SpaceX Falcon 9 First Stage Landing Prediction

Dr. Mohsen Mesgar

13.02.2022

OUTLINE

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

EXECUTIVE SUMMARY (Abstract)

- The task is to predict if Falcon 9 first stage lands successfully.
- What we do
  - Collect data from the wiki page of SpacX
  - Click to add text on the data to predict some trend
  - Train machine learning models (support vector machines (SVM), LogisticRegression, and Decsision Tree).
- Findings
  - The sucess rate since 2013 kept increasing till 2020
  - Different orbits have different success rates.
  - DecisionTree achieves 90% accuracy and outperforms the other examined ML model.

INTRODUCTION

- We aim at predicting if the Falcon 9 first stage lands successfully.
- 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. Therefore if we can determine if the first stage will land, 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.

METHODOLOGY

- Gathering the data including a class label to say if a landing is successful or not.
  - We do webscraping methods to collect such data from Wikipedia.
  - We deal with missing values of an attribute by replacing its NaN values with the mean of its values for other examples.
- We explore the data using some Exploratory Data Analysis (EDA) methods to find some patterns in the data.
  - We study these patterns using data visualization methods
- We conduct a feature engineering to understand what attribute in the data is distinctive between successful and unsuccessful landings.
- We build different types of machine learning methods such as (decision trees, logistic regression, and SVM) to predict if a landing is successful or not.

RESULTS

- We collected examples of landings represented with 16 different attributes and 1 class label that shows if a landing is successful or not.
- The values of PayloadMass and LandingPad attributes are missing for 5 and 26 examples, respectively. The rest of attributes have values for any example.
- Different launch sites have different success rates. CCAFS LC-40, has a success rate of 60 %, while KSC LC-39A and VAFB SLC 4E has a success rate of 77%.
- The VAFB-SLC launchsite there are no rockets launched for heavypayload mass (greater than 10000).
- Orbits SSO, HEO, GEO, and ES-L1 have high sucess rate > 0.80.
- There is no relationship between flight number when in GTO orbit.
- the sucess rate since 2013 kept increasing till 2020

Data collection and data wrangling

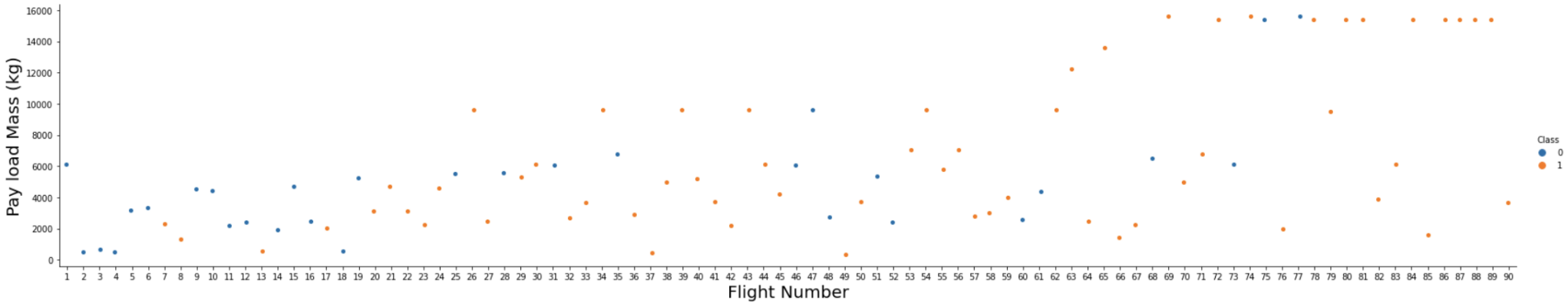
	FlightNumber	PayloadMass	Flights	Block	ReusedCount	Longitude	Latitude
count	94.000000	88.000000	94.000000	90.000000	94.000000	94.000000	94.000000
mean	54.202128	5919.165341	1.755319	3.500000	2.670213	-75.553302	28.581782
std	30.589048	4909.689575	1.197544	1.595288	3.412149	53.391880	4.639981
min	1.000000	20.000000	1.000000	1.000000	0.000000	-120.610829	9.047721
25%	28.250000	2406.250000	1.000000	2.000000	0.000000	-80.603956	28.561857
50%	52.500000	4414.000000	1.000000	4.000000	1.000000	-80.577366	28.561857
75%	81.500000	9543.750000	2.000000	5.000000	4.000000	-80.577366	28.608058
max	106.000000	15600.000000	6.000000	5.000000	10.000000	167.743129	34.632093

