SpaceX Falcon 9 Landing Prediction: A Data Science Approach

Create for Free

SpaceX Falcon 9 Landing Prediction: A Data Science Approach

Predicting First Stage Landing Success Using Machine Learning

IBRAHIMI MOHAMED

Executive Summary

SpaceX Falcon 9 rockets offer significant cost savings through **reusable first stages**. This project develops a machine learning pipeline to predict successful first stage landings.

Key question: **What factors determine successful rocket landings?**

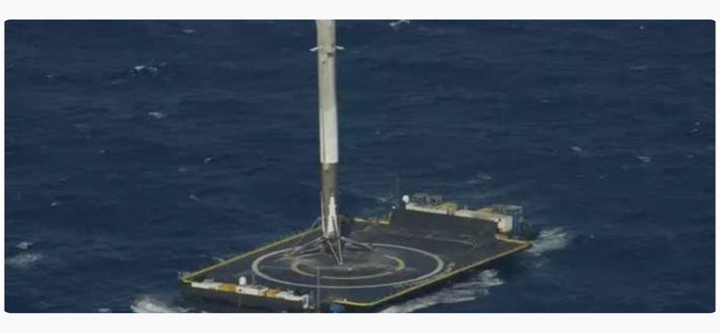
Methodology

- Data Collection via API & Web Scraping
- Data Wrangling & Preprocessing
- Exploratory Data Analysis (SQL & Visualization)
- Interactive Visual Analytics (Folium)
- Machine Learning Prediction Models

Key Findings

- Higher flight numbers correlate with increased success rates at launch sites
- Success rate has steadily increased from 2013 to 2020
- Orbits ES-L1, GEO, HEO, SSO, VLEO show highest success rates
- KSC LC-39A has the most successful launches of any site
- Decision Tree classifier outperforms other ML models
- Successful prediction of landings can inform competitive bidding against SpaceX

Introduction



SpaceX Falcon 9 rockets offer **significant cost advantages** through reusable first stages, revolutionizing space launch economics.

By accurately predicting first stage landing success, we can determine launch costs and enable competitive bidding against SpaceX for rocket launches.

Launch Cost Comparison

$62M	$165M+
SpaceX Falcon 9	Other Providers

Project Objectives

- Identify factors determining successful rocket landings
- Analyze interactions between features affecting success rates
- Determine optimal operating conditions for successful landings
- Develop machine learning pipeline to predict landing success

Methodology Overview

Data Collection — Data Wrangling — EDA — Interactive Analytics — Machine Learning

Data Collection

- SpaceX API requests with JSON parsing
- Web scraping Wikipedia with BeautifulSoup
- HTML table parsing to DataFrame

Data Wrangling

- Missing value handling
- One-hot encoding for categorical features
- Landing outcome label creation

EDA & Visualization

- PostgreSQL database integration
- Flight number vs. launch site analysis
- Success rate by orbit type

Interactive Analytics

- Folium map with launch site markers
- Proximity analysis to landmarks
- Plotly Dash dashboard creation

Machine Learning

- Hyperparameter tuning with GridSearchCV
- Multiple classification models
- Accuracy-based model evaluation

Data Collection & Wrangling

Data Collection Methods

SpaceX API

- GET requests to SpaceX API endpoints
- JSON response parsing with .json()
- DataFrame creation with .json_normalize()

```
response = requests.get(api_url)
data = response.json()
df = pd.json_normalize(data)
```

github.com/chuksoo/IBM-Data-Science-Capstone-SpaceX

Web Scraping

- BeautifulSoup for HTML parsing
- Extract Falcon 9 launch records
- HTML table to DataFrame conversion

Data Collection with Web Scraping.ipynb

Data Wrangling Process

- Missing value identification and handling
- One-hot encoding for categorical features
- Landing outcome label creation from outcome column
- Launch counts per site and orbit calculations

Identify Missing Values → Transform Features → Create Labels → Export to CSV

Data Wrangling.ipynb

Key Data Features

- Flight number and launch site
- Orbit type classification
- Payload mass information
- Booster version details
- Launch date and time
- Success/failure outcome

Exploratory Data Analysis

SQL Analysis

- Identified**unique launch sites**in space missions
- Calculated**total payload mass**by booster version
- Determined**average payload mass**for F9 v1.1
- Found first successful ground landing date

```
SELECT DISTINCT LaunchSite
FROM SpaceX;
```

