SpaceX Falcon 9 Launch Success Prediction

IBM Data Science Professional Certificate – Final Capstone Project

https://github.com/mustafatansel/spacex-capstone-project

Executive Summary

- This project explores the success classification of SpaceX Falcon 9 rocket launches.
  We collected data using the SpaceX API, performed data wrangling and exploratory data analysis, then built machine learning models to predict launch success.
  Our goal was to identify key factors influencing successful landings and create a model with high prediction accuracy.

Data Collection & Wrangling

- API Source – Used the SpaceX v4 Launches API (/v4/launches/past) to download raw JSON data.
- Normalization – Converted the nested JSON into a flat pandas DataFrame with pd.json_normalize.
- Filtering – Removed Falcon 1 records, keeping 94 Falcon 9 launches.
- Cleaning – Replaced PayloadMass null values with the column mean; kept LandingPad nulls for later one-hot encoding.
- Output – Saved the cleaned dataset as dataset_part_1.csv (94 rows × 14 columns).

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
0	1	2006-03-24	Falcon 1	20.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin1A	167.743129	9.047721
1	2	2007-03-21	Falcon 1	NaN	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2A	167.743129	9.047721
2	4	2008-09-28	Falcon 1	165.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin2C	167.743129	9.047721
3	5	2009-07-13	Falcon 1	200.0	LEO	Kwajalein Atoll	None None	1	False	False	False	None	NaN	0	Merlin3C	167.743129	9.047721
4	6	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857

	FlightNumber	Date	BoosterVersion	PayloadMass	Orbit	LaunchSite	Outcome	Flights	GridFins	Reused	Legs	LandingPad	Block	ReusedCount	Serial	Longitude	Latitude
4	1	2010-06-04	Falcon 9	NaN	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0003	-80.577366	28.561857
5	2	2012-05-22	Falcon 9	525.0	LEO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0005	-80.577366	28.561857
6	3	2013-03-01	Falcon 9	677.0	ISS	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B0007	-80.577366	28.561857
7	4	2013-09-29	Falcon 9	500.0	PO	VAFB SLC 4E	False Ocean	1	False	False	False	None	1.0	0	B1003	-120.610829	34.632093
8	5	2013-12-03	Falcon 9	3170.0	GTO	CCSFS SLC 40	None None	1	False	False	False	None	1.0	0	B1004	-80.577366	28.561857
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
89	86	2020-09-03	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	2	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1060	-80.603956	28.608058
90	87	2020-10-06	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	3	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	13	B1058	-80.603956	28.608058
91	88	2020-10-18	Falcon 9	15600.0	VLEO	KSC LC 39A	True ASDS	6	True	True	True	5e9e3032383ecb6bb234e7ca	5.0	12	B1051	-80.603956	28.608058
92	89	2020-10-24	Falcon 9	15600.0	VLEO	CCSFS SLC 40	True ASDS	3	True	True	True	5e9e3033383ecbb9e534e7cc	5.0	12	B1060	-80.577366	28.561857
93	90	2020-11-05	Falcon 9	3681.0	MEO	CCSFS SLC 40	True ASDS	1	True	False	True	5e9e3032383ecb6bb234e7ca	5.0	8	B1062	-80.577366	28.561857

90 rows × 17 columns

Exploratory Data Analysis (EDA)

- Display 5 records where launch sites begin with the string 'CCA’

cid	name	type	notnull	dflt_value	pk
0	Date	TEXT	0	None	0
1	Time (UTC)	TEXT	0	None	0
2	Booster_Version	TEXT	0	None	0
3	Launch_Site	TEXT	0	None	0
4	Payload	TEXT	0	None	0
5	PAYLOAD_MASS__KG_	INT	0	None	0
6	Orbit	TEXT	0	None	0
7	Customer	TEXT	0	None	0
8	Mission_Outcome	TEXT	0	None	0
9	Landing_Outcome	TEXT	0	None	0

Exploratory Data Analysis (EDA)

- Display the total payload mass carried by boosters launched by NASA (CRS)

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

Exploratory Data Analysis (EDA)

