SpaceX Falcon 9 Landings: A Data Science Analysis

Applied Data Science Capstone

Executive Summary

Project Overview:

- Analysis of Falcon 9 launch data to evaluate landing outcomes

Objectives:

- Explore launch trends and outcomes
- Predict success or failure of landings

Key Takeaways:

- Launch site and payload significantly impact outcomes

Predictive models achieved high accuracy

Introduction

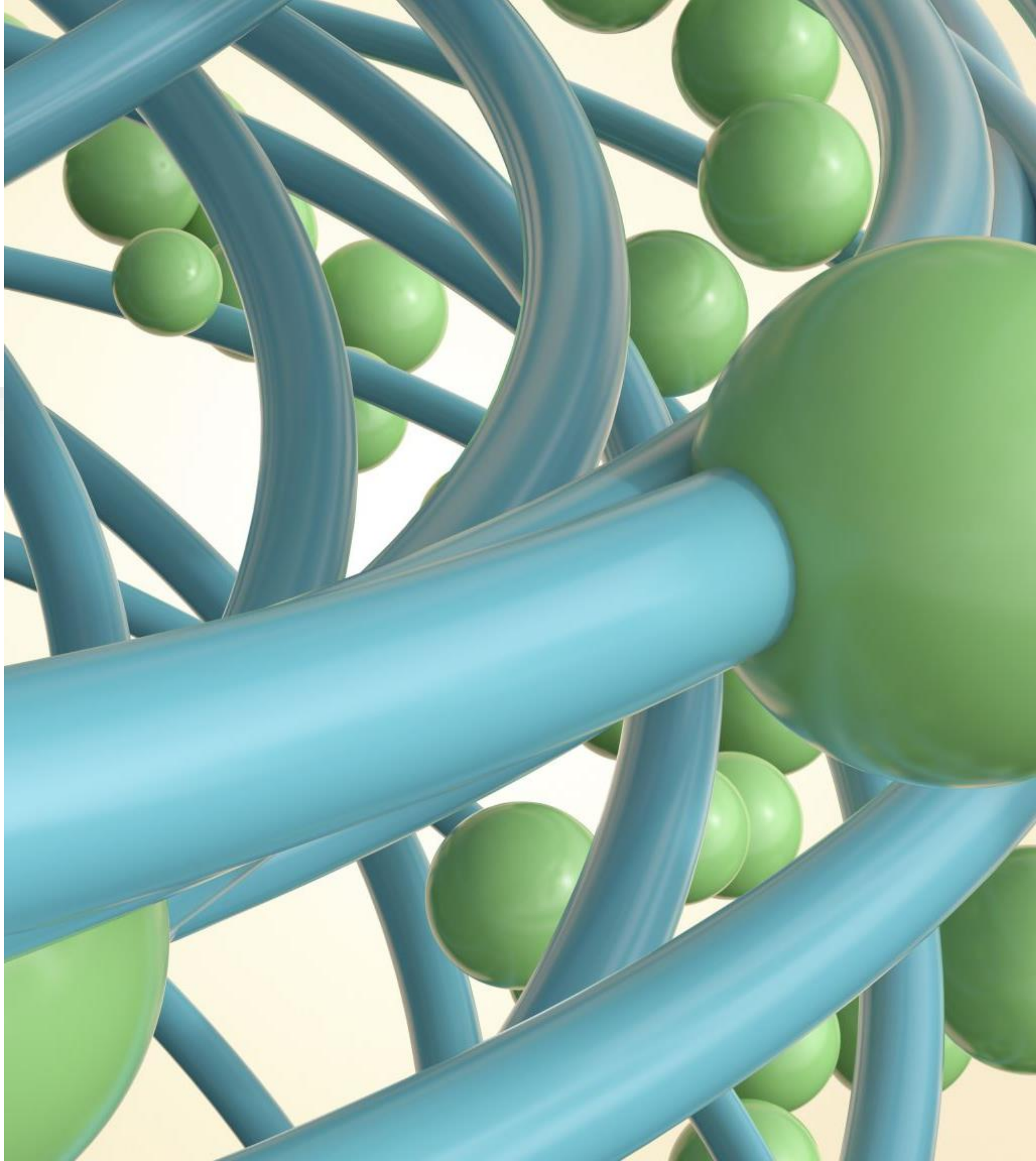
- Company: SpaceX
- Mission: Reusability of rockets via successful landings
- Falcon 9: Key reusable rocket launched since 2010
- Data Science Objective:
    - Analyze launches for patterns
    - Predict landing outcomes

Data Collection & Wrangling

- Sources:
  - SpaceX API
  - Wikipedia Falcon 9 Launch Table
  - External CSVs (spacex_launch_geo.csv, dataset_part_X.csv)
- Tools Used:
  - Pandas, SQLite, API requests
- Cleaning Steps:
  - Filtering Falcon 9 only
  - Handling missing/null values
  - Standardizing payload mass

Exploratory Data Analysis (EDA)

- Key Questions:
  - Which launch sites are most used?
  - Which orbits are most successful?
  - Is payload related to success?
- Libraries Used:
  - Seaborn, Matplotlib, Pandas



```
Number of launches at each site:
LaunchSite
CCSFS SLC 40    55
KSC LC 39A      22
VAFB SLC 4E     13
Name: count, dtype: int64
```

