FALCON-9 PROJECT REPORT

Prepared by Eng. Albaraa Mohammed
Aug-2025

OUTLINE

- Executive Summary
- Introduction
- Methodology
- Results
- Conclusion
- Appendix

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Executive Summary

- In this project, I've built a ML model to determine weather the SpaceX Falocn-9 rocket's first stage will land successfully or not.
- The ML model used for this project is Classification. And to determine the accuracy of the model I used Confusion Matrix method.
- Several algorithm for Classification used also , such: Logistic Regression, KNN, SVM, Decision Trees. And compare between them to determine the best option and that option was (either Logistic Regression and SVM).

Prepared by : Eng. Albaraa Mohammed

Aug - 2025

Introduction

- SpaceX has launched Falcon-9 rocket on their website, the rocket comes with a great price (62million$), comparing to other manufacturer which their cost nearby(165million$).

- The reason for this is that SpaceX can use the first stage of the rocket again.

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Data Collection Methodology

- The data used for this project from SpaceX official website.
- The methodology to get this data is API .
- To prepare the data well, I used some functions to get important features in consist way, like:(Booster Version, Payload Mass, Cores, and Launch Site).
- Also, since the ML is classification, I had to convert all object data into a numerical data to prepare it for analysis, I create a new feature columns called Class that represent weather the launch landed or not by (0 & 1).

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Exploratory Data Analysis (EDA) Methodology

- For EDA , visual plots were very helpful to discover more about the data and the relationships between features.
- I used various plots, such: Line plot , Cat plot, Bar charts, and more..
- Also, a new feature has been created (AVG success rate), Average Success Rate , to figure out the relationship between several features and landing average success rate.
- I also used SQL queries for EDA.

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

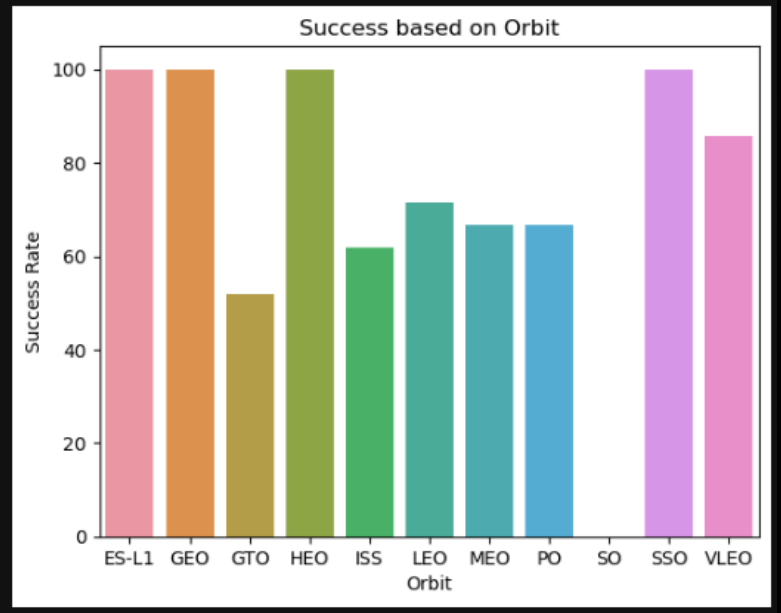
Predictive Model Methodology

- The predictive model was Classification. I've used several algorithm such: KNN, SVM, Decision Trees, and Logistic Regression.
- To compare between them, I've created a function that draw a confusion matrix to ease the process of analysis.
- Also, Cross-Validation was used using Grid Search to better overcome overfitting and underfitting.

