

DATA SCIENCE FINAL CAPSTONE ON SPACE-X

BY YI YANG



Introduction:



SPACE-X is an company established on 2002 by Elon Mask and focused on designs, manufactures and launches the world's most advanced rockets and spacecraft.

SpaceX advertises Falcon 9 rocket launches on its website, with a cost of 62 million dollars. To help estimating the cost, we will design several methods to estimate if the first land will be successful or not.



Proejct Overview/Summary

This project was designed to analysis the SPACE-X dataset and predict the result of FALCON 9 LAUNCH SITE. The CONTENT include:

1. Data Collection With API and Web Scraping.
2. EDA With SQL Analysis and Data Visualization.
3. Build an Interactive Visual Analytic Dashboard With Python,
4. Machine Learning method on Predicting If The First Stage Of the FALCON 9 Will Lauch successfully.
5. FINAL CONCLUSION AND RECOMMENDATION



Data Collection With API and Web Scraping.

DATASET URL:

https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json

METHOD:

JUPYTER-NOTEBOOK, PYTHON, PANDAS PACKAGE, NUMPY PACKAGE, REQUEST PACKAGE.

1. Request.get() function was used to read through URL
2. JSON(API) and BeautifulSoup(Web Scraping) functions were applied to extract the information
3. Pandas Package was used to recombine the DATASET
4. FALCON 9 SITE information was extracted

DATASET:

SIZE : 94 ROWS AND 17 COLUMNS (FALCON 9 After filtered:90*17)

17 Variables: FlightNumber, Date BoosterVersion, PayloadMass, Orbit, LaunchSite, Outcome, Flights
GridFins, Reused Legs, LandingPad, Block, ReusedCount, Serial, Longitude, Latitude



EDA With SQL Analysis and Data Visualization.

METHOD FOR EDA SQL ANALYSIS:

SQL PROGRAMMING, JUPYTER NOTEBOOK

PANDAS PACKAGE

METHOD FOR DATA VISUALIZATION;

PANDAS, MATPLOTLIB, SEABORN
PACKAGE, POTLY, DASH, VSCODE





EDA With SQL Analysis On SPACE-X RESULT

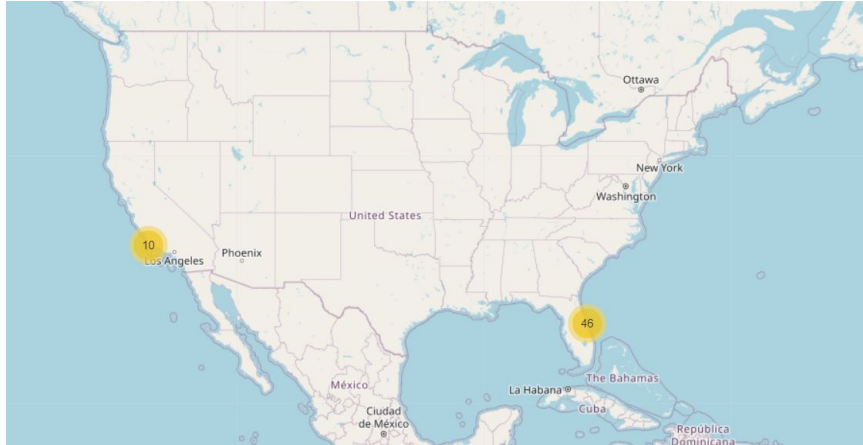
- **CRITICAL INFORMATION FROM SPACEX DATASET:**

1. **4 DISTINCT LAUNCH -STIES:** CCAFS LC-40, VAFB SLC-4E, KSC LC-39A, CCAFS SLC-40
2. **DATA TIME:** RANGE FROM 2010 TO 2020, TOTAL NUMBER, 101
3. **MISSION OUTCOME:** 90+ MISSION successfully completed
4. **CARRY CAPACITY:** F9 B5 Version shows the greatest payload mass
5. **FIRST SUCCESS LANDING OUTCOME:** Happened on 01-23-2013.

