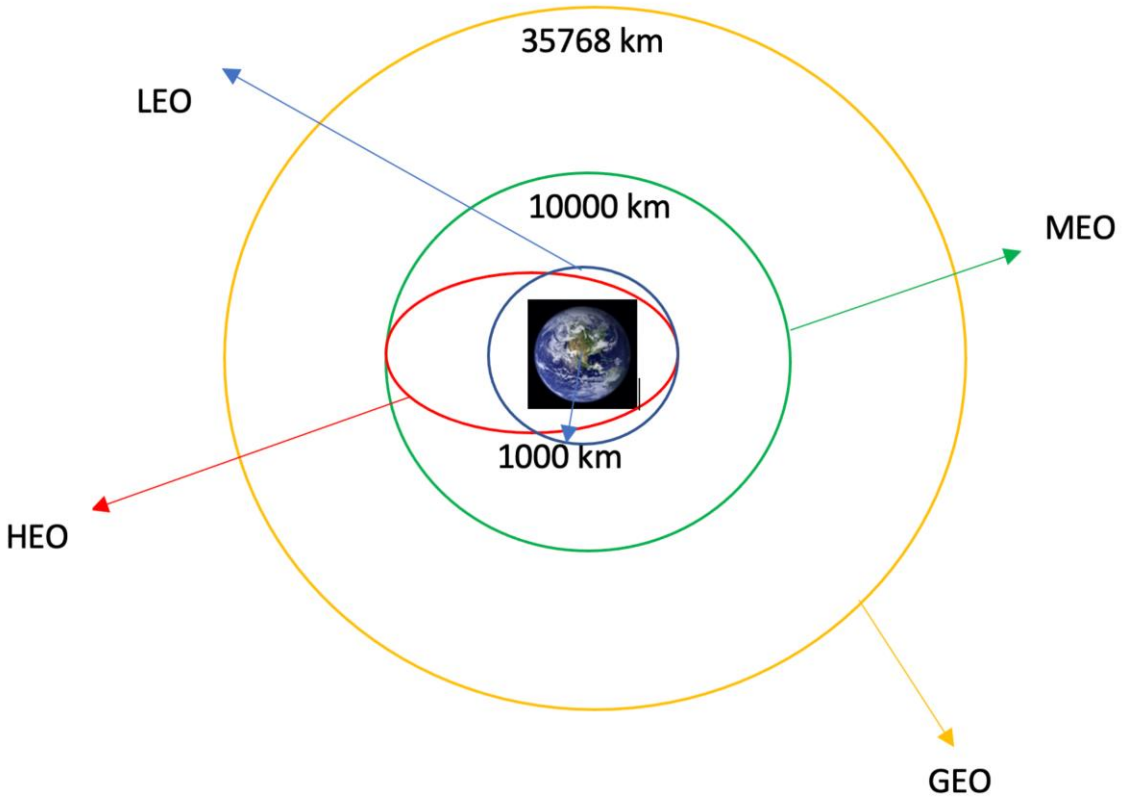
Class 0 = Failure – Class 1 = Success





