Capstone Project ML Classification

Credit Card Default Prediction

Capstone Project Summary

As we know in today's world no one is without a credit card, so it has become a challenge for the companies to keep record of all customers who make late payments and who makes it on time to decide whether their limits should be increased or not so we are trying to find customers who are defaulters.

This project is aimed at predicting the case of customers default payments in Taiwan. From perspective of risk management, the result of predictive of estimated probability of default will be more valuable than binary result of classification credible or not credible clients.

Model selection-

Logistic Regression.

SVC

Random Forest

XGBoost

Performance Metrics

Precision - Precision is a good metric to use when the costs of false positive (FP) is high.

Precision = TP / (TP + FP)

Recall- Recall is a good metric to use when the cost associated with false negative (FN) is high**.**

Recall = TP / (TP + FN)

F -1 Score- F1-score is a weighted average of precision and recall. Thus, it considers FP and FN. This metric is very useful when we have uneven class distribution, as it seeks a balance between precision and recall.

F1-score = 2 (precision recall) / (precision + recall)

Conclusion-

1. There are no null values and duplicate values in our dataset.
2. Defaulters are less than the Non defaulters. Approx. 78% are Non defaulters and 22% are defaulters.
3. No. of male credit holders is less than female.
4. Approx. 10% are defaulters from male,13% are defaulters from female.
5. We use box plot to detect outliers.
6. We implemented ML models and found that best accuracy, F1 score is obtained from random forest classifier, XGB classifier.
7. With RF classifier we get test accuracy of 81%.
8. With logistic regression test accuracy is 68%.
9. With SVC test accuracy is 77%.
10. With XG Boost test accuracy is 82%.

11. Therefore we can conclude based on test accuracy, F1 score, recall XG Boost classifier, RF Classifiers are best model to predict credit card default.

Team Member's Name, Email and Contribution:

Contributor Roles:

1. **Mihir Kulkarni**: mihirkulkarni50@gmail.com

    a. **EDA**
    b. **Handling null values.**
    c. **Handling outliers.**
    d. **Feature creation.**
    e. **Univariate and bivariate analysis.**
    f. **Feature encoding.**
    g. **Plotting heatmap.**
    h. **Implementing ML algorithm**
    i. **Logistic regression.**
    j. **Random forest classifier.**
    k. **SVC**
    l. **Matrix comparison.**

2. **Ram Bakale:** rambakale@gmail.com
    a. **EDA**
    b. **Feature creation.**
    c. **Feature encoding.**
    d. **One hot encoding.**
    e. **Dropping unwanted columns.**
    f. **Splitting dependent and independent features.**
    g. **Train test split.**
    h. **Handled imbalanced dataset using random over sampling.**
    i. **Implemented XGB classifier, SVC**
    j. **Hyperparameter tuning.**
    k. **Matrix comparison.**

Github link:- https://github.com/mihirkulkarni50/ML-Classification-capstone-project.git

Google Drive Link :
https://drive.google.com/drive/folders/1AbHQBl8mVIrI0DjCOHEzjlrC2Acu7X
ve?usp=sharing