IBM DATA SCIENCE CAPSTONE PROJECT

Niki Yaw

Sept 2023

TABLE OF CONTENTS

- EXECUTIVE SUMMARY

- INTRODUCTION

- METHODOLOGY

- RESULTS

- CONCLUSION

EXECUTIVE SUMMARY

- **GOAL:** IDENTIFY FACTORS THAT INFLUENCE THE SUCCESS OF FALCON9 ROCKET LANDING

- **METHODOLOGY:** STUDY RELATIONSHIPS BETWEEN VARIABLES THAT CONTRIBUTES TO THE SUCCESS OR FAILURE OF THE ROCKET LANDING

- **RESULTS:**
  - EXPLORATORY DATA ANALYSIS:
    - LANDING SITES WITH HIGHEST SUCCESS RATE -> KSC LC-39A
    - ORBITS WITH 100% SUCCESS RATE -> SSO, HEO, GEO, ES-L1
  - VISUALIZATION: LAUNCH SITES ARE CLOSE TO THE COAST, NEAR THE EQUATOR
  - PREDICTIVE ANALYSIS: DECISION TREE MODEL HAS THE HIGHEST ACCURACY

INTRODUCTION

- SpaceX advertises Falcon 9 rocket launches with a cost of 62 million dollars where other providers cost upwards of 165 million dollars each, mostly because SpaceX can reuse the first stage.

- But there are times where the first stage does not land, or it will crash mid-flight, or even being sacrificed due to the mission parameters like payload, orbit and customer.

- In this project, we will determine the price of each launch by gathering information and training a machine learning model to predict if SpaceX will be able to reuse the first stage.

METHODOLOGY

- DATA COLLECTION

- DATA WRANGLING

- EXPLORATORY ANALYSIS (EDA) WITH PANDAS, MATPLOTLIB & SQL

- INTERACTIVE VISUAL ANALYTICS (FOLIUM) & DASHBOARD (PLOTLY DASH)

- CLASSIFICATION PREDICTIVE ANALYSIS

DATA COLLECTION & DATA WRANGLING

- **DATA COLLECTION**
  - SPACEX REST API – ROCKET LAUNCH DATA (KEEP ONLY FALCON 9 DATA)
  - WIKIPEDIA WITH BEAUTIFULSOUP – FALCON 9 LAUNCH DATA

- **DATA WRANGLING**
  - UNDERSTAND THE DATASET
  - FOCUS ON LANDING OUTCOMES

EDA & INTERACTIVE VISUAL ANALYTICS

- **EXPLORATORY DATA ANALYSIS**
  - SQL – QUERIES TO LEARN AND UNDERSTAND THE DATASET
  - PANDAS + MATPLOTLIB – MULTIPLE VISUALIZATIONS TO STUDY RELATIONSHIPS