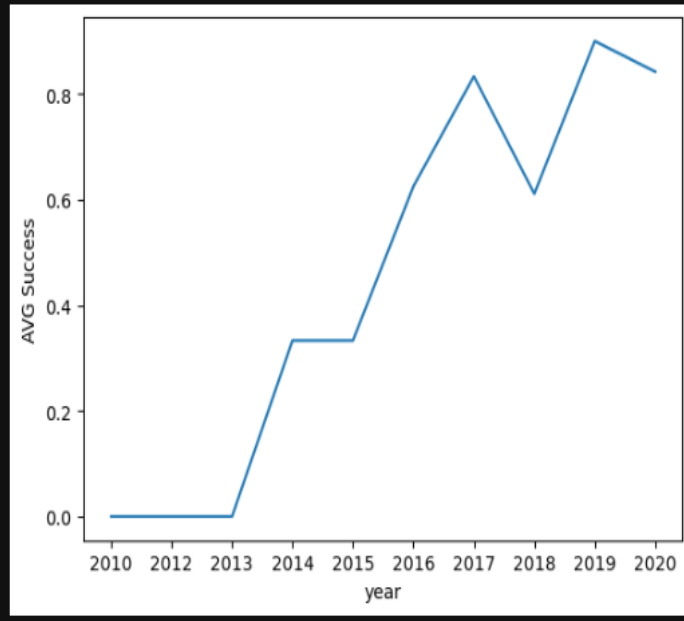
Prepared by : Eng. Albaraa Mohammed
Aug - 2025

EDA with Visualization Results

```
sns.barplot(data=success,x='Orbit',y='Success Rate')
plt.title('Success based on Orbit')
plt.xlabel('Orbit')
plt.ylabel('Success Rate')
plt.show()
```




```
sns.lineplot(data=suc,x='Year',y='AVG Success')
plt.xlabel('year')
plt.ylabel('AVG Success')
plt.show()
```



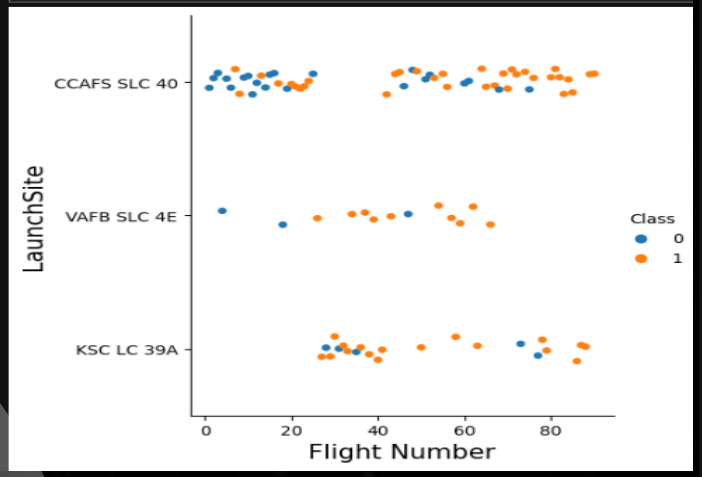
As we can see , Orbit has clear affect on the Success rate of the landing process And GTO is the least one.

From 2013 AVG success rate increased significantly.

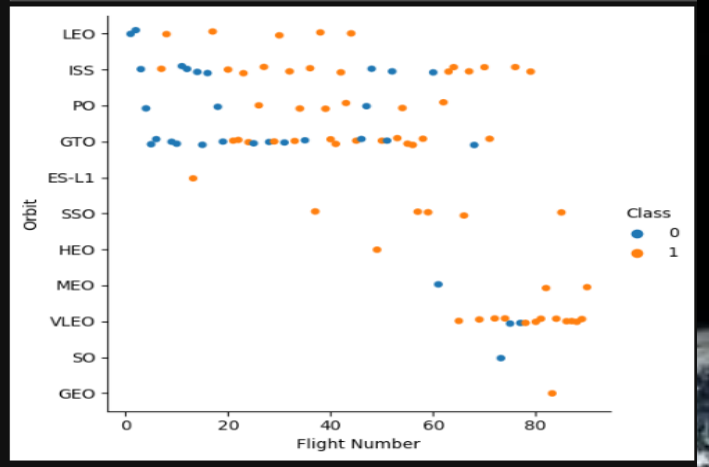
Prepared by : Eng. Albaraa Mohammed
Aug - 2025