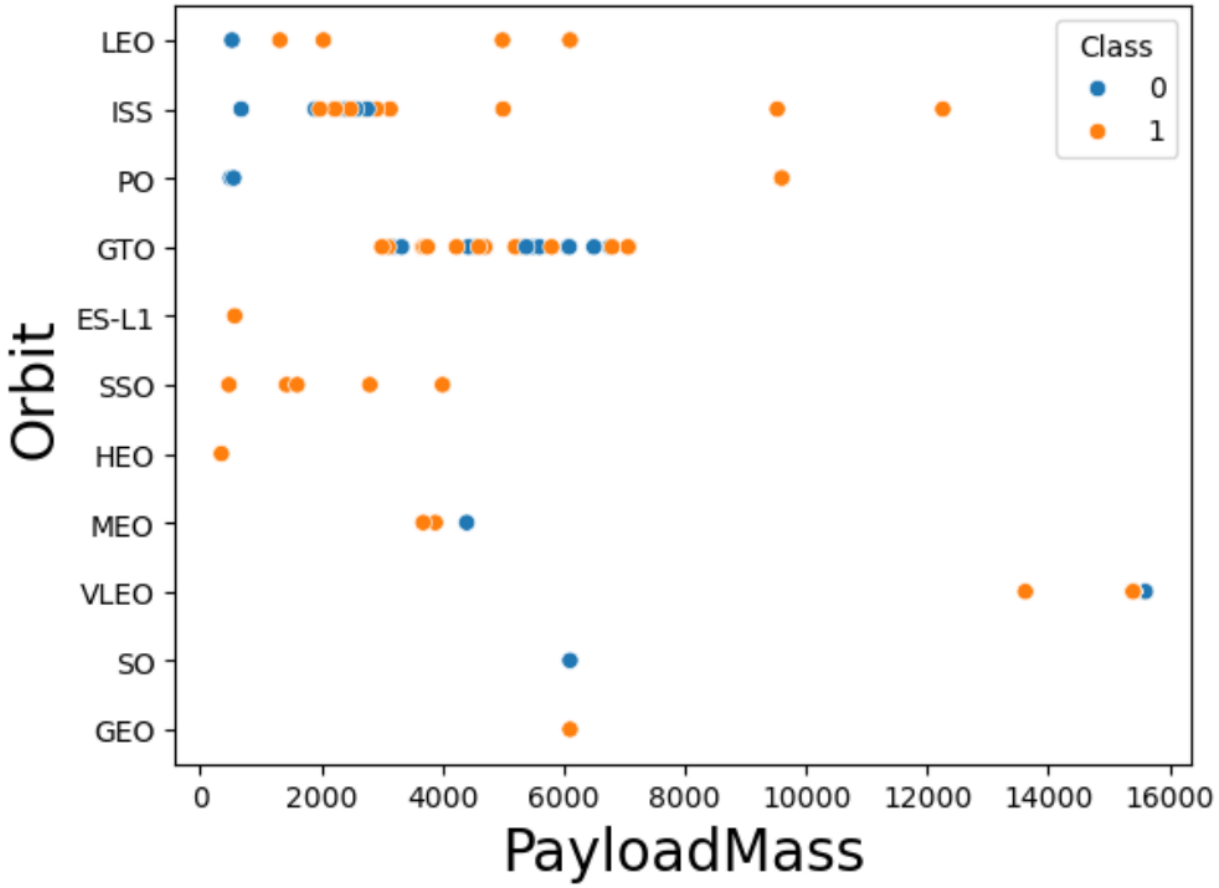
Number of Launches per Orbit

- Number of launches for each orbit type