Data collection and data wrangling

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False

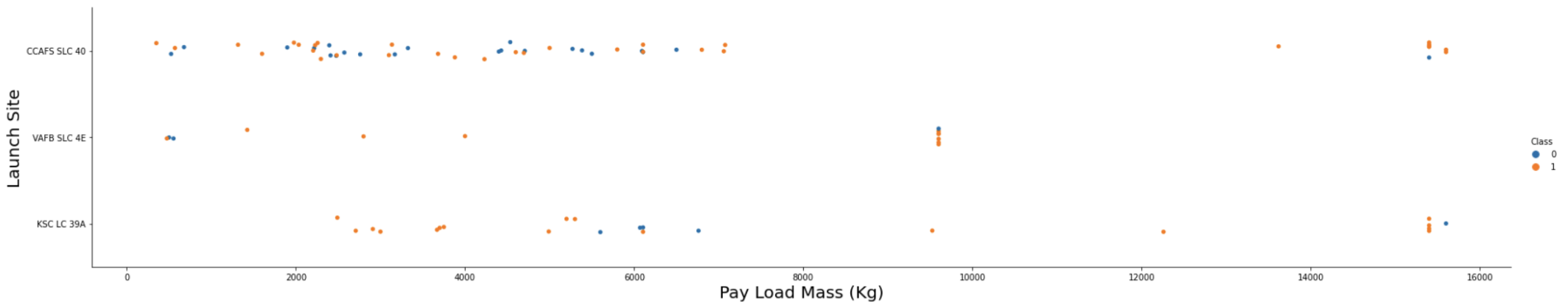
EDA and interactive visual analytics

EDA with visualization results



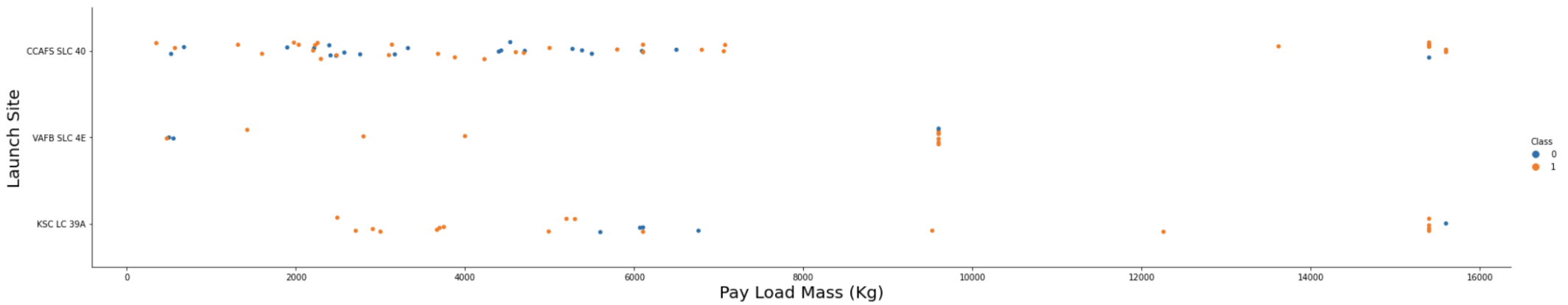
We see that different launch sites have different success rates. `CCAFS LC-40`, has a success rate of 60 %, while `KSC LC-39A` and `VAFB SLC 4E` has a success rate of 77%.

EDA with visualization results



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

EDA with visualization results



Now if you observe Payload Vs. Launch Site scatter point chart you will find for the VAFB-SLC launchsite there are no rockets launched for heavypayload mass(greater than 10000).

