SpaceX Falcon 9 Landings: A Data Science Analysis

ABSTRACT
This report presents a comprehensive data science analysis of SpaceX Falcon 9 rocket landings.

[Applied Data Science Capstone]

Table of Contents

Executive Summary

This report presents a comprehensive data science analysis of SpaceX Falcon 9 rocket landings. The primary objective was to examine patterns and trends across launch records, assess factors influencing landing outcomes, and develop predictive models for mission success. The project followed a structured methodology that involved data collection, preprocessing, exploratory analysis, interactive visualization, SQL analytics, machine learning, and dashboard development.

1.1 Data Collection & Wrangling:

In this capstone, we 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.

Data was collected from multiple sources: SpaceX API, Wikipedia Falcon 9 launch table, and external CSVs. The datasets were cleaned to retain only Falcon 9 launches, missing values were addressed, and numerical fields like PayloadMass were standardized. Categorical variables were encoded to enable model training.

1.2 Exploratory Data Analysis & Visualization:

Using Python libraries **(Pandas, Matplotlib, Seaborn),** the analysis explored launch distribution across sites, the influence of orbit type and payload on landing success, and temporal trends. Key insights include:

Identify and calculate the percentage of the missing values in each attribute

```
[3]: df.isnull().sum()/len(df)*100
```

```
[3]: FlightNumber       0.000000
     Date               0.000000
     BoosterVersion     0.000000
     PayloadMass        0.000000
     Orbit              0.000000
     LaunchSite         0.000000
     Outcome            0.000000
     Flights            0.000000
     GridFins           0.000000
     Reused             0.000000
     Legs               0.000000
     LandingPad        28.888889
     Block              0.000000
     ReusedCount        0.000000
     Serial             0.000000
     Longitude          0.000000
     Latitude           0.000000
     dtype: float64
```

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

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

1.2.1 SQL-Based EDA

SpaceX has gained worldwide attention for a series of historic milestones.

It is the only private company ever to return a spacecraft from low-earth orbit, which it first accomplished in December 2010. SpaceX advertises Falcon 9 rocket launches on its website with a cost of 62 million dollars wheras other providers cost upward of 165 million dollars each, much of the savings is because Space X 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.

This dataset includes a record for each payload carried during a SpaceX mission into outer space.

SQLite queries revealed:

- First successful ground landing date

```
%sql SELECT MIN(Date) AS FirstSuccessfulGroundPadLanding FROM SPACEXTABLE WHERE Landing_Outcome = 'Success (ground pad)';
 * sqlite:///my_data1.db
Done.
```

FirstSuccessfulGroundPadLanding
2015-12-22

- 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

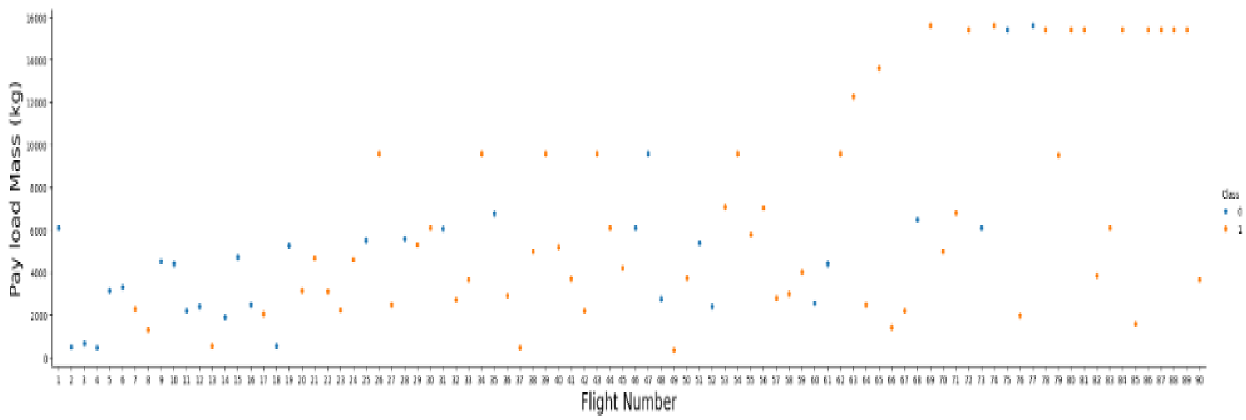
- Total counts of successful vs. failed missions.