- **INTERACTIVE VISUAL ANALYTICS**
  - FOLIUM – GEOSPATIAL MAPS TO LOCATE DIFFERENT LAUNCH SITES
  - PLOTLY DASH – INTERACTIVE WEB APPLICATION TO PRESENT FINDINGS
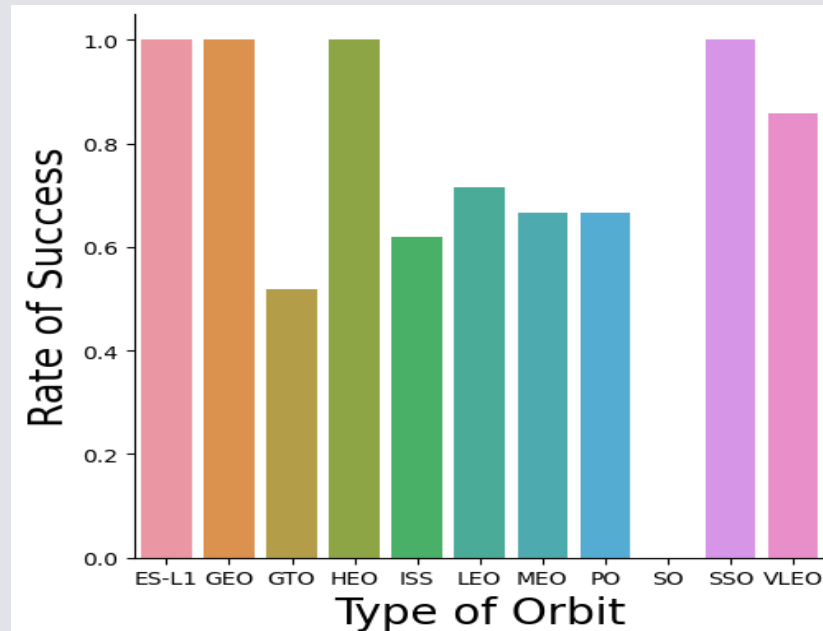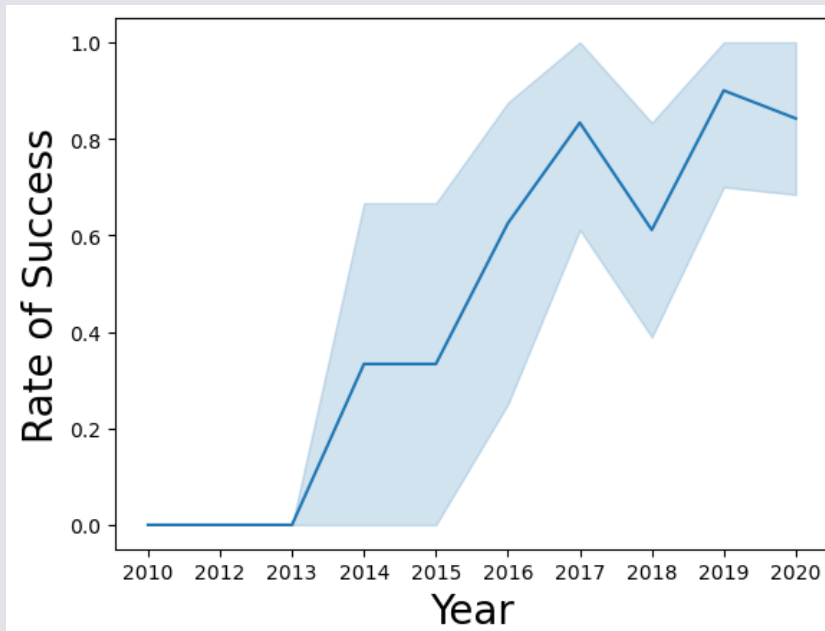
PREDICTIVE ANALYSIS

- MACHINE LEARNING MODELS WITH MULTIPLE CLASSFICATION METHODS
  - LOGISTIC REGRESSION
  - SUPPORT VECTOR MACHINE
  - DECISION TREE CLASSIFIER
  - K NEAREST NEIGHBOURS
- ACCURACY SCORE, JACCARD SCORE AND F1 SCORE FOR ALL

RESULTS - EDA WITH VISUALIZATION

2 IMPORTANT VISUALIZATIONS (6 IN TOTAL):

- LAUNCH SUCCESS YEARLY TREND

- SUCCESS RATE VS ORBIT TYPE

SCATTER PLOTS:
- FLIGHT NUMBER VS LAUNCH SITE
- PAYLOAD VS LAUNCH SITE
- FLIGHT NUMBER VS ORBIT TYPE
- PAYLOAD MASS VS ORBIT TYPE

RESULTS – EDA WITH SQL

3 IMPORTANT FINDINGS (10 QUERIES IN TOTAL):

- NAMES OF LAUNCH SITES

- TOTAL NUMBER OF SUCCESSFUL AND FAILURE MISSION OUTCOMES

- DATES FOR SUCCESSFUL LANDING OUTCOMES IN GROUND PAD

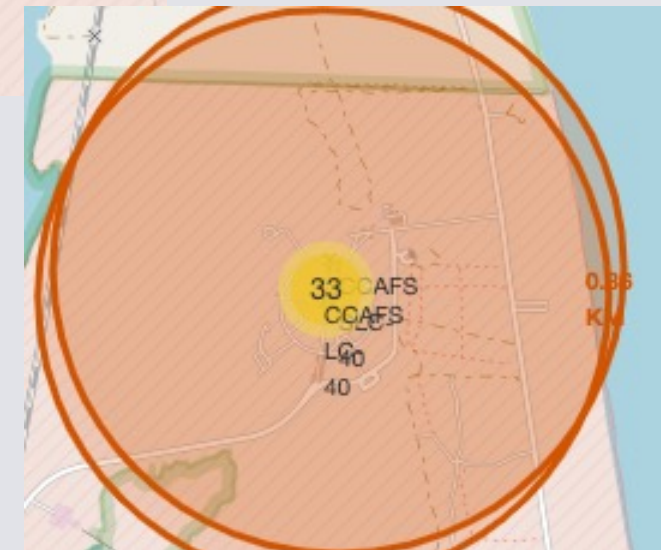| Launch_Site |
| --- |
| CCAFS LC-40 |
| VAFB SLC-4E |
| KSC LC-39A |
| CCAFS SLC-40 |
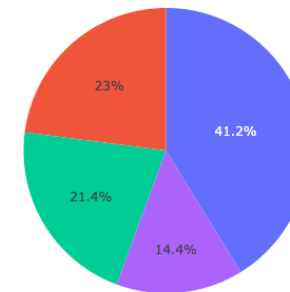
Mission_Outcome	COUNT(*)
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Date
2015-12-22
2016-07-18
2017-02-19
2017-01-05
2017-03-06
2017-08-14
2017-07-09
2017-12-15
2018-08-01