EDA with SQL results

```
%sql SELECT DISTINCT(launch_site) FROM SPACEXTBL;
```

```
 * ibm_db_sa://tzd23716:***@55fbc997-9266-4331-afd3-888
31929/bludb
Done.
```

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

Display 5 records where launch sites begin with the string 'CCA'

```
%sql SELECT * from SPACEXTBL WHERE launch_site like 'CCA%' LIMIT 5
```

```
 * ibm_db_sa://tzd23716:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.databases.appdomain.cloud:
31929/bludb
Done.
```

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)
2012-08-10	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
2013-03-12	22:41:00	F9 v1.1	CCAFS LC-40	SES-8	3170	GTO	SES	Success	No attempt

EDA with SQL results

```
%%sql SELECT booster_version,SPACEXTBL."Landing _Outcome", payload_mass__kg_ FROM SPACEXTBL
WHERE SPACEXTBL."Landing _Outcome" like '%Success (dron%'
AND payload_mass__kg_ between 4000 and 6000
;
```

```
 * ibm_db_sa://tzd23716:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lcg.database
31929/bludb
Done.
```

booster_version	Landing _Outcome	payload_mass__kg_
F9 FT B1022	Success (drone ship)	4696
F9 FT B1031.2	Success (drone ship)	5200

Display the total payload mass carried by boosters launched by NASA (CRS)

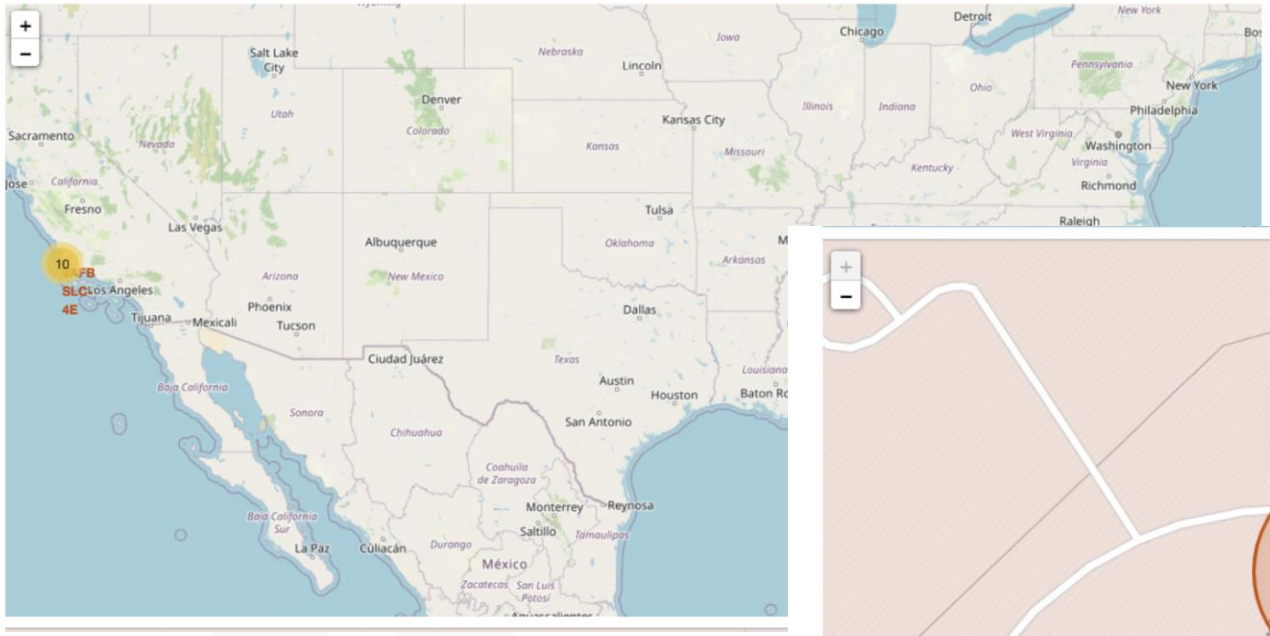
```
]:
%sql SELECT SUM(payload_mass__kg_) FROM SPACEXTBL WHERE  customer like '%NASA%';
```

