Predicting whether the Falcon 9 will land successfully

Digant Shetkar

17-01-2023

OUTLINE

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

EXECUTIVE SUMMARY

The objective of this study was to investigate the factors determining the outcome of a Falcon 9 launch. Impact of factors like payload mass, launch site, orbit and booster version on the outcome of the launch was investigated.

By creating an interactive dashboard, the impact of payload mass on the outcome of the mission and the successful launches from each launch site were visualised.

Various machine learning models were trained and tested on mutually exclusive data. By doing so, we aimed to find the best model and the best parameters for predicting the outcome of a Falcon 9 launch. We concluded that a decision tree model would be the best fit.

INTRODUCTION

SpaceX advertises the cost of launching a Falcon 9 rocket as $62 million while other competitors cost upwards of $165 million. SpaceX saves the odd $100 million by reusing the first stage of a Falcon 9. But, they can do so only if the launch is successful.

So, it is important to investigate which factors determine whether the first stage will land.

The knowledge gained through this study can be used by SpaceX for improving their first stage landing success rate or it can be useful for a competitor to establish a better program.

METHODOLOGY

- Data collection

The data on SpaceX launches was collected through the SpaceX REST API and the Wikipedia page on SpaceX launches.

- Pre-formatting data

Using the Pandas library and scikit learn library, the data was pre-formatted for the purposes of exploration and training and testing of machine learning models.

- Exploratory Data Analysis

Exploratory data analysis was conducted using SQL through Jupyter Notebooks and by creating an interactive dashboard using Plotly's libraries.

- Machine learning

Support vector machines, decision tree classifiers, logistic regression and K-nearest neighbors machine learning models were trained and tested on the data collected.

RESULTS

Data collection was done on two fronts:

1. Webscraping through the Beautiful Soup library
2. SpaceX REST API

Using the Pandas library, the data was pre-formatted for the purpose of exploratory data analysis. Exploratory data analysis was conducted using SQL queries and Plotly's interactive graphs.

Using GridSearchCV, the most appropriate model and its most appropriate parameters for predicting the launch outcome were selected.

Results from EDA (SQL)

- Launch sites

Display the names of the unique launch sites in the space mission

In [7]:
```
%sql SELECT DISTINCT Launch_site from SPACEX
```

```
 * ibm_db_sa://rbk38174:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

Out[7]:

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

- Date of the first successful launch

List the date when the first successful landing outcome in ground pad was acheived.

Hint:Use min function

In [23]:
```
%sql SELECT min(DATE) from SPACEX where mission_outcome = 'Success'
```

```
 * ibm_db_sa://rbk38174:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

Out[23]:

1
2010-06-04

IBM Developer

SKILLS NETWORK

Results from EDA (SQL)

- Boosters used for successful drone ship landings

List the names of the boosters which have success in drone ship and have payload mass greater than 4000 but less than 6000

In [25]:
```
1 SELECT DISTINCT booster_version from SPACEX where 4000 < payload_mass__kg_ < 6000 and landing__outcome = 'Success (drone ship)'
```

```
 * ibm_db_sa://rbk38174:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

Out[25]:

booster_version
F9 B4 B1042.1
F9 B4 B1045.1
F9 B5 B1046.1
F9 FT B1029.2
F9 FT B1021.1
F9 FT B1023.1
F9 FT B1038.1

- Launch sites and booster versions of failed drone ship landings

List the failed landing_outcomes in drone ship, their booster versions, and launch site names for in year 2015

In [61]:
```
%sql SELECT booster_version, launch_site from SPACEX where landing__outcome = 'Failure (drone ship)' and DATE like '2015%'
```

```
 * ibm_db_sa://rbk38174:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

Out[61]:

booster_version	launch_site
F9 v1.1 B1012	CCAFS LC-40
F9 v1.1 B1015	CCAFS LC-40

Results from EDA (SQL)

- Landing outcome tally

Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order

In [66]:
```
%sql SELECT landing__outcome, COUNT(landing__outcome) from SPACEX group by landing__outcome order by COUNT(landing__outcome) desc
```