EDA with Visualization Results

```
[99]: sns.catplot(y='LaunchSite',x='FlightNumber',data=df3,hue='Class')
      plt.xlabel('Flight Number',fontsize=15)
      plt.ylabel('LaunchSite',fontsize=15)
      plt.show()
```




```
[103]: sns.catplot(y='Orbit',x='FlightNumber',data=df3,hue='Class')
       plt.xlabel('Flight Number')
       plt.ylabel('Orbit')
       plt.show()
```



We can see here that GTO orbit has the most flights and we can say the success and failure were almost equal.
Flights to SSO orbit always succeed landing.

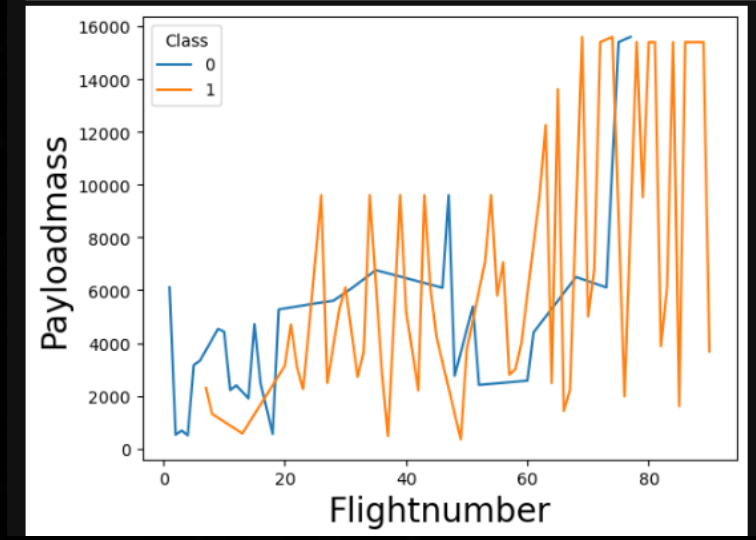
CCAFS SLC 40 has the most flights and after 50 flights the success rate increased.

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

EDA with Visualization Results

```
# FlightNumber and Payloadmass VS Outcome
```

```
sns.lineplot(data=df3,x='FlightNumber',y='PayloadMass',hue='Class')
plt.xlabel('Flightnumber',fontsize=20)
plt.ylabel('Payloadmass',fontsize=20)
plt.show()
```



The effect of payload to success rate was clear if the payload above 12 Ton and flights above 75.

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

SQL Results for EDA

```
[78]: %sql SELECT * from SPACEXTBL WHERE (Launch_Site) LIKE 'CCA%' LIMIT 5;
 * sqlite:///my_data1.db
Done.
```

[78]:

Date	Time (UTC)	Booster_Version	Launch_Site	Payload	PAYLOAD_MASS__KG_	Orbit	Customer	Mission_Outcome	Landing_Outcome
2010-06-04	18:45:00	F9 v1.0 B0003	CCAFS LC-40	Dragon Spacecraft Qualification Unit	0	LEO	SpaceX	Success	Failure (parachute)
2010-12-08	15:43:00	F9 v1.0 B0004	CCAFS LC-40	Dragon demo flight C1, two CubeSats, barrel of Brouere cheese	0	LEO (ISS)	NASA (COTS) NRO	Success	Failure (parachute)
2012-05-22	7:44:00	F9 v1.0 B0005	CCAFS LC-40	Dragon demo flight C2	525	LEO (ISS)	NASA (COTS)	Success	No attempt
2012-10-08	0:35:00	F9 v1.0 B0006	CCAFS LC-40	SpaceX CRS-1	500	LEO (ISS)	NASA (CRS)	Success	No attempt
2013-03-01	15:10:00	F9 v1.0 B0007	CCAFS LC-40	SpaceX CRS-2	677	LEO (ISS)	NASA (CRS)	Success	No attempt

```
[92]: %%sql SELECT DISTINCT Launch_Site
FROM SPACEXTBL;
 * sqlite:///my_data1.db
Done.
```