- Rank the count of landing outcomes (such as Failure (drone ship) or Success (ground pad)) between the date 2010-06-04 and 2017-03-20, in descending order.¶

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

[60]:

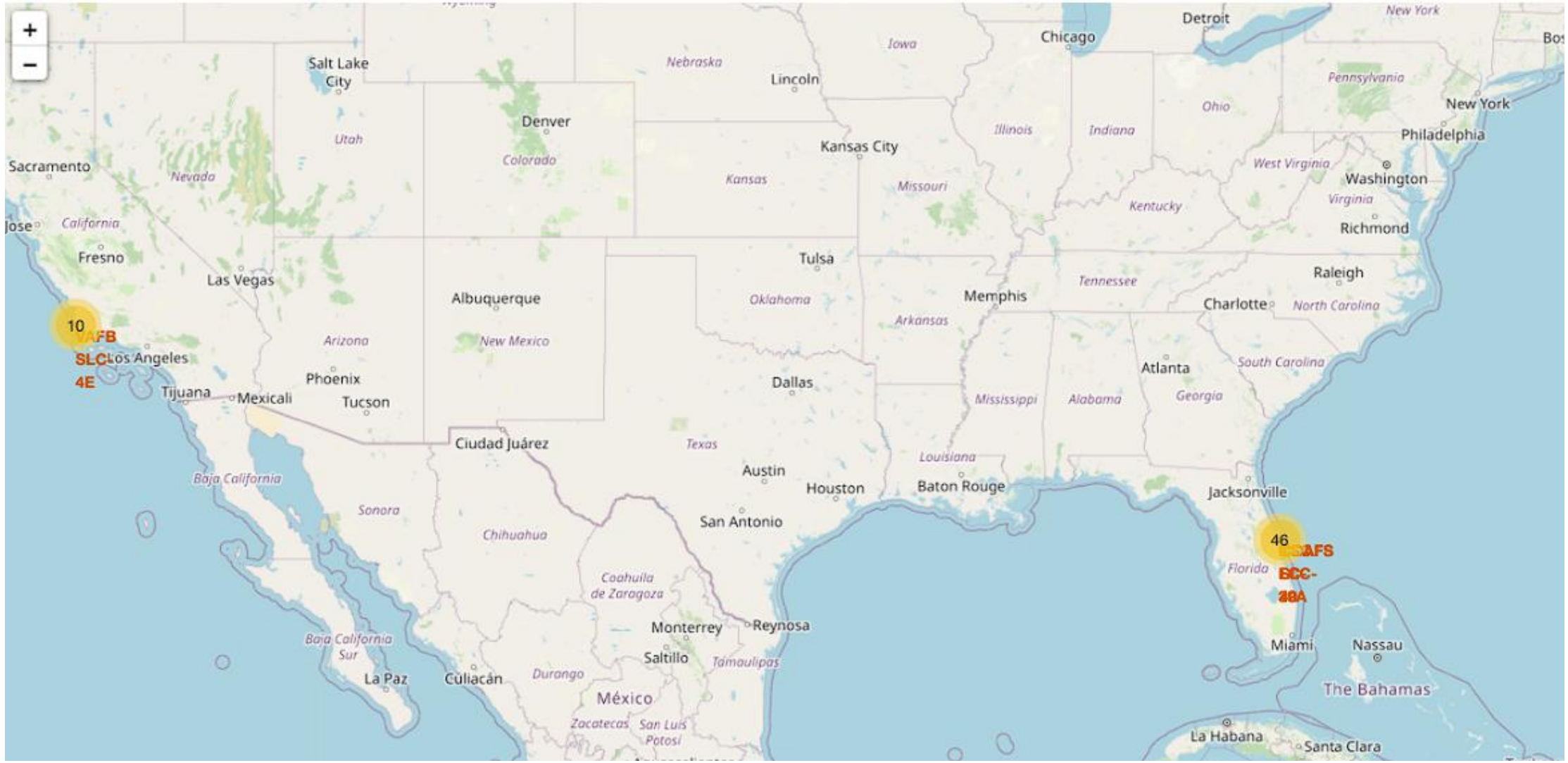
Landing_Outcome	count
Success (drone ship)	5
Failure (drone ship)	5
Success (ground pad)	3
Precluded (drone ship)	1