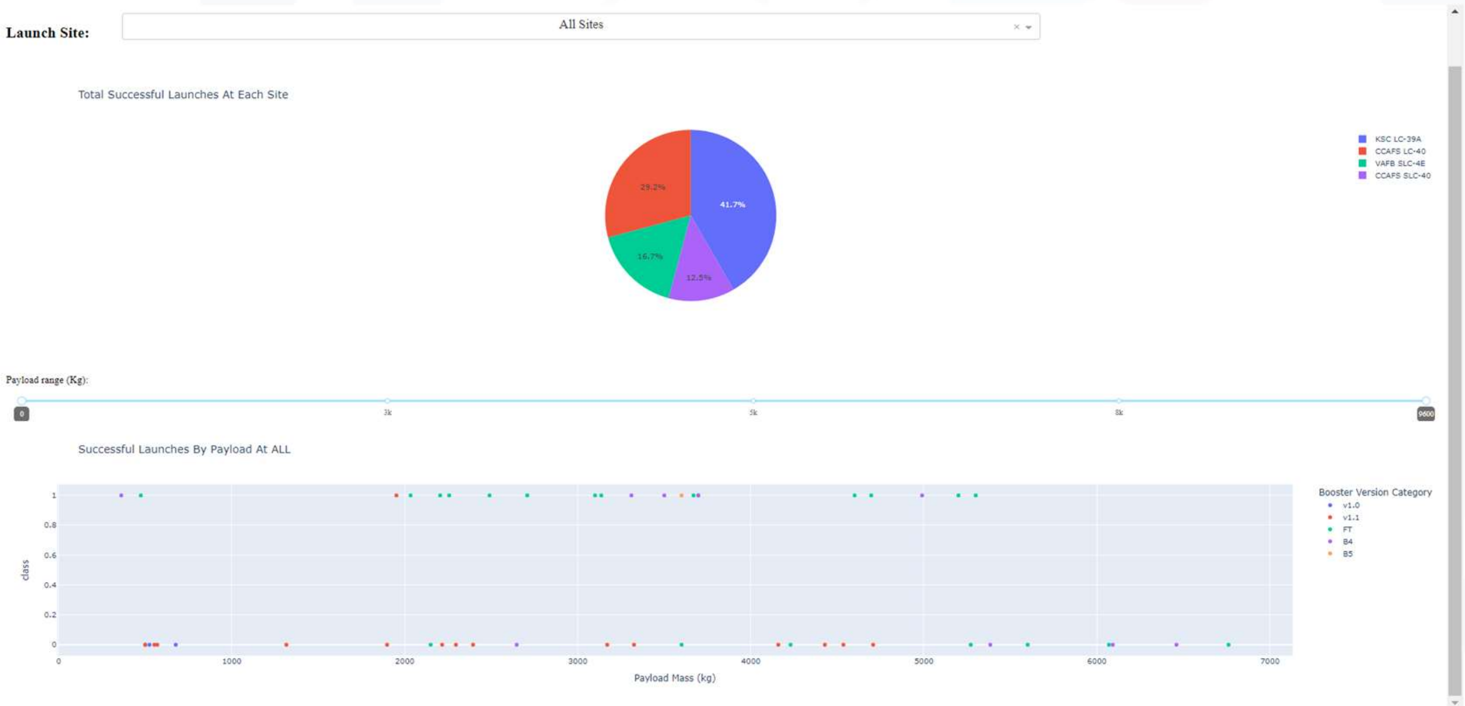
```
 * ibm_db_sa://rbk38174:***@b1bc1829-6f45-4cd4-bef4-10cf081900bf.c1ogj3sd0tgtu0lqde00.databases.appdomain.cloud:32304/bludb
Done.
```

Out[66]:

landing__outcome	2
Success	38
No attempt	22
Success (drone ship)	14
Success (ground pad)	9
Controlled (ocean)	5
Failure (drone ship)	5
Failure	3
Failure (parachute)	2
Uncontrolled (ocean)	2
Precluded (drone ship)	1