The Most Utilized Launch Site

- Cape Canaveral SLC-40 is the most utilized launch site.

```
%sql SELECT MIN(Date) AS FirstSuccessfulGroundPadLanding FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';
```

```
 * sqlite:///my_data1.db
Done.
```

FirstSuccessfulGroundPadLanding
2015-12-22

SQL-Based EDA

- First successful ground landing date

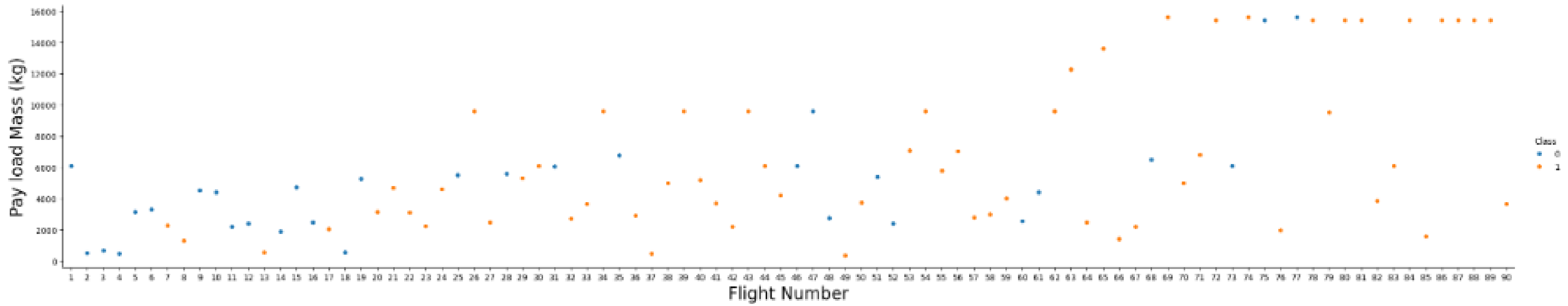
Rankings of landing outcomes from 2010 to 2017

Landing_Outcome	OutcomeCount
Uncontrolled (ocean)	2
Success (ground pad)	3
Success (drone ship)	5
No attempt	10
Failure (parachute)	2
Failure (drone ship)	5
Controlled (ocean)	3
Precluded (drone ship)	1

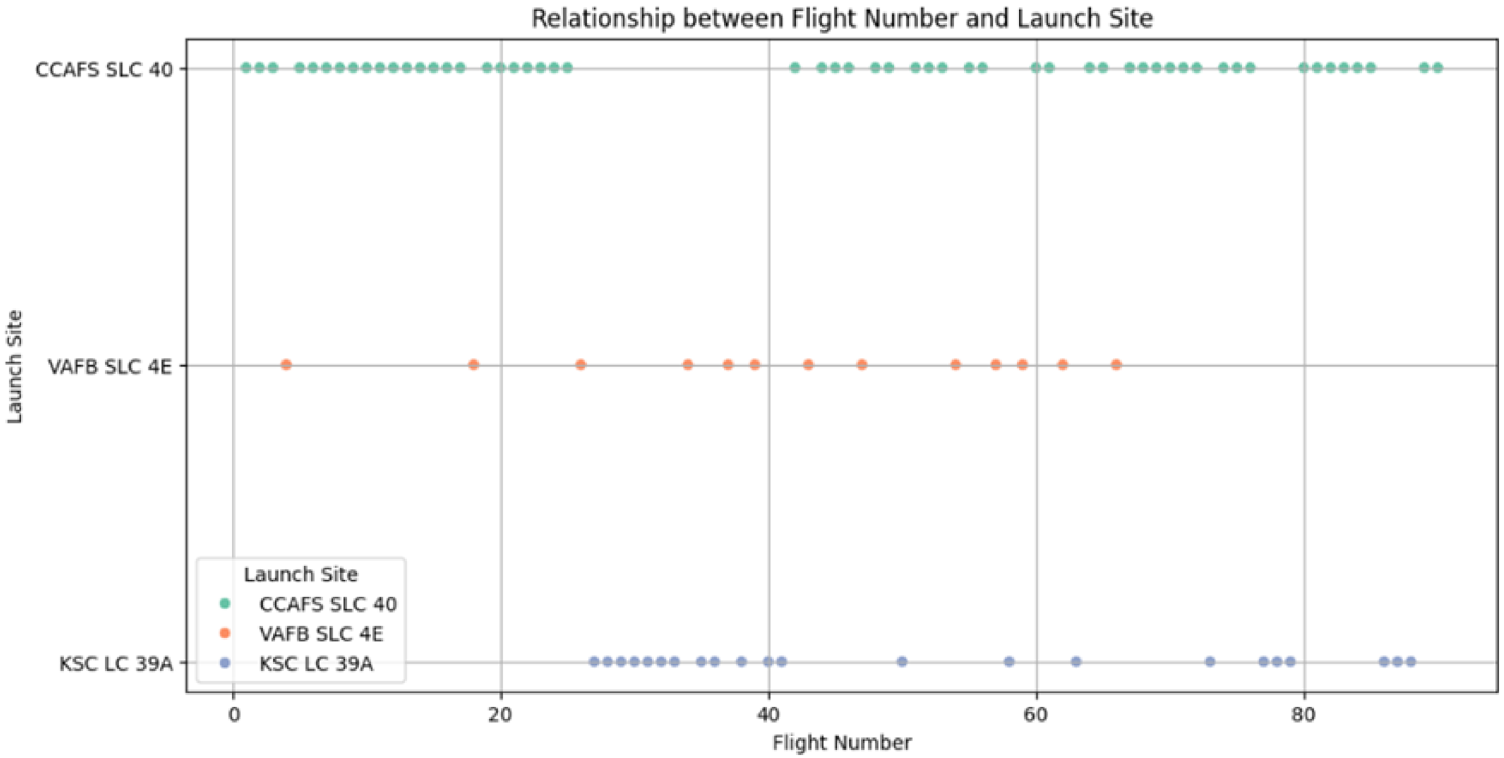
Mission_Outcome	MissionOutcome
Failure (in flight)	1
Success	98
Success	1
Success (payload status unclear)	1

Total counts of successful vs. failed missions.

