Capstone Project Submission

Instructions:

i) Please fill in all the required information.

ii) Avoid grammatical errors.

Team Member's Name, Email and Contribution:

1) Chandan Kumar Raxit

E-mail: chandanraxit2468@gmail.com

- Exploratory Data Analysis
- SMOTE (Synthetic Minority Oversampling Technique)
- Feature Engineering
- One Hot Encoding
- Implementing Logistic Regression
- Implementing SVC

2) Deepak Kumar Jena

E-mail: jdeepakumar986@gmail.com

- Implementing Decision Tree
- Implementing Random Forest
- Implementing XGBoost
- Hyperparameter Tuning
- Evaluating the models
- Plotting ROC AUC for all the models

This paper proposes a comprehensive way by using *k*-means SMOTE and BP neural network algorithms for data imbalance. We find that the improved version of the smote algorithm (*k*-means SMOTE) not only effectively solves the problem of data imbalance but also improves the prediction performance of the model. In addition, we also find that using the feature importance calculated by the random forest as the initial weight of the hidden layer of the BP neural network can slightly improve the prediction performance of the model to a certain extent. However, this change is not obvious. On the one hand, it may be because the credit card default data has many influencing factors and is more complicated. We cannot take all such influencing factors into account, which may indirectly affect the calculation results of feature importance. On the other hand, we think that the amount of sample data may not be enough, the model of BP neural network is relatively simple, and there is no better interpretation of these data for predictive analysis.

In addition, with the gradual increase in the penetration rate of credit cards in our country, the research on its default risk has the following suggestions. On the one hand, we should further improve the construction of the credit indicator system. A good credit index system is conducive to better assessment of personal credit, and a risk prediction model with better classification performance can be established. Specifically, methods such as Delphi expert method, analytic hierarchy process, and regression analysis can be used to find the most representative individual credit indicators, then determine the weight of each indicator, and finally dynamically manage the evaluation system. On the other hand, we should strengthen risk management and control. Since credit card loan default involves personal moral issues, it is highly subjective and uncontrollable. Although major financial institutions are committed to developing the best methods for credit card loan risk avoidance, they have not been able to completely resolve the problem of credit defaults. Therefore, financial institutions should focus on controlling and avoiding risks and try their best to reduce risk losses. Based on the idea of machine learning integration methods, they can comprehensively use each superior classifier to develop a more versatile risk control model.

Please paste the GitHub Repo link.
GitHub Link: - https://github.com/Deepak-Kumar-jena/Credit_Card_Default_Prediction#credit_card_default_prediction
Please write a short summary of your Capstone project and its components. Describe the problem statement, your approaches and your conclusions. (200-400 words)