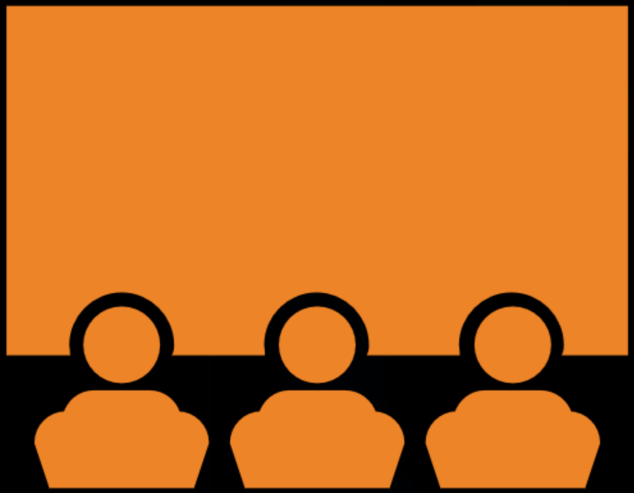
FALCON 9 LANDING PREDICTION

Emilija Stanković

14.03.2023.

SKILLS NETWORK

OUTLINE



- Executive Summary
- Introduction
- Methodology
  - Data Collection
  - Data Wrangling
  - Visualization
  - Machine Learning
- Results
  - EDA - Tables
  - Visualization – Charts, Folium maps
  - Dashboard
  - Machine Learning Models
- Conclusion

EXECUTIVE SUMMARY

Data Collection

- Falcon 9 launch data collected from SpaceX API and Wikipedia contains comprehensive data on Falcon 9 launches and landings.

Data analysis

- After initial data collection and data wrangling, visual data analysis was performed.

Machine Learning

- Landing success prediction was performed with machine learning algorithms.

INTRODUCTION

SpaceX is an American aerospace company that develops commercial spaceflight. It was the first private company which successfully launched and returned a spacecraft from Earth orbit. SpaceX focuses on making reusable rockets.

In 2010 SpaceX first launched its Falcon 9, a spacecraft that was designed so that its first stage could be reused.

The purpose of this project is to predict whether Falcon 9 will land successfully.

Falcon 9 lifting off from LC-39A
Source: https://en.wikipedia.org/wiki/Falcon_9

IBM Developer

SKILLS NETWORK

METHODOLOGY – DATA COLLECTION

Data was collected from SpaceX API and Wikipedia.

- Data is requested from provided URL with request.get() method. In collected data, a lot of the data are IDs so we select some variables, store their values in lists and make a dictionary from lists. Then, we make a dataframe from dictionary using pd.DataFrame() method. Data is filtered to get only Falcon 9 data and missing values in PayloadMass column are replaced with mean.
- HTML table from Wikipedia is requested from provided URL and BeautifulSoup object was created. Then, we created dictionary by parsing HTML table and converted dictionary into a dataframe.

METHODOLOGY – DATA WRANGLING

Exploratory data analysis (EDA) is performed to make insights from data and prepare it for modelling.

- For EDA we used Pandas package.
- Different launch sites have different numbers of launches, examined with DataFrame.value_counts() method.
- Column Orbit contains different orbits where each launch aims.
- Landing outcomes can be Success (label: 1) or Unsuccess (Label: 0).
- Column Class was created based on landing outcome.

METHODOLOGY – VISUALIZATION

Data visualization was performed using Matplotlib and Seaborn package. Also, we used Folium for creating interactive maps.

- With sns.catplot() we created categorical plots of different variables to show relationship between them.
- sns.barplot() was used for visual representation of success rate in each orbit type.
- Matplotlib was used to plot success rates by years.
- Folium was used to mark all launch sites on map and add different color markers based on landing success.
- Also, Folium helped calculate distance from launch site to the closest coast, railway and city.

METHODOLOGY – MACHINE LEARNING

For predictive analysis we use logistic regression, support vector machine, decision tree classifier and k nearest neighbours.

- First, predictor and response variables are defined and predictor variables are standardized.
- Data is split into training and test data.
- We fit training and test data.
- For every model we find the best parameters using attribute best_params_ and accuracy of the validation data using best_score_ attribute of GridSearchCV object.
- Then, we calculate test data accuracy with method score() and make a prediction with predict() method.
- Finally, we plot the confusion matrix.

RESULTS - EDA

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

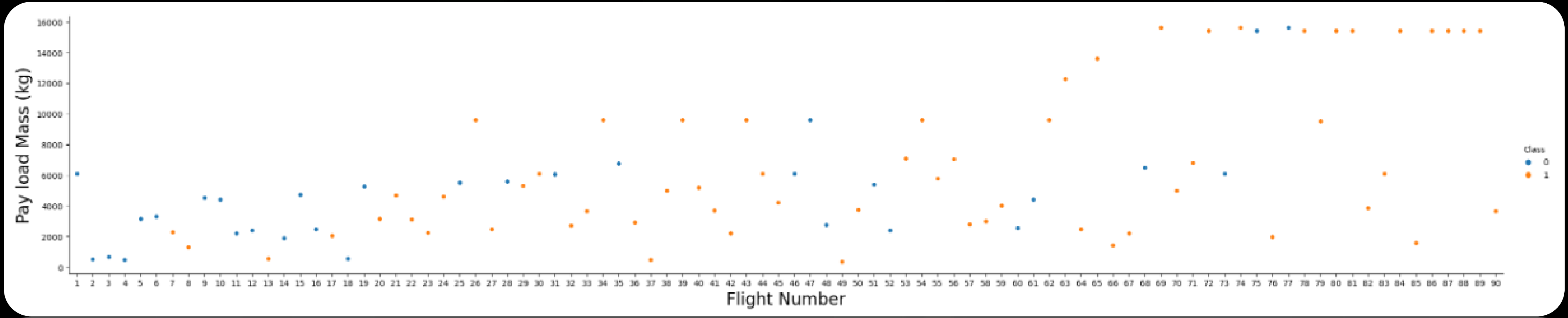
Every launch aims at specific orbit. Table shows number of launches per orbit.

Landing outcome can be successful (True) or unsuccessful (False). Landing site can be ocean, ground pad (RTLS) or drone ship (ASDS). None ASDA and None None represent unsuccess to land.

Outcome	Site	Count
True	ASDS	41
None	None	19
True	RTLS	14
False	ASDS	6
True	Ocean	5
False	Ocean	2
None	ASDS	2
False	RTLS	1

RESULTS – VISUALIZATION

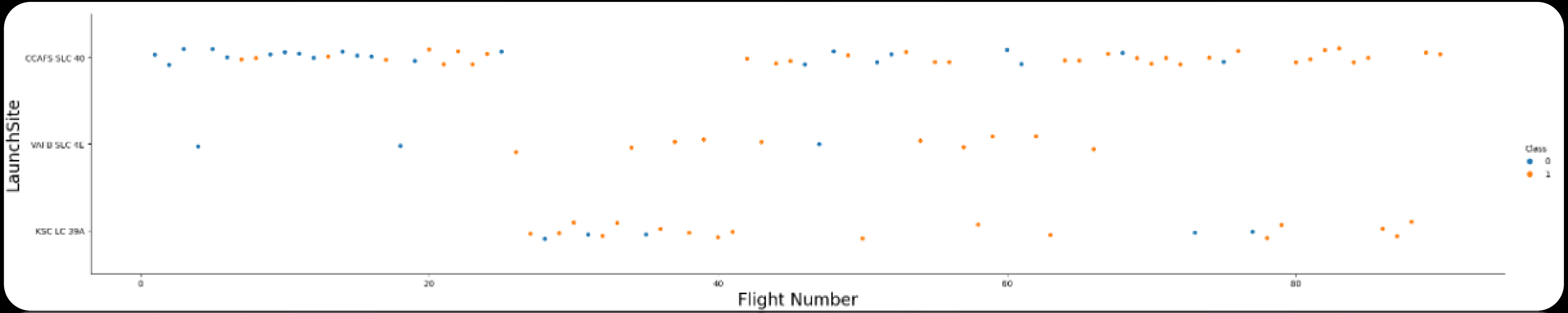
Categorical plot: Flight Number vs Pay load Mass



From the plot above we can see that with increase in flight number payload mass is increasing meaning that newer spacecraft have higher payload mass. Flights with higher number have higher success rate.