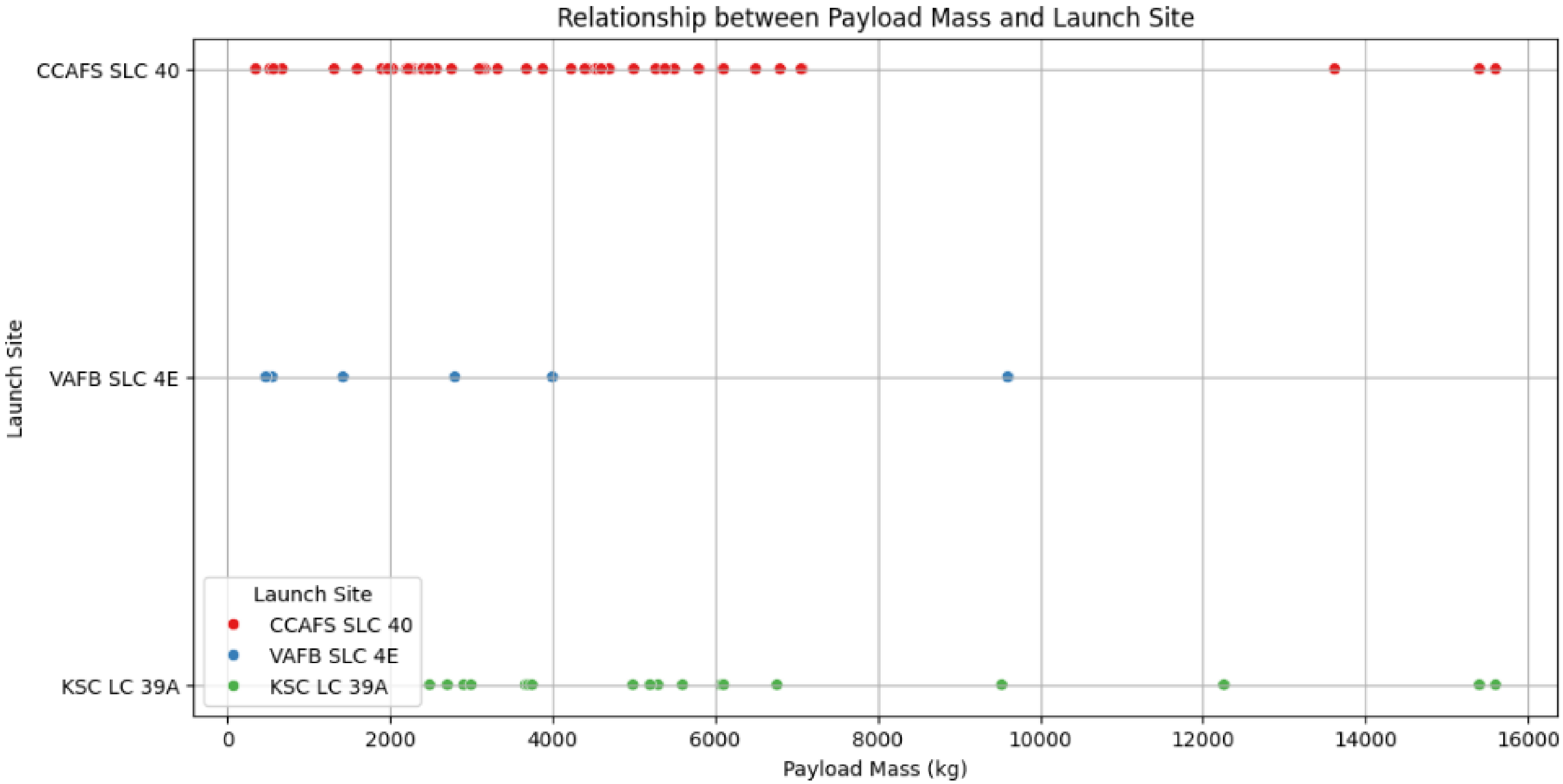
FlightNumber vs. PayloadMass and Overlay