Key Finding
First successful ground landing occurred on December 22, 2015

EDA with SQL.ipynb

Data Visualization

- Flight number vs. launch site relationship
- Success rate by orbit type
- Payload mass vs. orbit type analysis
- Launch success yearly trend

Flight Number vs. Launch Site | Success Rate by Orbit Type

Key Finding
Higher flight numbers correlate with increased success rates

EDA with Data Visualization.ipynb

Interactive Visual Analytics

Folium Map Visualization

- Marked all launch sites with color-coded markers
- Calculated distances to nearby landmarks
- Added map objects for launch outcomes

Key Finding
Launch sites maintain strategic distances from cities and are positioned near coastlines

Build an Interactive Map with Folium.ipynb

Plotly Dash Dashboard

- Pie charts showing launch distribution by site
- Scatter plots of payload vs. outcomes
- Interactive filters for booster versions

SpaceX Launch Analytics

Launch Site Filter | Payload Range Slider

app.py - Plotly Dashboard

Machine Learning Prediction

Classification Models

- Data split into training and testing sets
- Hyperparameter tuning with GridSearchCV
- Accuracy as primary evaluation metric

Decision Tree	Decision Tree (Tuned)	Random Forest
88.5%	94.7%	86.2%

Best Performing Model
Decision Tree classifier with optimized hyperparameters achieved the highest accuracy

Machine Learning Prediction.ipynb

Model Evaluation

Confusion Matrix

28 True Positive	2 False Positive
3 False Negative	7 True Negative

Key Metrics

Precision	Recall	F1 Score	Accuracy
93.3%	90.3%	91.8%	94.7%

Key Finding
Main challenge is minimizing false positives (unsuccessful landings marked as successful)

Key Results

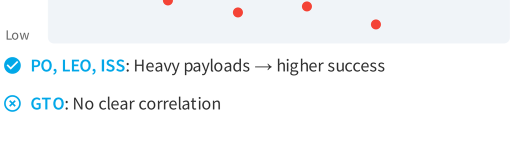
Success Rate Trends

Steady increase from 2013 to 2020

40% improvement · Peak at 95% in 2020

Launch Site Performance

KSC LC-39A leads in success rate



Higher flight count → higher success

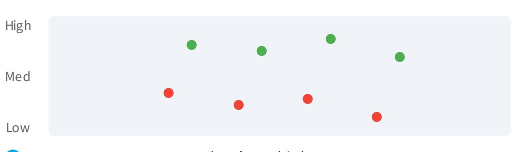
Orbit Type Success Rates

ES-L1, GEO, HEO, SSO, VLEO lead

ES-L1	GEO	HEO
100%	100%	100%
SSO	**VLEO**	**LEO**
100%	100%	75%
ISS	**MEO**	**GTO**
65%	60%	50%

Payload vs. Success Rate

Heavy payloads show higher success in PO, LEO, ISS



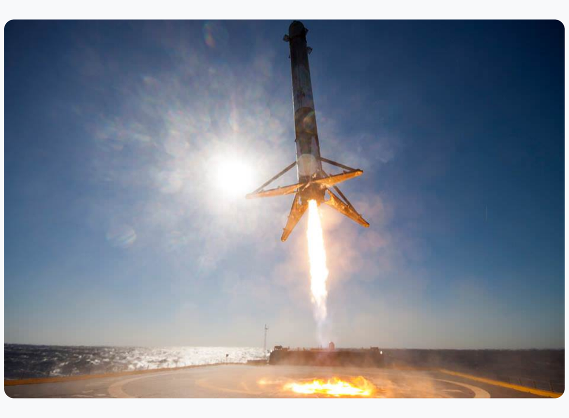
- **PO, LEO, ISS**: Heavy payloads → higher success
- **GTO**: No clear correlation

Conclusion

Key Takeaways

- Higher flight numbers correlate with **increased success rates** at launch sites
- Success rate has **steadily increased** from 2013 to 2020
- Orbits **ES-L1, GEO, HEO, SSO, VLEO** show highest success rates
- **KSC LC-39A** has the most successful launches of any site
- **Decision Tree classifier** outperforms other ML models

Industry Implications
Successful prediction of landings enables competitive bidding against SpaceX, potentially disrupting the space launch market



Future Work

- Integrate real-time weather data
- Explore deep learning approaches
- Compare with other launch providers
- Optimize booster maintenance schedules