[92]:

Launch_Site
CCAFS LC-40
VAFB SLC-4E
KSC LC-39A
CCAFS SLC-40

```
[89]: %%sql
SELECT
  substr(Date, 6, 2) AS Month,
  Booster_Version,
  Launch_Site,
  Landing_Outcome
FROM SPACEXTBL
WHERE Landing_Outcome = 'Failure (drone ship)'
  AND substr(Date, 0, 5) = '2015';
 * sqlite:///my_data1.db
Done.
```

[89]:

Month	Booster_Version	Launch_Site	Landing_Outcome
01	F9 v1.1 B1012	CCAFS LC-40	Failure (drone ship)
04	F9 v1.1 B1015	CCAFS LC-40	Failure (drone ship)

```
[84]: %%sql SELECT DISTINCT(Booster_Version) FROM SPACEXTBL
WHERE Landing_Outcome='Success (drone ship)'
AND PAYLOAD_MASS__KG_ BETWEEN 4000 AND 6000;
 * sqlite:///my_data1.db
Done.
```

[84]:

Booster_Version
F9 FT B1022
F9 FT B1026
F9 FT B1021.2
F9 FT B1031.2

We notice here that CCAFSLC-40 have the biggest customer such, SpaceX and NASA.

And, there are several launch sites as shown above.

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

SQL Results for EDA

```
[82]: %sql select sum(PAYLOAD_MASS__KG_) from spacextbl where Customer='SpaceX';
 * sqlite:///my_data1.db
Done.
[82]: sum(PAYLOAD_MASS__KG_)
185220
```

```
[83]: %%sql SELECT MIN(Date) AS First_Success_landing_pad
FROM SPACEXTBL
WHERE Landing_Outcome ='Success (ground pad)';
 * sqlite:///my_data1.db
Done.
[83]: First_Success_landing_pad
2015-12-22
```

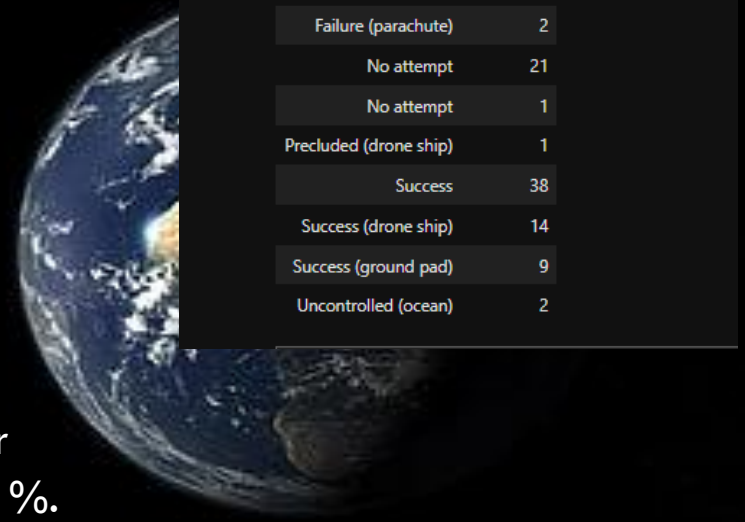
```
80]: %sql select sum(PAYLOAD_MASS__KG_) from spacextbl where Customer='NASA (CRS)
 * sqlite:///my_data1.db
Done.
80]: sum(PAYLOAD_MASS__KG_)
45596
```

F9 FT B1031.2

```
[85]: %%sql SELECT Landing_Outcome,COUNT(*)
FROM SPACEXTBL
GROUp BY Landing_Outcome;
 * sqlite:///my_data1.db
Done.
```

[85]:

Landing_Outcome	COUNT(*)
Controlled (ocean)	5
Failure	3
Failure (drone ship)	5
Failure (parachute)	2
No attempt	21
No attempt	1
Precluded (drone ship)	1
Success	38
Success (drone ship)	14
Success (ground pad)	9
Uncontrolled (ocean)	2