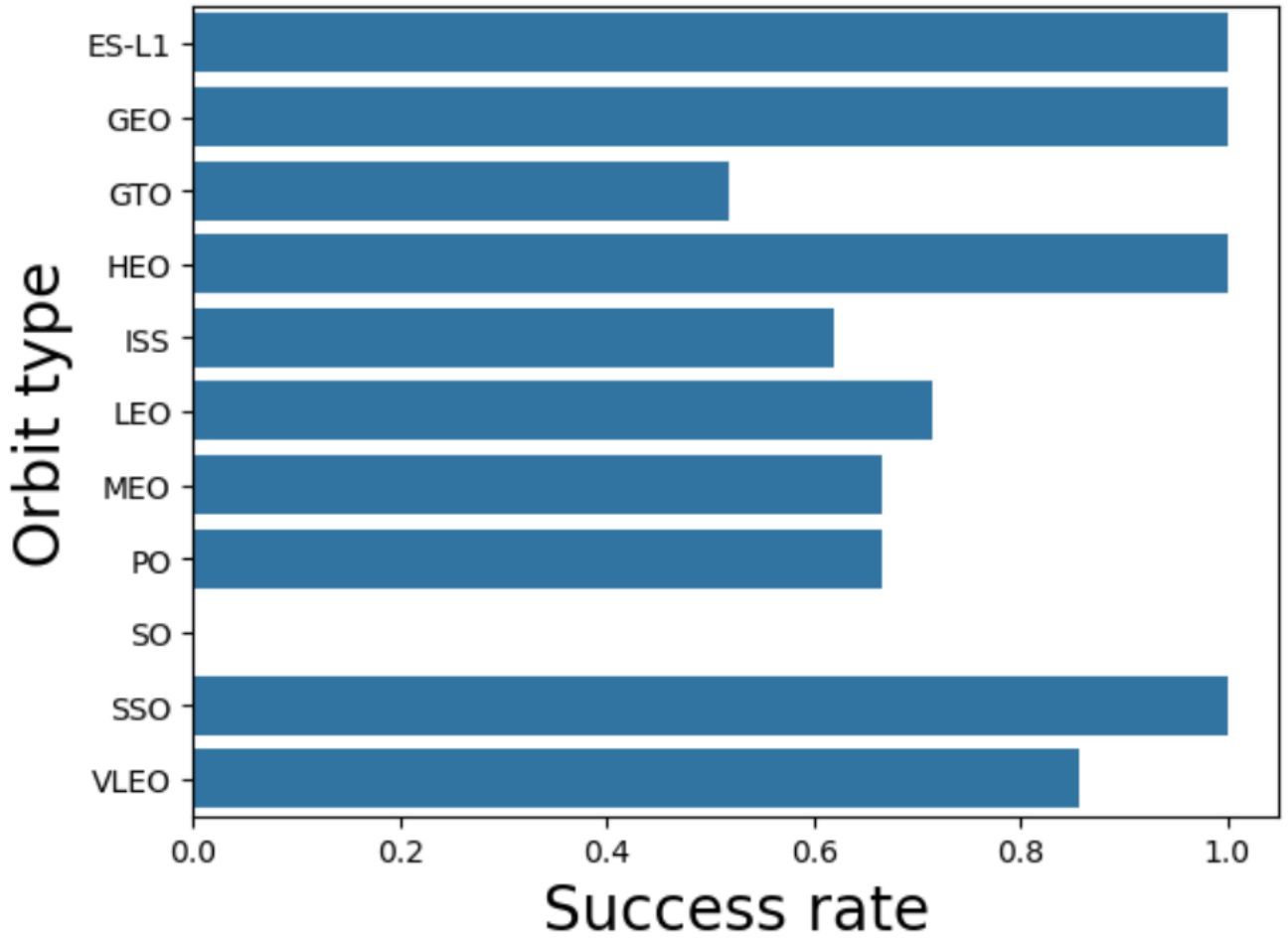
GTO	27
ISS	21
VLEO	14
PO	9
LEO	7
SSO	5
MEO	3
ES-L1	1
HEO	1
SO	1
GEO	1

