



WINNING SPACE RACE WITH DATA SCIENCE

Mr. K. Rupendra Singh¹, Reshma Doddi², Nalla Niharika³, Tejaswini Sai Srija Malla⁴, Yeddu Chinni⁵
1,2,3,4,5 Department of Computer Science and Engineering, Vignan's Institute of Engineering for
Women, Visakhapatnam, Andhra Pradesh, India

ABSTRACT

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, it is determined that the first stage will land, cost of a launch can be determined. This information can be used if an alternate company wants to bid against SpaceX for a rocket launch. By using data science methodologies to define and formulate a real-world business problem data analysis tool is used to load a dataset, clean it, and find out interesting insights from it. An API to extract information from a web service is used. To guide the modeling process data is visualized as scatter plots and bar charts and meaningful patterns are extracted. Launch site proximity is analyzed to build an interactive map with folium to calculate distances. Machine learning predictive analysis is used to determine if the first stage of Falcon 9 will land successfully. Data is split into training data and test data to find the best Hyperparameter for SVM, Classification Trees, and Logistic Regression. Comparative analysis is performed using the algorithm and the one with the best accuracy is implemented on the real-time data.

Keywords: Web Scraping, Data Wrangling, Exploratory Data Analysis, Predictive Analysis.

INTRODUCTION

This capstone will predict if the Falcon 9 first stage will land successfully. 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. In this project, I am a data scientist working for a new rocket company named Space Y that would like to compete with SpaceX. My job is to determine the price of each launch by gathering information about SpaceX and creating dashboards. I also determine if SpaceX will reuse the first stage by training a machine learning model and use public information to predict if SpaceX will reuse the first stage.

LITERATURE SURVEY

A literature survey is focused on utilizing data science methodologies to address a real-world business problem in the space launch industry. It begins by highlighting the cost savings achieved by SpaceX through the reuse of the Falcon 9 rocket's first stage, compared to other providers. The central objective is to predict the successful landing of the first stage, which directly impacts launch costs and competitiveness. The methodology involves leveraging data analysis tools, such as Python's Pandas library, to load, clean, and analyze datasets. Additionally, an API is utilized to extract data from web services, and SQL queries are employed for data manipulation within a database. Exploratory data analysis techniques, including visualization through scatter plots and bar charts, are employed to identify meaningful patterns. Furthermore, the study incorporates geographical analysis by using folium to calculate distances and build interactive maps of launch site proximity. Machine learning



techniques, including Support Vector Machines (SVM), Classification Trees, and Logistic Regression, are applied for predictive analysis. The dataset is split into training and test sets to optimize hyperparameters and select the best-performing model.

EXISTING SYSTEM

The commercial space age is here, companies are making space travel affordable for everyone. Perhaps the most successful is SpaceX.

SpaceX's accomplishments include:

- Sending manned missions to Space.
- One reason SpaceX can do this is the rocket launches are relatively inexpensive. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars.
- Other providers cost upwards of 165 million dollars each, much of the savings is because SpaceX can reuse the first stage. SpaceX's Falcon 9 launch like regular rockets.
- Unlike other rocket providers, SpaceX's Falcon 9 can recover the first stage. Sometimes the first stage land successfully. Sometimes it will crash. Therefore, if we can analyse it, we can provide the analysis to another company that would like to compete with SpaceX.

DISADVANTAGES

1. Data bias and inaccuracies: Data science projects rely heavily on data, and if the data used is biased or inaccurate, it can lead to flawed results and decisions.
2. Dependency on technology: Relying too heavily on data science projects may lead to a dependency on technology, which could be problematic if the technology fails or malfunctions.
3. Ethical concerns: Data science projects may raise ethical concerns related to privacy, data security, and the potential misuse of information collected during space exploration.
4. Environmental impact: Data science projects may have environmental impacts, such as increased energy consumption for data processing and storage.

PROPOSED SCHEME

Data collection methodology:

Data was collected using SpaceX API and web scraping from Wikipedia.

Perform data wrangling:

One-hot encoding data fields for Machine Learning and data cleaning of null values and irrelevant columns.

Perform exploratory data analysis (EDA) using visualization and SQL. Perform interactive visual analytics using Folium and Plotly Dash.

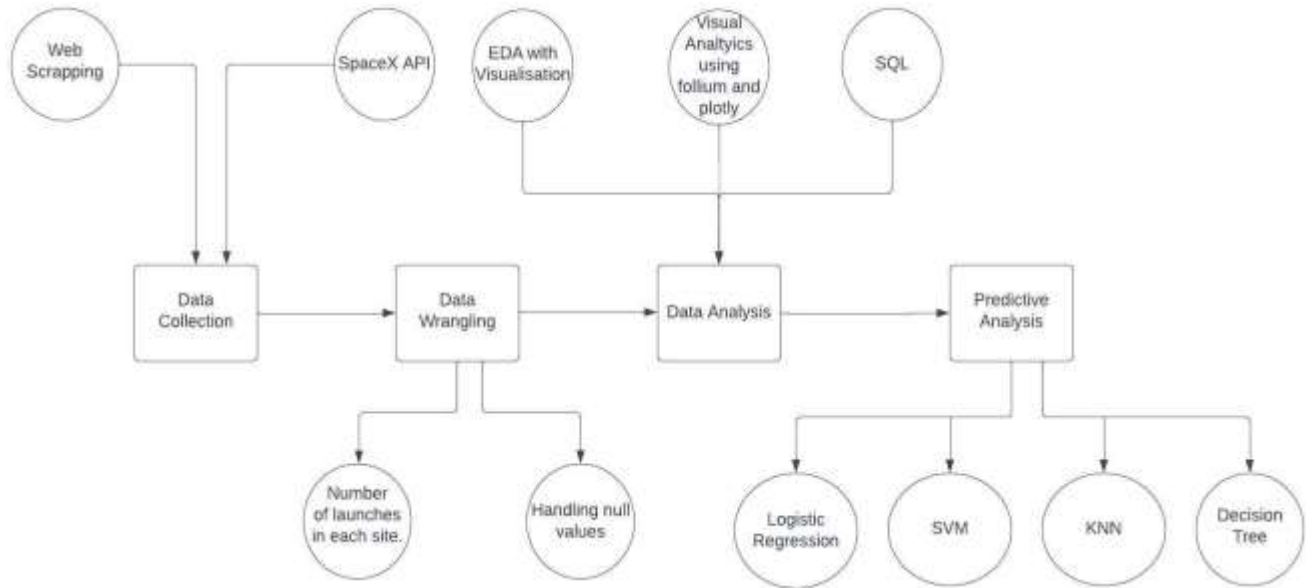
Perform predictive analysis using classification models.