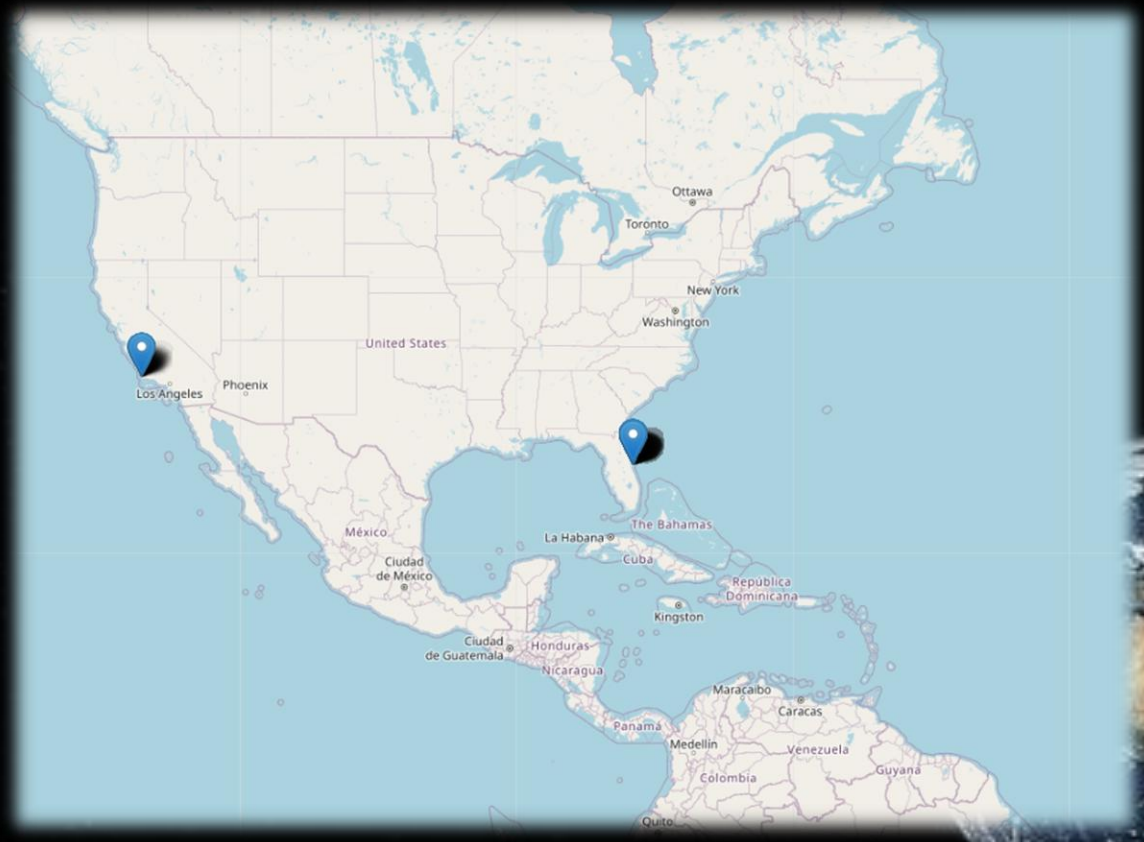
We notice that SpaceX have much higher payload comparing to NASA, about 121%.

Also, some landing outcomes for several locations

First success ground pad location was in 2015

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Folium Results



We notice from this map, that launch sites are explicitly located at eastern and western coasts of US.

