Successful Rocket Landing to Increase Revenue

Executive summary report for the Sales Project Team Prepared by Data Science

ISSUE / PROBLEM

SpaceX seeks to improve its proposal bid with succcesfuly landing the rockets. The following question arise:

What's likely to successfully land a rocket?

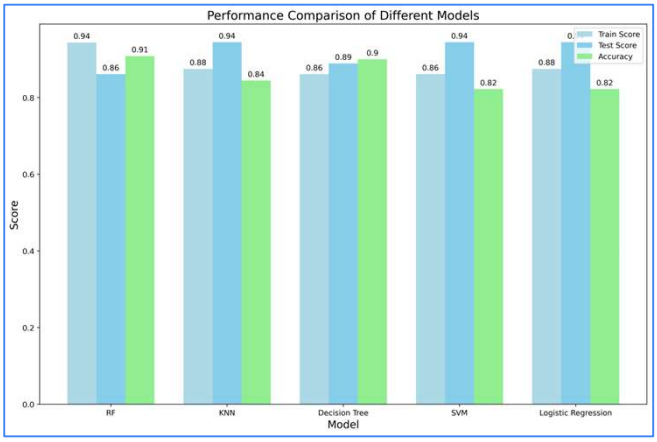
RESPONSE

Since the variable we are seeking to predict is categorical, the team could build either a logistic regression or a tree-based machine learning model or a XGboost.

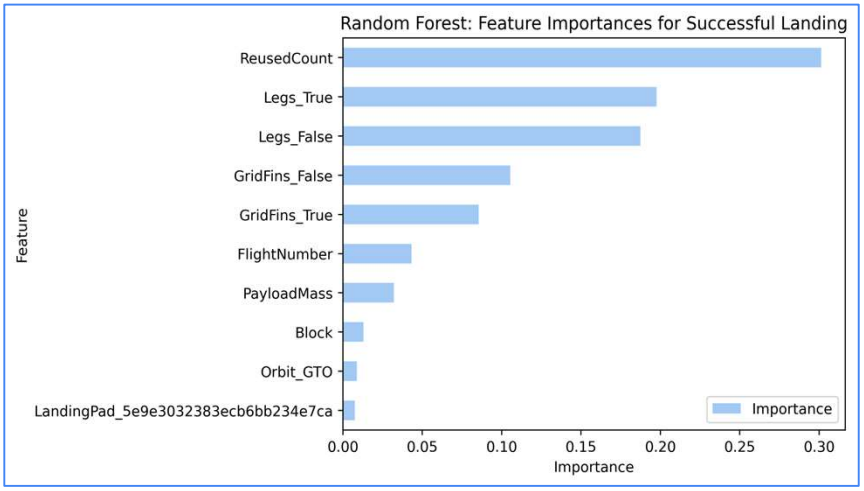
The models perform on the test set 94% with the highest accuracy in the random forest.

IMPACT

This model helps predict whether the launch will be successful or fail to launch which factors are most influential. These insights can help make decisions to win the bid proposal.



Barplot above shows KNN, SVM and Logistic Regression performs reasonably well on the test set with a high score. However, the accuracy is slightly lower compared to some other models. Decision Treaa and Random Forest have balanced performance.



In theRandom Forest model above, *'ReusedCount', 'Legs', 'GridFins", 'FlightNumber', 'PayloadMass', `Block` , `Orbit_GTO` and `LandingPad with 3 serial 7ca`* have the highest importance. These variables are most helpful in predicting the outcome variable, *`successful landing`.*

INSIGHTS/NEXT STEPS

- We used different data across the assignments. The number of successful landings increase with the flight number increases
- The data in SQL we acquired 4 launch sites. Space X start operating in 2010 and first successful landing in 2015.
- The location site, KSC LC-39A, has the highest success rate, yet it is not accounted for in the feature importance results of the Random Forest Model.
- Considering XGBoost and Kmeans
- Analyze the feature importance from random forest further
- Review the hyperparameter tuning