RESULTS – VISUALIZATION

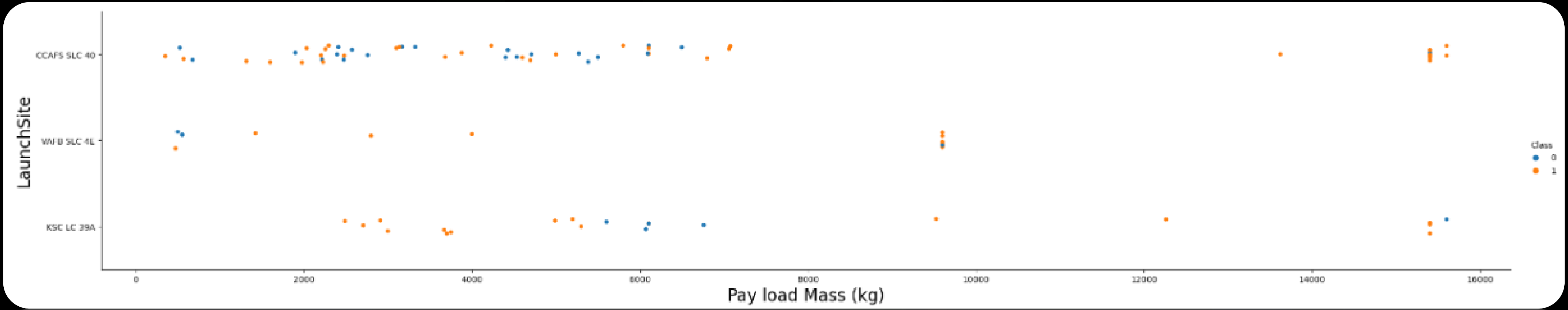
Categorical plot: Flight Number vs Launch Site



Plot shows relationship between Flight Number and Launch Site. We can conclude that some of the first launches were carried out from CCAFS SLC-40 site and many of them were unsuccessful. Later launches were carried out the other two launch sites and rate of success increased.

RESULTS – VISUALIZATION

Categorical plot: Payload Mass vs Launch Site



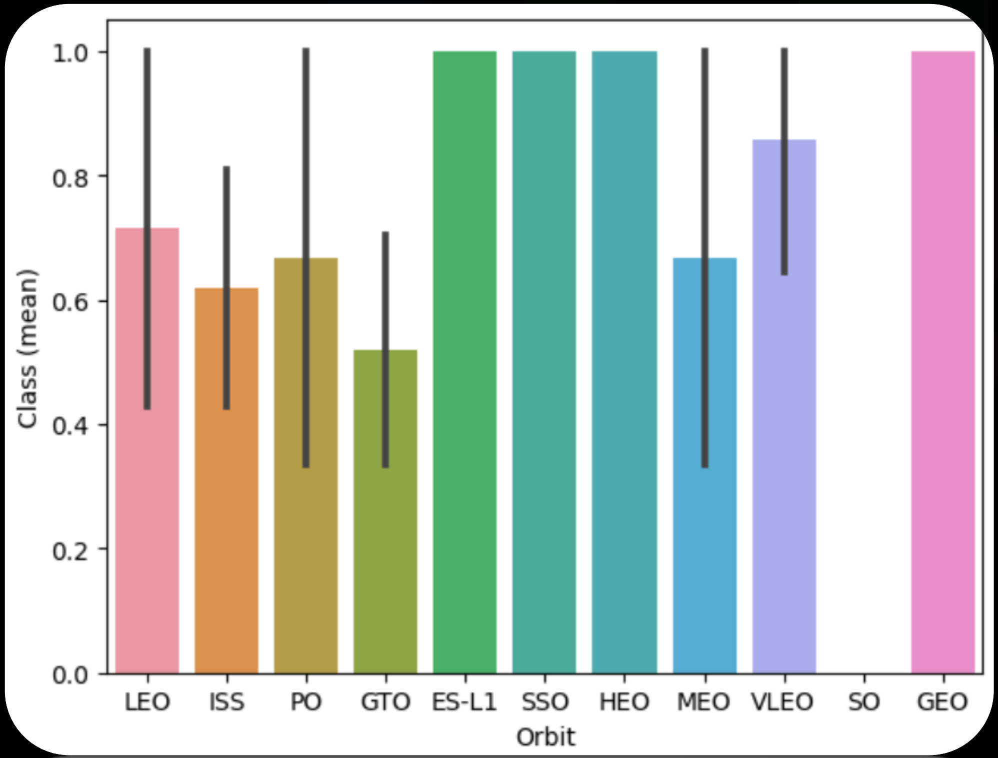
Graph shows distribution of payload mass by launch site. We can see that rate of success depends mainly on launch site, while crafts with higher payload mass were mostly successful.

IBM Developer

SKILLS NETWORK

RESULTS – VISUALIZATION

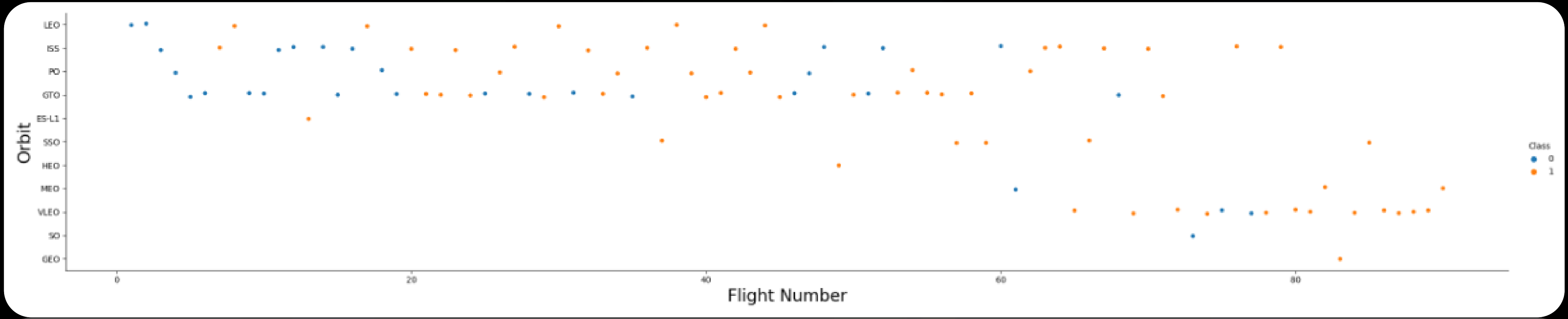
Bar plot: Orbit vs Class



We can conclude that rate of success is different based on orbit the flight aims at. Some orbits like ES-L1, SSO, HEO and GEO have 100% success rate, while some other orbits like GTO have success rate of about 50%.