Launches per Orbit vs Payload Mass

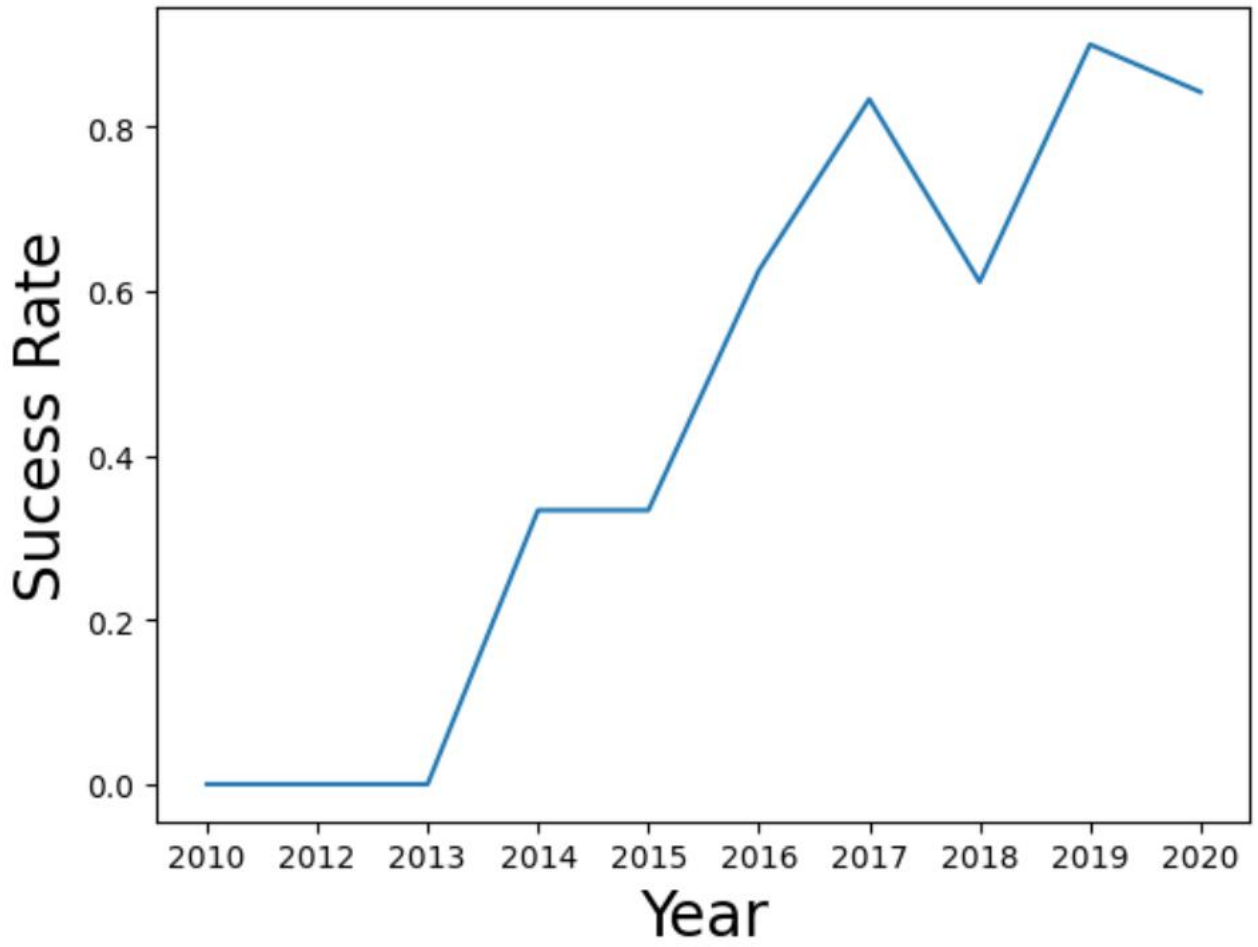


Class 0 = Failure – Class 1 = Success

Launch Success Rate per Orbit Type



Launch Success vs Years



Predictive Analysis - Goals

- Perform exploratory Data Analysis and determine Training Labels
- 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

Predictive Analysis – Feature Engineering

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude	Class
0	1	2010-06-04	Falcon 9	6104.959412	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0003	-80.577366	28.561857	0
1	2	2012-05-22	Falcon 9	525.000000	LEO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0005	-80.577366	28.561857	0
2	3	2013-03-01	Falcon 9	677.000000	ISS	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B0007	-80.577366	28.561857	0
3	4	2013-09-29	Falcon 9	500.000000	PO	VAFB SLC 4E	False Ocean	1	False	False	False	NaN	1.0	0	B1003	-120.610829	34.632093	0
4	5	2013-12-03	Falcon 9	3170.000000	GTO	CCAFS SLC 40	None None	1	False	False	False	NaN	1.0	0	B1004	-80.577366	28.561857	0

