Predicting landing outcomes for the Falcon 9 by SpaceX

By Nandini Tejas Mehta

Background

- Significance of the Falcon 9:
  - Reusability
  - Cheaper space travel
  - More sustainable

- Project objective:
  - Objective: Using Falcon 9's historical data to predict the landing outcome of the first stage.

- Deliverables:
  - EDA on the Falcon 9 dataset.
  - Interactive map to view launch sites.
  - Robust machine learning model.

Previous Findings Vs Current Findings

- Attributes analyzed:

Launch outcome and land outcome correlation	Launch success rates
Launch sites	Orbit types
Payload types	Payload mass range

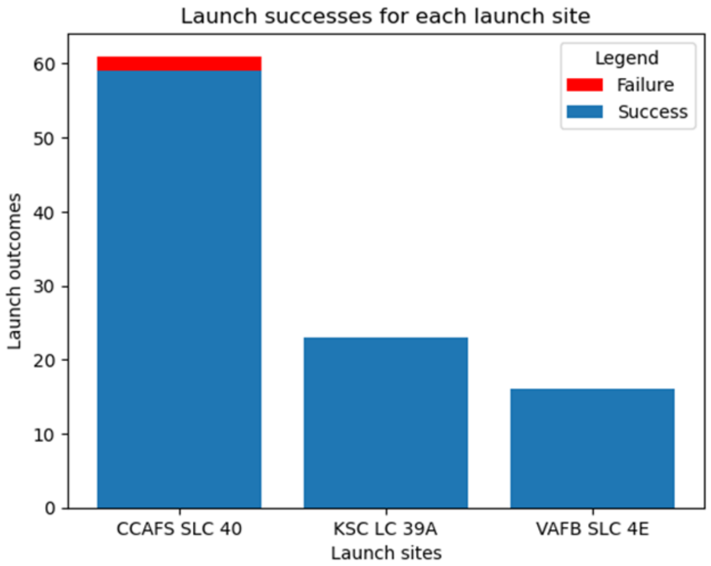
- Key Findings:
  - 86.49% launch successes= landing successes
  - Most used site: CCAFS
  - Positive success trend
  - Most used payload: Satellite
  - Most frequent orbits: GTO, ISS & VLEO
  - Payload mass has no drastic effect on success

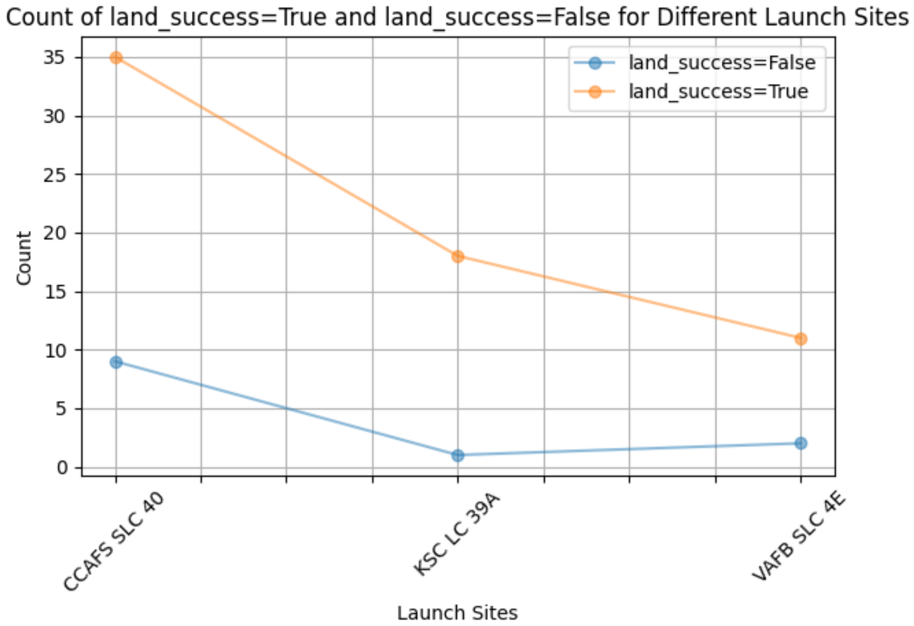
- Attributes analyzed:

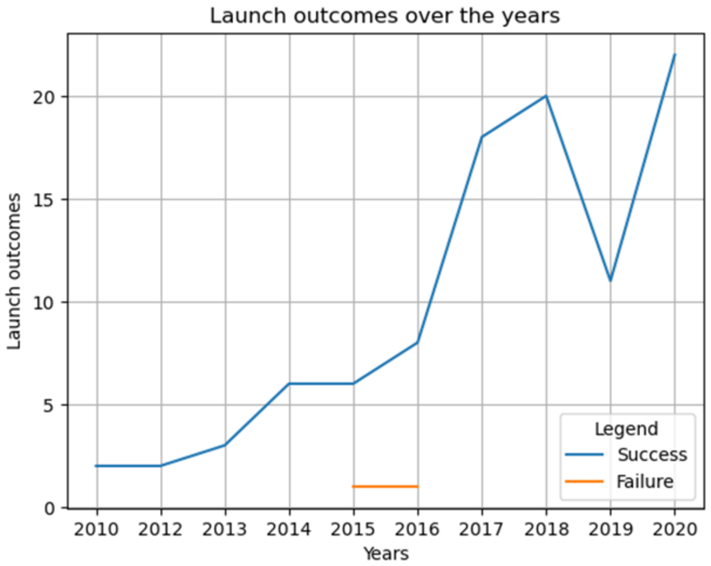
Land success and landing vehicle	Grid fins and legs
Launch sites	Orbit types
Payload type and mass effect	Reuse and reuse count