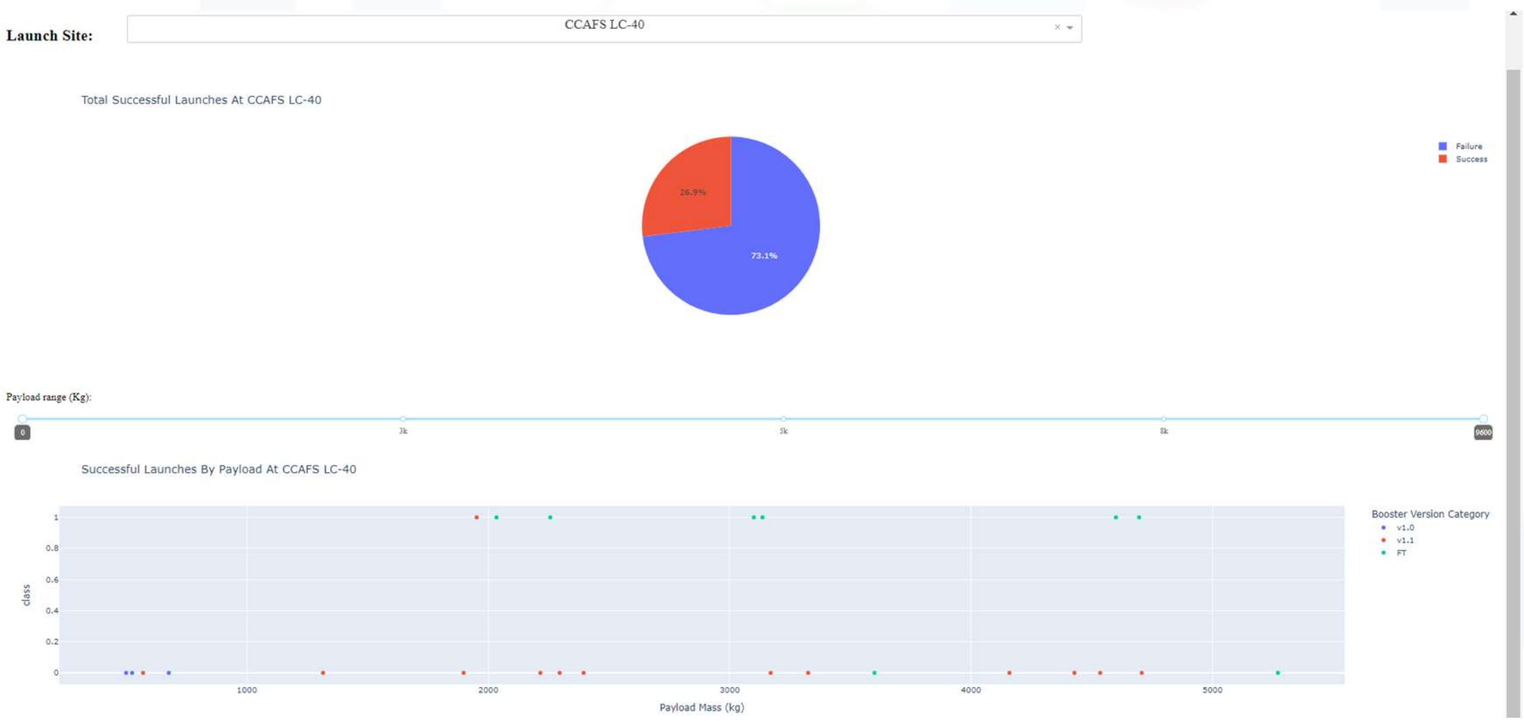
Results from Dashboard

- Outcomes from all sites



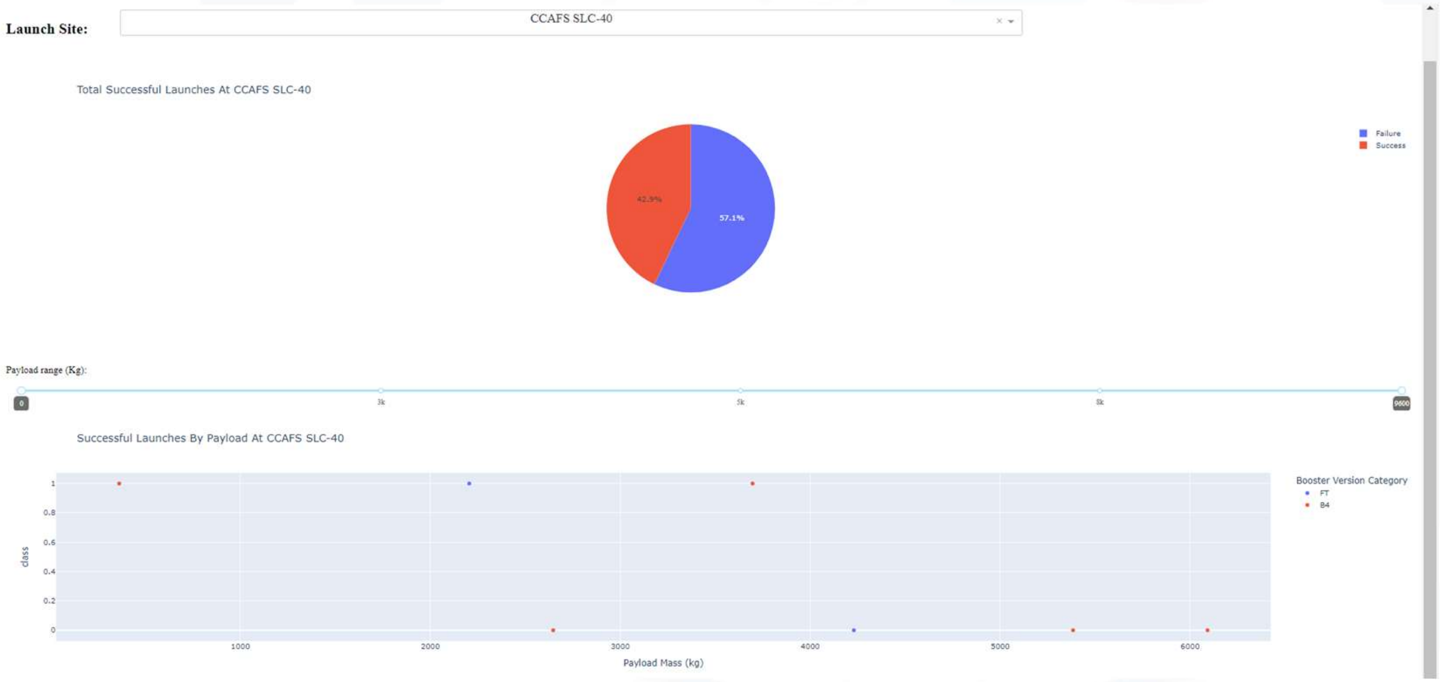
Results from Dashboard

- Outcomes from site – CCAFS LC-40



Results from Dashboard

- Outcomes from site – CCAFS SLC-40

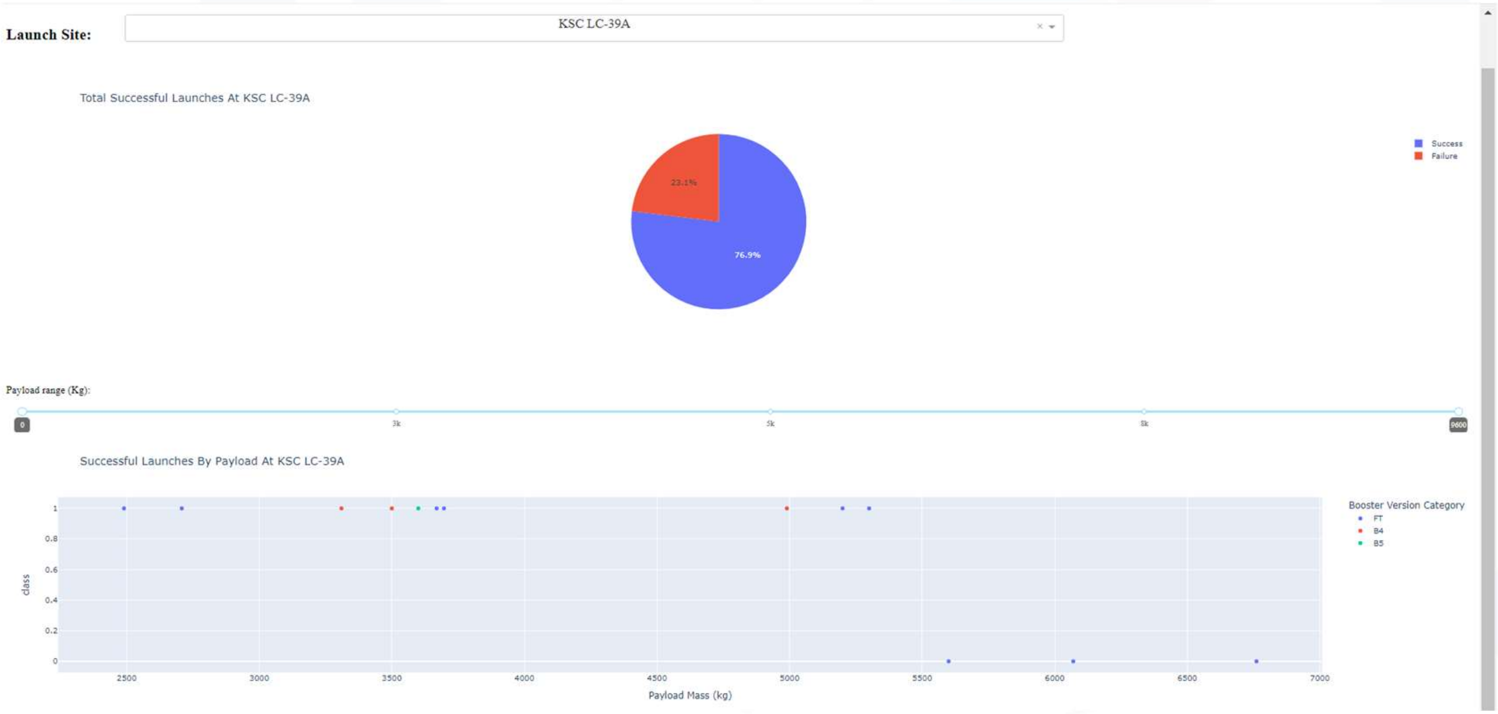


IBM Developer

SKILLS NETWORK

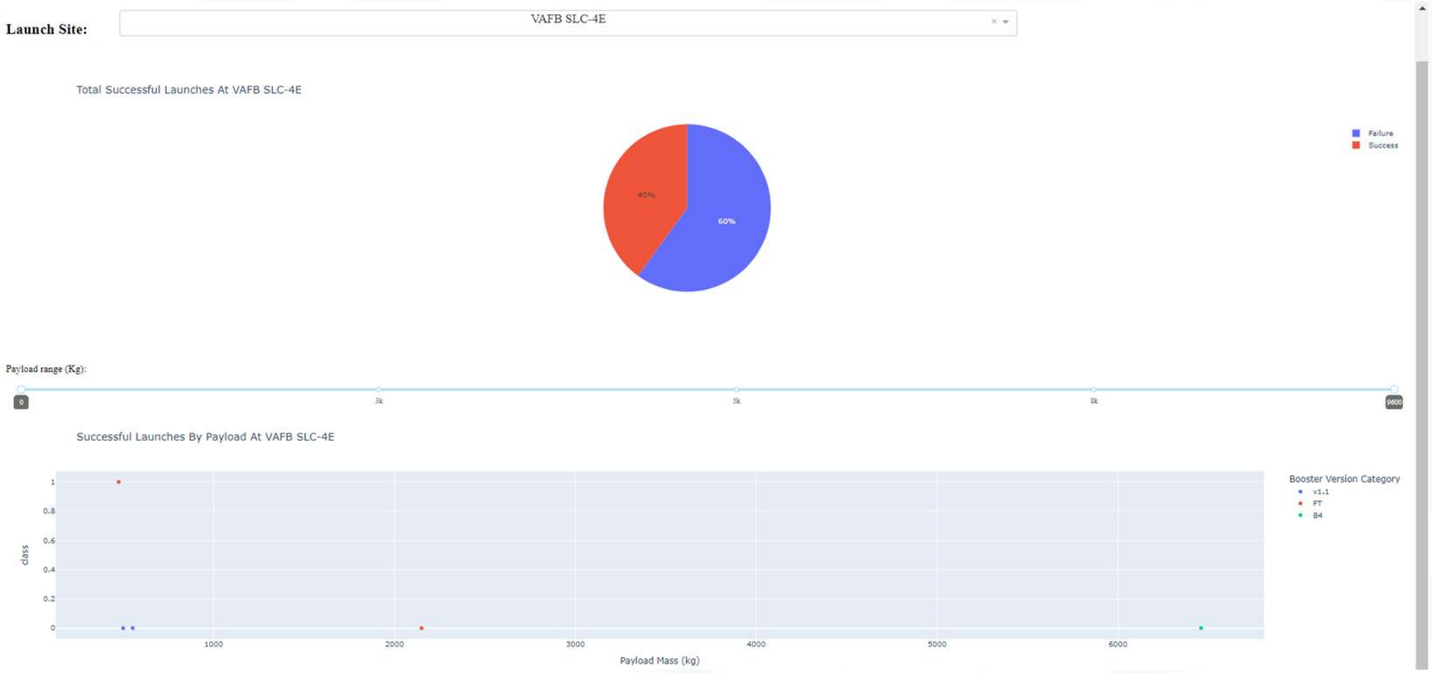
Results from Dashboard

- Outcomes from site – KSC LC-39A



Results from Dashboard

- Outcomes from site – VAFB SLC-4E