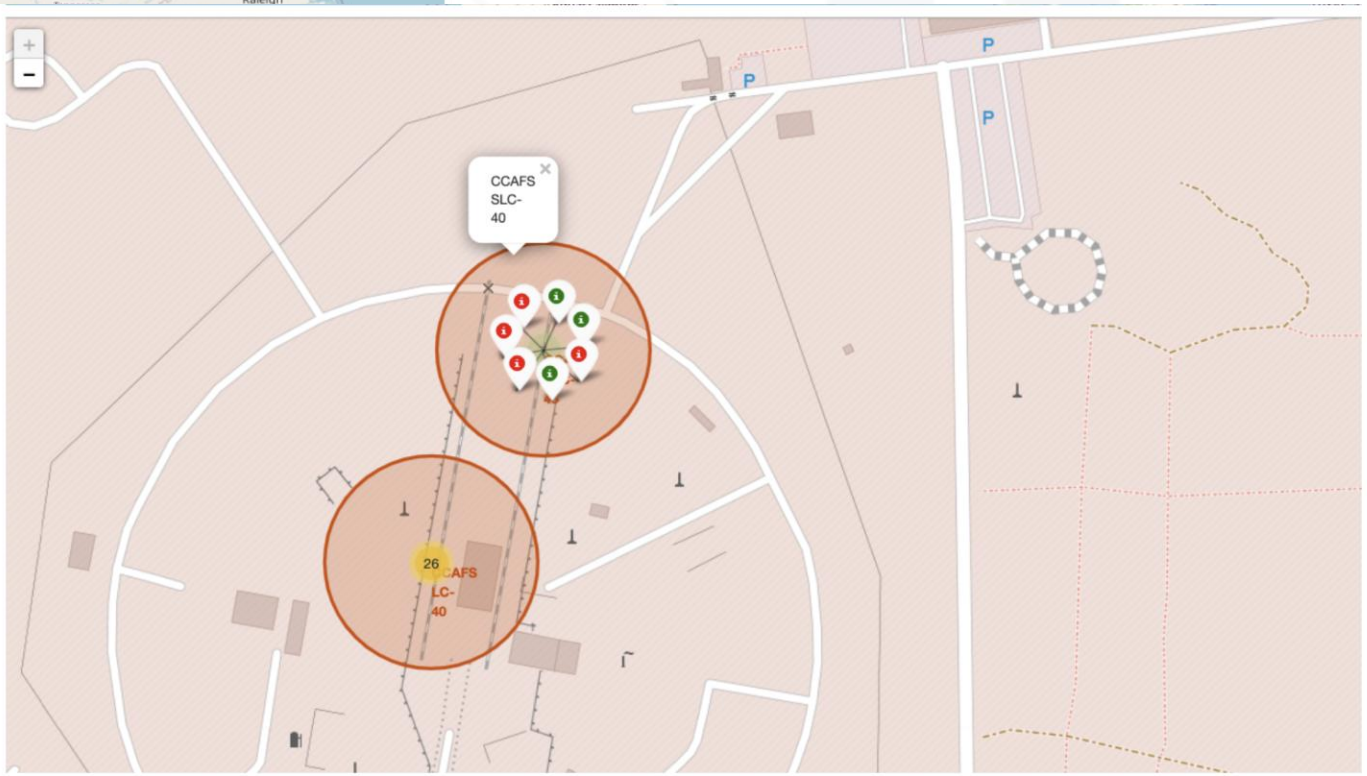
```
 * ibm_db_sa://tzd23716:***@55fbc997-9266-4331-afd3-888b05e734c0.bs2io90l08kqb1od8lc
31929/bludb
Done.
```

```
]:
```