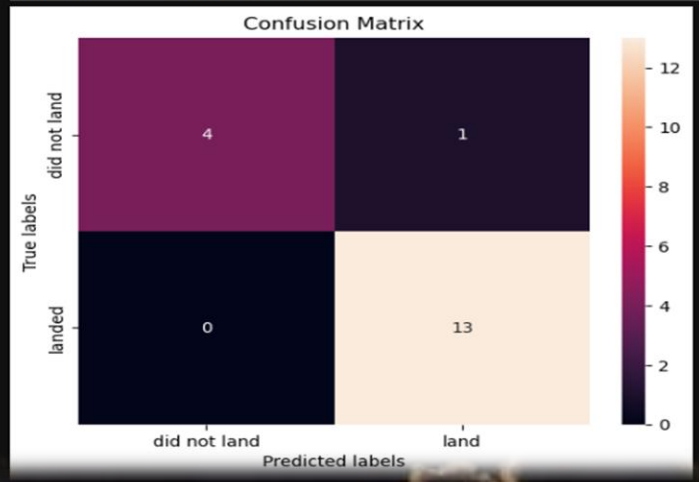
Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Predictive Model Results

Logistic Regression

The model performs very well and give overall accuracy 94%, also since the problem is False Positive the confusion matrix did well as well.

```
195]: GridSearchCV
      estimator: LogisticRegression
        LogisticRegression

196]: print('the parameters of this model ',logreg_cv.best_params_)
      print('the accuracy of these parmaeters ',logreg_cv.best_score_)

      the parameters of this model  {'C': 0.1, 'penalty': 'l2', 'solver': 'lbfgs'}
      the accuracy of these parmaeters  0.8285714285714286

197]: print('The Accuracy of test data for this model (logistic regression) is ',logreg_cv.score(Xtest,Ytest))

      The Accuracy of test data for this model (logistic regression) is  0.9444444444444444

198]: yhat=logreg_cv.predict(Xtest)

199]: plot_confusion_matrix(Ytest,yhat)
```



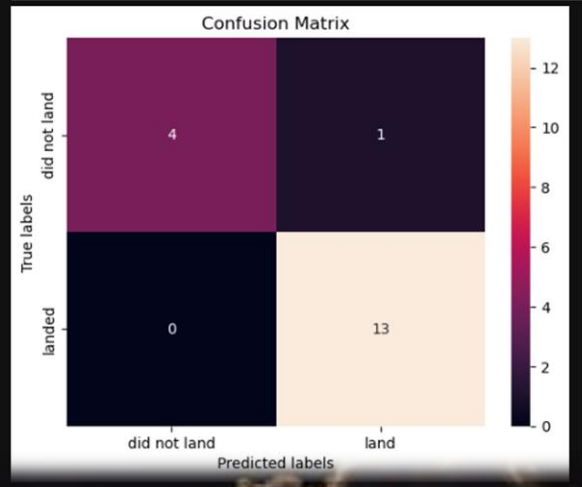
Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Predictive Model Results

SVM

The model performs very well and give overall accuracy 94%, also since the problem is False Positive the confusion matrix did well as well.

```
217]: ▸ GridSearchCV
      ▸ estimator: SVC
          ▸ SVC

218]: print('the best hyperparameters ',svm_cv.best_params_)
      print('the accuarcy of hyperparmeters ',svm_cv.best_score_)

      the best hyperparameters  {'C': 0.03162277660168379, 'gamma': 0.001, 'kernel': 'linear'}
      the accuarcy of hyperparmeters  0.8285714285714286

219]: print('the accuarcy of the test data for this model (SVM) is ',svm_cv.score(Xtest,Ytest))

      the accuarcy of the test data for this model (SVM) is  0.9444444444444444

220]: yhat2=svm_cv.predict(Xtest)

221]: plot_confusion_matrix(Ytest,yhat)
```



Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Predictive Model Results

Decision Tree

The model performs good and give overall accuracy 83%, also since the problem is False Positive the confusion matrix did well as well.

```
estimator: DecisionTreeClassifier
  DecisionTreeClassifier
```

```
print('the best parameters here are ',tree_cv.best_params_)
print('the accuracy here ',tree_cv.best_score_)
```