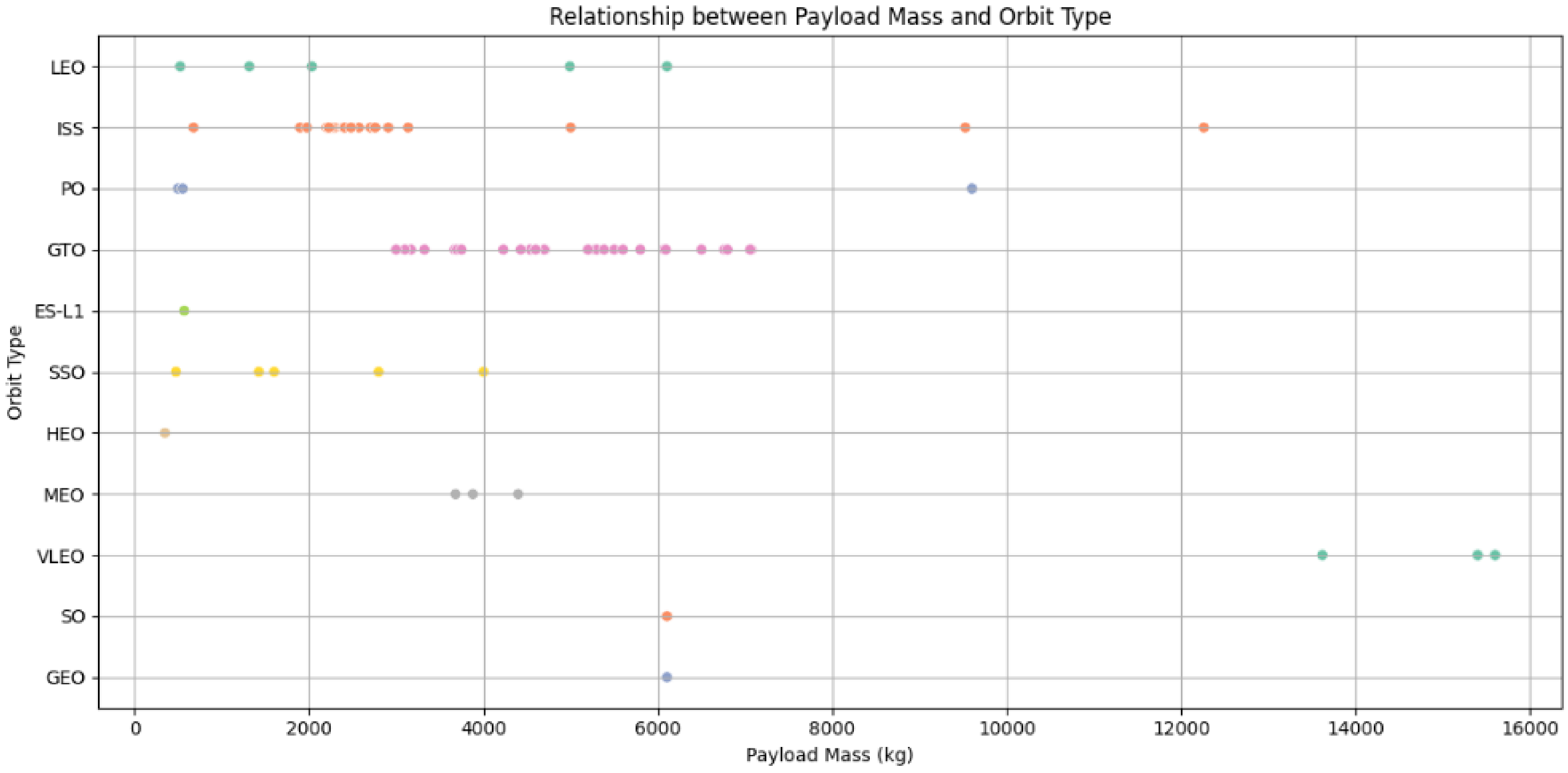
Visualize the relationship between Flight Number and Launch Site