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

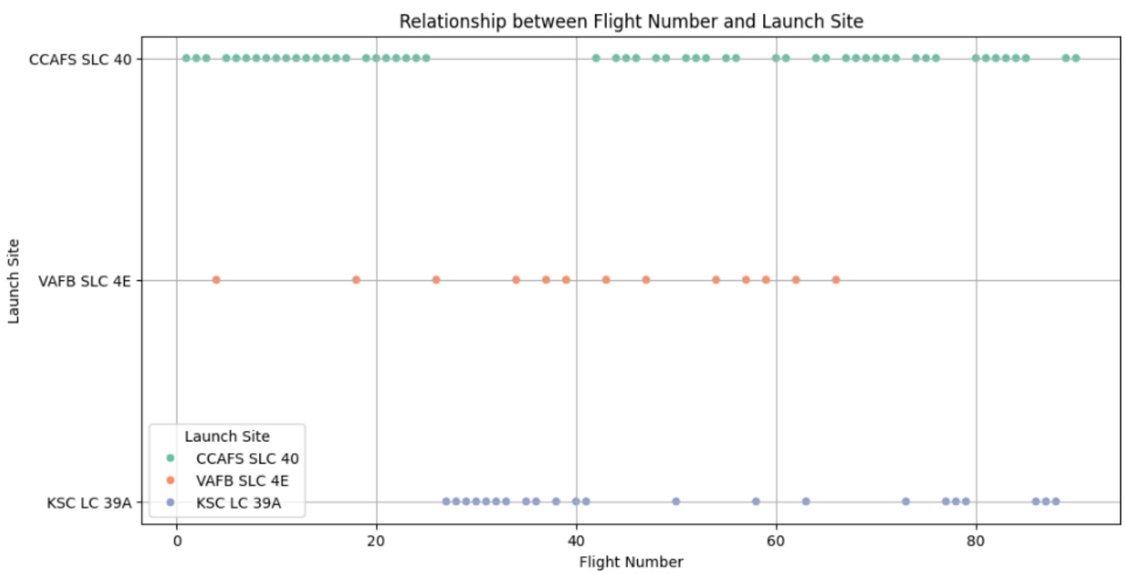
1.2.2 Interactive Visual Analytics

Folium maps plotted launch sites, overlaid with success/failure markers using MarkerCluster. The map also allowed measurement of site distances from coastlines. Plotly Dash dashboard enabled dynamic exploration with dropdown-based filtering, showing launch success rate and payload vs. outcome plots.

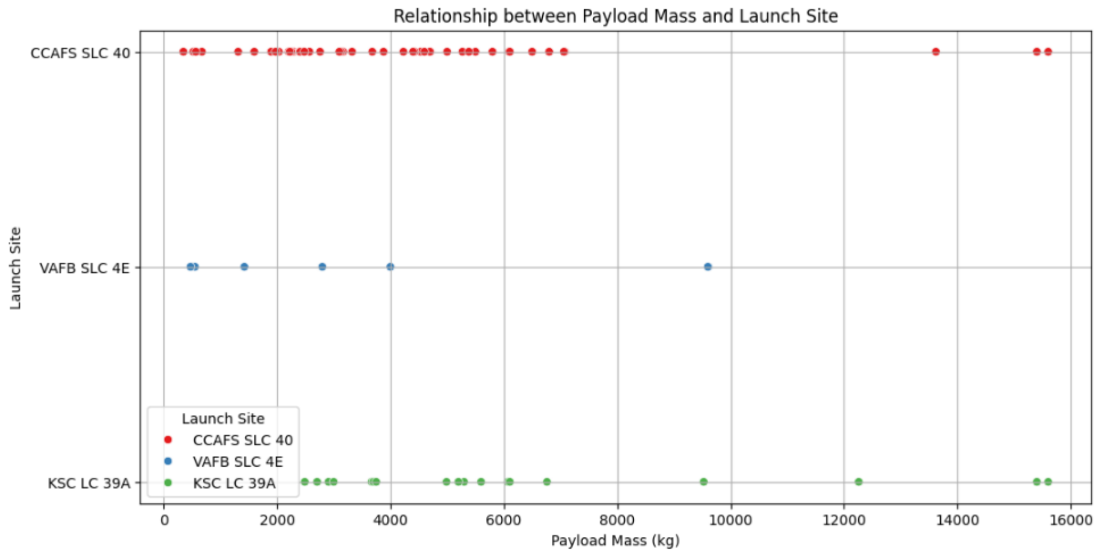
We can plot out the FlightNumber vs. PayloadMassand overlay the outcome of the launch. We see that as the flight number increases, the first stage is more likely to land successfully. The payload mass is also important; it seems the more massive the payload, the less likely the first stage will return.