RESULTS – VISUALIZATION

Categorical plot: Flight Number vs Orbit



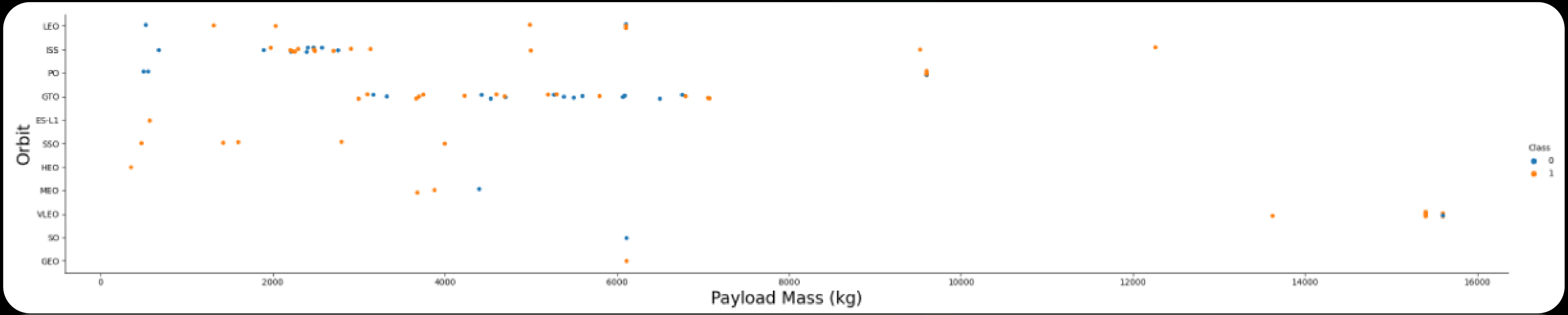
More recent flights aim at more distant orbits, nevertheless, there are still flights to closer orbits. Launch success rate is higher in more recent launches.

IBM Developer

SKILLS NETWORK

RESULTS – VISUALIZATION

Categorical plot: Payload Mass vs Orbit



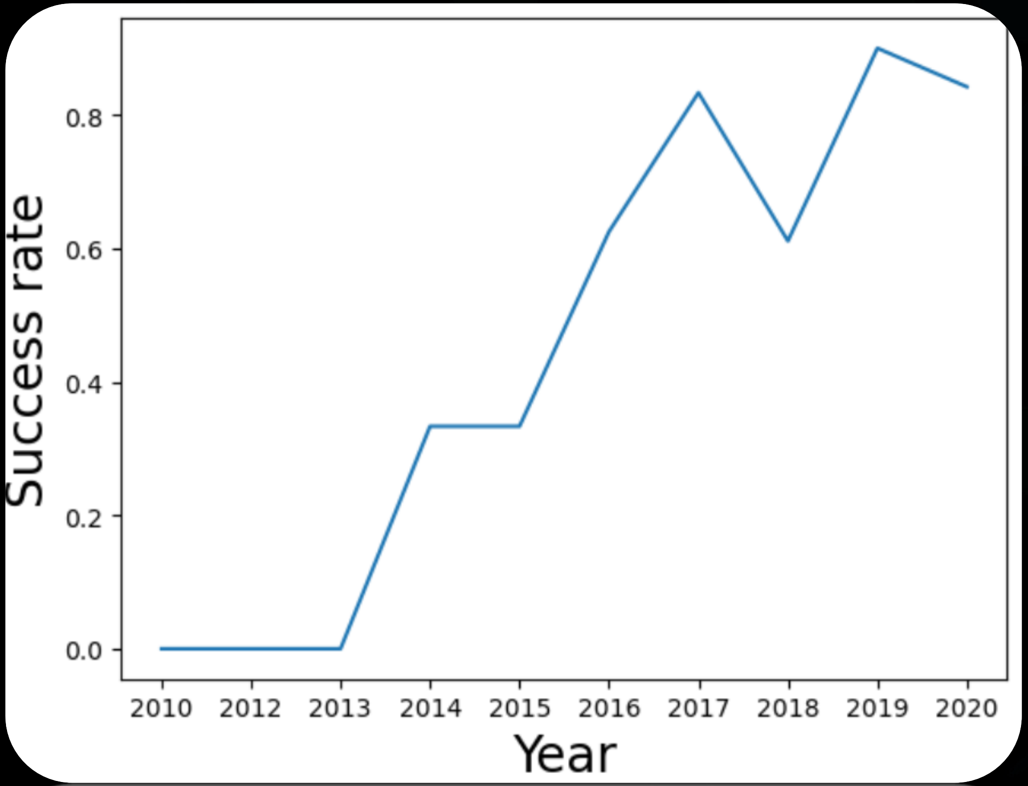
We can see that spacecraft with the highest payload masses aims at VLEO, ISS and PO orbits. GTO is an orbit with the most middle payload mass values.

IBM Developer

SKILLS NETWORK

RESULTS – VISUALIZATION

Line plot: Year vs Success Rate



Success rate increases with years. Technological progress since 2013 has been significant and made strong influence on landing success. We can assume that this trend will continue.

RESULTS – EDA WITH SQL

Launch sites
CCAFS LC-40
CCAFS SLC-40
KSC LC-39A
VAFB SLC-4E

SQL magic was used within Jupyter notebook to perform exploratory data analysis. Output of SQL queries is shown in next few slides.
Table on the left shows distinct launch sites and bottom table displays first five records where launch sites begin with 'CCA'.

aunch_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

RESULTS – EDA WITH SQL

45596

- total payload mass carried by boosters launched by NASA (CRS)

2534

- average payload mass carried by booster version F9 v1.1

2015-12-22

- date when the first succesful landing outcome in ground pad was acheived

Boster version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

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

IBM Developer

SKILLS NETWORK

RESULTS – EDA WITH SQL

Mission Outcome	
Failure (in flight)	1
Success	99
Success (payload status unclear)	1

Total number of successful and failure mission outcomes

Landing outcome	Booster version	Launch site	Month
Failure (drone ship)	F9 v1.1 B1012	CCAFS LC-40	5
Failure (drone ship)	F9 v1.1 B1015	CCAFS LC-40	5

Booster version and launch site for failure landing outcomes on drone ship in year 2015

Landing Outcome	Total
Success (drone ship)	5
Success (ground pad)	3

Count of successful landing outcomes between dates 04.06.2010. and 20.03.2017.