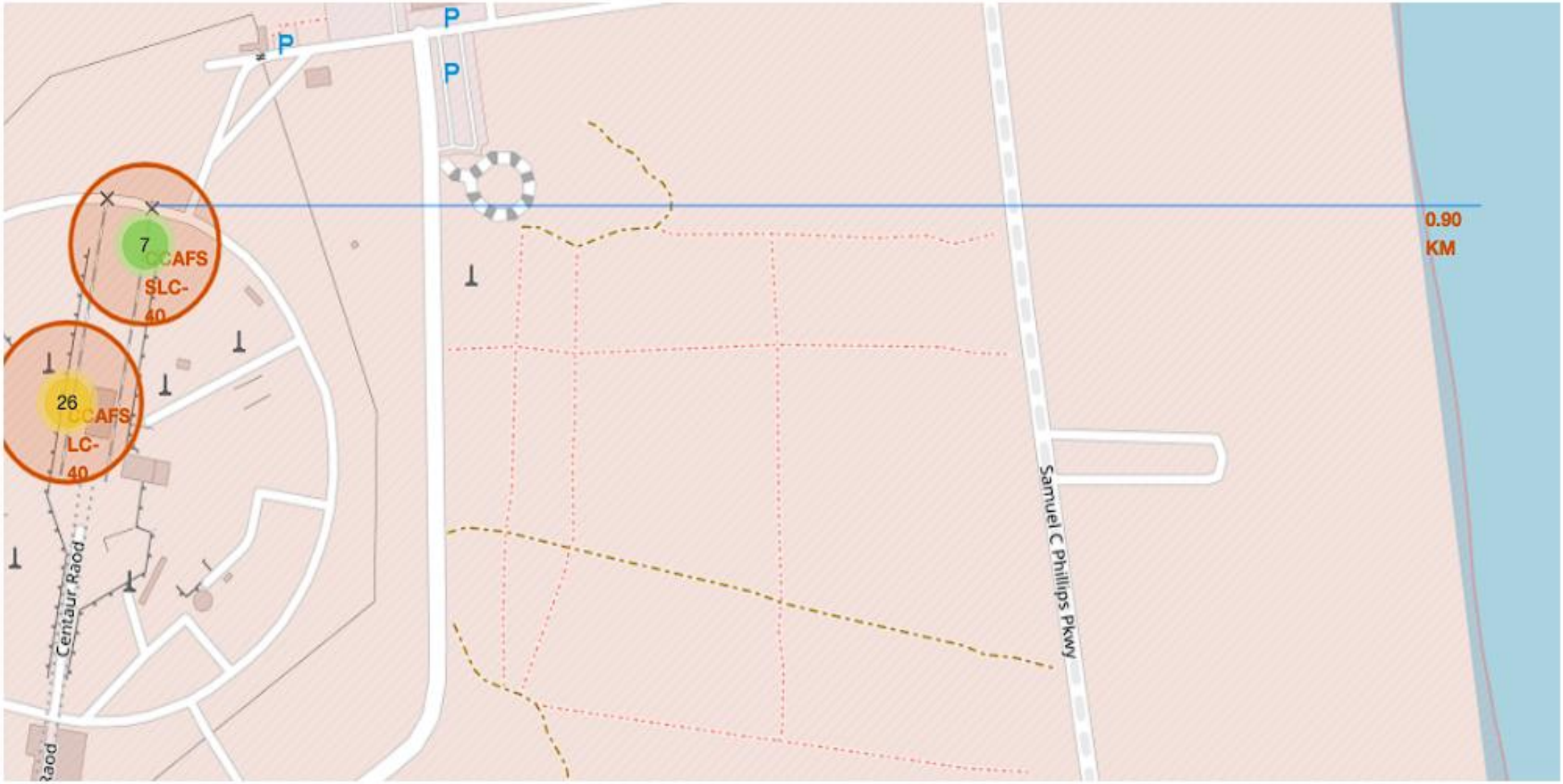
Interactive Map

- Mark the success/failed launches for each site on the map

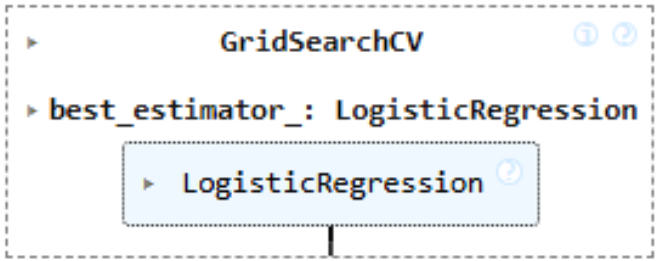
[18]:

	Launch Site	Lat	Long	class
46	KSC LC-39A	28.573255	-80.646895	1
47	KSC LC-39A	28.573255	-80.646895	1
48	KSC LC-39A	28.573255	-80.646895	1
49	CCAFS SLC-40	28.563197	-80.576820	1
50	CCAFS SLC-40	28.563197	-80.576820	1
51	CCAFS SLC-40	28.563197	-80.576820	0
52	CCAFS SLC-40	28.563197	-80.576820	0
53	CCAFS SLC-40	28.563197	-80.576820	0
54	CCAFS SLC-40	28.563197	-80.576820	1
55	CCAFS SLC-40	28.563197	-80.576820	0


10
AFB
SLC
4E
46
AFS
ECC-
39A
Los Angeles
Tijuana
Mexicali
Phoenix
Tucson
Las Vegas
Fresno
Sacramento
Salt Lake City
Denver
Albuquerque
Ciudad Juárez
Dallas
Austin
Houston
San Antonio
Tulsa
Kansas City
Lincoln
Chicago
Detroit
New York
Philadelphia
Washington
Richmond
Raleigh
Charlotte
Memphis
Atlanta
Jacksonville
Baton Rouge
Miami
Nassau
The Bahamas
La Habana
Santa Clara
Monterrey
Reynosa
Saltillo
Culiacán
La Paz
México
Nevada
Utah
Colorado
Kansas
Nebraska
Iowa
Missouri
Illinois
Indiana
Ohio
Pennsylvania
West Virginia
Virginia
Kentucky
Tennessee
North Carolina
South Carolina
Georgia
Alabama
Mississippi
Arkansas
Louisiana
Oklahoma
Texas
New Mexico
Arizona
California
Baja California
Baja California Sur
Sonora
Chihuahua
Coahuila de Zaragoza
Durango
Zacatecas
San Luis Potosí
Tamaulipas
Florida


P
P
P
7
CCAFS
SLC-
40
26
CCAFS
LC-