LR, KNN, SVM, DT models have been built and evaluated for the best classifier.

ADVANTAGES

The Falcon 9 is a reusable, two-stage rocket designed and manufactured by SpaceX for safely transporting people and payloads into Earth orbit and beyond. As the world's first orbital class reusable rocket, this rocket can rely its most expensive parts, which drives down the cost of space access.

BLOCK DIAGRAM



OUTPUT SCREENS

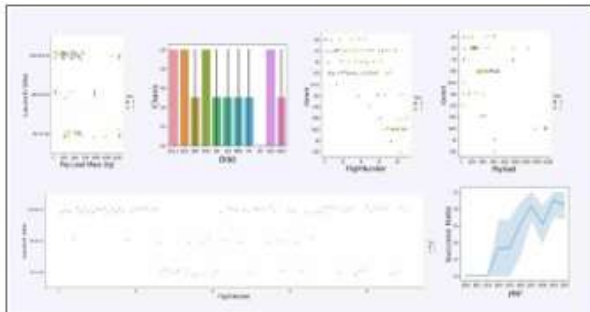


Fig1: EDA with Data Visualization



Fig2: Build an Interactive Map with Folium



Fig3: Markers showing launch sites with color labels



Fig4: Launch site distance to landmarks



Fig5: Build a Dashboard with Plotly Dash

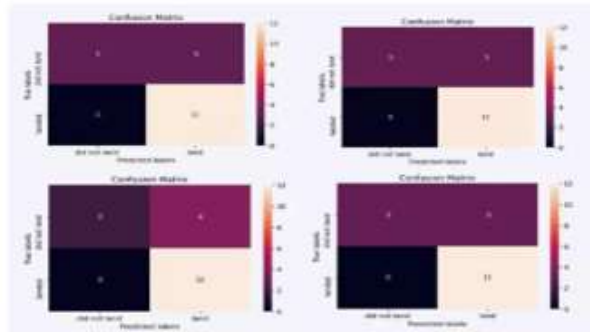
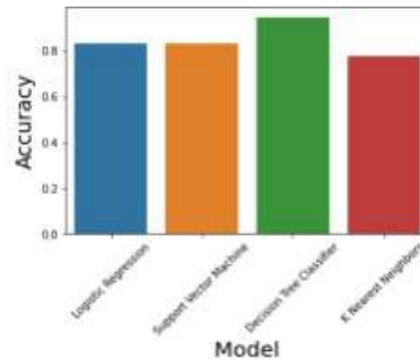
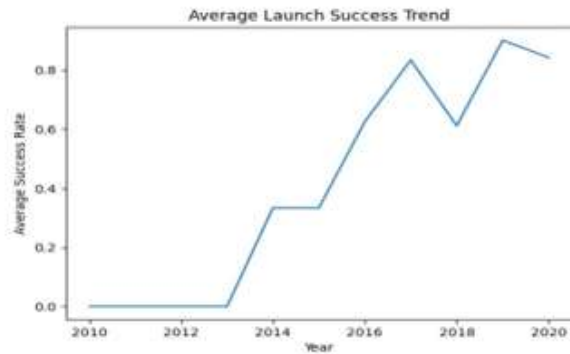


Fig6: Predictive Analysis



CONCLUSION

- The larger the flight amount at a launch site, the greater the success rate at a launch site.
- Launch success rate started to increase in 2013 till 2020.
- Low weighted payloads perform better than the heavier payloads.
- Orbits ES-L1, GEO, HEO, SSO had the most success rate.
- KSC LC-39A had the most successful launches of any sites.
- The Decision tree classifier is the best machine learning algorithm for this task.

REFERENCES

- 1.SpaceX API (JSON):https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/API_call_spacex_api.json
- 2.Wikipedia(Webpage):https://en.wikipedia.org/w/index.php?title=List_of_Falcon_9_and_Falcon_Heavy_launches&oldid=1027686922
3. SpaceX(CSV):https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321ENSkillsNetwork/labs/module_2/data/Spacex.csv?utm_medium=Exinfluencer&utm_source=Exinfluencer&utm_content=000026UJ&utm_term=10006555&utm_id=NA-SkillsNetworkChannel-SkillsNetworkCoursesIBMDS0321ENSkillsNetwork26802033-2022-01-01
4. Launch Geo (CSV):https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_geo.csv
5. Launch Dash (CSV):https://cf-courses-data.s3.us.cloud-object-storage.appdomain.cloud/IBM-DS0321EN-SkillsNetwork/datasets/spacex_launch_dash.csv