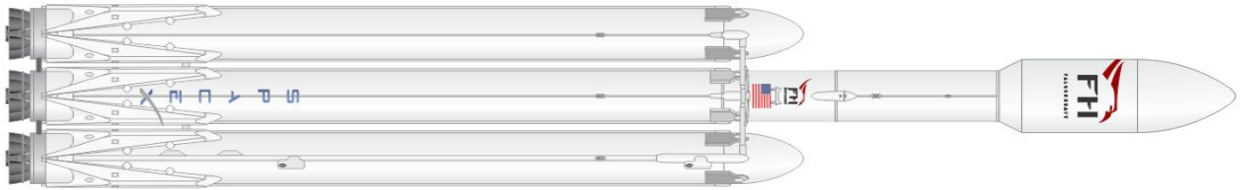

# SpaceX Falcon 9 landing prediction

An analysis of the Falcon 9's landing history and potential future

GitHub:
https://github.com/ifnamartin/Falcon_9_projections.git

# Executive Summary

The main focus of this project is the study of the success rate for a SpaceX's partially reusable launcher, the Falcon 9. One of the main reasons this rocket is so well-known is the fact that it is partially reusable, which makes the cost of each launch reduce considerably. For the study, we will take the published data of the launches of the rocket, which includes orbit type, launch site, etc. With this data, our main objective will be to determine how the likelihood of the rocket landing successfully is related to the different variables and circumstances of the launch.



# Table of contents

Introduction

Methodology

Results and discussion

Conclusion

Appendix

# Introduction

**WHY**

The reason why I felt the need to carry this analysis is the importance of the rocket landing correctly. SpaceX is now one of the leaders in terms of space expeditions, and one of the reasons for it is the not-seen-before affordability they offer.

**HOW**

The database containing launch details of the Falcon 9 has multiple variables, such as the landing pad, orbit, etc. Here we will study correlation between them and draw potential conclusions about landing prediction.

**ANALYSIS**

- How can we treat the data in order to carry out the research in the best way possible?
- Which variables show apparent relation to the success rate?
- What conclusions can we draw from data visualization?
- What machine learning algorithm will give the best response to our problem?

# Methodology

1. Collecting the data from SpaceX
2. Webscraping Falcon 9 data from Wikipedia
3. Data Wrangling
4. Exploratory Data Analysis
5. Data Visualization
6. Machine Learning Prediction

# Main Results

# Data Collection

The data was obtained in two different ways:

1. **Request to the SpaceX API**
   Data obtained directly from the SpaceX database and then filtered to only get the relevant information

2. **Webscraping from Wikipedia data**
   Obtained the relevant information found in the Falcon 9 Wikipedia by scraping

# Data wrangling

Once we have obtained the data and selected the information we will consider for the study, it is time to manipulate the data for future handling and visualization. The main tunings made were as follows

1. **Handling missing values**
2. **Creating an outcome column with a boolean representing landing success (1) or failure (0)**
3. **Giving the boolean variables values of (0 or 1 instead of False or True)**
4. **One-hot encoding categorical values**

| | FlightNumber | Date | BoosterVersion | PayloadMass | Orbit | LaunchSite | Outcome | Flights | GridFins | Reused | Legs | LandingPad | Block | ReusedCount | Serial | Longitude | Latitude |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 2010-06-04 | Falcon 9 | 6104.959412 | LEO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0003 | -80.577366 | 28.561857 |
| 1 | 2 | 2012-05-22 | Falcon 9 | 525.000000 | LEO | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0005 | -80.577366 | 28.561857 |
| 2 | 3 | 2013-03-01 | Falcon 9 | 677.000000 | ISS | CCAFS SLC 40 | None None | 1 | False | False | False | NaN | 1.0 | 0 | B0007 | -80.577366 | 28.561857 |

| | FlightNumber | PayloadMass | Flights | GridFins | Reused | Legs | Block | ReusedCount | ES-L1 | GEO | ... | B1048 | B1049 | B1050 | B1051 | B1054 | B1056 | B1058 | B1059 | B1060 | B1062 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 6104.959412 | 1 | False | False | False | 1.0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 2 | 525.000000 | 1 | False | False | False | 1.0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 3 | 677.000000 | 1 | False | False | False | 1.0 | 0 | 0 | 0 | ... | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

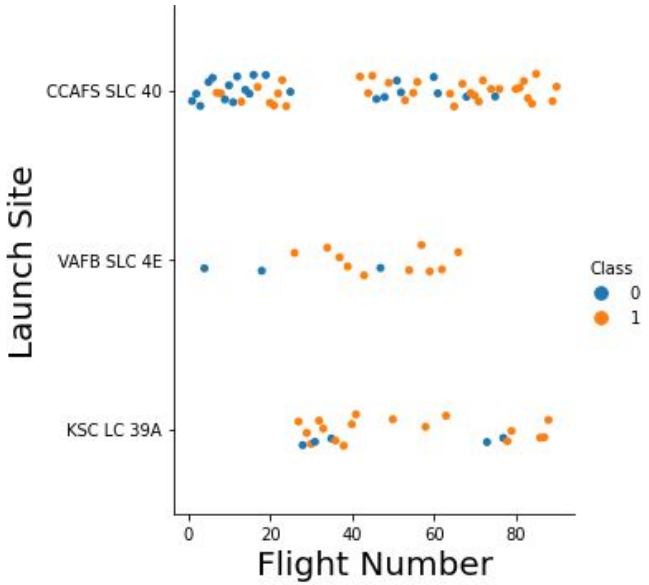
# Data Analysis

Now the next natural step to take is to explore this data, studying how many different values variables take, possible relation between this values (e.g. launch site vs rate) that may help draw some conclusions.

