Prediction of People with Good and Bad Credits

Goh Yong Xuan, Han Weihang, Lim Shu Yau, Meng Siong Hwee, Tan Xuan Huan

National University of Singapore

e0857383@u.nus.edu, e0726704@u.nus.edu, e0774074@u.nus.edu, e0910147@u.nus.edu, e0775000@u.nus.edu

Introduction

Our project aims to provide a tool to classify the 'good' and 'bad' customers to facilitate in more efficient credit card approval process through the use of machine learning. Around three-quarters of Singaporeans own at least one credit card, with 7.3% owning 6 or more credit cards (Binsted, 2022). With large number of credit card ownerships among people, it is important to find accurate ways to speed up the process for credit card approval. We hope to do so with machine learning methods.

In this project, we will adopt K-Means Clustering and K-Nearest Neighbors (KNN), as they are able to identify the clusters each individual belongs to, hence labeling them effectively. We will also be using Random Forests (RF) and Neural Network (NN) to identify the complex relationships between each variable.

This project has provided us with an opportunity to practice the use of ML algorithms learned in the module and explore the ways that they can be applied to real-world contexts. The project problem that we are solving is also applicable and relatable to our daily lives.

Related Works

One of the key factors in credit card approval processes is credit risk, which measures the risk that may arise from borrowers failing to meet contractual obligations. Probability of default (PD), which is the likelihood a borrower fails to pay, and early warning signals, which help identify borrowers exposed to a higher risk of default before the default occurs, are integral parts of credit risk management. Traditionally, PD models rely heavily on logistic regression, whose models are relatively easy to interpret. However, they are unable to capture complex relationships that may be present in actual data. Currently, systems for early warning signals require a large number of experimentally defined indicators and rely heavily on expert judgment, which is manpower heavy. Incorporating ML algorithms in credit risk management has allowed for more sensitive credit risk assessment that takes real-time factors into account. Hence, there is more predictive power in the data arising from ML algorithms.

Dataset

The Credit Card Approval Dataset consists of clients' applications and credit records.

The application records contain information such as clients' gender, education level, employment length, age, income type, and job that are crucial in helping us to determine if the clients are 'good' or 'bad'.

We have discovered cases where clients are unemployed for more than 1000 years and this would definitely affect the feature scaling of employment length. As such, we propose the retirement age to be the median of "Birthday" (60 y/o) and their length of unemployment would be adjusted from there on.

There are records with missing data values, specifically the "Job" column. Among those with missing job labels, we assigned clients who are not employed but with an income to be "Pensioner". For the rest, we predicted their job using Random Forest.

The credit records show the status of each client on whether they have paid off their loan for that specific month. We attempted to use K Means to cluster the labels using a simple average and exponential weighted moving average, however, it did not work. We have labeled clients with no loan, or paid within 1 month as 'good' and the rest as 'bad'.

Methods

We used KNN, RF, NN, and K-Means Clustering.

KNN was chosen as it allowed us to predict classes based on the data's closest neighbors. KNN is also a good baseline for the proximity of bad and good label clusters in our datasets' vector space.

RF uses ensemble learning (bagging), which reduces the overfitting problem in decision trees and lessens the variance (H20.ai, n.d.) so we used it instead of decision trees. RF also allows us to discover the important features that differentiate people with good and bad credit.

We used NN as multilayer feedforward networks are universal approximators (Hornik et al., 1989). NN can make generalizations and uncover hidden relationships. (H20.ai, n.d.).

Our labels are imbalanced (88% good) so we trained our models on our cleaned and augmented datasets.

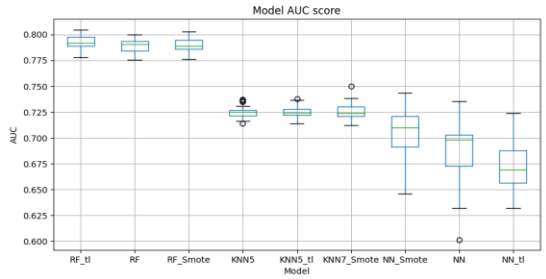
Augmentation	Reason
Smote	Smote balances class distribution using generated data
Tomek links	Tomek-Link can reduce class overlap on data (Dai et al., 2022).

K-Means clustering is used to find patterns and form clusters among the data present and determine a new data's 'status' label by assigning it to the nearest cluster.
The "status" label for each id was redefined to scores and the mean was taken for k-means. We adopted the elbow method to determine the suitable value of k.

Results & Discussions

Model	Finetune	Reason
KNN	Different K	Find the best K value as the square root size dataset does not work to detect the minority class
RF	Hyperparameters	We used class weights = balanced subsample to deal with the imbalanced data.
NN	Dropout layers	Reduce overfitting

Provost and Fawcett (1997) advised using the ROC curve instead of accuracy in the case of imbalanced data learning. We used AUC as a comparison metric.
We used stratified K Folds with 5 splits and repeated this 5 times. We present only KNN with the best K value for each data augmentation. Our models' validation AUC scores are shown in the boxplot.



Across all data augmentations, the performance of the models was similar so the data augmentation did not do much to help the performance except for smote for NN.
As a rule of thumb, 0.7< AUC score is poor discrimination, 0.7-0.8 is acceptable discrimination, and >0.8 is excellent discrimination (Hosmer et al., 2013).