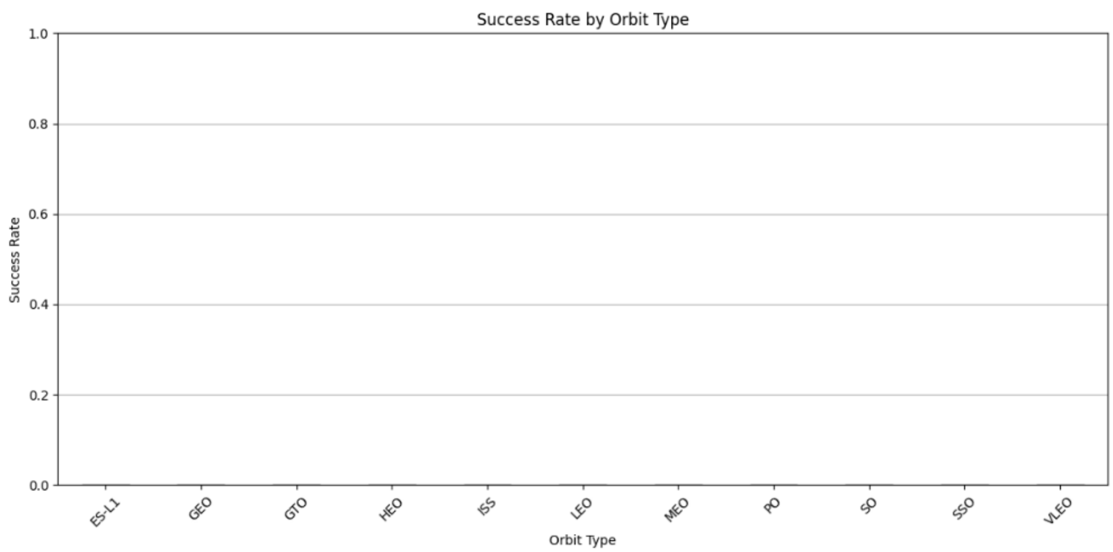
Visualize the relationship between Flight Number and Launch Site