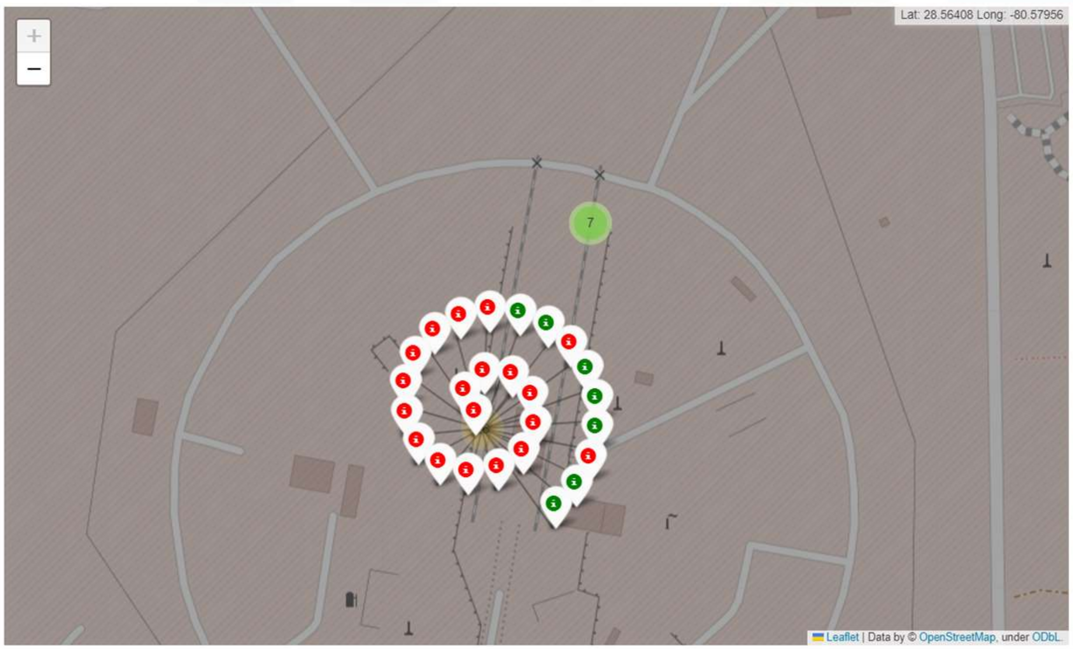


IBM Developer

SKILLS NETWORK

Results from Folium

- Outcomes from site – CCAFS LC-40

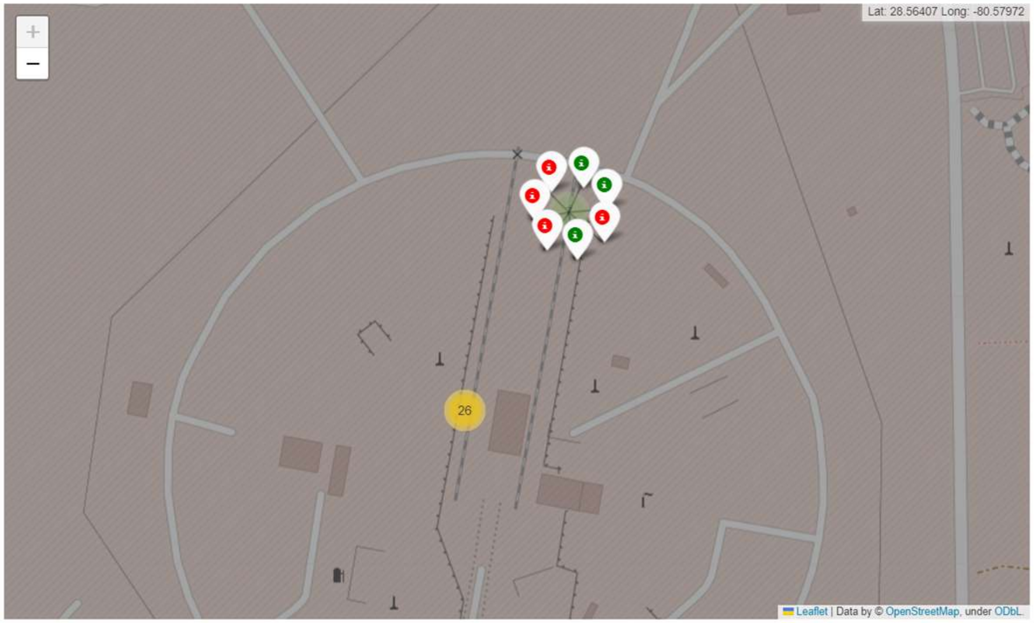


IBM Developer

SKILLS NETWORK

Results from Folium

- Outcomes from site – CCAFS SLC-40

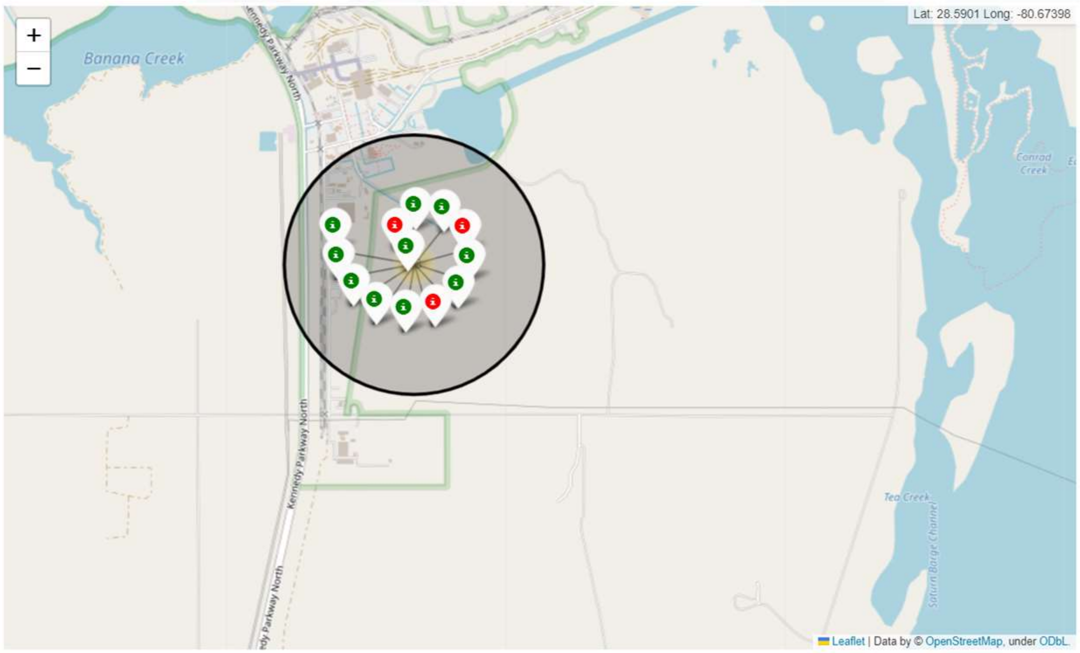


IBM Developer

SKILLS NETWORK

Results from Folium

- Outcomes from site – KSC LC-39A

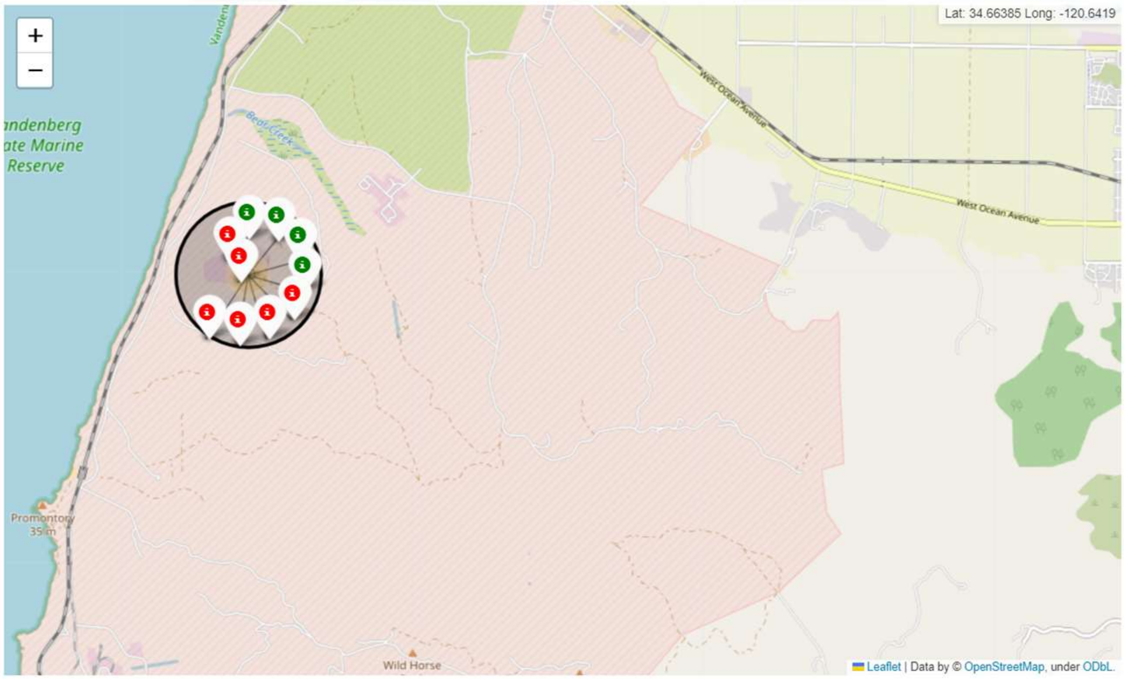


IBM Developer

SKILLS NETWORK