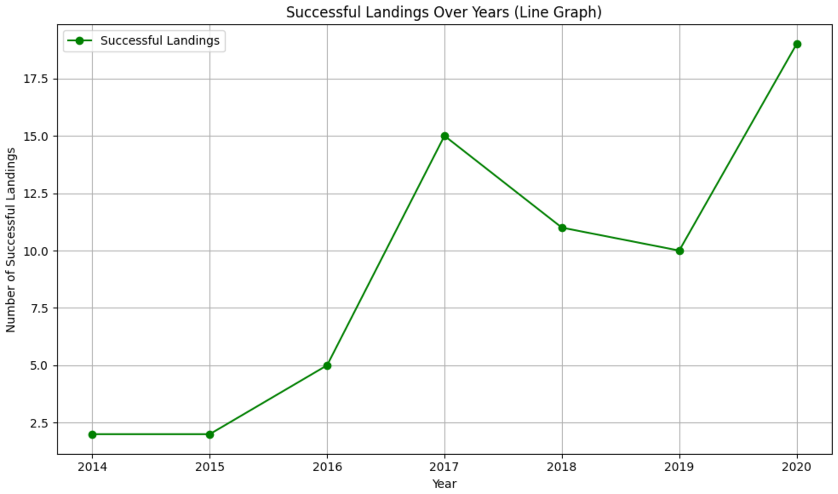
- Key Findings:
  - Most used site: CCAFS
  - Positively increasing landing success trend
  - Most common kind of landing: ASDS
  - Most used landing vehicle: OCISLY
  - Most used payload: Satellite
  - Most frequent orbits: GTO, ISS & PO