- **All Variables:**

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS_KG_	Orbit	Customer	Mission_Outcome	Landing _Outcome
------	---------------	-----------------	-------------	---------	------------------	-------	----------	-----------------	---------------------

EDA WITH DATA VISUALIZATION FOLIUM RESULT



LAUNCH SITE FOR SPACEX:

-10 IN CA

-46 IN FL

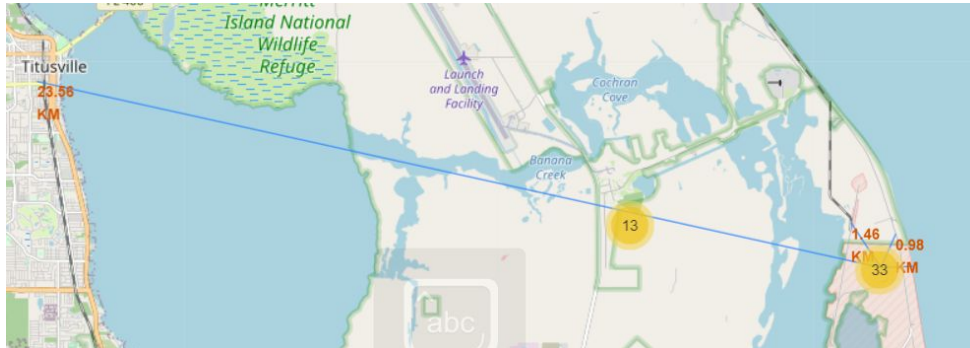
LAUNCH SITE IN FL:

**-0.9790313877747014 km,
HIGH_WAY**

**-1.4576743542071646 km,
RAILWAY**

**-23.556319091273405 km,
NEARBY City**

**-0.8627671182499878 km,
COASTLINE**





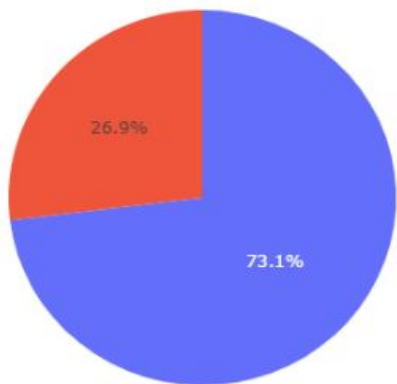
EDA WITH DATA VISUALIZATION RESULT SUCCESS/FAILURE RATIO on 4 SITES