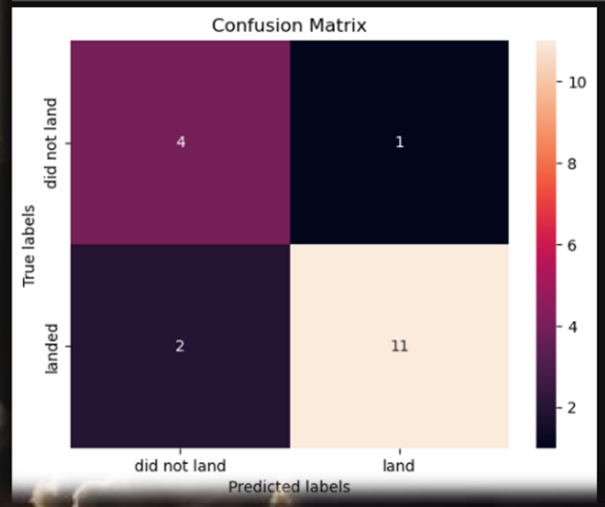
```
the best parameters here are  {'criterion': 'entropy', 'max_depth': 18, 'max_features': 'sqrt', 'min_samples_leaf': 2, 'min_samples_split': 10, 'splitter': 'best'}
the accuracy here  0.8857142857142858
```

```
print('the accuracy of test data for this model (Tree) is ',tree_cv.score(Xtest,Ytest))
```

```
the accuracy of test data for this model (Tree) is  0.8333333333333334
```

```
yhat3=tree_cv.predict(Xtest)
```

```
plot_confusion_matrix(Ytest,yhat3)
```



Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Predictive Model Results

KNN

The model performs good and give overall accuracy 83%, also since the problem is False Positive the confusion matrix did well as well.

```
GridSearchCV
▸ estimator: KNeighborsClassifier
    ▸ KNeighborsClassifier
```

```
print('best parameters for this model ',knn_cv.best_params_)
print('accuracy for these parameters',knn_cv.best_score_)
```

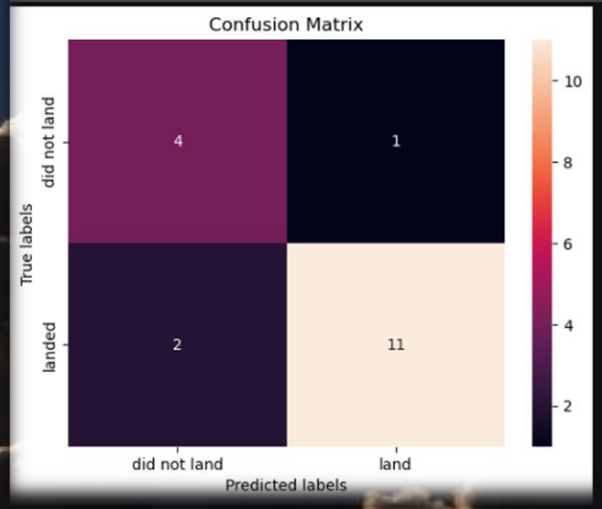
```
best parameters for this model  {'algorithm': 'auto', 'n_neighbors': 3, 'p': 1}
accuracy for these parameters 0.8160714285714284
```

```
print('the accuracy for the test data for this model (KNN) is ',knn_cv.score(Xtest,Ytest))
```

```
the accuracy for the test data for this model (KNN) is  0.8333333333333334
```