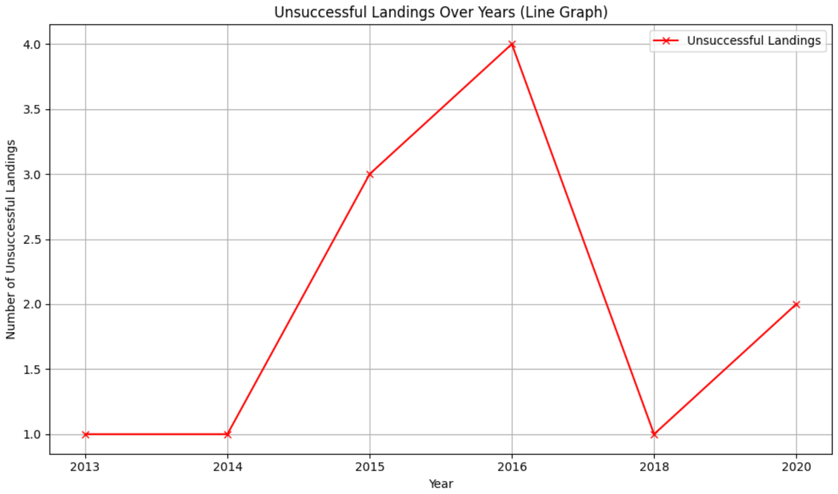
Previous Findings Vs Current Findings



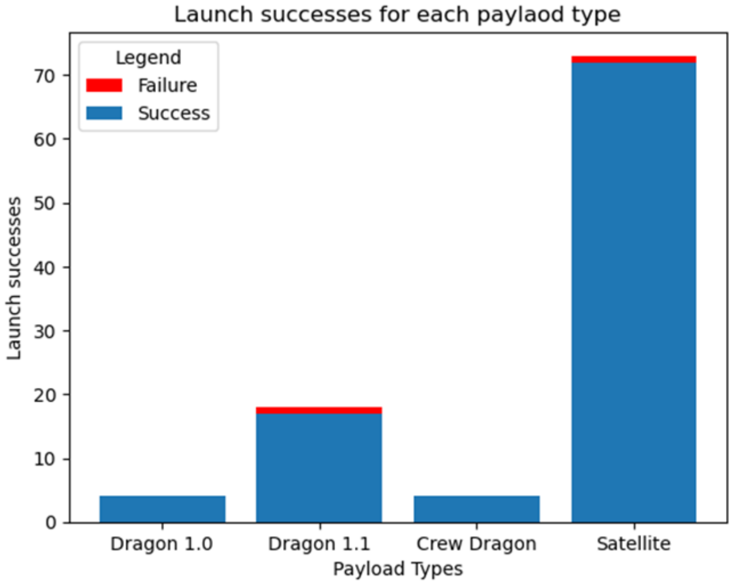


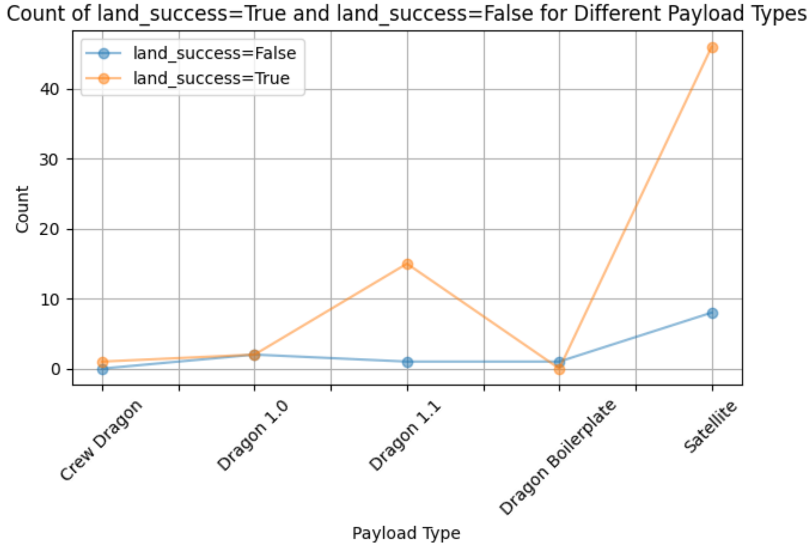


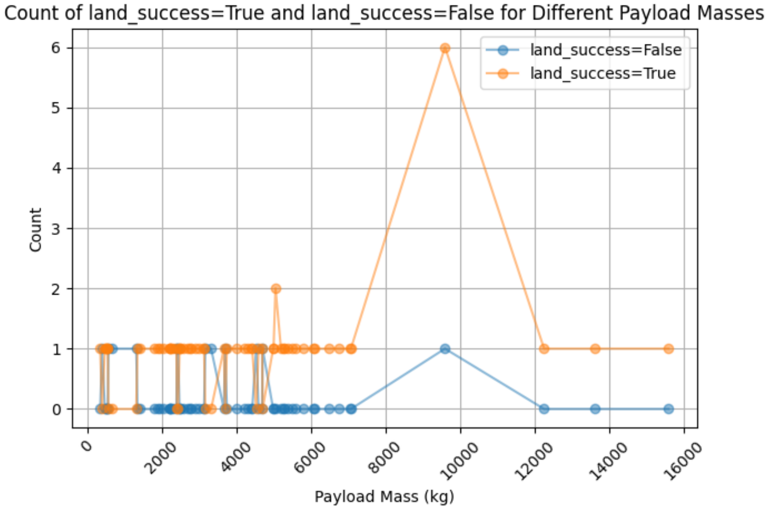


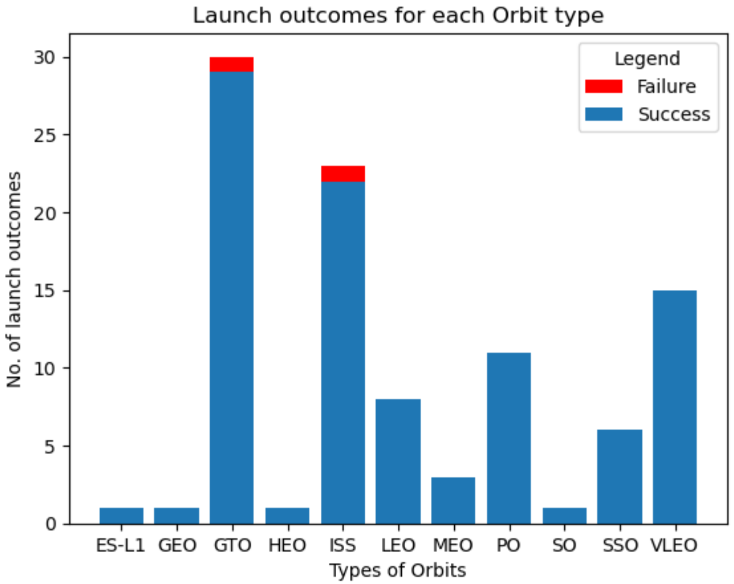


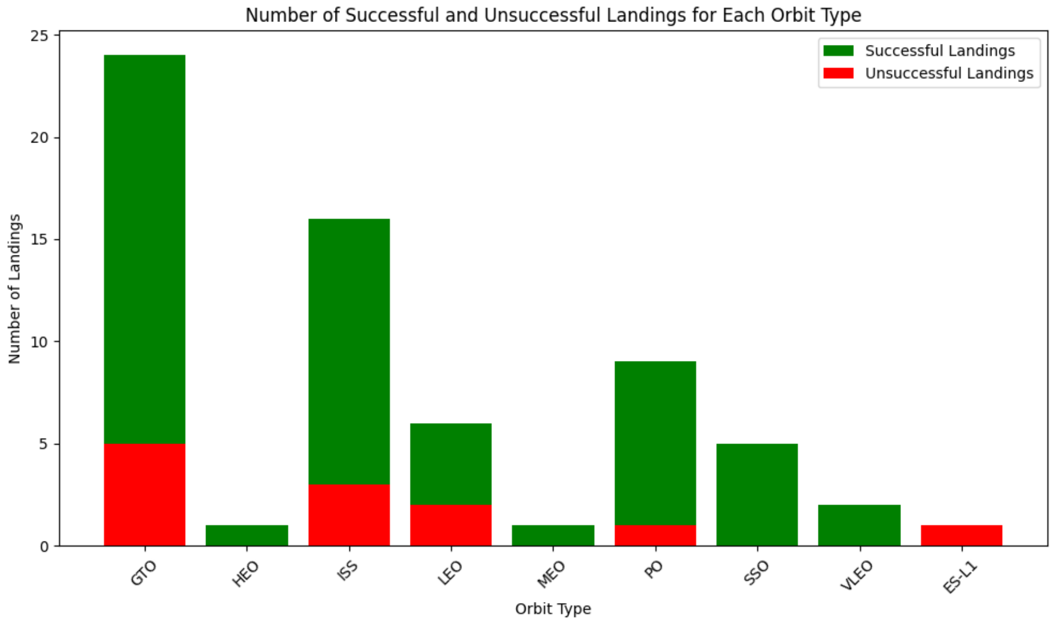
Previous Findings Vs Current Findings



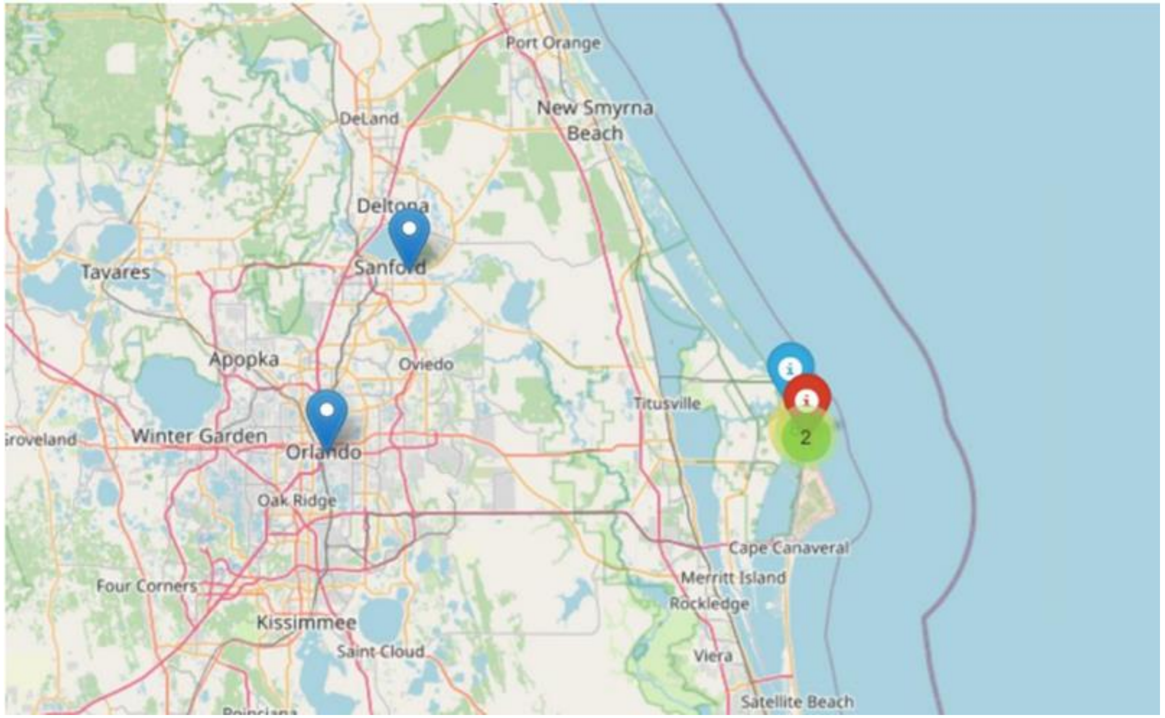


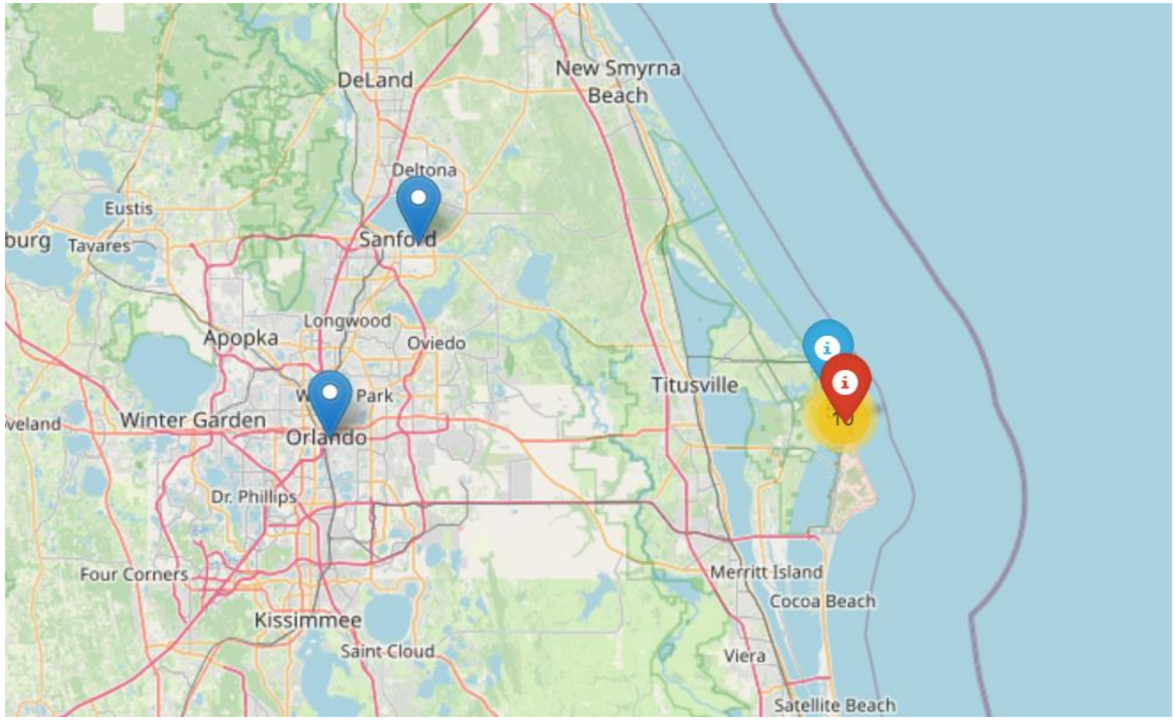