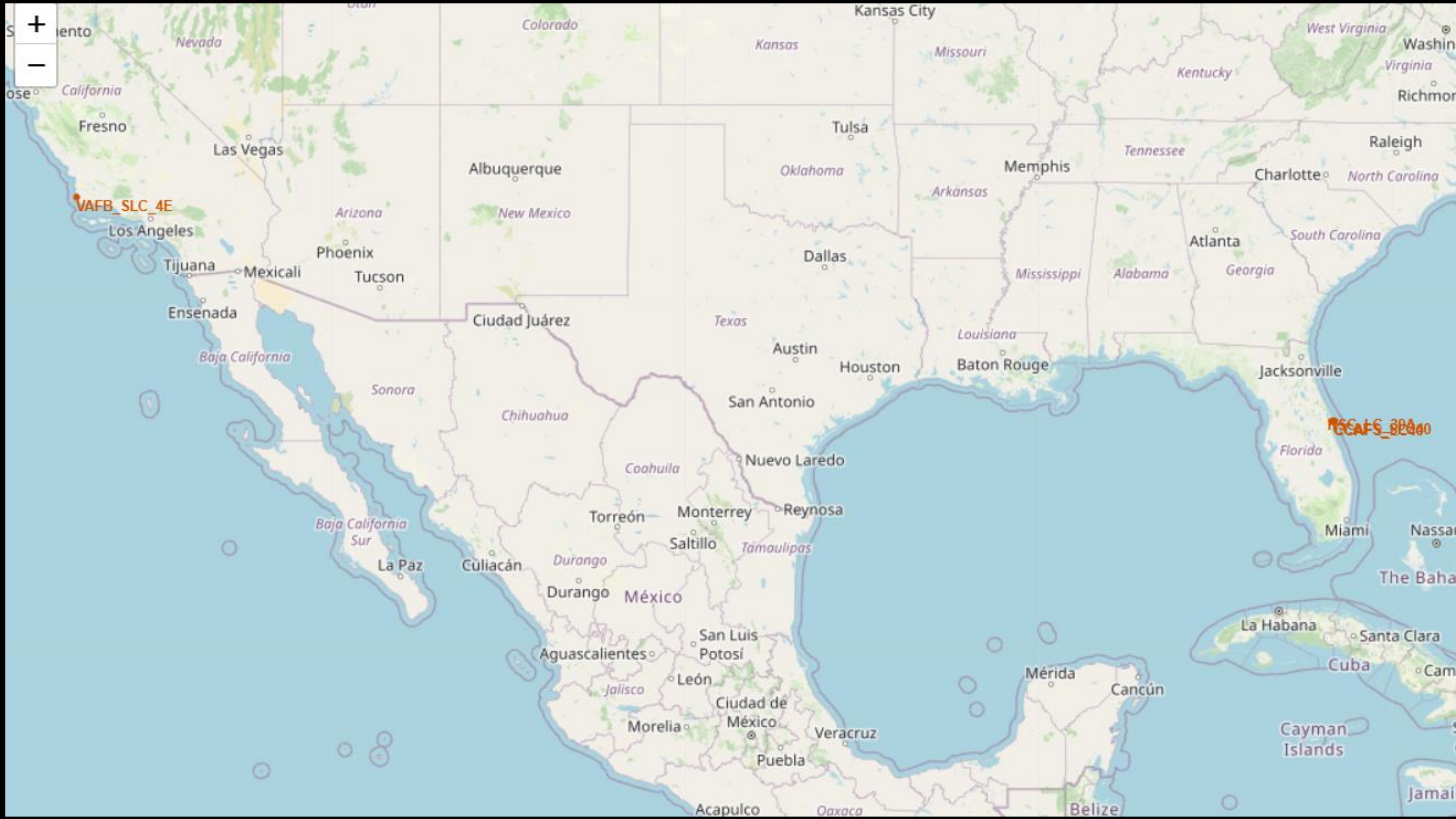
Buster version
F9 B5 B1048.4
F9 B5 B1049.4
F9 B5 B1051.3
F9 B5 B1056.4
F9 B5 B1048.5
F9 B5 B1051.4
F9 B5 B1049.5
F9 B5 B1060.2
F9 B5 B1058.3
F9 B5 B1051.6
F9 B5 B1060.3
F9 B5 B1049.7

Booster_versions which have carried the maximum payload mass

RESULTS – FOLIUM MAPS

Map displays launch site markers with exact locations based on coordinates that we can see in the table. There is some overlap due to spatial proximity of CCAFS LC-40, CCAFS SLC-40 and KSC LC-39A launch sites.

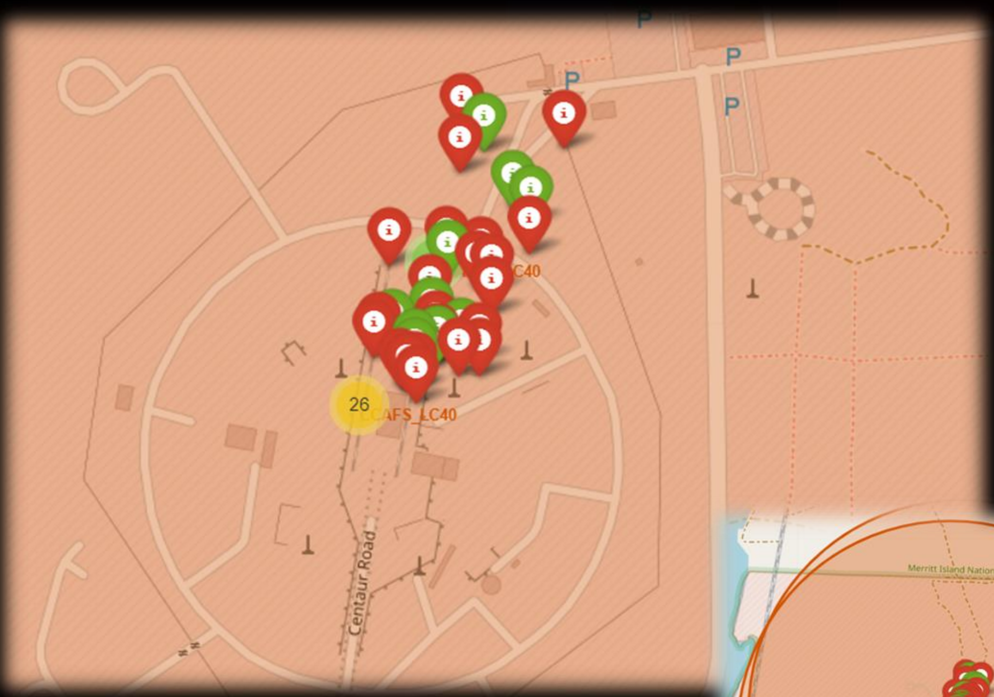
Launch Site	Latitude	Longitude
CCAFS LC-40	28.562302	-80.577356
CCAFS SLC-40	28.563197	-80.576820
KSC LC-39A	28.573255	-80.646895
VAFB SLC-4E	34.632834	-120.610745