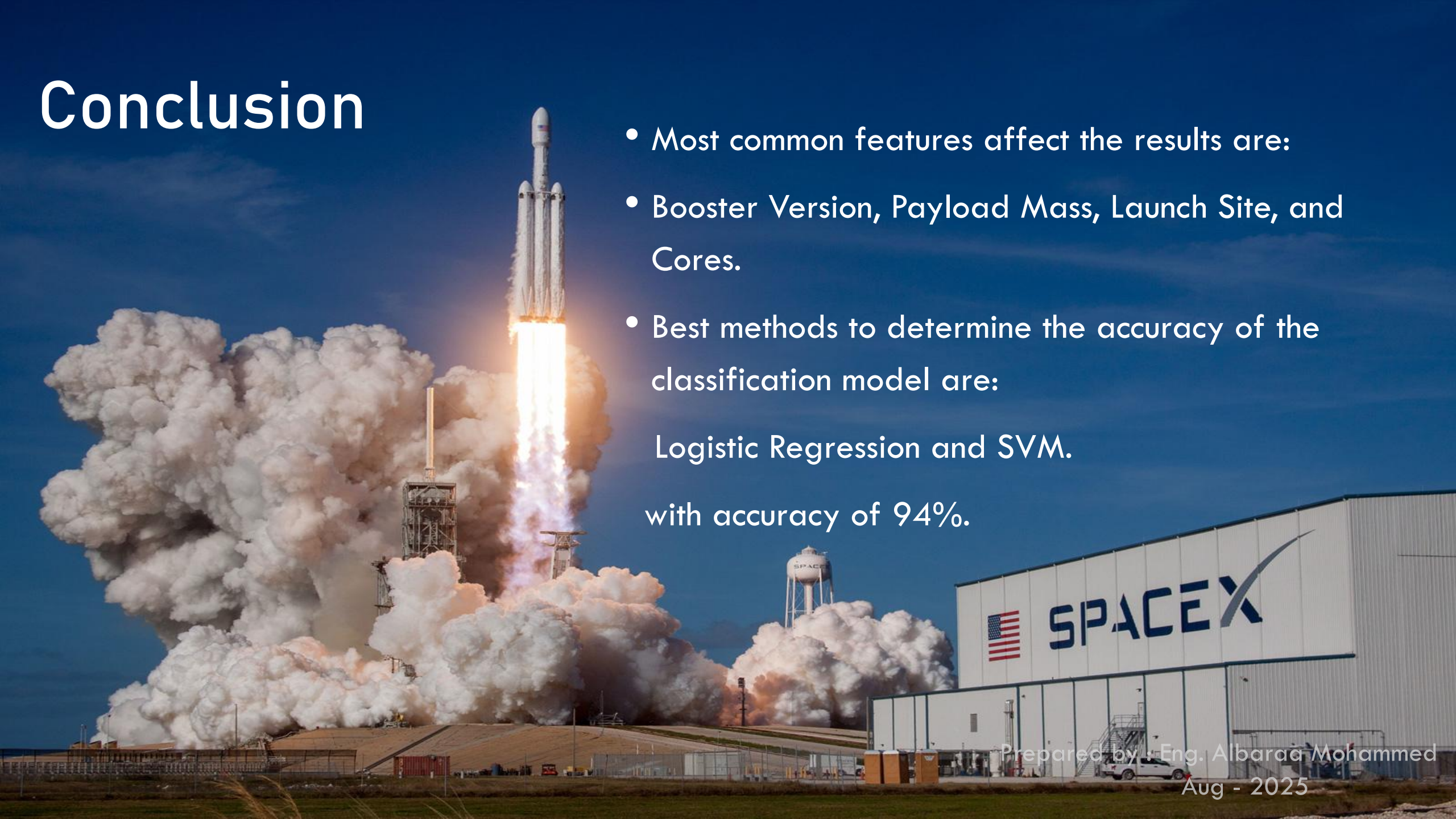
```
yhat4=knn_cv.predict(Xtest)
```

```
plot_confusion_matrix(Ytest,yhat4)
```



Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Conclusion



- Most common features affect the results are:
- Booster Version, Payload Mass, Launch Site, and Cores.
- Best methods to determine the accuracy of the classification model are:

  Logistic Regression and SVM.

  with accuracy of 94%.

Prepared by : Eng. Albaraa Mohammed

Aug - 2025

Appendix

- https://github.com/Albaraaakarman/Falcon-9-First-Stage-Classifier-

Prepared by : Eng. Albaraa Mohammed
Aug - 2025

Thank You !

Prepared by : Eng. Albaraa Mohammed
Aug - 2025