Previous Findings Vs Current Findings





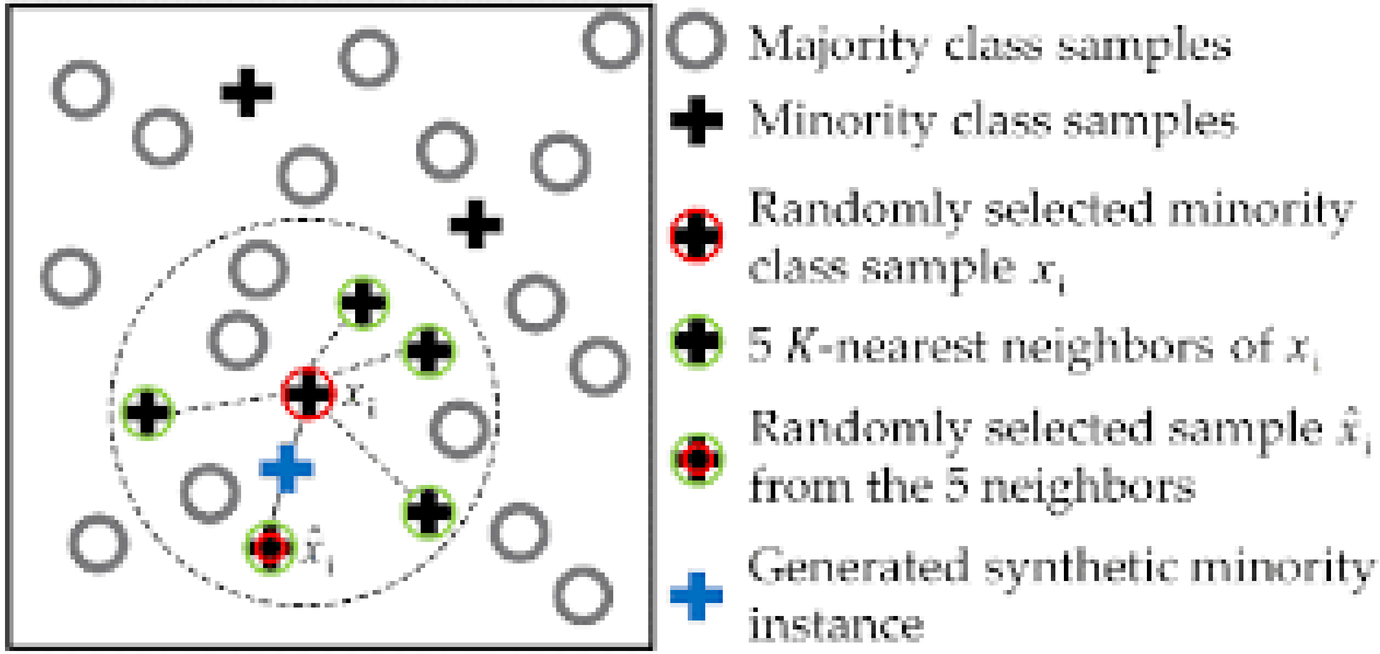
ML dataset preparation

- Cleaning the data:
  - Remove all rows where the landing intent was False, as if the launch had no intent to land the outcome was going to be False regardless of the characteristics of the first stage.
  - The objective of the ML model is to predict the outcome of a first stage whose mission is to land successfully to be reused.