As an example, we can study the launch sites in relation to the flight number, and we can conclude that the launch site VAFB is not being used recently, or that the success rate for KSC seems to be higher than the CCAFS

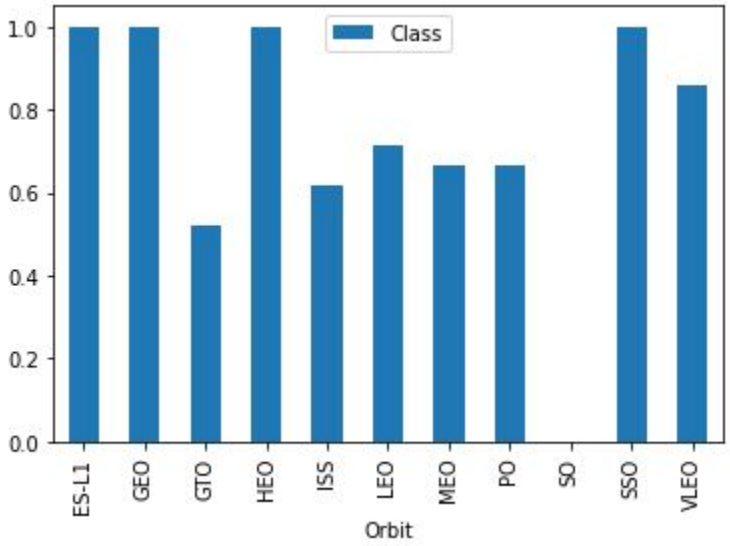
Different launch sites

| | |
|---|---|
| CCAFS SLC 40 | 55 |
| KSC LC 39A | 22 |
| VAFB SLC 4E | 13 |

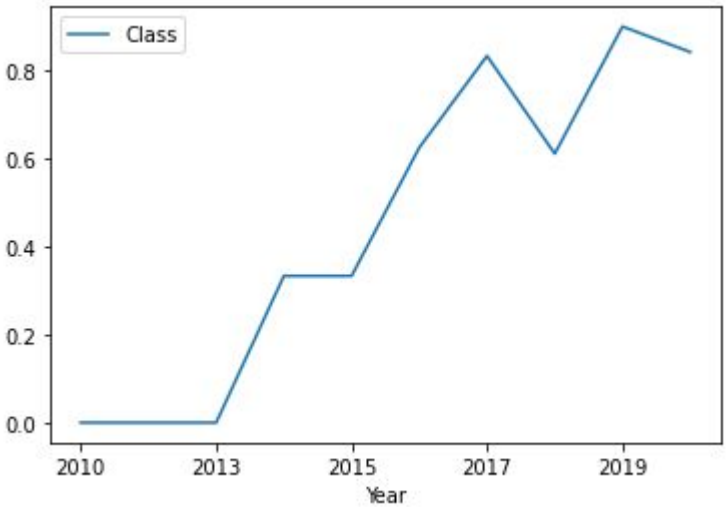


Another aspect which might be interesting to study  is the relationship between the orbit type and the success rate for the landings. Here we can appreciate that, for example, all five with orbit type SSO have landed successfully, in contrast with the GTO, that appears to land successfully only half the times.

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



Also, something that I was able to see from the start, thanks to graphs such as Flight no. vs Launch Site figure is that the probability of success has increased notably from launch to launch. This made me study the yearly trend of the success rate, and you can clearly see that the success rate generally increases through the years, excepting 2018.



# Models and prediction

To try to predict the landing outcome for launches, we will study the accuracy and scorings of the following machine learning algorithms

01 | K - Nearest Neighbors

02 | Decision Tree

03 | Support Vector Machines

04 | Logistic Regression

We obtain really similar scores for all of them, as we can see in the figure. However, based on the In Sample accuracy, we can presume that **Decision Tree** is the algorithm that shows best performance.

| Algorithm | IS acc. | OOS acc. | Jaccard | F1-score | LogLoss |
|---|---|---|---|---|---|
| KNN | 0.860952 | 0.833333 | 0.800000 | 0.814815 | NA |
| Decision Tree | 0.875238 | 0.833333 | 0.800000 | 0.814815 | NA |
| SVM | 0.834286 | 0.833333 | 0.800000 | 0.814815 | NA |
| LogisticRegression | 0.834286 | 0.833333 | 0.800000 | 0.814815 | 5.756596 |

# Conclusions

- The Falcon 9 landing success probability has grown notably throughout the years
- The selection of the Launch Site is crucial for the success of the landing
- Depending on what kind of orbit the rocket is going to follow, the probability of landing successfully varies notably
- Whereas all of the models used to predict the probability of landing successfully for a certain launch gave good results, the Decision Tree algorithm appears to be the best option


# Thanks for reading

GitHub:
https://github.com/ifnamartin/Falcon_9_projections.git