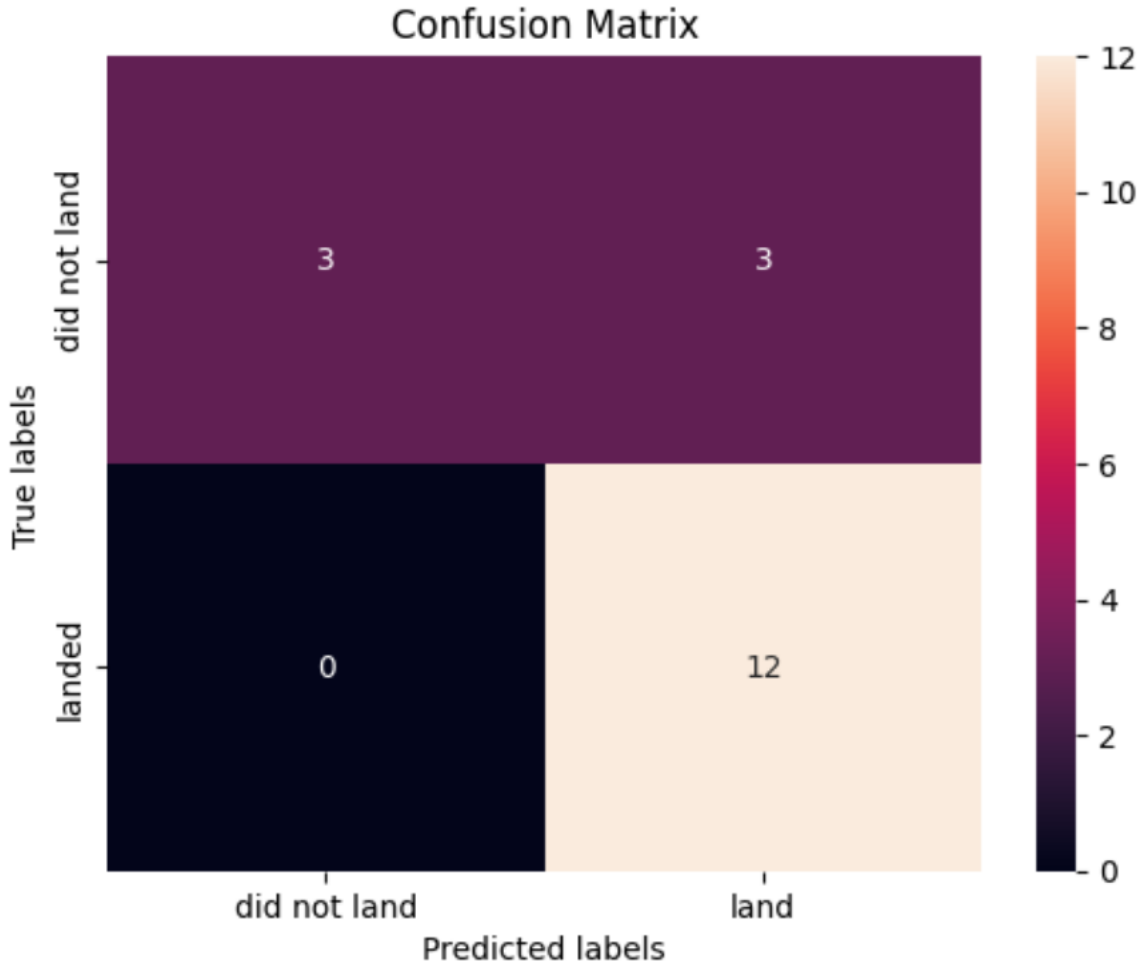
Feature Selection :

'FlightNumber', 'PayloadMass', 'Orbit', 'LaunchSite', 'Flights', 'GridFins', 'Reused', 'Legs', 'LandingPad', 'Block', 'ReusedCount', 'Serial'

	FlightNumber	PayloadMass	Flights	Block	ReusedCount	Orbit_ES-L1	Orbit_GEO	Orbit_GTO	Orbit_HEO	Orbit_ISS	...	Serial_B1058	Serial_B1059	Serial_B1060	Serial_B1062	GridFins_False	GridFins_True	Reused_False	Reused_True	Legs_False	Legs_True
0	1.0	6104.959412	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
1	2.0	525.000000	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
2	3.0	677.000000	1.0	1.0	0.0	0.0	0.0	0.0	0.0	1.0	...	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
3	4.0	500.000000	1.0	1.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
4	5.0	3170.000000	1.0	1.0	0.0	0.0	0.0	1.0	0.0	0.0	...	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
85	86.0	15400.000000	2.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0
86	87.0	15400.000000	3.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	...	1.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0
87	88.0	15400.000000	6.0	5.0	5.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0
88	89.0	15400.000000	3.0	5.0	2.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	1.0	0.0	0.0	1.0	0.0	1.0	0.0	1.0
89	90.0	3681.000000	1.0	5.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	1.0	0.0	1.0	1.0	0.0	0.0	1.0