- This left us with 76 rows out of which only 12 belonged to the land_success= False class.
- This was a problem as the dataset was small and imbalanced. If used to make a machine learning model, it would not be a robust or well-generalized model.
- Bias towards the majority class

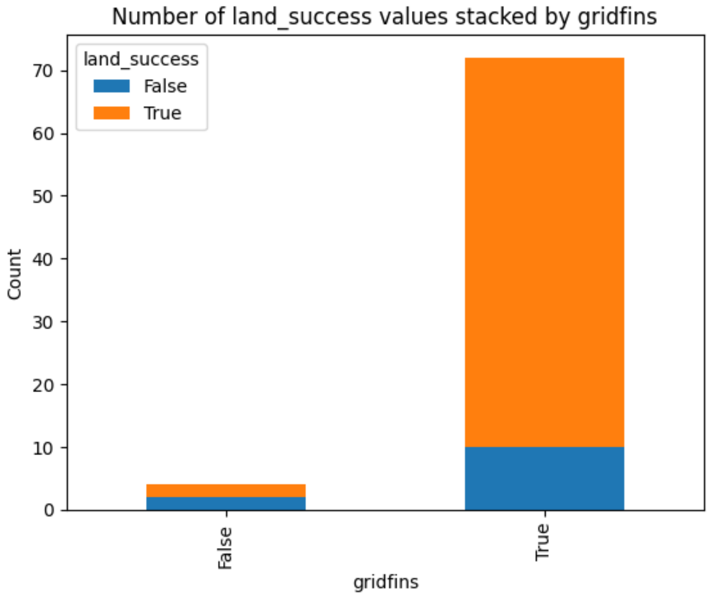
SMOTE on the dataset

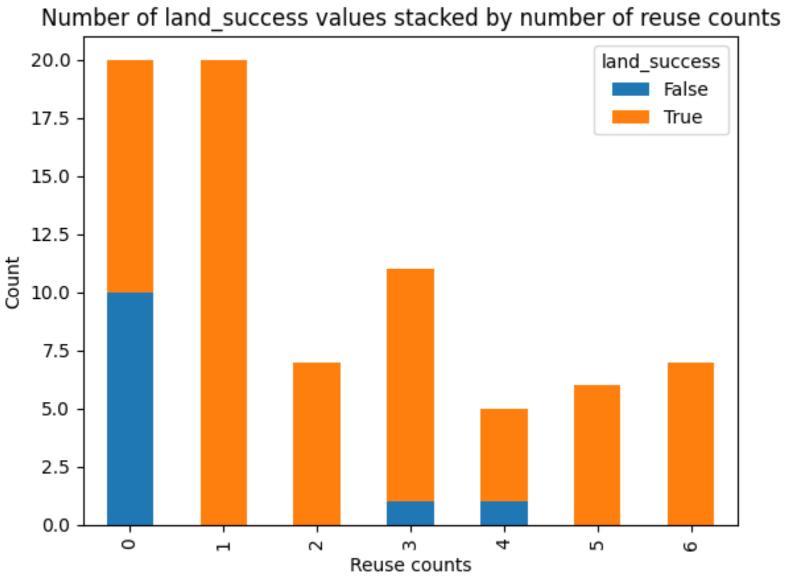
- To deal with the dataset imbalance we can use SMOTE techniques.

