SPACEX Race With DATA Science

Linyuan Zou

2024/05/02

Outline

Executive Summary

+ Summary of methodologies

+ The research has the objective to identified the success or failure factors of the last rocket landing from Space X. The methodologies where used is the following:

+ •Collect data using SpaceX REST API and web scraping techniques.

+ •Wrangle data to create success/fail outcome variable.

+ •Explore data with data visualization techniques.

+ •Analyze the data with SQL, calculating important metrics like total payload, payload range for successful launches, etc.

+ •Explore launch site success rates and proximity.

+ •Visualize the launch sites with the most success and successful payload.

+ •Build Models to predict landing outcomes using logistic regression, support verctor machine (SVM), decision tree and K-nearest neightboar (KNN)

+ Summary of all results

+ •Launch success has improved over time

+ •KSC LC-39A has the highest success rate among landing sites

+ •Orbits ES-L1, GEO, HEO and SSO have a 100% success rate

+ •All models performed similarly on the test set.

Introduction

+ **Background** SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars; other providers cost upward of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. Therefore if we can determine if the first stage will land, we can determine the cost of a launch. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch **Explore**

+ •How payload mass, launch site, number of flights, and orbits affect first-stage landing success

+ •Rate of successful landing over time

+ •Best predictive model for successful landing (binary classification)

Methodology