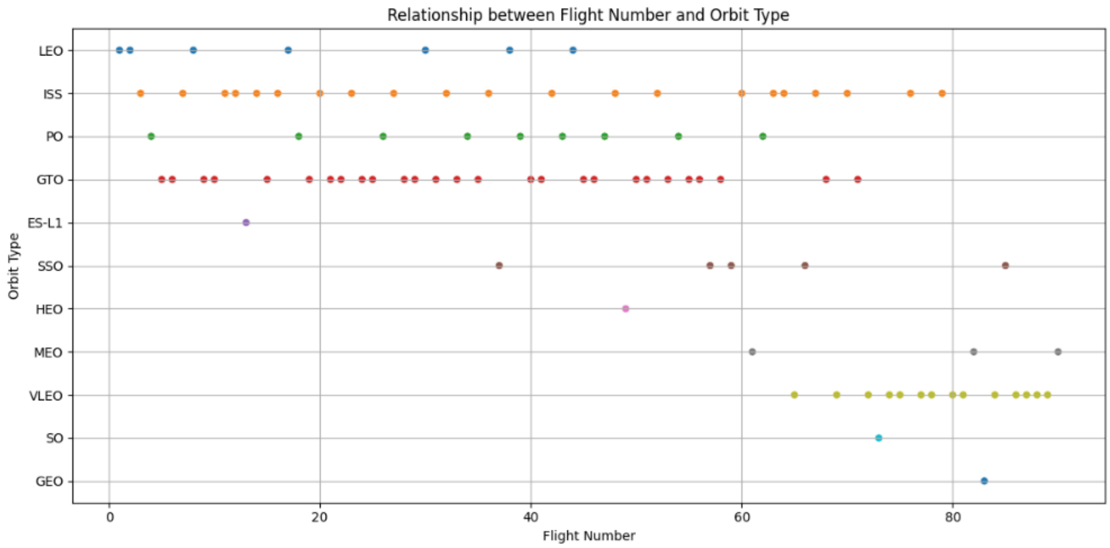
Visualize the relationship between Payload and Launch Site



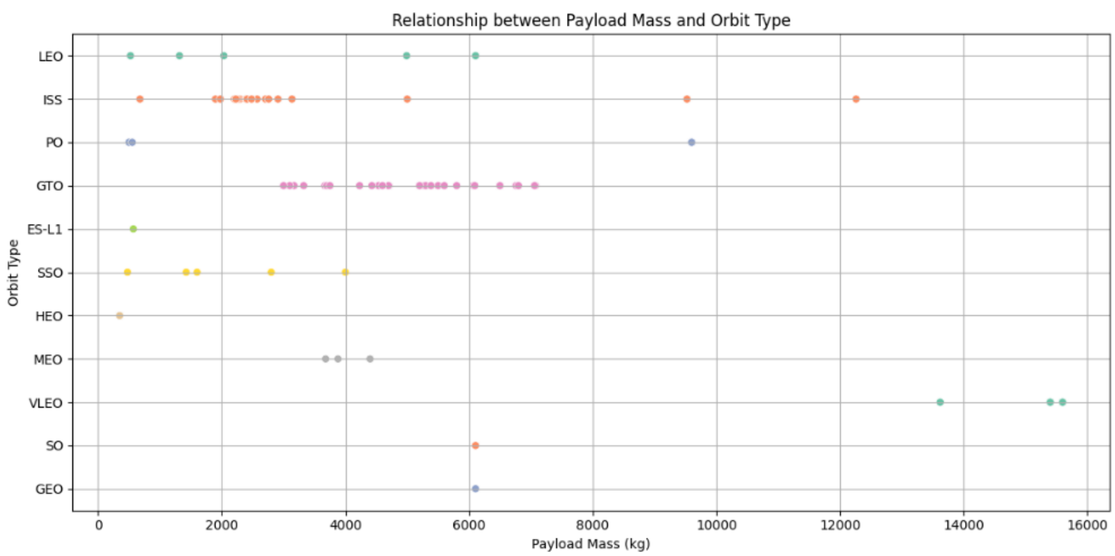
Visualize the relationship between success rate of each orbit type