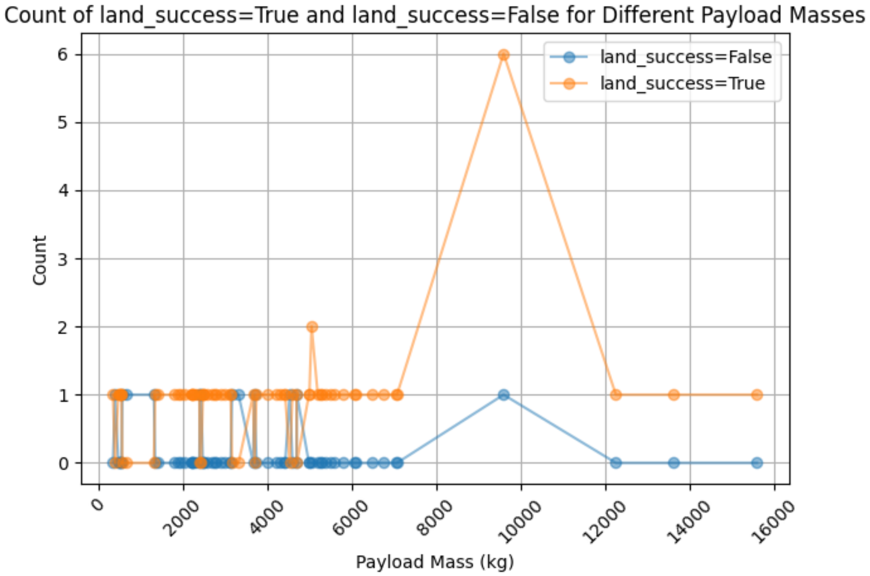
- Synthetic Minority Over-sampling Technique (SMOTE)
- Over sample the minority class, in this case land_success= False.