1
37249

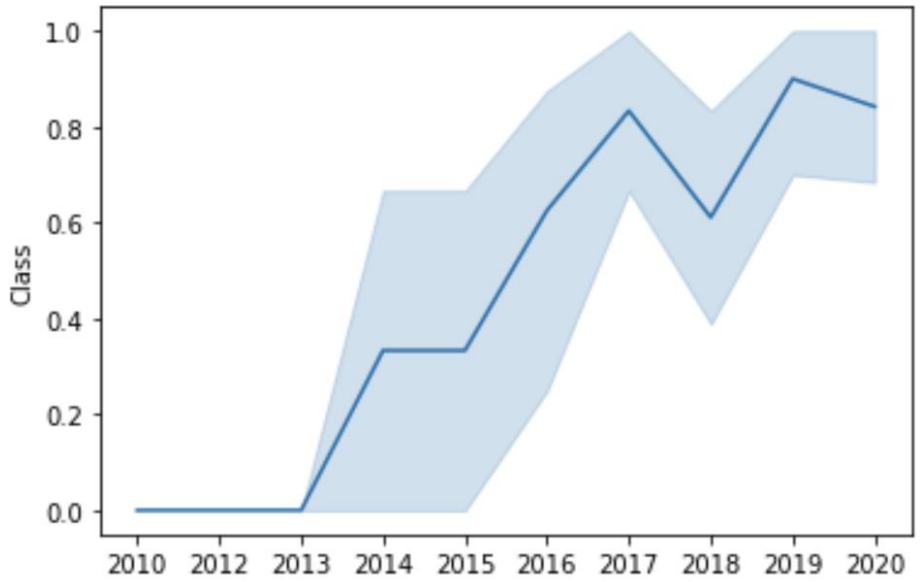
interactive map with Folium



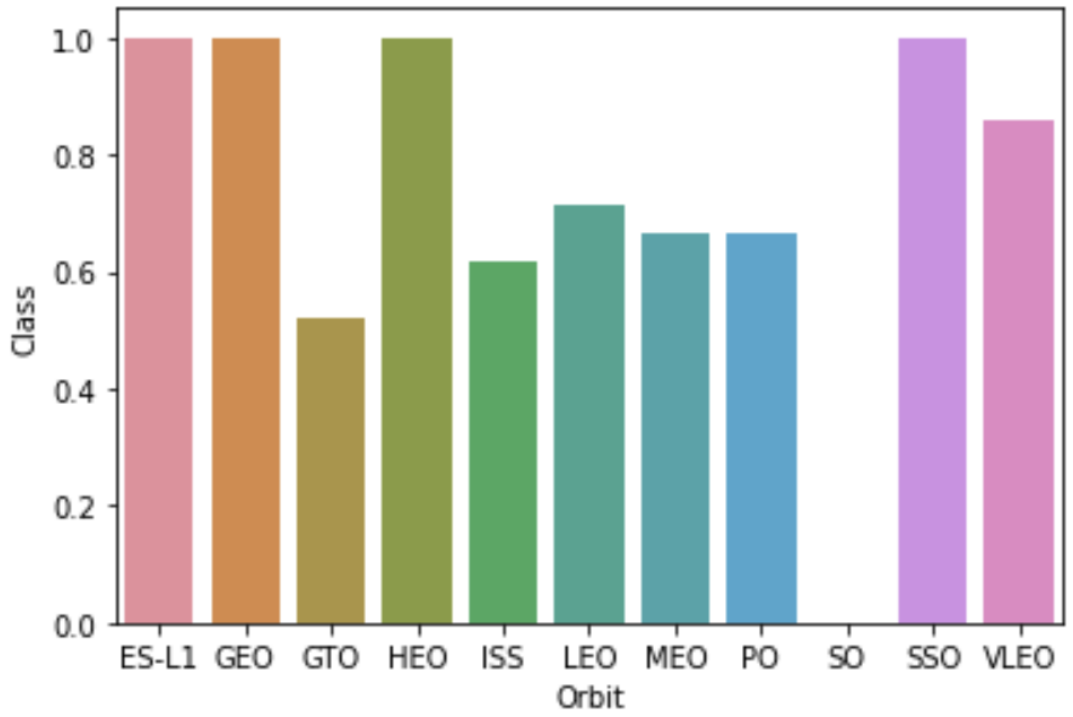


From the color-labeled markers in marker clusters, you should be able to easily identify which launch sites have relatively high success

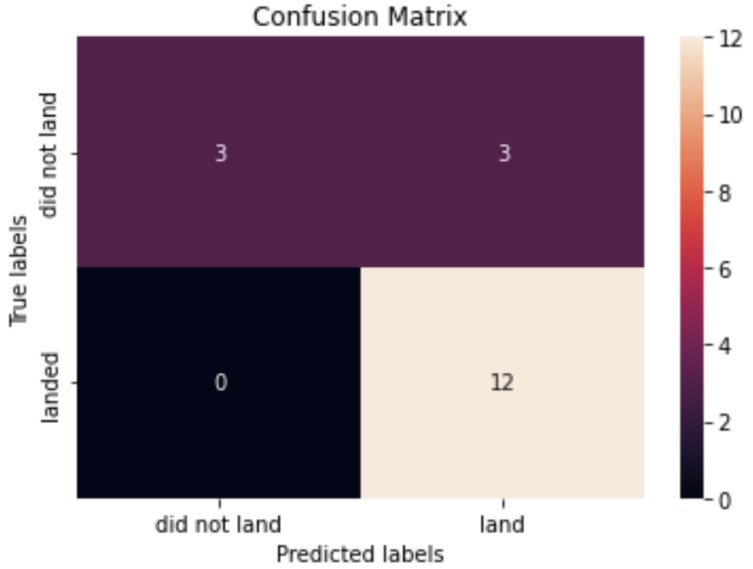
The sucess rate since 2013 kept increasing till 2020