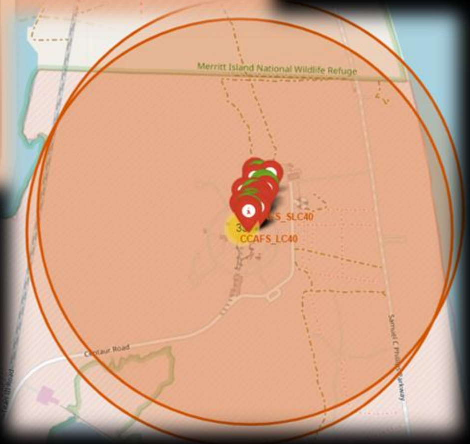


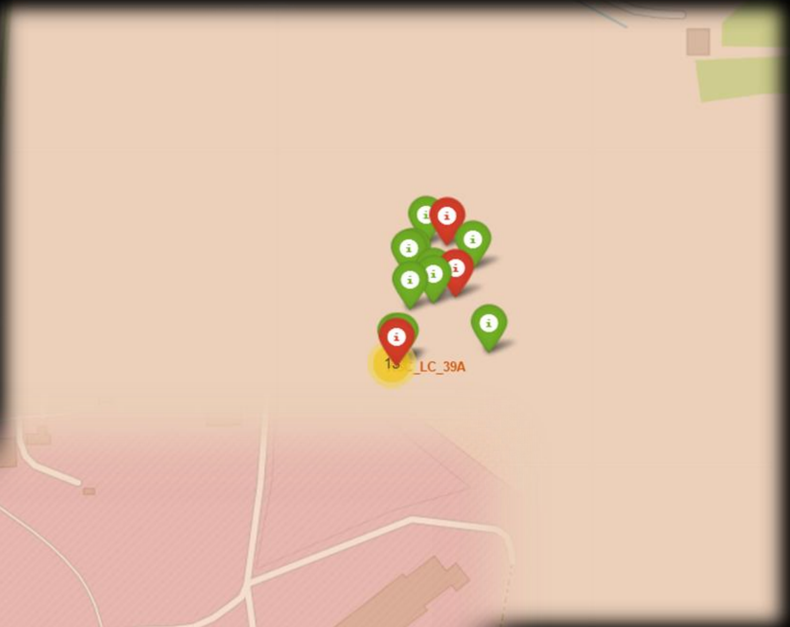
IBM Developer

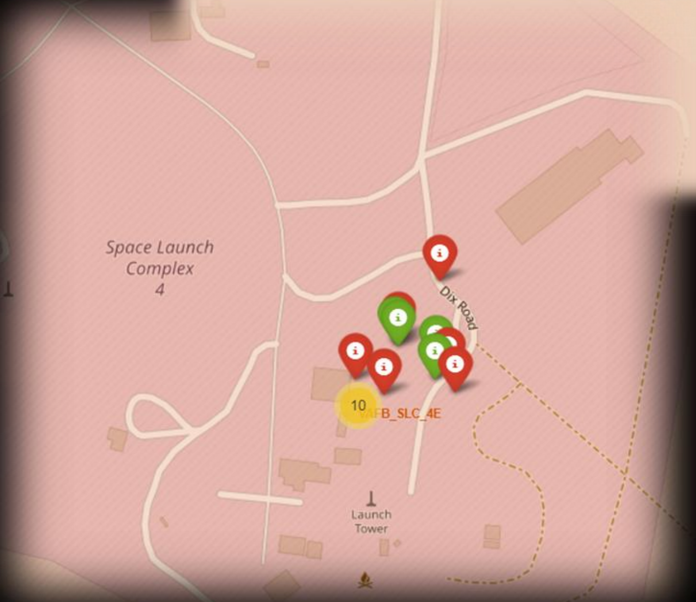
SKILLS NETWORK

RESULTS – FOLIUM MAPS









Jittered markers of successful (green) and unsuccessful (red) landings for CCAFS LC-40 and CCAFS SLC-40 launch sites.