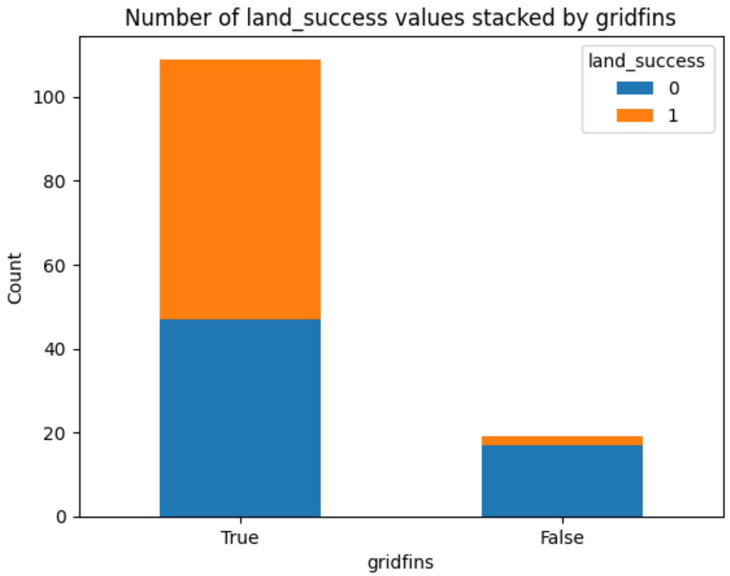


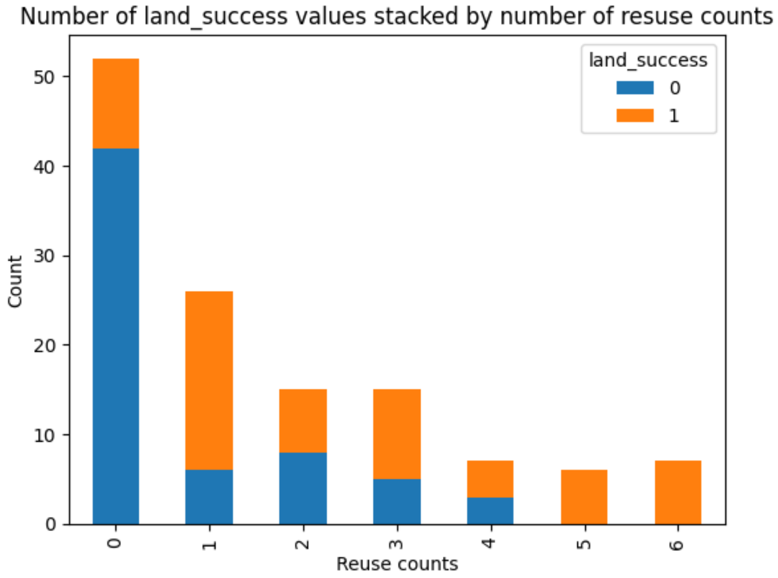
Before and After SMOTE

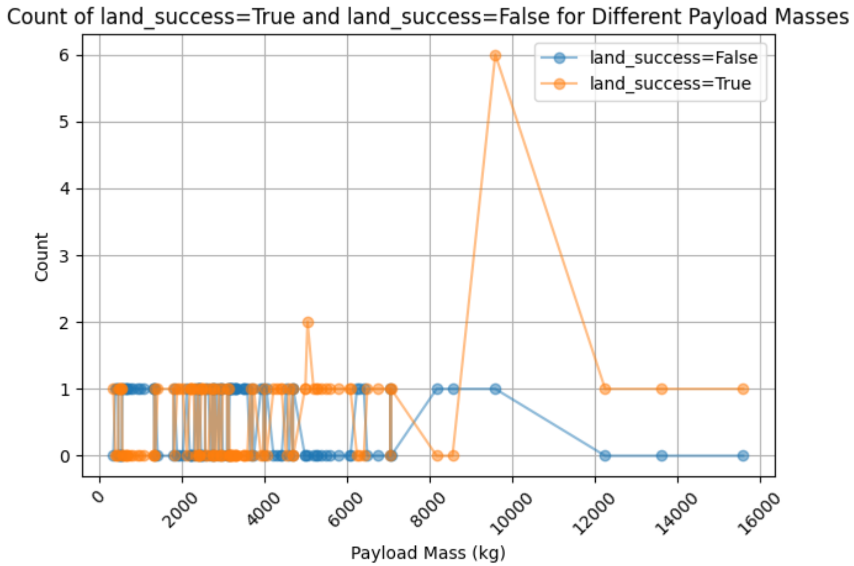












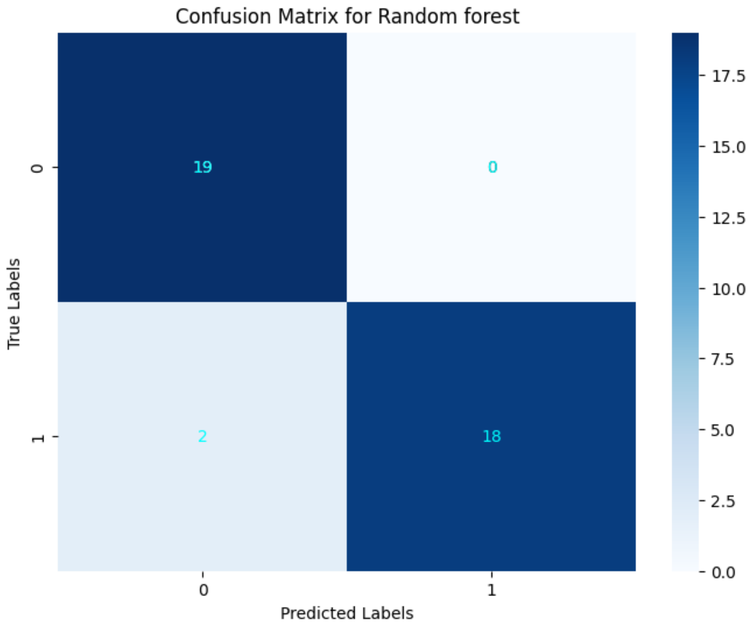
ML model selection

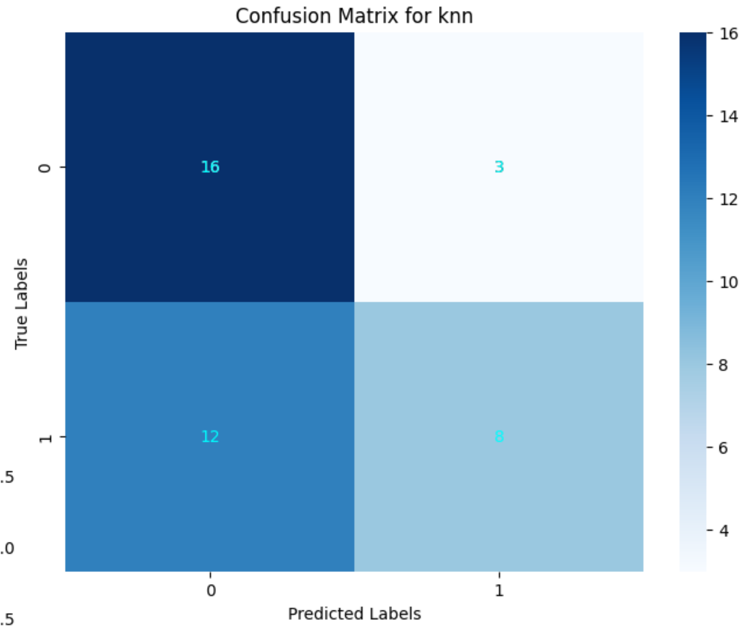
- Knn Model:
- Best Parameters: {'n_neighbors': 20}
- Accuracy: 0.6153846153846154
- F1 Score: 0.5161290322580645

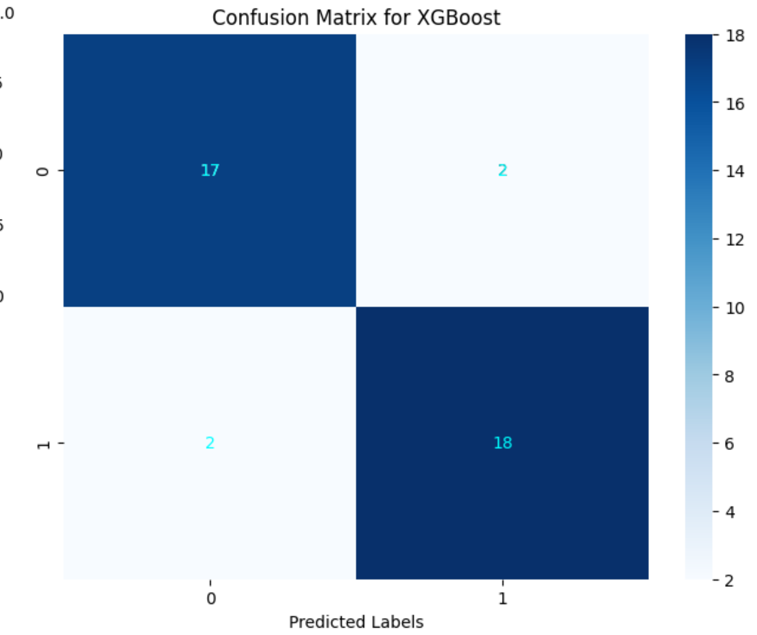
- Random Forest:
- Best Parameters: {'n_estimators': 250}
- Accuracy: 0.9487179487179487
- F1 Score: 0.9473684210526316

- Gradient Boost:
- Accuracy: 0.8974358974358975
- F1 Score: 0.9

- All were with a train-test split of 70-30 and 10 fold cross validation.