Predictive Analysis – Logistic Regression

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

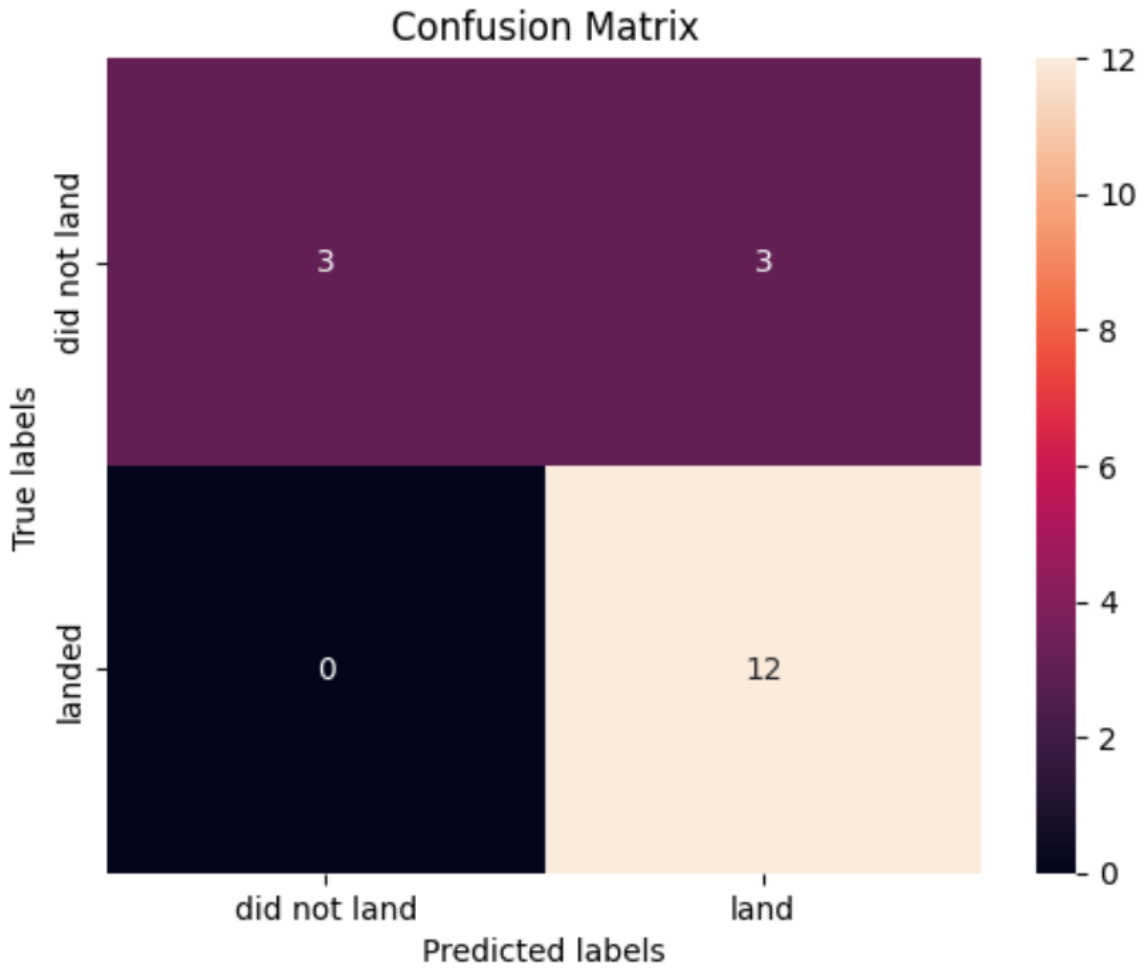


Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

SKILLS NETWORK

Predictive Analysis – Support Vector Machine

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