40
0.90
KM
Samuel C Phillips Pkwy
Centaur Raod

Predictive Analysis



[64]:
```
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import GridSearchCV

parameters = {
    'criterion': ['gini', 'entropy'],
    'splitter': ['best', 'random'],
    'max_depth': [x for x in range(1, 10)],
    'max_features': ['auto', 'sqrt'],
    'min_samples_leaf': [1, 2, 4],
    'min_samples_split': [2, 5, 10]
}
tree = DecisionTreeClassifier()
tree_cv = GridSearchCV(tree, parameters, cv=10)
tree_cv.fit(X_train, Y_train)

print("Tuned hyperparameters: (best parameters) ", tree_cv.best_params_)
print("Accuracy :", tree_cv.best_score_)
```

```
Tuned hyperparameters: (best parameters)  {'criterion': 'gini', 'max_depth': 4, 'max_features': 'sqrt', 'min_samples_leaf': 1, 'min_samples_split': 10, 'splitter': 'random'}
Accuracy : 0.8910714285714286
```

Conclusion

- In this project, we explored SpaceX Falcon 9 launch records to predict the likelihood of a successful mission.
- We performed data collection, wrangling, and feature engineering to prepare a clean dataset.
- We applied and compared several machine learning models including Logistic Regression, Decision Tree, KNN, and SVM.
- After hyperparameter tuning, the Decision Tree Classifier achieved the highest test accuracy of 89.1%.
- Our analysis demonstrates that machine learning can be a valuable tool to support space mission planning and reliability assessment.