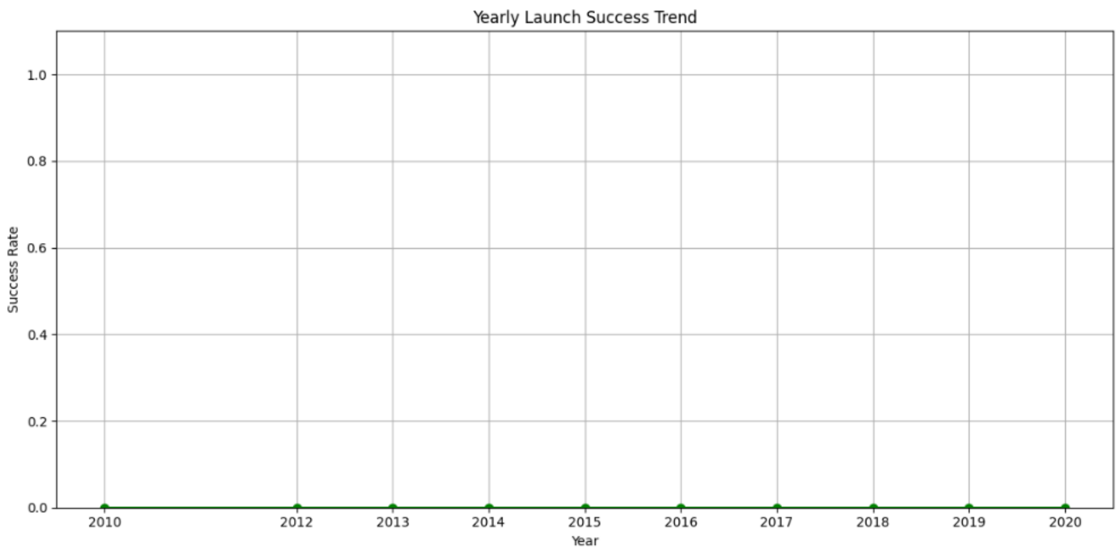
Visualize the relationship between FlightNumber and Orbit type



Visualize the relationship between Payload and Orbit type



Visualize the launch success yearly trend

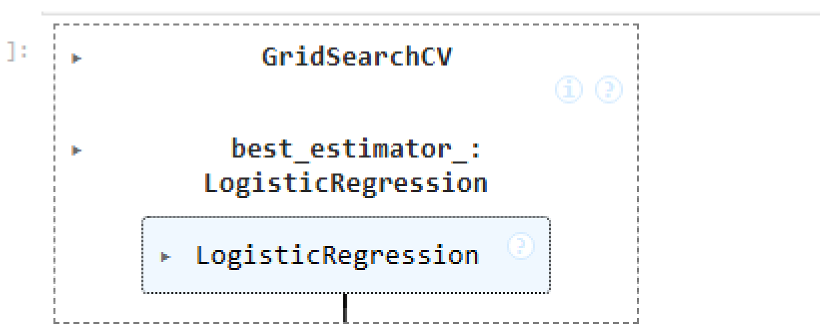


1.3 Predictive Modeling

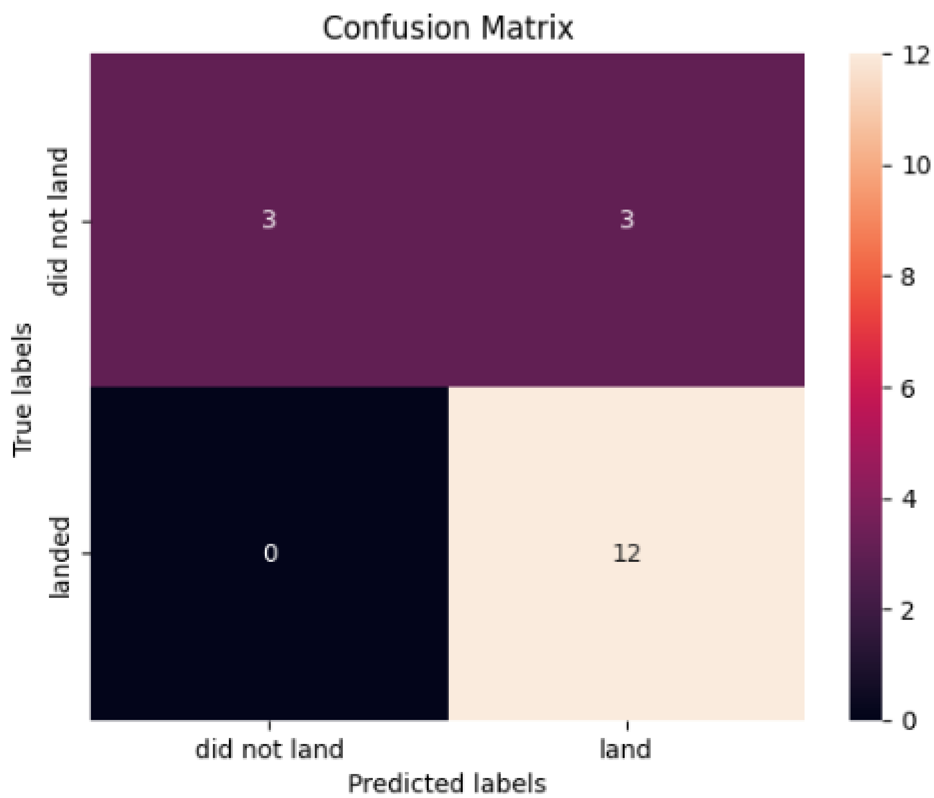
Four machine learning models—Logistic Regression, SVM, Decision Tree, and KNN—were trained on processed data. Using **GridSearchCV** and **standardization,** models were optimized. Accuracy was evaluated on test data, identifying the best performing algorithm.