Results from Folium

- Outcomes from site – VAFB SLC-4E



Results from Machine Learning

- Scores of machine learning models

```
In [40]: models = [['Logistic Regression', logreg_cv.best_score_, logreg_cv.score(X_test, Y_test)],
                   ['Support Vector Machines', svm_cv.best_score_, svm_cv.score(X_test, Y_test)],
                   ['Decision Tree', tree_cv.best_score_, tree_cv.score(X_test, Y_test)],
                   ['K-Nearest Neighbors', KNN_cv.best_score_, KNN_cv.score(X_test, Y_test)]]
         mt_df = pd.DataFrame(models, columns=['Method', 'Training Score', 'Testing Score'])

In [41]: mt_df
```

Out[41]:

	Method	Training Score	Testing Score
0	Logistic Regression	0.846429	0.833333
1	Support Vector Machines	0.848214	0.833333
2	Decision Tree	0.889286	0.833333
3	K-Nearest Neighbors	0.848214	0.833333

```
In [42]: tree_cv.best_params_

Out[42]: {'criterion': 'entropy',
          'max_depth': 14,
          'max_features': 'sqrt',
          'min_samples_leaf': 2,
          'min_samples_split': 2,
          'splitter': 'best'}
```

DISCUSSION

Exploratory data analysis reveals that the highest payload mass successfully landed was 5,300 kg.

The worst performing launch sites with success rates of 26.9% and 23.1% were CCAFS LC-40 and KSC LC-39A. Other sites had a success rate of close to 40%. So, it would be better to avoid CCAFS LC-40 and KSC LC-39A and to avoid having a payload higher than 5,300 kg.

CONCLUSION

To boost the success probability, payload mass should be kept lower than 5,300 kg and launch sites CCAFS LC-40 and KSC LC-39A should be avoided.

The best machine learning model for predicting the outcome of the first stage was the decision tree classifier.

There was no difference in performance on testing data, so the model was chosen based on its fit on training data.

APPENDIX

Training data size = 80%

Testing data size   =20%

Cross-validation splitting strategy = 10