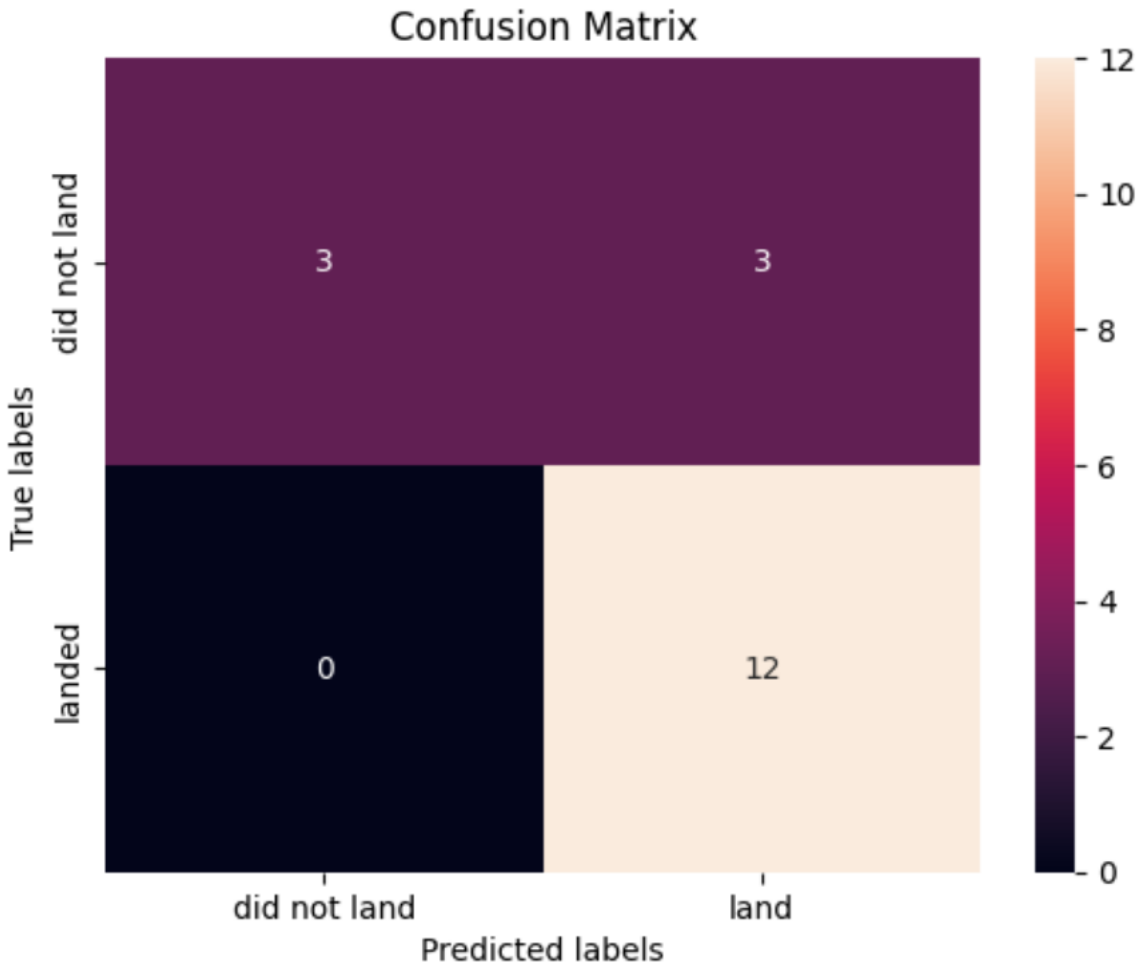


Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

SKILLS NETWORK

Predictive Analysis – Decision Tree

```
tuned hpyerparameters :(best parameters)  {'criterion': 'entropy', 'max_depth': 2, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 2, 'splitter': 'random'}
accuracy : 0.875
```