Upper: KSC LC-39A

Left: VAFB SLC-4E

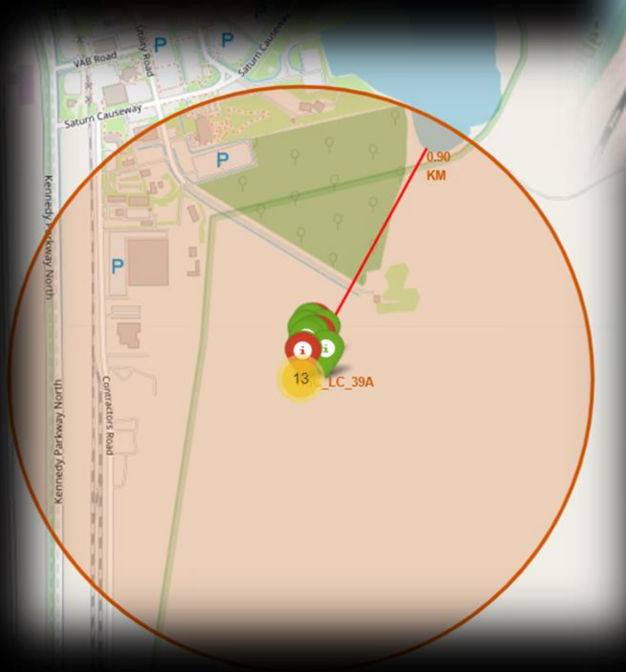
IBM Developer

SKILLS NETWORK

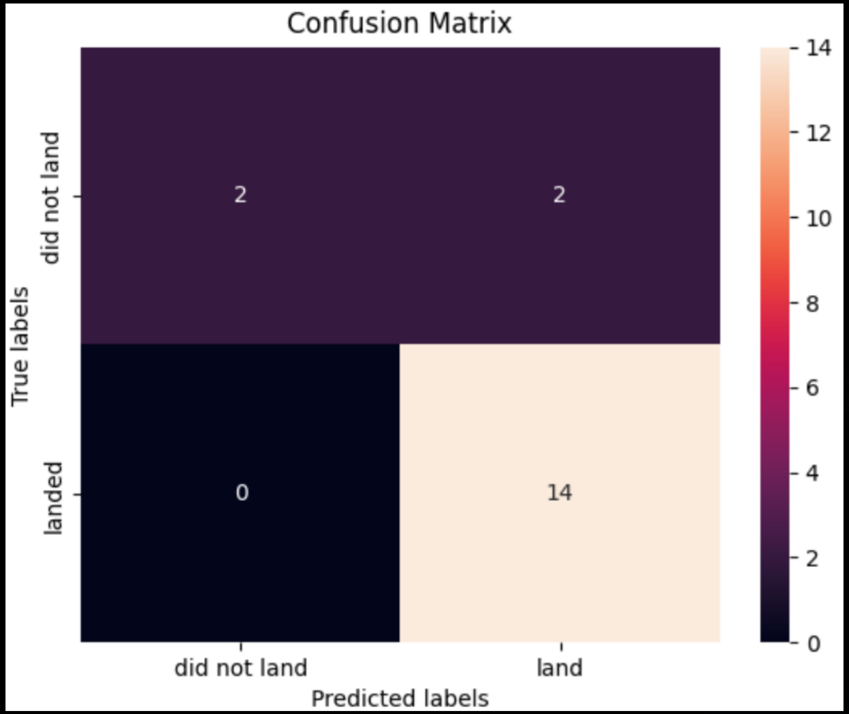
RESULTS – FOLIUM MAPS

Distance from the coast line

Left: CCAFS LC-40 and CCAFS SLC-40
Right: VAFB SLC-4E
Bottom: KSC LC-39A

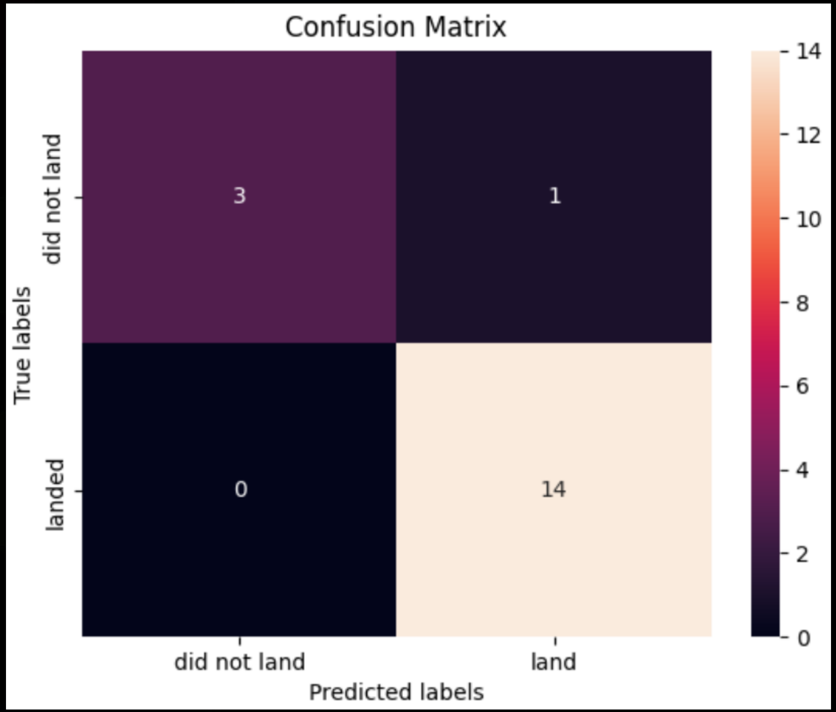


RESULTS – MACHINE LEARNING



Left: logistic regression confusion matrix shows high true positive rate. True negative and false positive are equal, while false negative is zero.

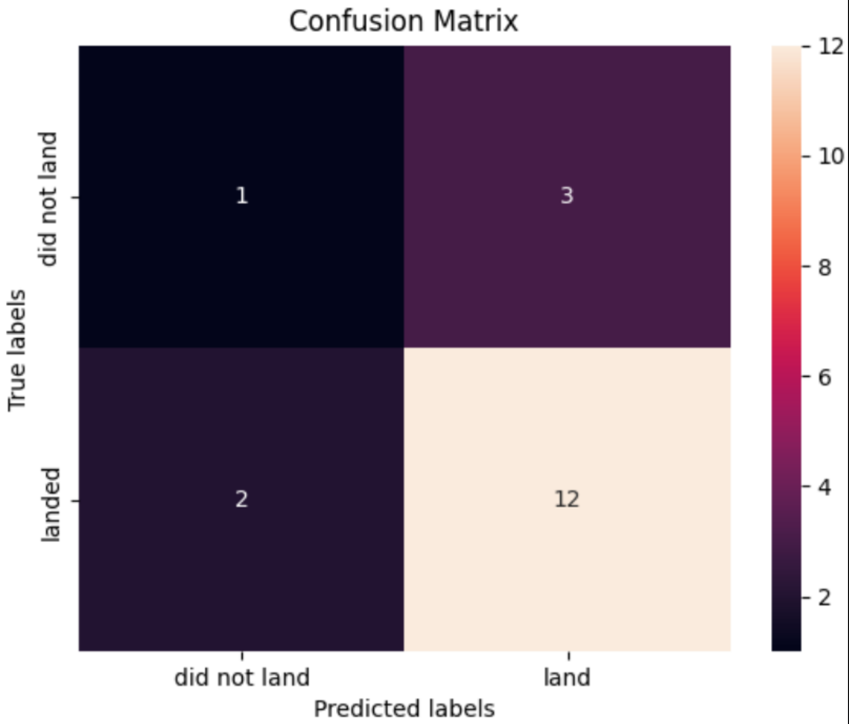
Right: support vector machine also has high number of true positives, but it has one less false positive and one more true negative. False negative is zero.



IBM Developer

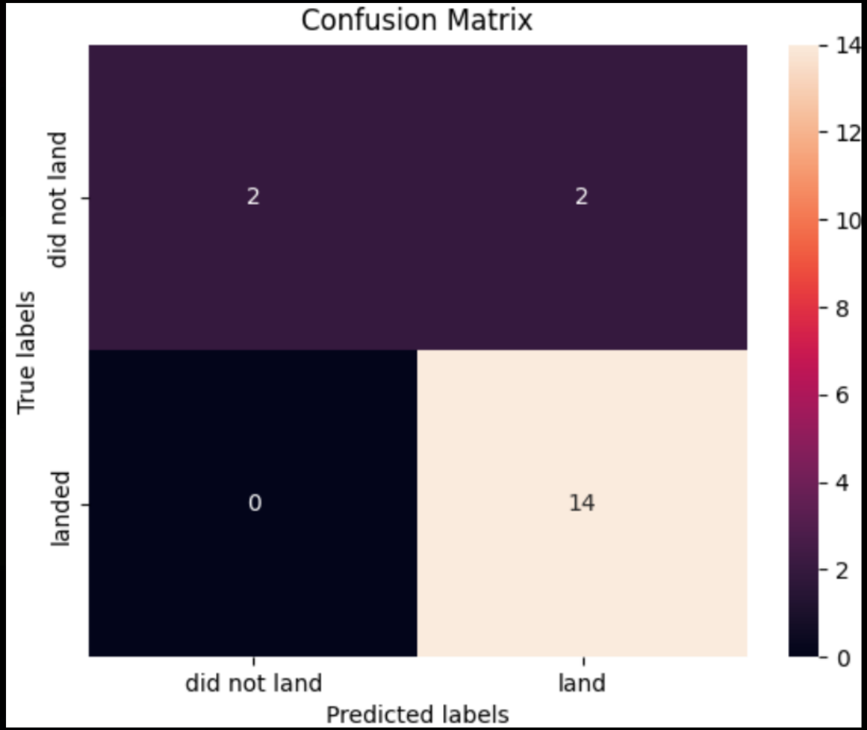
SKILLS NETWORK

RESULTS – MACHINE LEARNING



Left: decision tree classifier confusion matrix shows a bit lower true positive rate than previous models and we can see some false negatives.

Right: k nearest neighbors confusion matrix is the same as for logistic regression.