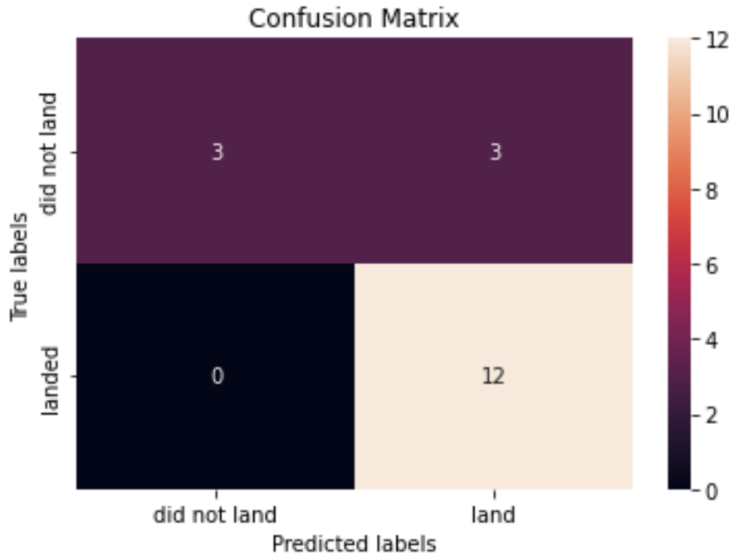
The success from different orbits



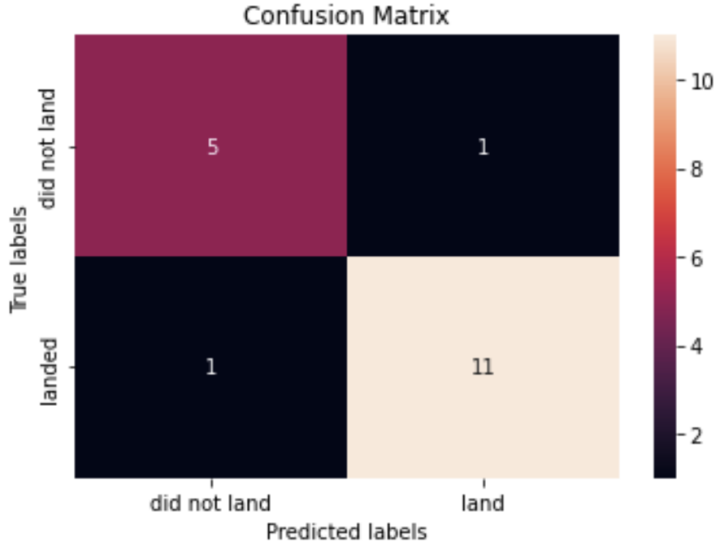
Confusion Matrices of Different Models



LogistcRegression



SVM



Decision Tree

Accuracy of Different Models on 18 unseen examples

Model	Accuracy (%)
LogisticRegression	84.64
Support Vector Machines (SVM)	84.82
Decision Tree	**90.18**

DISCUSSION

- The DecisionTree model achieves the best accuracy compared with LogisticRegression and Support Vector Machines.
- So if we use our DecsisionTree model, we can predict if a future landing can be successful with about 10% error rate.

CONCLUSION

- Our goal was to predict if the Falcon 9 first stage lands successfully.
- We collected data from SpaceX wiki page.
- We then extracted several distinctive features.
- We find out the decision tree is the best performing model for achieving our goal.

APPENDIX

- The repository of this project is here:
- https://github.com/MMesgar/tutorial-data-science/tree/main/capstone