“**SpaceX Launch Outcome prediction and Reusability Impact Analysis**”

Definition:

It aims to analyze and predict the outcome of SpaceX rocket launches using combination of statistical analysis and machine learning with growing interest in space commercialization and reusable rocket technology, the study investigates role of reusability in mission success and payload performance.

Description Of Dataset:

1. SpaceX Launch Data

* The dataset comprises **186 rows** and **10 columns**, each representing a unique SpaceX launch with details on the mission, rocket, payload, and outcome.
* Key columns include mission\_name, Rocket Type, Reusability (True/False), Payload Mass (kg), Payload Type, Orbit, Country, Launch Success, launch\_date, and launch\_year.
* The Launch Success field is binary (1.0 for success, 0.0 for failure), and Reusability indicates whether the rocket was reused in the mission.
* Launches span across multiple years (launch\_year) and are associated with various countries (Country), allowing for both time-based and location-based analysis.

Steps Performed:

1. Data Pre-Processing:
   * Verified no missing values and identified class imbalance.
   * Analyzed the impact of rocket reusability on launch success.
2. Feature Engineering and Analysis:

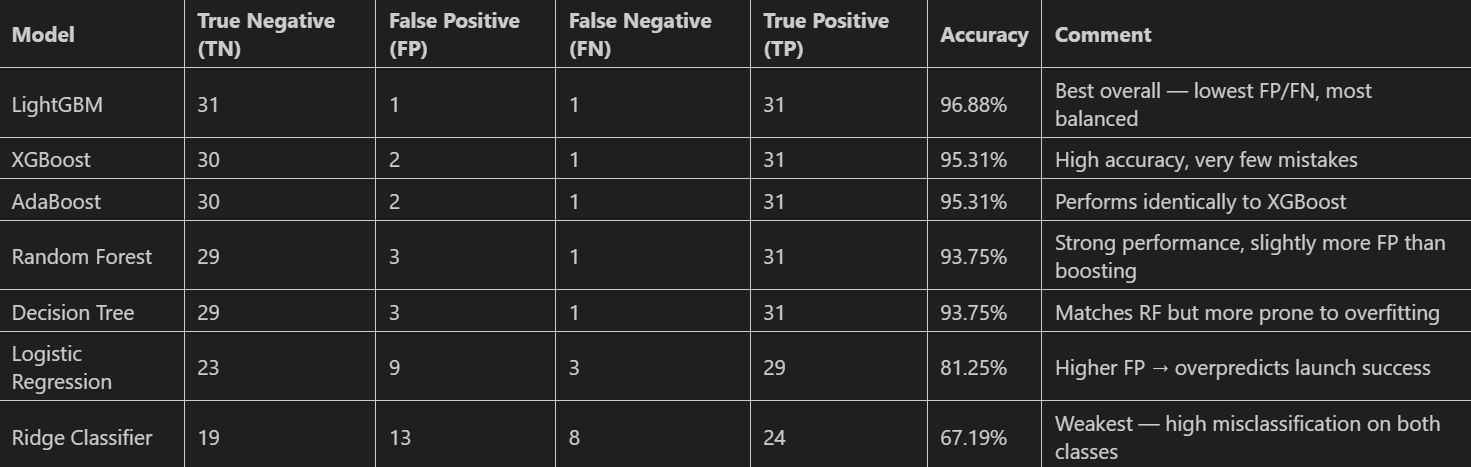
* Explored categorical features like Reusability, Orbit, and Rocket Type.
* Visualized launch outcomes and feature distributions.

1. Model Training and Evaluation:

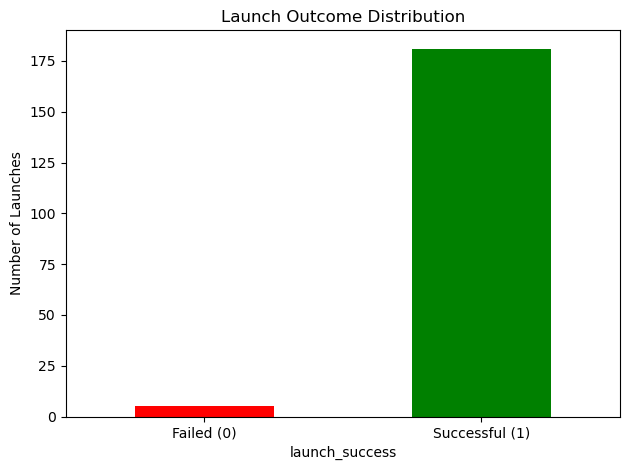
* Split data into training/testing sets.
* Applied multiple ML models and evaluated with ROC curves and AUC.

1. Model Comparison:

* Compared confusion matrices and accuracy of each model.



Techniques used:



The dataset is highly imbalanced as the Failed launches are very ‘minor’ and Successful launches falls in ‘majority’ class.

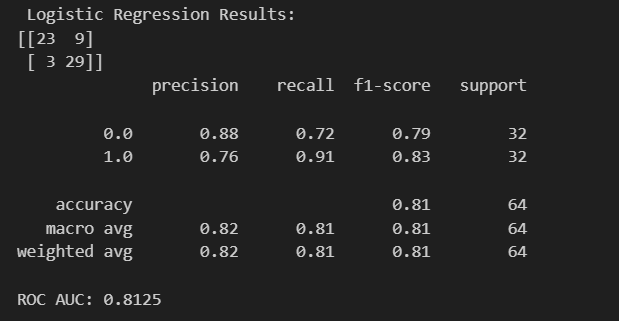
To balance the distribution SMOTE and SMOTETomek is used. SMOTE because it synthetically generates samples but as it produces ambiguity when the minority and majority classes are overlapping. So, to cover this, SMOTETomek is used when overlapping or fuzzy boundaries. But as, SMOTETomek needs k-neighbours = 5, and one more to generate synthetic points.

