CREDIT RISK ANALYTICS

❖ Problem Definition

The main objective of this project is to:

(1) identify the factors that influence a loan getting repaid successfully or closed on account of default,

(2) based on those factors, build a model that can predict whether a borrower will eventually default on his loan or not, and

(3) build separate model that will, in the event of a default, predict the net impact of the default on the lender.

This will help individual lenders or lending institutions minimize the risk of losing money due to bad loans.

Such loan default prediction and loss calculation scenarios are part of *Credit Risk Modelling* and based on the current attributes collected from the borrower's loan application form a classic data science problem. This also provides an opportunity to apply various supervised and unsupervised methodologies of data science to solve the business problem.

However, we must build the model using a conservative approach using rigorous evaluations since rejecting too many prospective borrowers can hamper the business.

❖ Literature Survey

Loans are very important for any financial institution and are one of the primary sources of income. When a loan goes bad, it can be very fatal for the institution's financial stability. Authors in [1] have shown how bad loans (also known as NPA or Non-performing assets) negatively affects the profitability of banks. All these have given rise to the need of Credit Risk Modelling which is the process of using data models to find out:

(1) the probability of the borrower defaulting on a loan, and

(2) the effect of this default on the financial stability of the lender.

Our current computational capabilities, emerging data analytics techniques, along with the recent need of automating the loan approval process in the lending sector, have greatly improved the century old practise of predicting the risk of default in the lending process. Numerous papers have been published and research has been done to accurately model *Credit Risk* using various data science techniques.

One of the main factors on which Credit Risk Modelling is dependent is *Probability of Default (PD)*. Authors in [5], [6] and [10] show us a whole range of modelling techniques using data science for determining PD. Research is also done, e.g. [3] and [4], that tell us how different factors affect loan default. However, with so many algorithms now available with us, there is an obvious question regarding which is the most suitable technique for predicting default. [9] Tells us about a few metrics that can be used for comparing the performances of different algorithms for a given dataset.

Another important factor for Credit Risk Modelling is *Loss Given Default (LGD)*. There have been many studies like [13], [14] and [17] with a focused approach on modelling LGD by considering different attributes of the borrower's personal information, credit history and loan information.

Nowadays, we also have different data science tools like *Weka*, as mentioned in [7] that can be used for the classification process. Also, there are new fields of research that are further improving the modelling process of Credit Risk like *Forensic Analytics* where electronic data is being used reconstruct or detect financial fraud.

❖ Sample Data

For this proposed project, we are looking into data published by various Peer to Peer (P2P) lending platforms since it is very difficult to obtain real bank data. P2P platforms, such as Prosper, Lending Club, Kiva, Fynanz provide various online services, thus enabling individuals and small businesses to get hassle-free loans from interested lenders. Amongst them, Lending Club is the largest P2P lending platform in the world and the dataset released by them is currently available in Kaggle (https://www.kaggle.com/wordsforthewise/lending-club), containing 2 million+ loan records issued between the years 2007 and 2018.

Also, the dataset has 151 variables which include information such as borrower's credit history, personal information (e.g. annual income, years of employment, zip code), loan information (e.g. description, type interest rates, grade), current loan status and, etc. It also contains some variables like when the last repayment was done which is a knowledge of the future and needs to be handled accordingly.

❖ Tentative List of Algorithms

For Credit Risk Modelling, a wide range of algorithms and techniques can be used.

Some of the prospective supervised learning algorithms are,

- Linear Regression
- Logistic Regression
- k-Nearest Neighbors
- Random Forest
- Support Vector Machine
- Artificial Neural Network

These supervised learning algorithms can be associated with various unsupervised learning techniques, like k-Means Clustering and Hierarchical Clustering as well as with different dimensionality reduction techniques like Linear Discriminant Analysis and Principal Component Analysis.

By using the above techniques, we propose to present our subject project.

References

[1] B. Senthil Arasu, P. Sridevi, P. Nageswari, R. Ramya, "A Study on Analysis of Non-Performing Assets and its Impact on Profitability", International Journal of Scientific Research in Multidisciplinary Studies, Volume-5, Issue-6, pp.01-10, June (2019) (d)

[2] Yang Yang, "An empirical study on P2P loan default prediction model", Financial Engineering and Risk Management (2020) 3: 14-22, Clausius Scientific Press, Canada

[3] Michal Polena, Tobias Regner, "Determinants of Borrowers' Default in P2P Lending under Consideration of the Loan Risk Class", Games 2018, 9, 82 (d)

[4] Rajkamal Iyer, Asim Ijaz Khwaja Erzo, F. P. Luttmer, Kelly Shue, "Screening in New Credit Markets: Can Individual Lenders Infer Borrower Creditworthiness in Peer-to-Peer Lending?", SSRN-id1570115 (d)

[5] Don Carmichael, "Modeling Default for Peer-to-Peer Loans", SSRN-id2529240

[6] Alan Zhang, "DEVELOPMENT OF LOGISTIC REGRESSION MODEL TO PREDICT DEFAULT PROBABILITIES OF LOAN ACCOUNTS", International Journal of Information, Business and Management, Vol. 12, No.2, 2020, pg 95-115 (d)

[7] Anand Motwani, Goldi Bajaj, Sushila Mohane, "Predictive Modelling for Credit Risk Detection using Ensemble Method", International Journal of Computer Sciences and Engineering, June 2018, Vol-6, Issue-6 (d)

[8] Zakaria Alomari, Dmitriy Fingerman, "Loan Default Prediction and Identification of Interesting Relations between Attributes of Peer-to-Peer Loan Applications", New Zealand Journal of Computer-Human Interaction ZJCHI 2,2 (2017)

[9] Pedro G. Fonseca and Hugo D. Lopes, "Calibration of Machine Learning Classifiers for Probability of Default Modelling", James Finance (CrowdProcess Inc.), October 24th, 2017 (d)

[10] Rising Odegua, "Predicting Bank Loan Default with Extreme Gradient Boosting" (d)

[11] Jose A. Lopez, Marc R. Saidenberg, "Evaluating Credit Risk Models", Journal of Banking & Finance, Volume 24, Issues 1-2, January 2000, pgs:151-165

[12] Jeremy Turiel, Tomaso Aste, "P2P Loan Acceptance and Default Prediction with Artificial Intelligence", https://www.researchgate.net/publication/334223307

[13] Guangyou Zhou, Yijia Zhang, Sumei Luo, "P2P Network Lending, Loss Given Default and Credit Risks" (d)

[14] Felix Martinson, "Exotic Approaches for Modelling Loss Given Default" (d)

[15] Shunpo Chang, Simon Dae-Oong Kim, Genki Kondo, "Predicting Default Risk of Lending Club Loans"

[16] Carlos Eduardo Canfield Rivera, "Determinants of Default in P2P Lending: The Mexican Case", Independent Journal of Management & Production (IJM&P), v.9, n.1, January - March 2018

[17] Christophe Hurliny, Jérémy Leymariez, Antoine Patinx, "Loss functions for Loss Given Default Model Comparison"

[18] Peter Martey Addo, Dominique Guegan, Bertrand Hassani, "Credit Risk Analysis Using Machine and Deep Learning Models", Risks 2018, 6, 38

[19] Til Schuermann, "What Do We Know About Loss Given Default?", Forthcoming in D. Shimko (ed.), Credit Risk Models and Management 2nd Edition, London, UK: Risk Books, February 2004

[20] Han Sheng Sun and Zi Jin, "Estimating credit risk parameters using ensemble learning methods: an empirical study on loss given default", Journal of Credit Risk 12(3), pgs 43–69