IBM Developer

SKILLS NETWORK

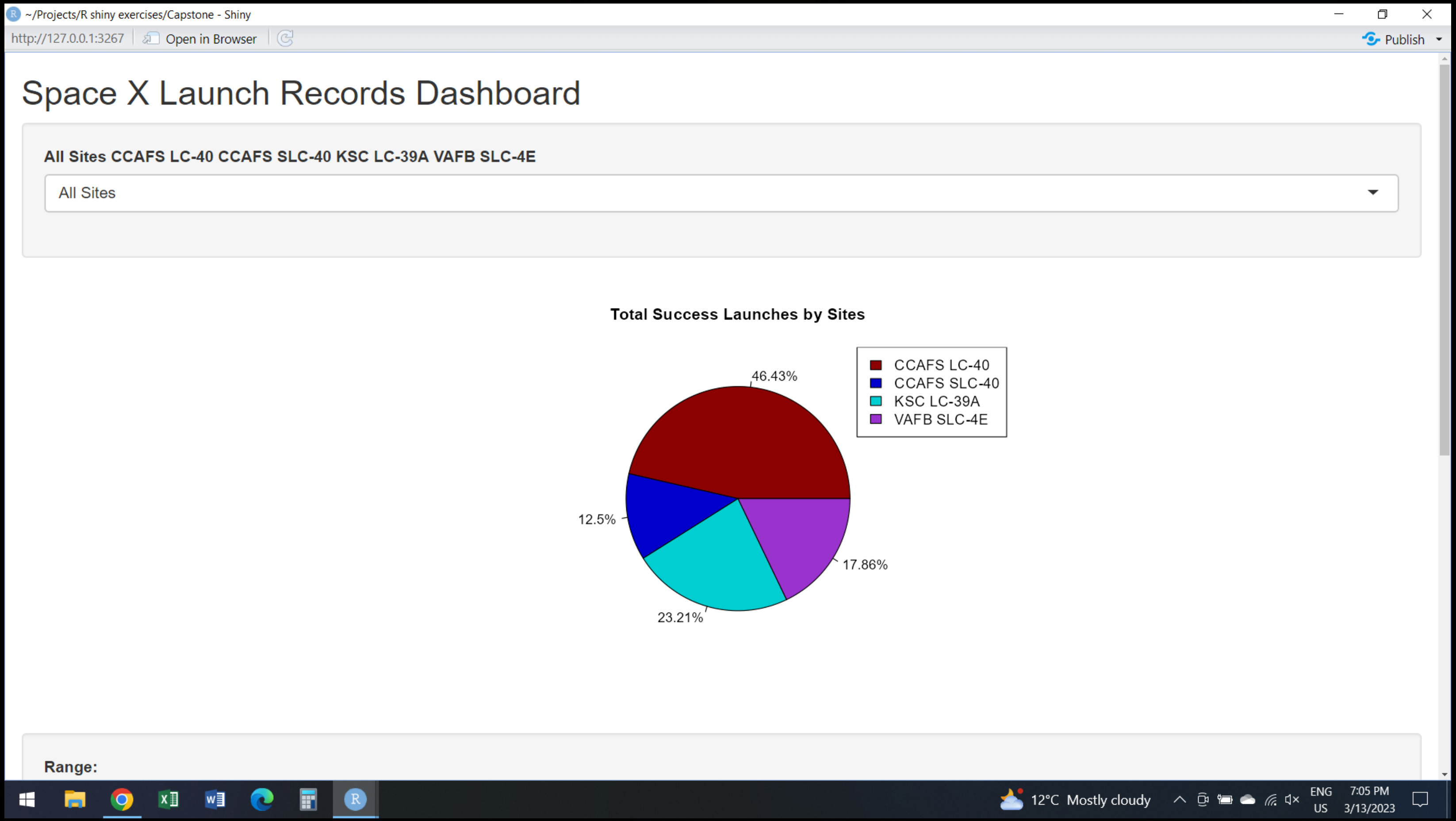
RESULTS – MACHINE LEARNING

Best parameters for different models:

- Logistic regression
  - C: 0.01
  - Penalty: l2
  - Solver: lbfgs
- SVM
  - C: 1.0
  - Gamma: 0.032
  - Kernel: sigmoid
- Decision tree
  - Criterion: gini
  - Max_depth: 10
  - Max_features: auto
  - Min_samples_leaf: 4
  - Min_samples_split: 5
  - Splitter: random
- KNN
  - Algorithm: auto
  - N_neighbors: 10
  - P: 1

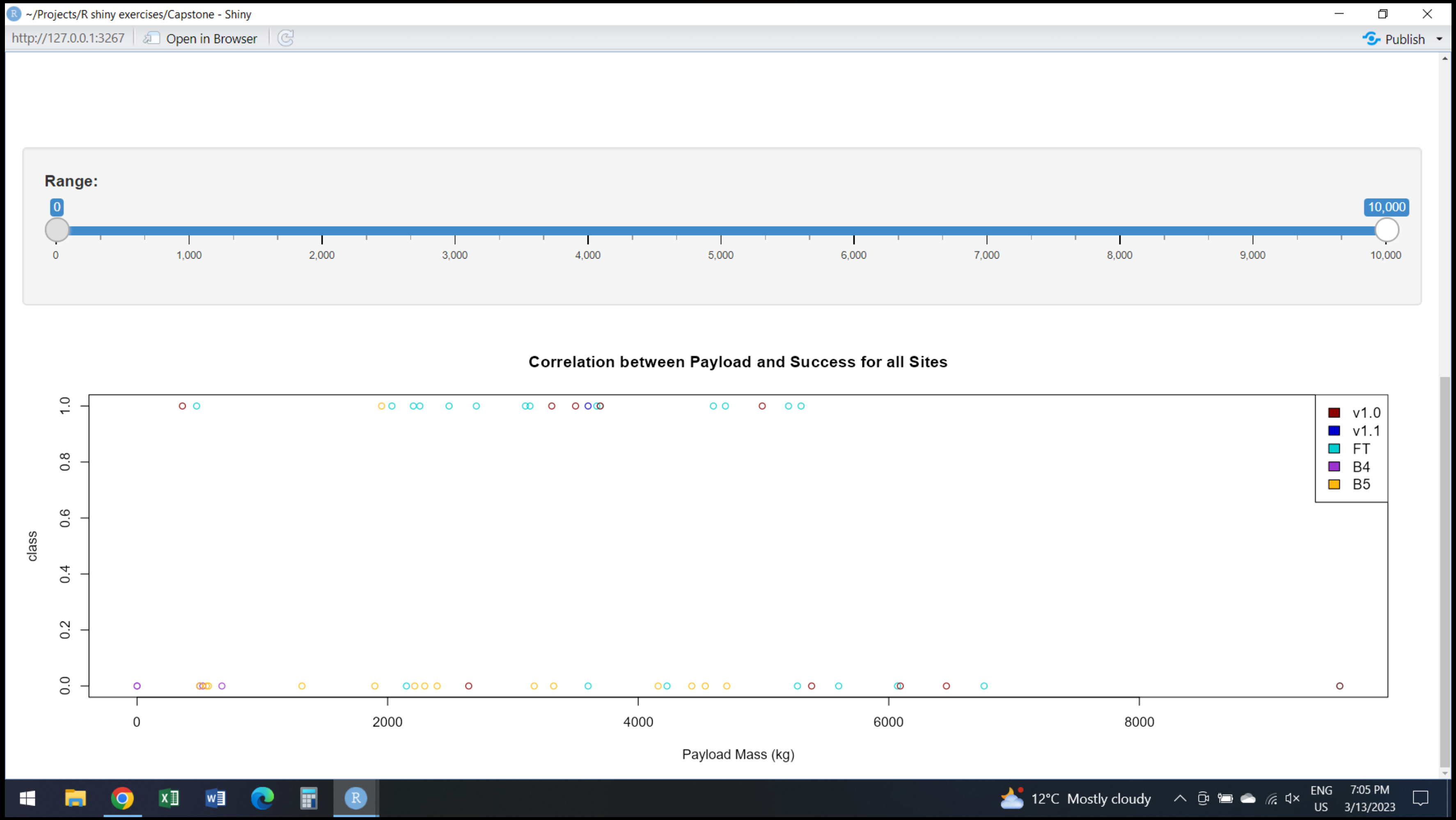
	Model accuracy	Test score
Log. regression	0.805357	0.888889
SVM	0.817857	0.944444
Decision tree	0.887500	0.722222
KNN	0.832143	0.888889