Model	Performance	Evidence
KNN	Acceptable	K values of 5/7 are susceptible to noise as it is small
RF	Good	RFs can pinpoint important features like employment length and birthday as seen from the feature importance plot in Fig 1 (In annex). RF being good is explained by specific features of tabular data: irregular patterns in the target function and uninformative features (Grinsztajn et al., 2022).
NN	Bad	All NN are overfitted as the training loss, AUC is vastly lower than validation loss, AUC in Fig 2, 3 (In annex). NN had the greatest variance in performance (from boxplot).
K-means	Bad	K-means clustering did not cluster well as the credit record data was skewed shown in Fig 4 (In annex).

Conclusion

By using ML algorithms, we managed to build a classifier that predicts the approval of a credit card application. We were able to learn how to perform common preprocessing steps such as feature scaling, handling of missing values and labeling of data, as well as optimizing the hyperparameters and evaluating the performance of the different ML models implemented. While the models were limited due to the presence of noisy data, the random forests classifier showed the most desirable performance across all the models. In the future, when a larger dataset is available, more work can be done to improve the classifier's accuracy and performance. Overall, this project provided us with a good opportunity to explore and implement ML algorithms learned in the module.

References

Artificial Intelligence for Credit Risk Management. Deloitte. Retrieved October 24, 2022 from https://www2.deloitte.com/cn/en/pages/risk/articles/artificial-intelligence-for-credit-risk-management.html

Binsted, S. (2022, October 2). *Credit Card Statistics Singapore*. Finder Singapore. Retrieved October 24, 2022, from https://www.finder.com/sg/credit-card-statistics

Credit Scoring Using Machine Learning. Datrics. Retrieved October 24, 2022 from https://datrics.ai/credit-scoring-using-machine-learning

Dai, Q., Liu, J., & Liu, Y. (2022). *Multi-granularity relabeled under-sampling algorithm for imbalanced data*. arXiv. https://doi.org/10.48550/arXiv.2201.03957

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). *Why do tree-based models still outperform deep learning on tabular data?* arXiv. https://doi.org/10.48550/arXiv.2207.08815

Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks, 2*(5), 359-366.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (1st ed.). Wiley. https://doi.org/10.1002/9781118548387

H20.ai. (n.d.). *What is a neural network used for?* Retrieved October 24, 2022, from https://h2o.ai/wiki/neural-network

Provost, F., & Fawcett, T. (1997). Analysis and visualization of classifier performance with nonuniform class and cost distributions. In *Proceedings of AAAI-97 Workshop on AI Approaches to Fraud Detection & Risk Management* (pp. 57-63).

Annex

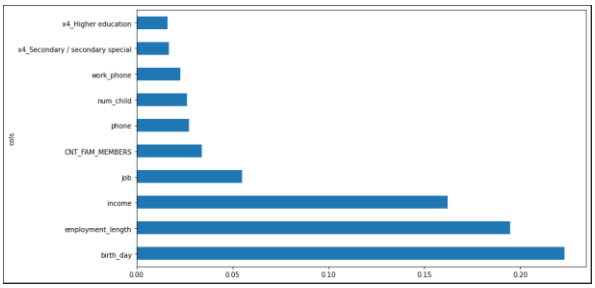


Fig 1: Random Forest feature importance

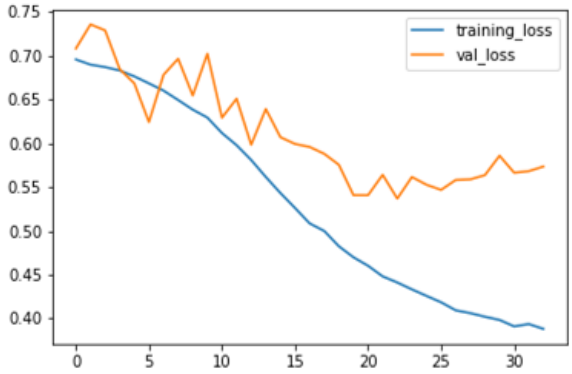


Fig 2: NN loss

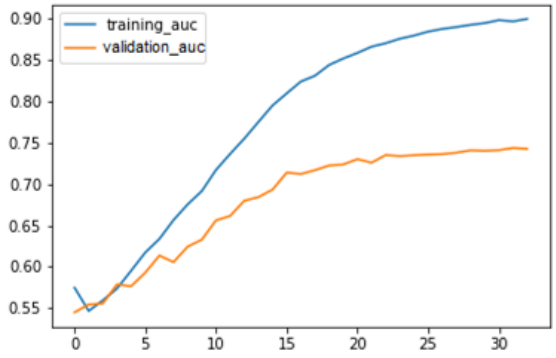


Fig 3: NN auc scores

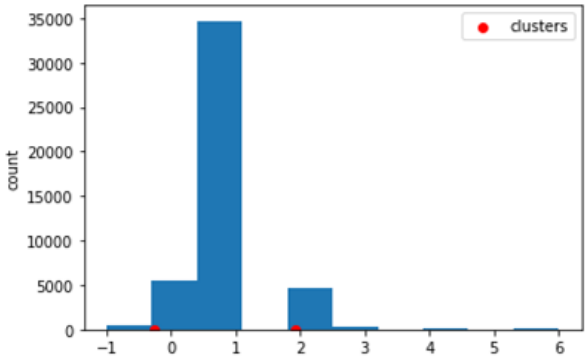


Fig 4: kMeans clustering.
(if only >2 status are in the 2nd cluster, the 2nd cluster centre should be >2)