Visualize the relationship between Payload and Launch Site

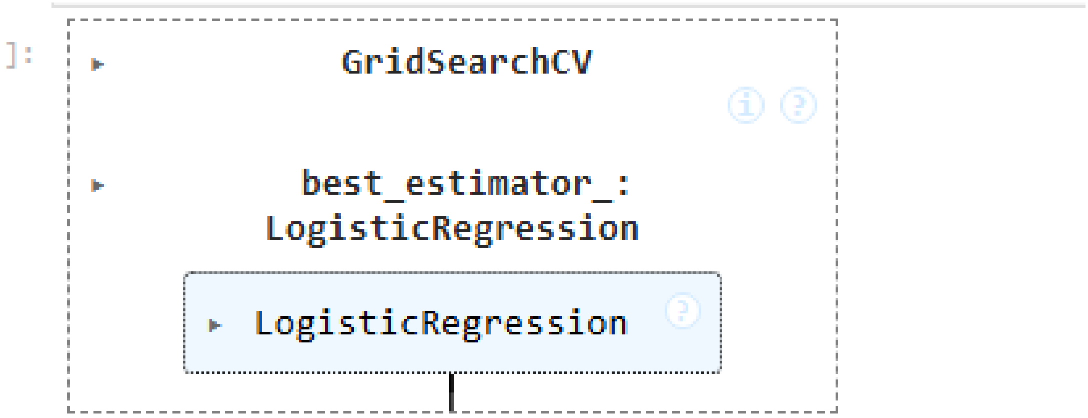


Visualize the relationship between Payload and Orbit type



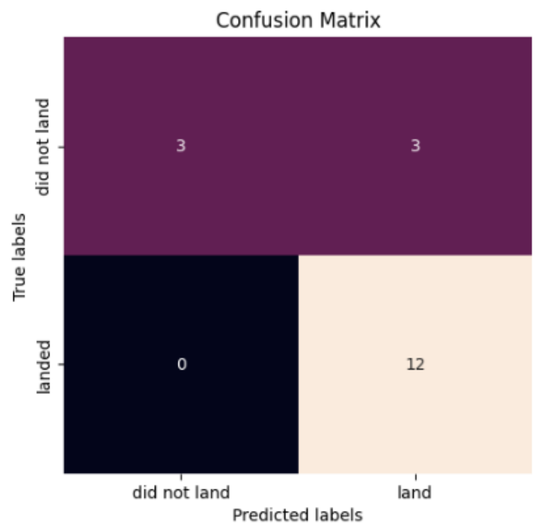
Predictive Analysis - Methodology

- **Logistic Regression**



Parameters

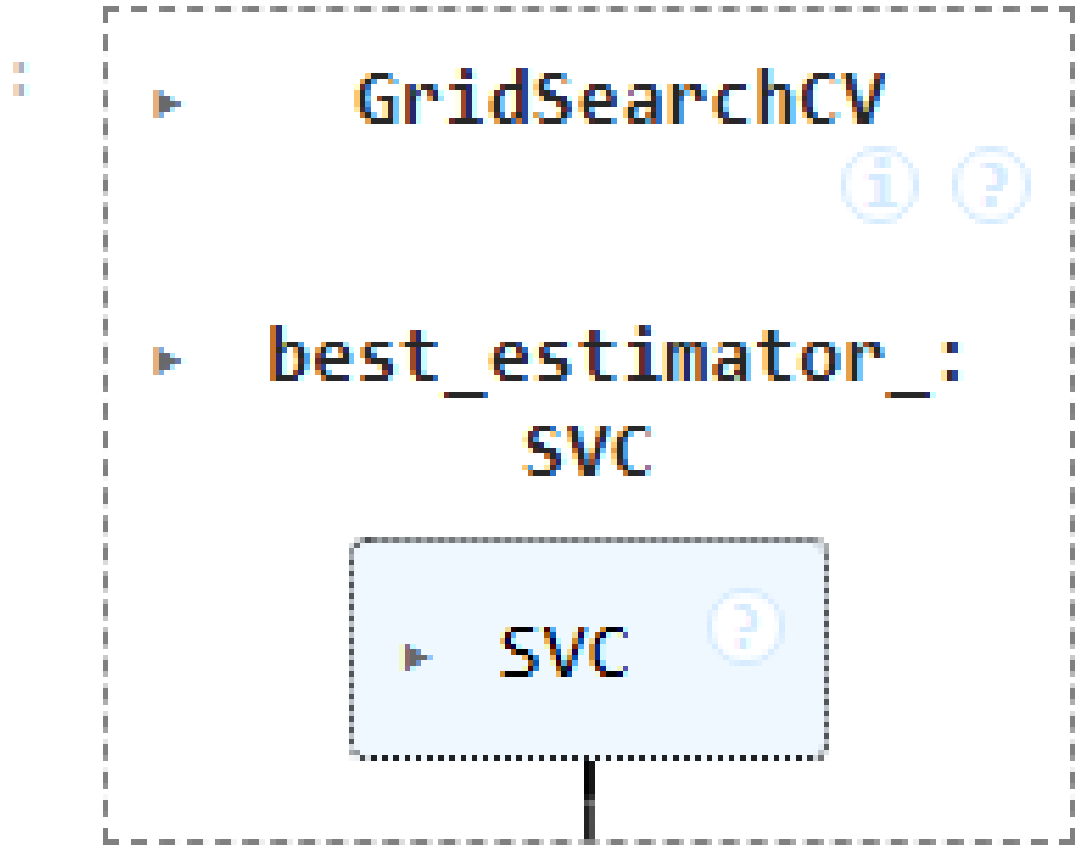
- Best parameters: {'C': 0.01, 'penalty': 'l2', 'solver': 'lbfgs'}
- Best cross-validation score: 0.8464285714285713
- Test set accuracy: 0.8333333333333334



True Postive - 12 (True label is landed, Predicted label is also landed)