BLUE: FAIL

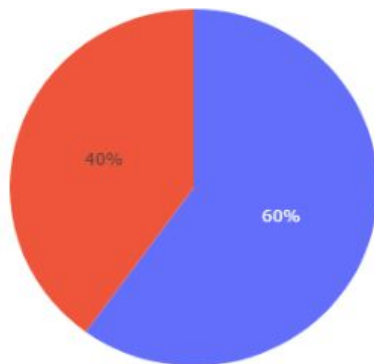
RED: SUCCESS

CONCLUSION: CCAFS SLC-40 SHOWS THE HIGHEST SUCCESSFUL RATE, WHILE KSC LC-39 A HAS THE LOWEST

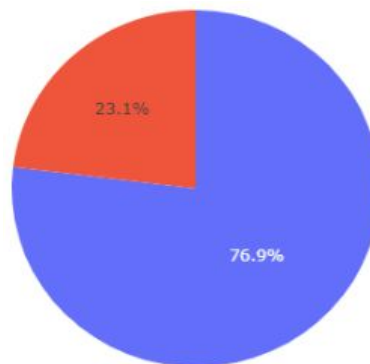
CCAFS LC-40



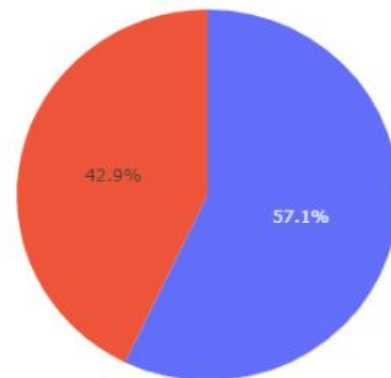
VAFB SLC-4E



KSC LC-39A



CCAFS SLC-40

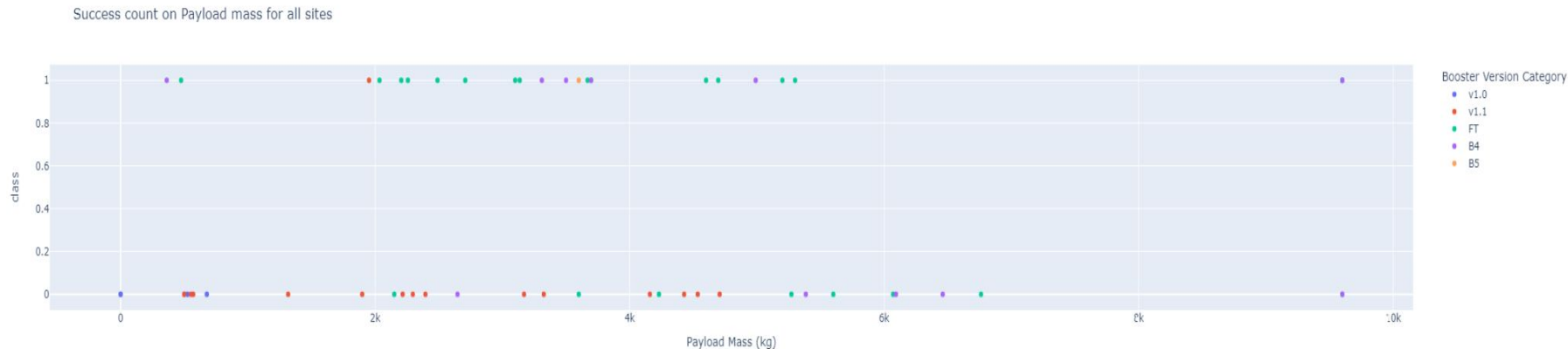




EDA WITH DATA VISUALIZATION RESULT

PLAYLOAD MASS on ALL SITES

ACCORDING TO THE SCATTER PLOT. B4 BOOSTER VERSION SHOWS THE HIGHEST PAYLOAD MASS COMPARING TO ALL OTHER BOOSTER VERSION





ML ON FALCON 9 SITE PREDICTION SUCCESS OR FAILURE

PRE-PROCESSING:

DATASET SPLITTING:

60% For TRAINING, 40% For TESTING

LOGISTIC REGRESSION, SVM, KNN, REGRESSION TREE WERE APPLIED ON TRAINING

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split( X, Y, test_size=0.4, random_state=4)
print ('Train set:', X_train.shape, y_train.shape)
print ('Test set:', X_test.shape, y_test.shape)
```

Train set: (54, 83) (54,)

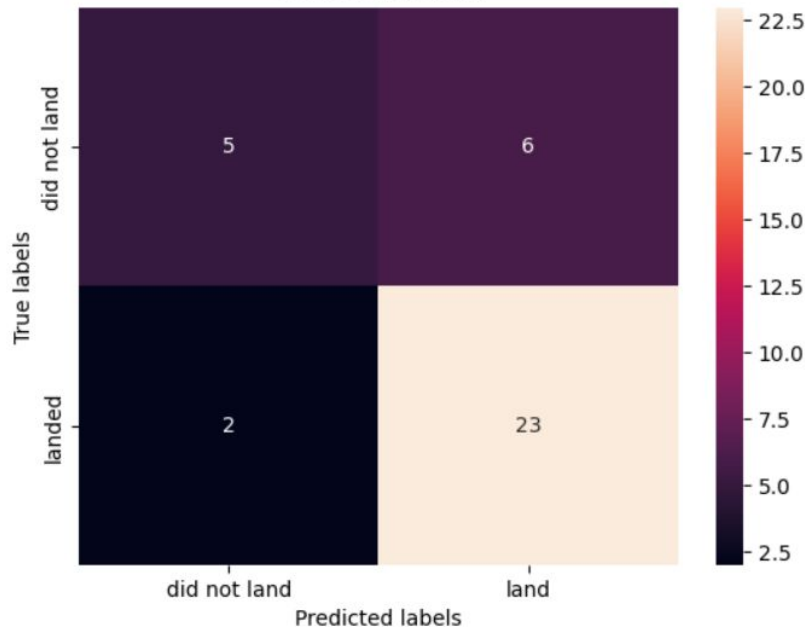
Test set: (36, 83) (36,)



ML ON FALCON 9 SITE PREDICTION SUCCESS OR FAILURE (LOGISTIC REGRESSION AND SVM)

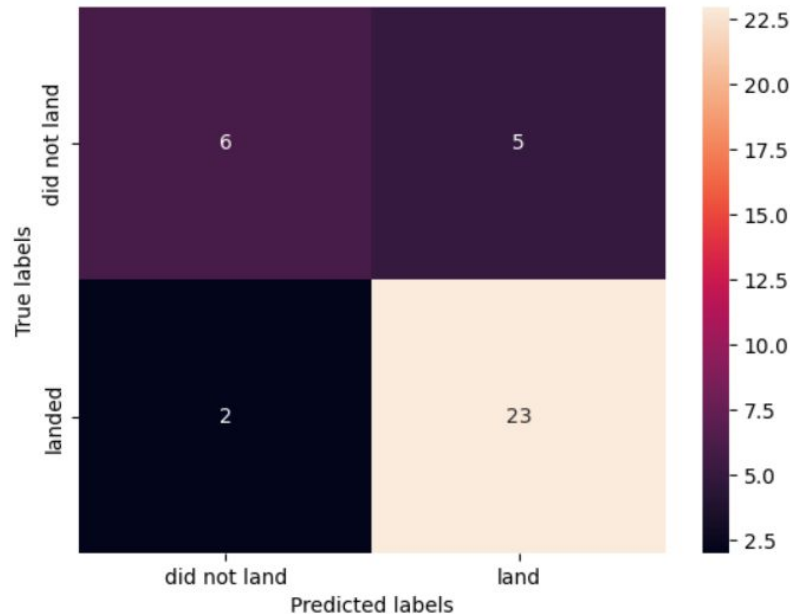
ACCURACY: 78%

Confusion Matrix



ACCURACY: 81%

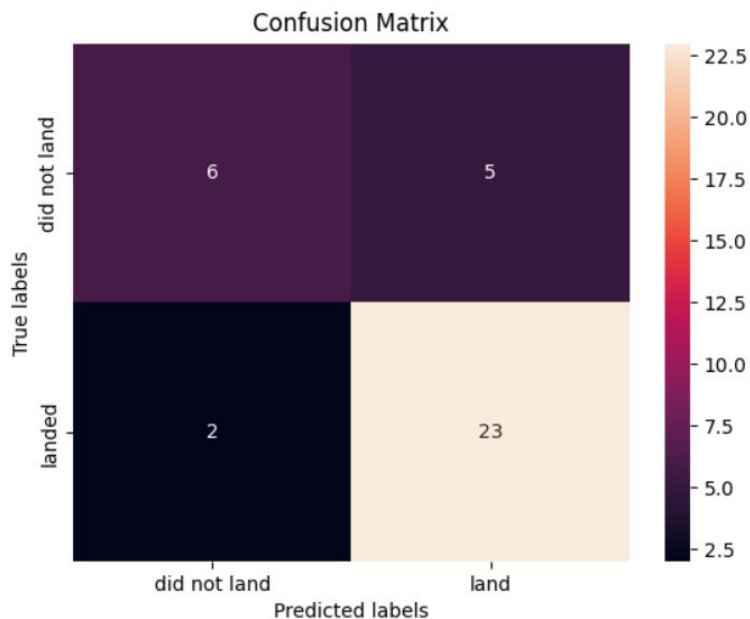
Confusion Matrix



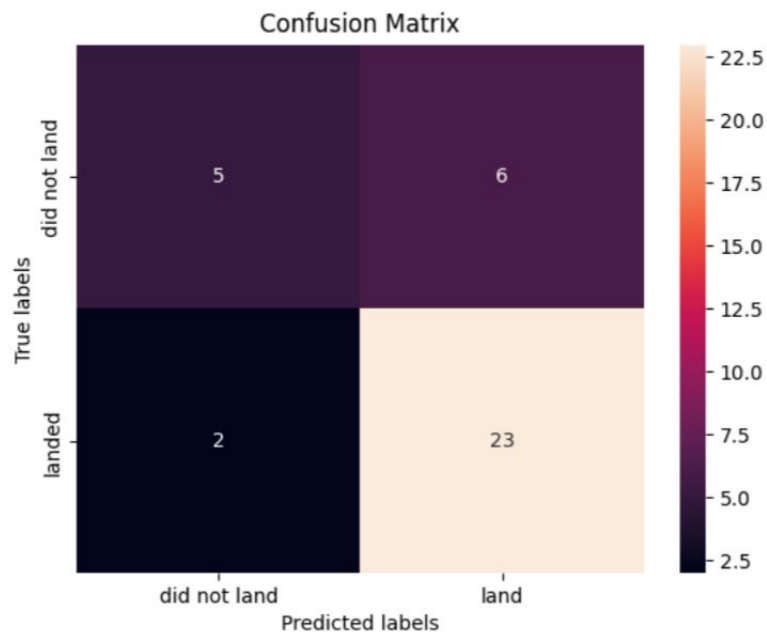


ML ON FALCON 9 SITE PREDICTION SUCCESS OR FAILURE (REGRESSION TREE AND KNN)

ACCURACY: 81%



ACCURACY: 78%





CONCLUSION AND RECOMMENDATION

- According to the data analysis result that the mission that Launch successfully in the Flacon 9 region is all filed is relatively satisfied, and the CCAFS SLC-40 were the site shows the highest successful rate.
- B4 Booster Version Shows THE HIGHEST PAYLOAD MASS. Therefore, this was recommended to apply much more widely
- The final result on SVM and Regression tree shows the best accuracy on the predict the mission outcome of the rocket launch, which is nearly 81%, which shows a good tendency to shows success ratio. However, there can be more method on future applied. For example, RNN, LSTM, even CNN can also be one of the possible model to show good result.

Thank you