1.3.1 Logistic Regression

We have Created a logistic regression object then create a GridSearchCV object logreg_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters



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



Examining the confusion matrix, we see that logistic regression can distinguish between the different classes. We see that the problem is false positives.

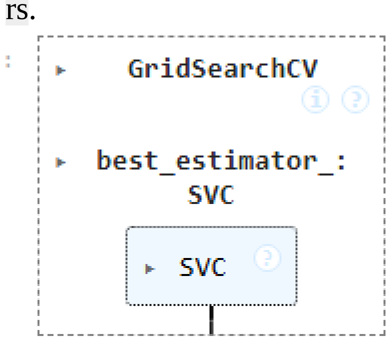
Overview:

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

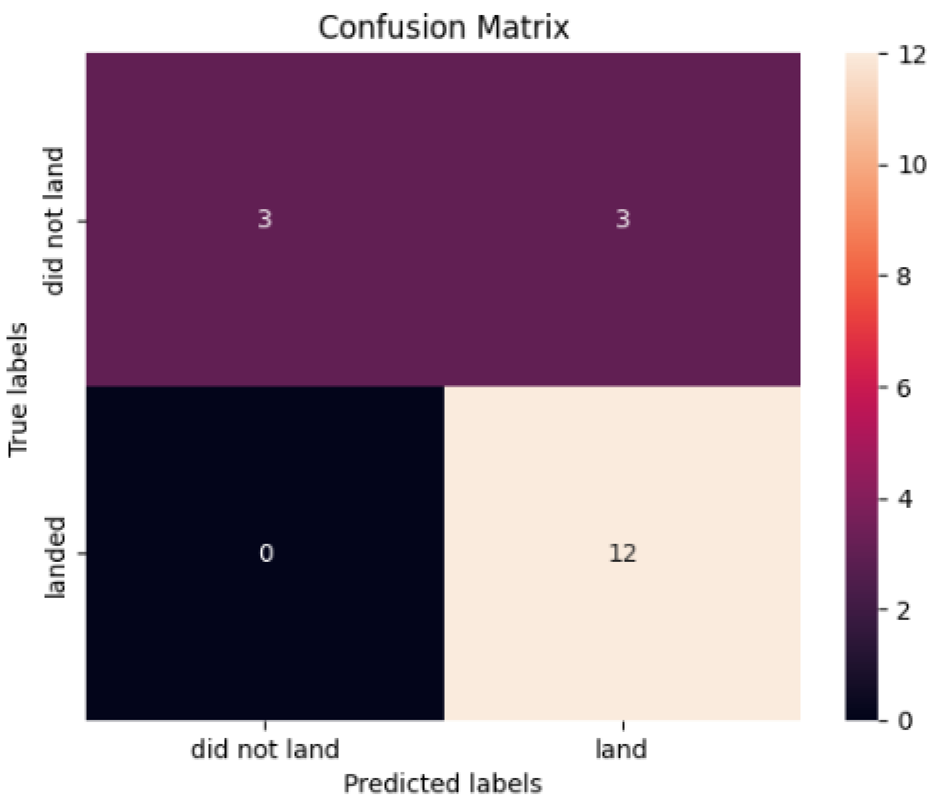
False Postive - 3 (True label is not landed, Predicted label is landed)

1.3.2 Support Vector Machine

We have Created a support vector machine object then create a GridSearchCV object svm_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters.

