

INDIVIDUAL CONTRIBUTION-REPORT

210170G

W.A.R.T. Fonseka

Task Overview:

Our team embarked on the challenge presented by the Kaggle competition, focusing on devising a credit risk model with a distinctive emphasis on stability over time. Targeting individuals with limited or no credit history, our primary aim was to empower consumer finance providers with a refined tool for evaluating default risks with precision. By meticulously dividing tasks among team members, we executed each aspect of the competition with unwavering diligence and expertise, determined to excel in this competitive arena.

Individual Contribution:

- Data Encoding:
 - I spearheaded the data encoding process, ensuring that categorical variables were appropriately transformed into numerical representations suitable for model training. By encoding the data effectively, I facilitated the incorporation of categorical features into the predictive framework, enhancing the model's overall performance.
- Up sampling and Down sampling:
 - Given the substantial size of our dataset, I implemented both upsampling and downsampling techniques(because upsampling techniques are not suitable for this scenario) to address class imbalance effectively. Upsampling involved increasing the representation of minority classes, while downsampling reduced the dataset's size without compromising its integrity. By balancing the class distribution, I mitigated the risk of biased model predictions and improved overall performance.
- Random Forest Classifier Model Training:
 - I led the training of the Random Forest Classifier model, leveraging its ensemble learning capabilities to build a robust credit risk assessment model. By harnessing the collective strength of multiple decision trees, I aimed to develop a predictive framework capable of accurately identifying default risks. The Random Forest algorithm's versatility and resilience made it an ideal choice for handling our complex dataset and delivering reliable predictions.
- Prediction and Performance Analysis:
 - After training the model, I carefully analyzed its performance by generating predictions and conducting thorough evaluations, including AUC, Gini coefficient, confusion matrix, precision, recall, F1 score, and accuracy. Through a comprehensive analysis of these metrics, I scrutinized the model's accuracy and effectiveness in predicting credit risk. By interpreting the feedback on performance, I iteratively improved the model to boost its predictive capabilities, guaranteeing its reliability and suitability.