

Capstone Project Submission

Instructions:

- i) Please fill in all the required information.
- ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:
Kollana Harish, harishkollana@gmail.com : <ul style="list-style-type: none">1. Importing Libraries2. Loading the dataset3. Data Analysis on Columns4. Imbalanced Target Column5. Feature Engineering on Columns6. Fitting the models and Hyper Parameter Tuning<ul style="list-style-type: none">1. Logistic Regressor2. Random Forest Classifier3. XGB Classifier7. Model Comparison8. Conclusion
Please paste the GitHub Repo link.
Github Link:- https://github.com/harishkollana/Credit-Card-Default-Prediction-ML-Classification-Project
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)

The Taiwan Credit card issuer issues credit limits to the customer and in that there will be defaulters and non-defaulters. Based on the limit the issuer provided, Age, Education, Gender and other features the limit is provided.

We were provided with one such already classified label in our data set containing 30,000 observations with 25 columns.

Our experiments can help the issuer have a better understanding of their current and potential customers, which would inform their future strategy, including their planning of offering targeted credit products to their customers.

After Importing and libraries and Data Set, we had gone with Data Analysis on each and every column and tried to find out the relation between independent variables and dependent variables. On the second Step we had done Feature engineering on each and every column to find out the categorical and numerical analysis. We had applied smote technique to balance the dataset. We had done label encoding. One hot encoding and created age intervals for defaulters

After that we fit three models for training namely Logistic Regressor, Random Forest classifier and XGBOOST classifier. For Hyper parameter tuning we had used Randomized Search cv and Gridsearch Cv for accuracy we had used f1score.

The Conclusion Was XGBOOST Classifier is the best model with highest recall rate f1 score taken into account as our model is finance and recall is the best approach.