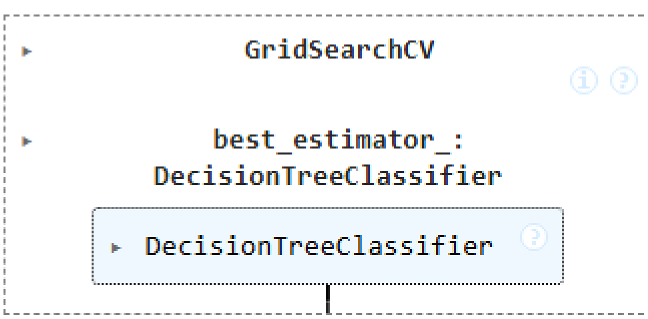


tuned hpyerparameters :(best parameters) {'C': 1.0, 'gamma': 0.03162277660168379, 'kernel': 'sigmoid'}
accuracy : 0.8482142857142856
Test set accuracy: 0.8333333333333334

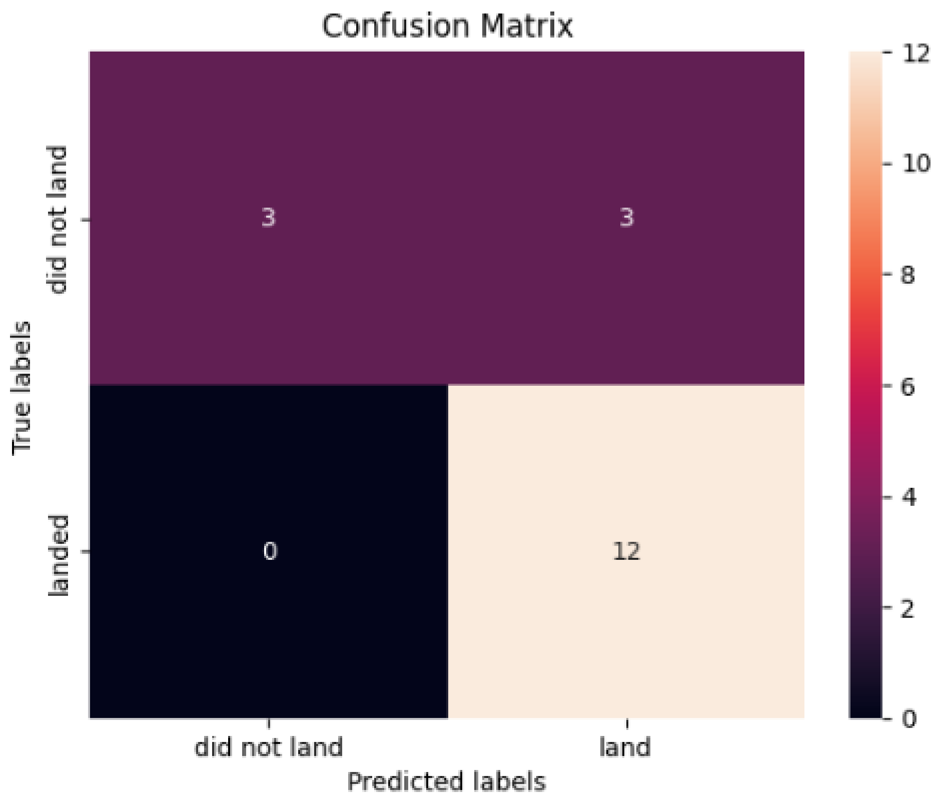


1.3.3 Decision Tree Classifier

Create a decision tree classifier object then create a GridSearchCV object tree_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters.