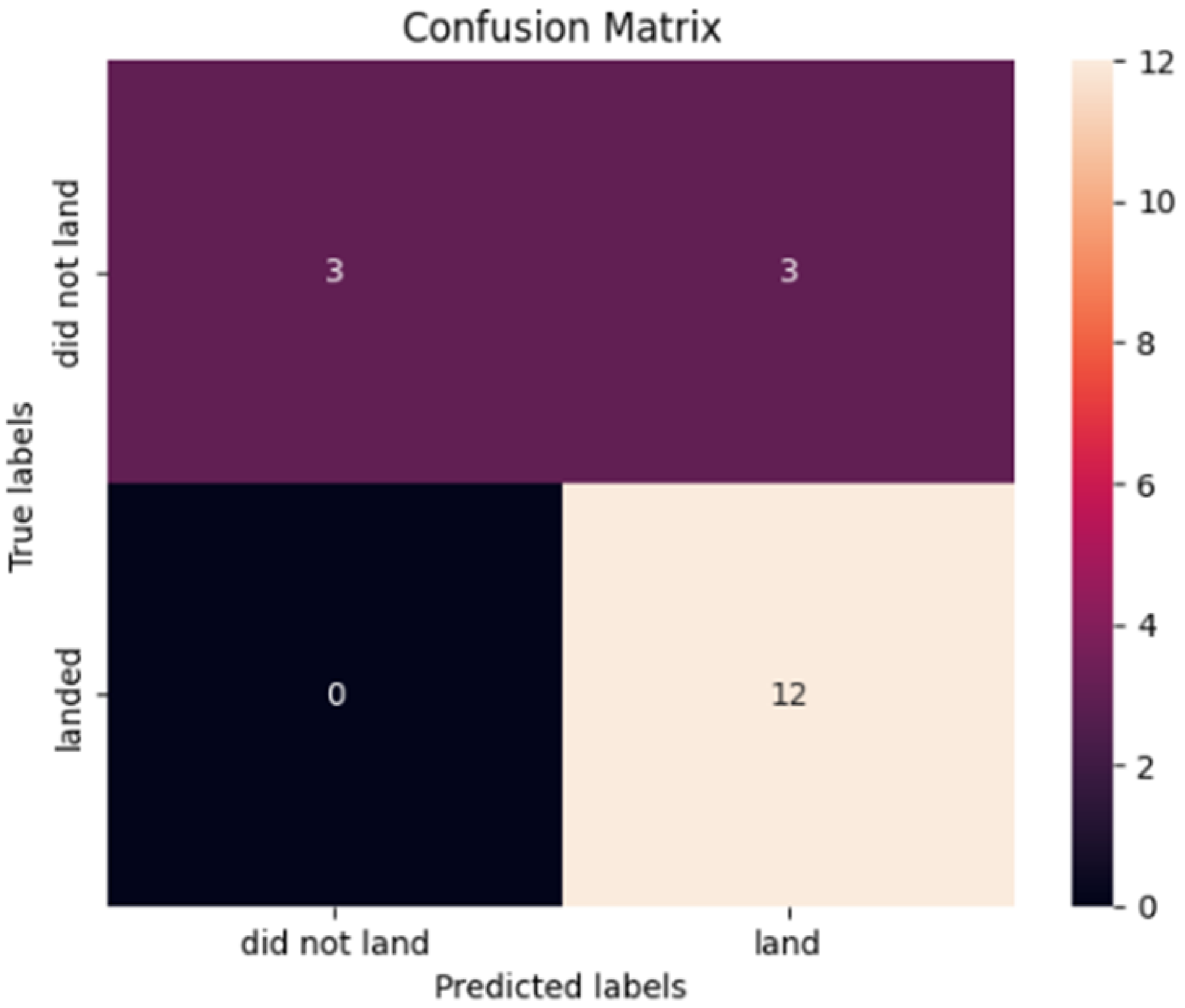
False Postive - 3 (True label is not landed, Predicted label is landed)