RESULTS – INTERACTIVE MAPS WITH FOLIUM

3 VISUALIZATIONS:

- MARKERS FOR LAUNCH SITES
- COLORED MARKERS FOR LAUNCH OUTCOMES
- DISTANCE FROM LAUNCH SITES

RESULTS - DASHBOARD WITH PLOTLY DASH
WEB-BASED APPLICATION DASHBOARD WITH DROPDOWN OPTIONS & RANGE SLIDER
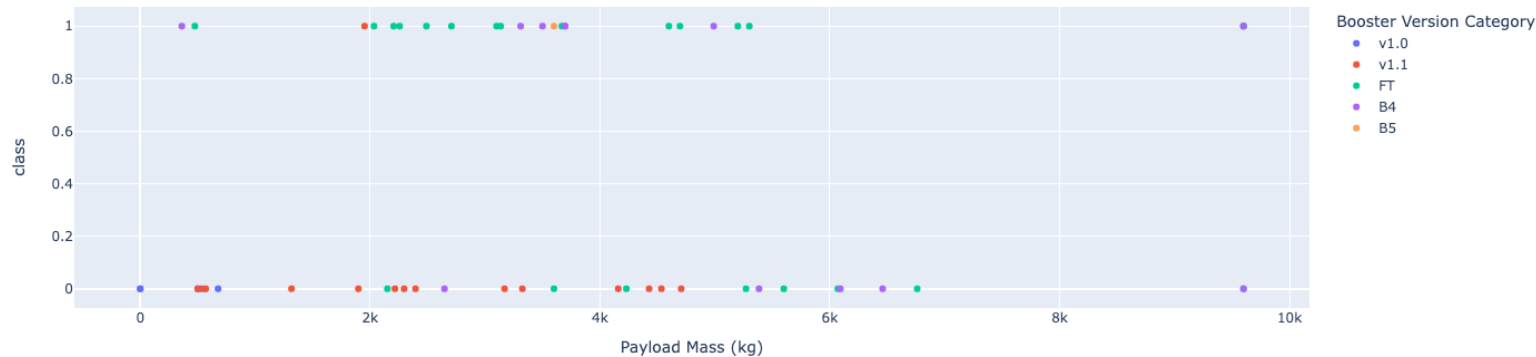
RESULTS – PREDICTIVE ANALYSIS

- LOGISTIC REGRESSION
  - BEST SCORE: 84.64%

- SUPPORT VECTOR MACHINE
  - BEST SCORE: 84.82%

- DECISION TREE ***

```
The best model is DecisionTree with an accuracy score of 0.9017857142857144
Best parameters are : {'criterion': 'gini', 'max_depth': 4, 'max_features': 'sqrt', 'min_
samples_leaf': 4, 'min_samples_split': 2, 'splitter': 'random'}
```

  - **BEST SCORE: 90.18%**

- K NEAREST NEIGHBOUR
  - BEST SCORE: 84.82%

CONCLUSION

IMPORTANT FINDINGS

- **OVER THE YEARS:** SUCCESS RATE HAS BEEN INCREASING STEADILY

- **TYPES OF ORBIT:** GEO, SSO, HEO AND ES-L1 ALL HAVE A SUCCESS RATE OF 100%

- **PAYLOAD MASS:** THE HIGHER THE PAYLOAD MASS, THE HIGHER THE SUCCESS RATE

- **LOCATION:** ALL LAUNCH SITES ARE NEAR COASTAL AREAS

- **MACHINE LEARNING ALGORITHM:** DECISION TREE SEEMS TO OUTPERFORM THE REST BY JUST SLIGHTLY