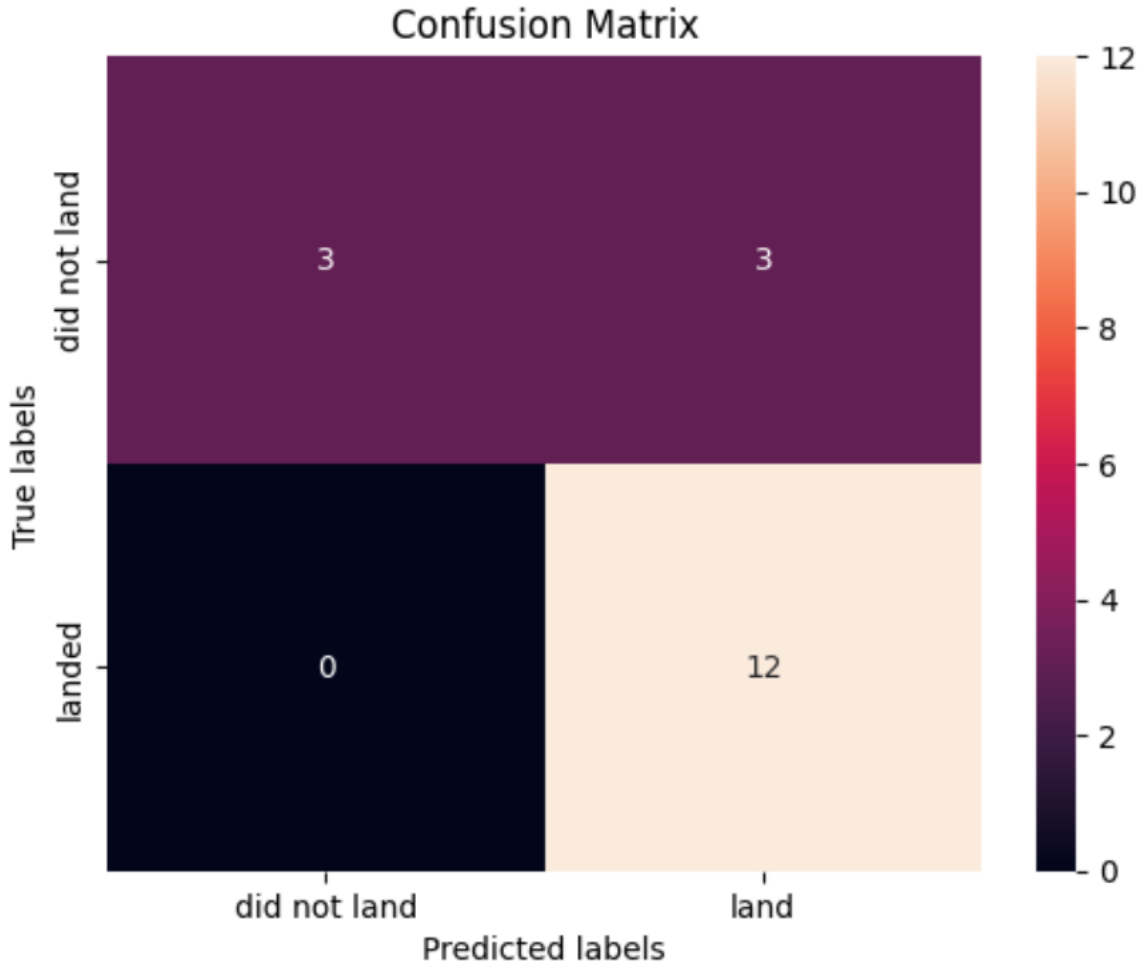


Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

SKILLS NETWORK

Predictive Analysis – k Nearest Neighbors

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



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the major problem is false positives.

SKILLS NETWORK

Predictive Analysis – Model Selection

Model	Accuracy
logistic regression	0.846429
support vector machine	0.848214
decision tree classifier	0.875000
k nearest neighbors	0.848214

Decision Tree Classifier

Falcon 9 Launches – Facts

- Total payload mass carried by boosters launched for NASA (CRS): **45596 kg**
- Average payload mass carried by booster version F9 v1.1: **2928.4 kg**
- Date when the first successful landing outcome in ground pad was achieved: 22.12.2015
- Names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000: **F9 FT B1022, F9 FT B1026, F9 FT B1021.2, F9 FT B1031.2**
- Total number of successful mission outcomes: **100**
- Total number of failure mission outcomes: **1**

CONCLUSION

- Mission success rate of Falcon 9 for launches after 2019 is above 95 %
- VAFB-SLC launchsite: no rockets launched for heavy payload
- SSO, HEO, GEO and ES-L1 orbits lanches have 100 % success rate

Decision Tree Classifier gives the best prediction accuracy for the lanch success/failure of Falcon 9 (above 87 %)

APPENDIX

- Data-set used for this report:

https://cf-courses-data.s3.us.cloud-objectstorage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/dataset_part_2.csv

Thank You!