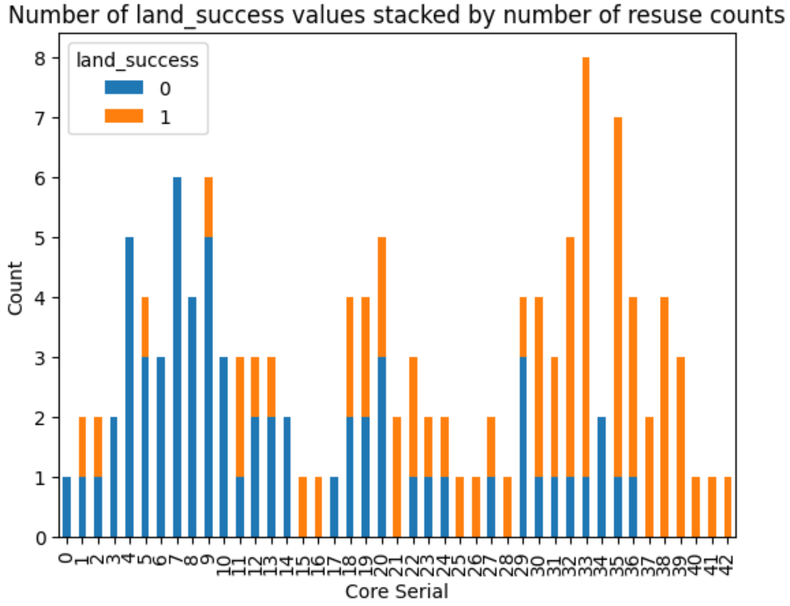


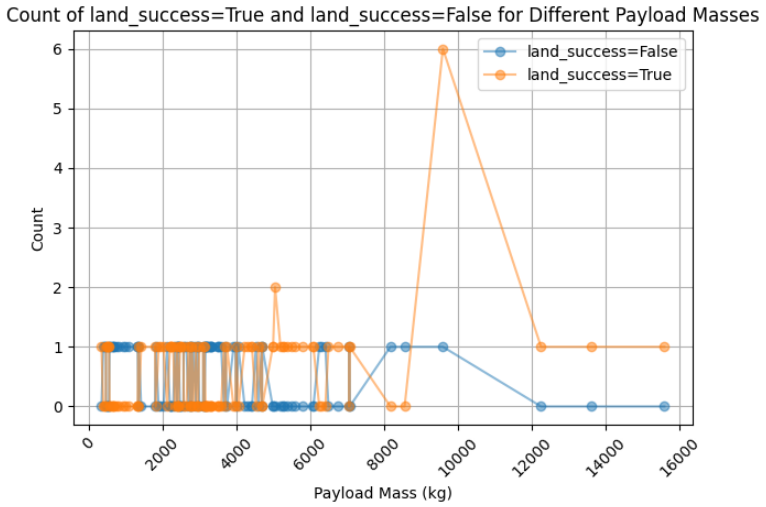
ML model selection

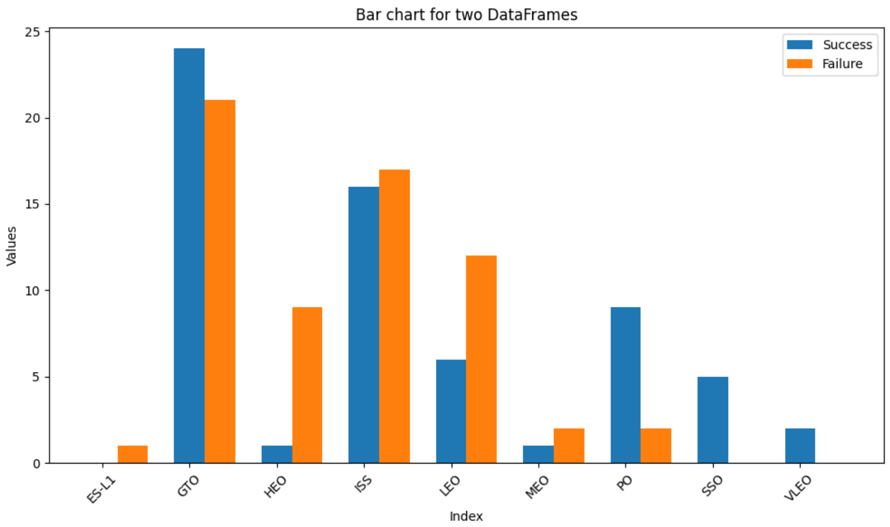
- Random Forest Model and the importance it weighs for each feature:

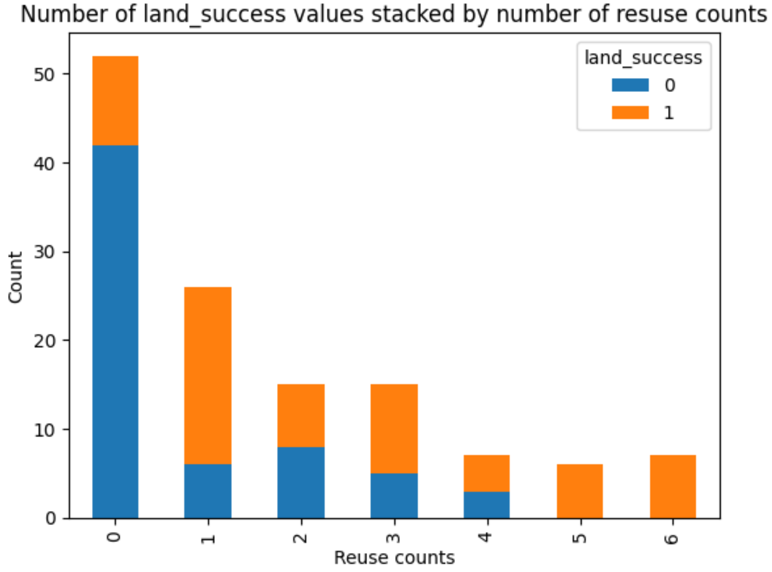
Feature Importances:

	Feature	Importance
5	core_serial	0.216627
2	payload_mass_kg	0.152619
1	reuse_count	0.124570
8	orbit	0.100794
3	launch_year	0.098141
9	payload_type	0.068327
0	flight	0.067720
7	landing_vehicle	0.049187
4	site_names	0.040426
6	landing_type	0.030714
12	legs	0.024591
10	gridfins	0.011701
11	reused	0.010354
13	landing_intent	0.004229









Improvements and conclusion

- Improvements:
  - Can use GANS to create more synthetic data, to have bigger training and testing sets, this will help the model learn better and be more robust towards unseen data.
  - With a much larger dataset it would also enable us to use a more complex model such as a neural net.

- Conclusion:
- Using historical data and SMOTE we were able to create a robust RandomForest Classifier which predicts the landing outcome with an accuracy of 0.949.
- This model mainly considered the core_serial, payload_mass, reuse_count and orbit type to be important factors that affect the landing outcome of the first stage.