Models Used:

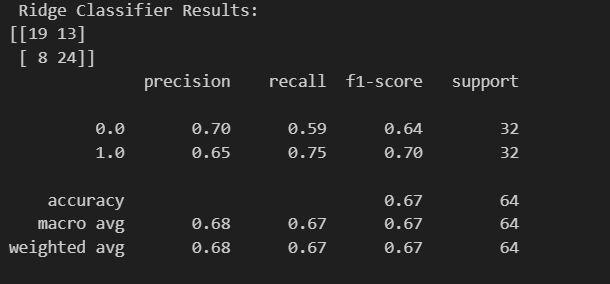
There are three categories of models: Linear Models, Tree-Based Models, Boosting Models.

1. Linear Models:

Logistic Regression:

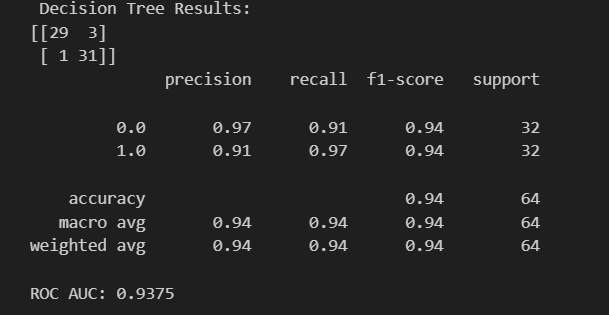
* It's a baseline classification model ideal for imbalanced binary problems with interpretable coefficients.

Ridge Classifier:

* Ridge Classifier is a regularized version of Logistic Regression, used to prevent overfitting by penalizing large coefficients—especially useful when features are correlated.

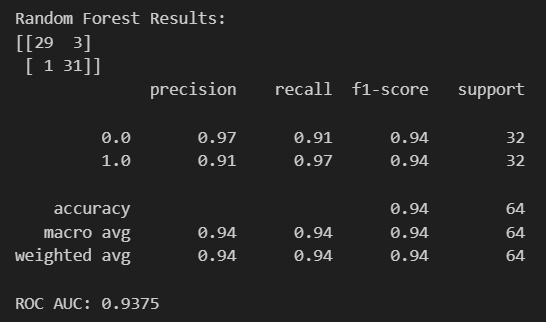
1. Tree-Based Models:

Decision Tree:

* Offers rule-based decisions, easy visualization, and handles non-linear relationships well.

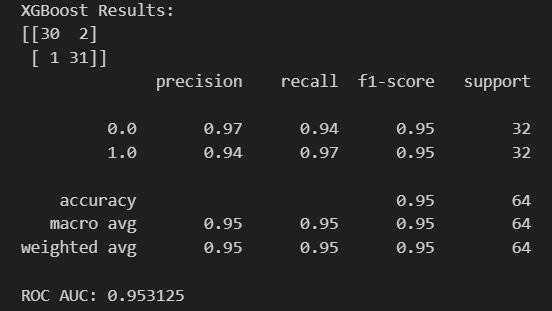
Random Forest:

* Reduces variance and improves accuracy by combining multiple decision trees.

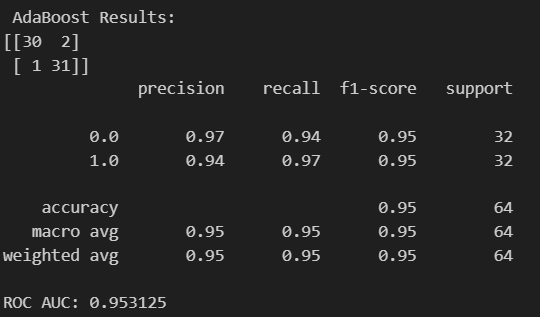


1. Boosting Models:

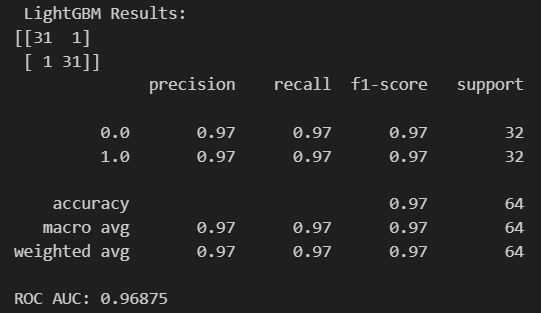
XG Boost: Highly efficient gradient boosting method with built-in handling of class imbalance.



AdaBoost: Simple boosting algorithm that improves weak learners by focusing on errors iteratively.



LightGBM: Extremely fast gradient boosting framework optimized for speed and memory usage.



Reusability Impact Analysis

* **Reusability as a Key Feature**: The dataset includes a Reusability column indicating whether a rocket was reused (True) or not (False), making it a critical factor in analyzing launch success trends.
* **100% Success for Reusable Rockets**: All launches marked as reusable showed a **100% success rate**, indicating a strong correlation between reusability and mission reliability.
* **Lower Success in Non-Reusable Rockets**: Non-reusable rockets had a mixed performance with both successful and failed launches, contributing to model learning on what affects failure.
* **Imbalance in Reusability Distribution**: The majority of rockets in the dataset were non-reusable, emphasizing the technological shift SpaceX made over time toward reusable systems.
* **Impact on Predictive Modeling**: Reusability emerged as a highly influential feature across all ML models, contributing significantly to higher AUC scores and improved classification accuracy.