RESULTS – DASHBOARD APP



Dashboard app – pie chart

RESULTS – DASHBOARD APP

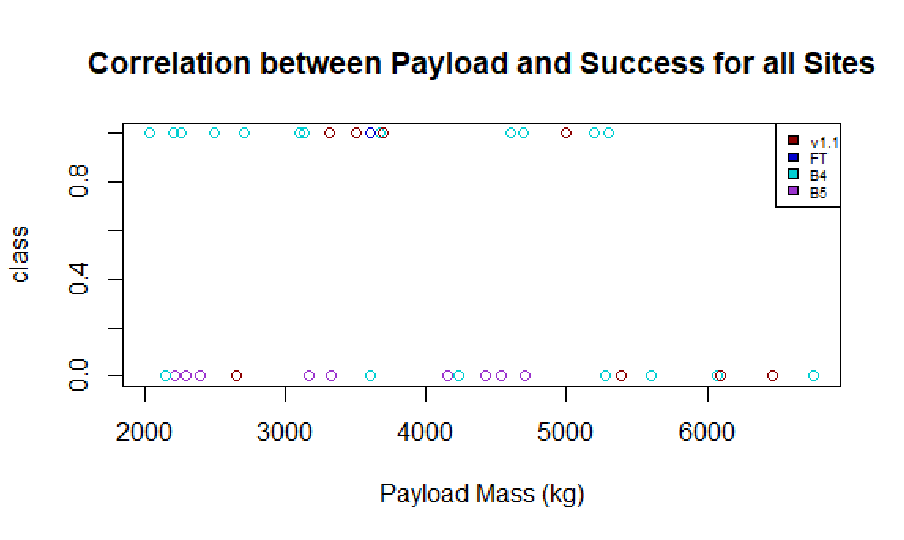


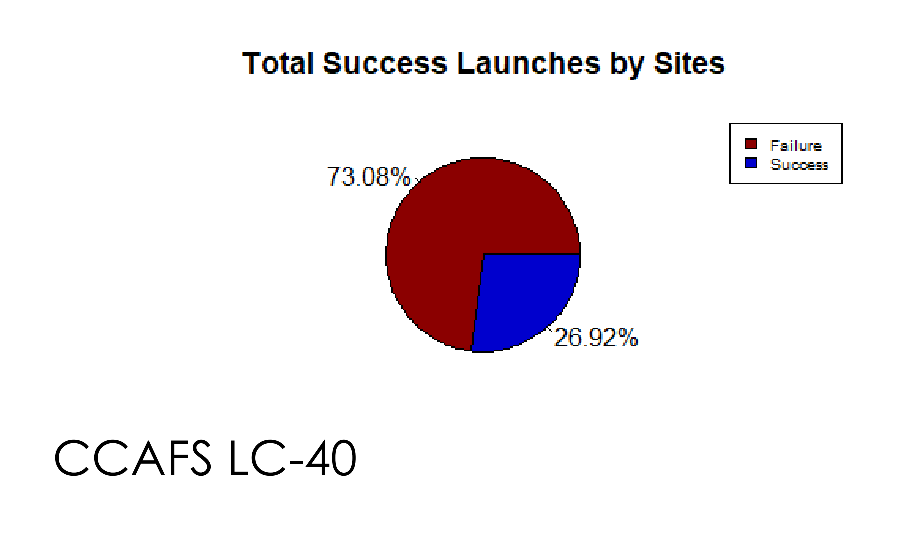
Dashboard app – scatter plot

IBM Developer

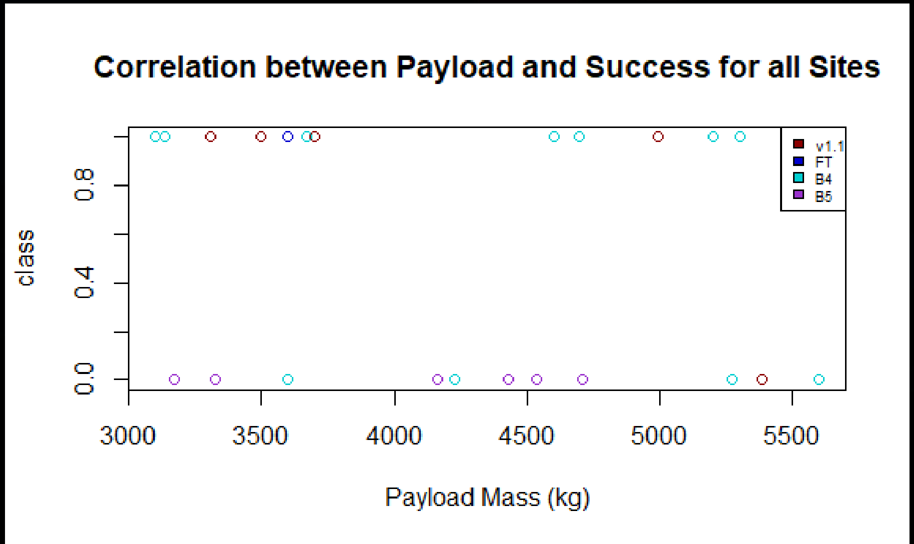
SKILLS NETWORK

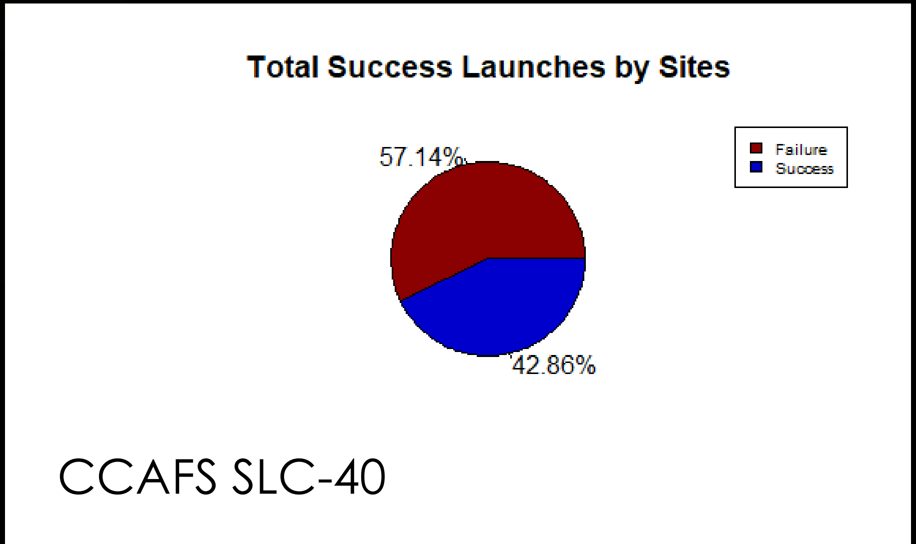
AND SOME MORE EXAMPLES…





* due to technical difficulties dashboard app was built in RStudio





IBM Developer

CONCLUSION

- This project aimed to predict whether Falcon 9 first stage will land successfully.
- From exploratory data analysis we can conclude that data was well recorded with small amount of missing data. Collected data was comprehensive, so there is solid base for quality analysis. At this point, we can conclude that most launches aims at lower orbits.
- Visual analysis showed that newer spacecraft have higher payload mass, and some orbits have a significantly high success rate. Also, we can conclude that more recent flights aim at more distant orbits and that success rate increases with time. At Folium maps, we can see that while some launch sites are geographically close success rate still varies significantly and that all launch sites are close to coastline and roads.
- Data was randomly split into training and test dataset, and different machine learning models were fitted. Decision tree has high model accuracy, but low test score which means that there is overfitting. According to same parameters, support vector machine is the best model for given data.