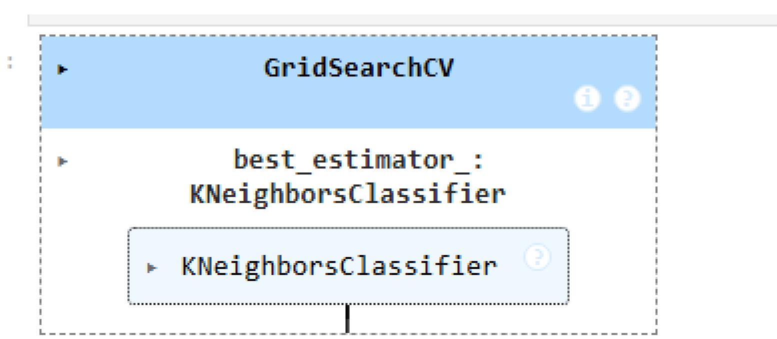



```
Decision Tree Test Accuracy: 0.8333333333333334
```

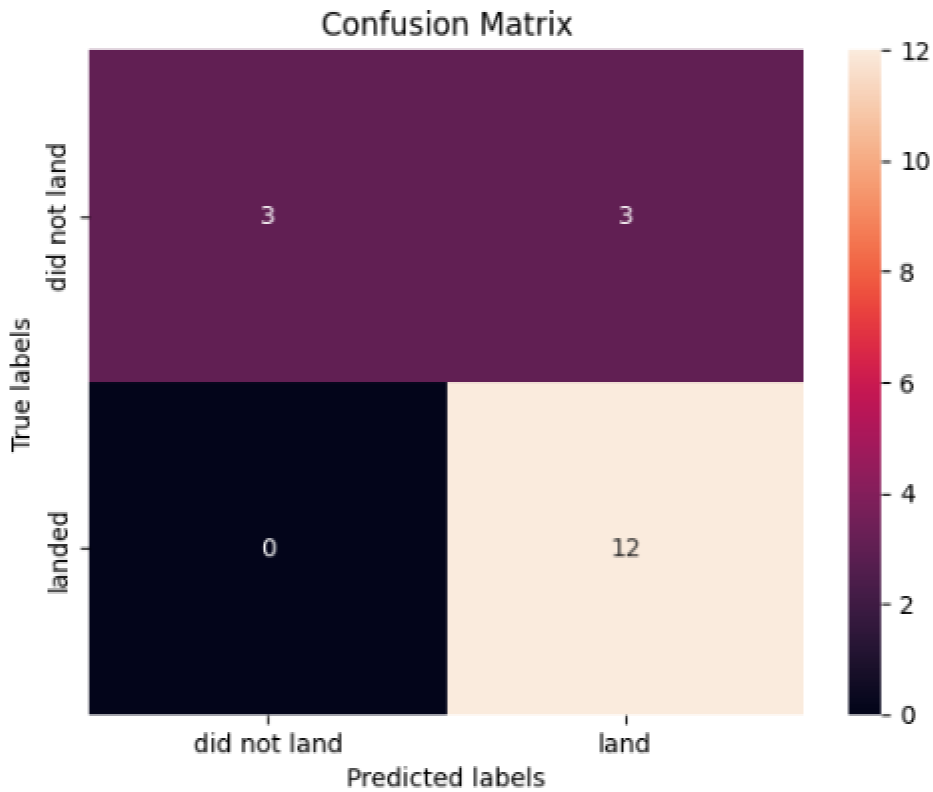


1.3.4 K Nearest Neighbors

Create a k nearest neighbors object then create a GridSearchCV object knn_cv with cv = 10. Fit the object to find the best parameters from the dictionary parameters.



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 Regression 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.

1.4 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.

1.5 Submission

- GitHub URL: [https://github.com/jagatbhai/Data_Sc_Pro/commits?author=jagatbhai]

- PDF Presentation URL: [Insert your PDF link]