Support Vector Machine

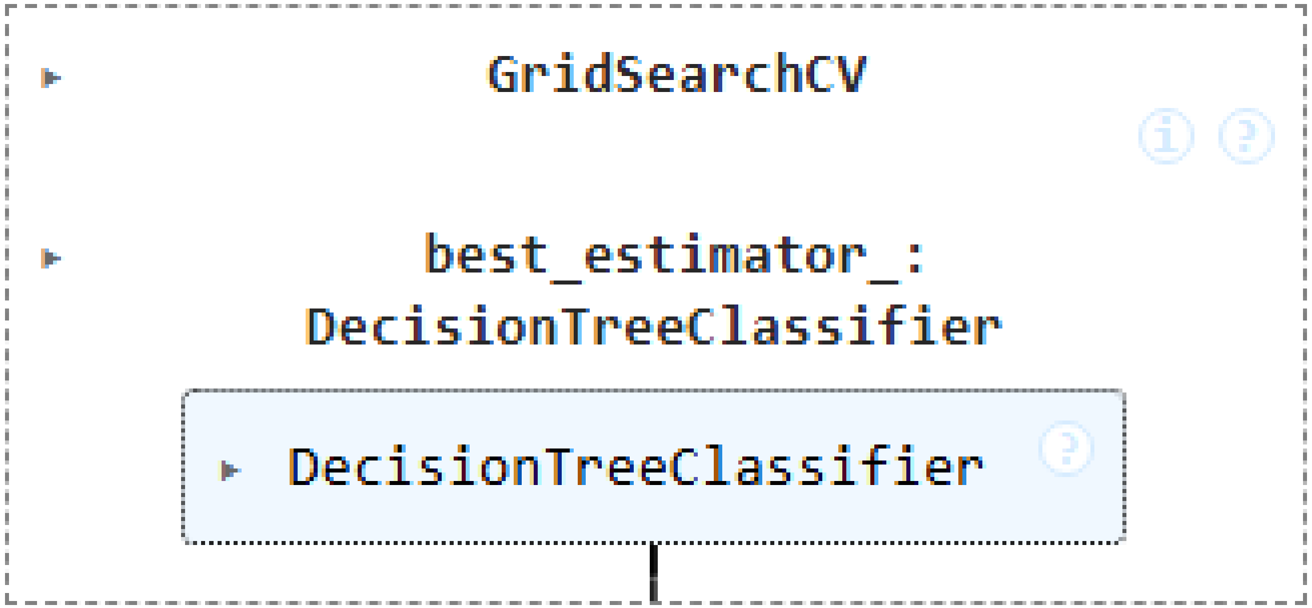


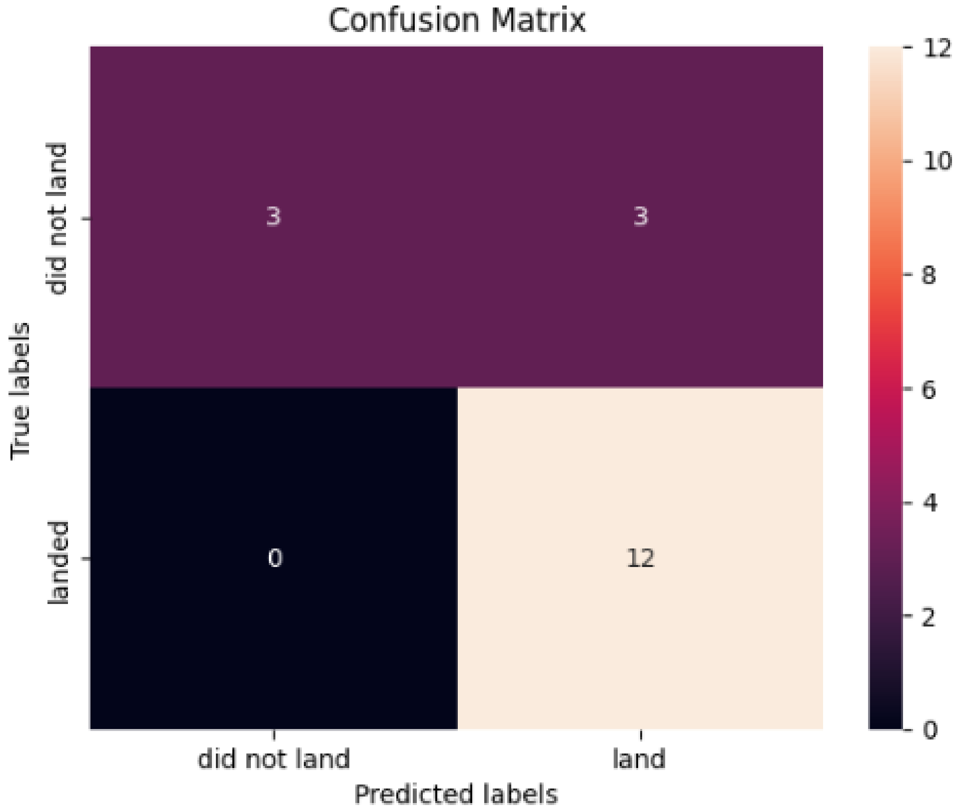
Model Prediction

- tuned hpyerparameters :(best parameters) {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid’}
- accuracy : 0.8482142857142856
- Test set accuracy: 0.8333333333333334



