Using Data Science to Predict SpaceX Launch Outcomes

How predictive analytics is shaping the future of spaceflight

Denis Ibragimov

March 2025

Agenda

- Background & Problem
- Data Collection
- Data Wrangling & Exploration
- Interactive Visualizations
- Predictive Modeling
- Results & Insights
- Key Takeaways

Background

- Falcon 9 launches cost $62M — much cheaper than competitors.
- Reusability of the first stage significantly cuts costs.
- Predicting successful landings = predicting profitability.

The Problem

- Can we predict whether Falcon 9's first stage will land successfully?
- Why solve it? To enable smarter launch bids and strategic decisions.

Data Collection

- SpaceX API – Launch metadata, specs, outcomes
- Wikipedia – Historical launch tables via web scraping

Data Wrangling

- Cleaned & merged datasets
- Handled missing values
- Created binary target column: Class (1 = success, 0 = failure)
- Identified categorical and numeric features

Exploratory Data Analysis

- Used Pandas, SQL, Matplotlib, Seaborn
- Key insights into launch site success, orbit performance
- KSC LC-39A emerged as the most reliable site

Launch Sites & Geolocation

- Mapped all launch pads using Folium
- Color-coded markers based on outcomes
- Visualized proximity to coasts, roads, and urban zones

Interactive Dashboard

- Built using Plotly Dash
- Launch site and payload filters
- Pie chart and scatterplot views
- Live exploration of success distribution

Predictive Modeling

- Tested Logistic Regression, SVC, KNN, Decision Tree
- Used GridSearchCV, train-test split, and StandardScaler
- Decision Tree was the top performer

Model Evaluation

- Best model: Decision Tree Classifier
- Accuracy: High
- Confusion matrix showed minimal false outcomes

Success Trends

- Landing success rate improved post-2015
- KSC LC-39A had the highest success rate
- GTO missions were riskier than LEO/ISS ones

Key Takeaways

- Landing prediction models are highly accurate
- Success tied to orbit, booster version, and site
- SpaceX has dramatically improved landing reliability

Data Sources

- SpaceX API
- Wikipedia Launch Tables (2021 snapshot)
- Processed datasets:
- - dataset_part_1.csv
- - dataset_part_2.csv
- - spacex_launch_geo.csv
- - spacex_launch_dash.csv

Thank You

- Questions?
- Let’s launch some insights 🚀