Decision Tree Classifier

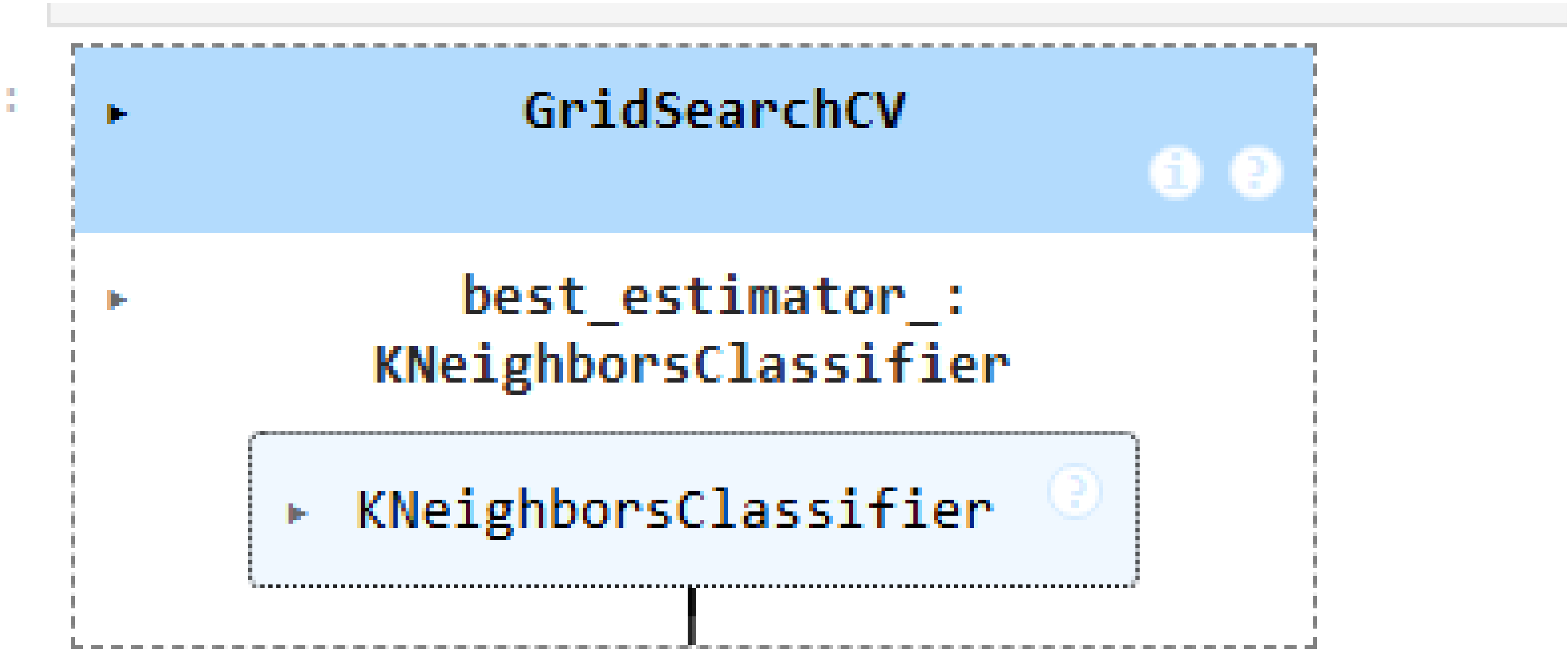




Model Performance

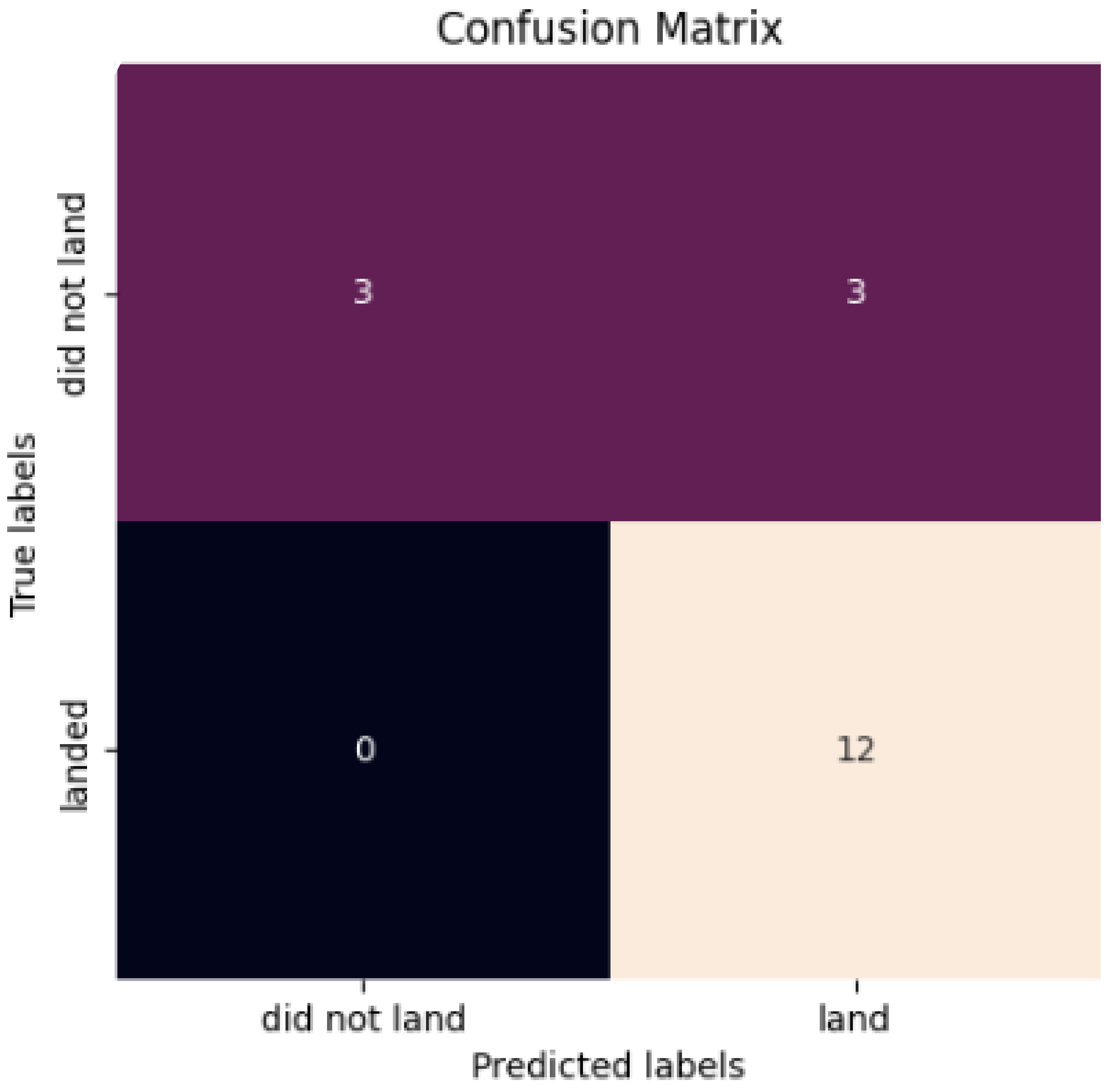
- Decision Tree Test Accuracy: 0.8333333333333334

K Nearest Neighbors



Model Performance

- tuned hpyerparameters :(best parameters) {'algorithm': 'auto', 'n_neighbors': 10, 'p': 1}
- accuracy : 0.8482142857142858
- KNN Test Accuracy: 0.8333333333333334



Best Performance Method

- Logistic Regression: 0.8333
- Support Vector Machine: 0.8333
- Decision Tree: 0.8333
- K-Nearest Neighbors: 0.8333
-
- ✅ Best Performing Model: Logistic Regres sion with accuracy 0.8333

Key Outcomes

LAUNCH SITE AND PAYLOAD STRONGLY AFFECT SUCCESS.

REUSE COUNT AND SPECIFIC ROCKET SERIALS ARE INFLUENTIAL.

MODELS ACHIEVED HIGH PREDICTIVE PERFORMANCE, SUPPORTING RELIABILITY ANALYSIS.

Conclusion

- The analysis showcases how SpaceX’s operational success can be decoded using data science. Visual and predictive tools contribute to understanding and enhancing mission reliability. Future work could involve deep learning, real-time API pipelines, or integrating weather and telemetry data.

Thanks For Your Time