Minh Q. Vu

Date-Month-Year

Econ 297 Honors

**Honor Thesis Idea**

Research Question: What are the determinants of the loan default rate on micro scale?

**Motivation**

I have always been fascinated with data and its application in our modern world. My enthusiasm for data arose when I found out about the game of poker - a game of pure math and statistics (arguably also a little bit of psychology). Data is the new fuel of the 21st century; it has improved a wide range of industries, such as medical, computer science, finance, etc. As a student double majoring in Computer Science and Economics and minoring in Data Science, I am curious to know more about the application of data, or Machine Learning to be more exact, in those two fields. I want to know more about loan and to gain more insight in different factors that might affect the default rate in retail banking industry and microfinance institutions.

Loan default can be defined as the inability of a borrower to fulfil his or her loan obligation as at when due (Balogun, E.D. & Alimi, A, 1988). In “Loan Default Rate and its Impact on Profitability in Financial Institutions” by Evans Brako Ntiamoah, Emmanuel Oteng, Beatrice Opoku and Anthony Siaw, the authors describe the cause to this problem as “multi-dimensional” and “non-uniform among different literatures”. Microfinance perceived as a financially sustainable instrument meant to reach significant number of poor people of which most are not able to access financial service because of the lack of strong retailing financial intermediaries (Ntiamoah, B.E. , Oteng, E., Opoku, B., Siaw, A., 2014). The role of microfinance in giving the poor population access to small loans cannot be emphasized enough. Microfinance provides people with capital to start or expand their businesses. Small businesses with microfinance support have grown into medium enterprises, creating employment opportunities for others. Nonetheless, some of the loans given out by the lending institutions, unfortunately, are not paid back and eventually result in bad debts with adverse consequences for the overall financial performance of the institutions (Ntiamoah, B.E. , Oteng, E., Opoku, B., Siaw, A., 2014). The costs of loan delinquencies would be felt by both the lenders and the borrowers. The lender has costs in delinquency situations, including lost interest, opportunity cost of principal, legal fees and related costs. For the borrower, the decision to default is a trade-off between the penalties in lost reputation from default versus the opportunity cost of forgoing investments due to working out the current loan (Ntiamoah, B.E. , Oteng, E., Opoku, B., Siaw, A., 2014). In a recent study of Mortgage loan defaults, the most frequently cited causes of defaults were curtailment of income (36%), excessive obligations (19%), unemployment (8%), illness of principal mortgagor or family member (6%) and marital difficulties (3%) (Merritt, 2009). Other critical factors associated with loan delinquencies are: type of the loan, term of the loan, interest rate on the loan, poor credit history, borrowers’ income and transaction cost of the loans (Ntiamoah, B.E. , Oteng, E., Opoku, B., Siaw, A., 2014).

Loan default also happens even on institutional level. Corporate loan default has been found to increase as real GDP decline and exchange rate depreciation directly affects the repayment ability of borrowers (Hurt and Fesolvalyi, 1998). The following were identified to be the main causes of default of loans from industrial sector: improper selection of an entrepreneur, deficient analysis of project viability, inadequacy of collateral security/equitable mortgage against loans, unrealistic terms and schedule of repayment, lack of follow up measures and default due to natural calamities (Ntiamoah, B.E. , Oteng, E., Opoku, B., Siaw, A., 2014).

In this thesis, I will only focus on loan default rate on microfinance level.

**Relevant sources**

1. **“Loan Default Rate and its Impact on Profitability in Financial Institutions” by Evans Brako Ntiamoah, Emmanuel Oteng, Beatrice Opoku, Anthony Siaw (School of Management & Economics, University of Electronic Science and Technology of China)**

This peer-reviewed article discusses the positive impact that microfinance has on low-income households and small and medium sized enterprises, how loan default is affecting these types of financial institutions, determinants of default rates and different ways of reducing loan defaults. In the end, the authors conclude that there is a strong relationship between problem of recovery and overdue of loans and profitability with a correlation coefficient of 0.593 at the 0.01 level (2-tailed).

This is relevant to my thesis as it offers a great background of the topic. It discusses many important concepts of loan default like what is it, why it is important and how to deal with it.

1. **“Determinants of Microcredit Repayment in Malaysia: The Case of Agrobank” by A.H. Roslan and Mohd Zaini Abd Karim (Department of Economics, College of Arts and Sciences, University Utara Malaysia)**

This peer-reviewed paper examines the determinants of loan repayment among borrowers of Agrobank – a commercial bank in Malaysia. The data used in this study is gathered through a survey on 2630 respondents that is carried out in 86 branches of Agrobank throughout Malaysia. The authors employ the probit and logit models to identify the main determinants that influence microcredit repayment. They also divide determinants into three categories – characteristics of the borrowers, characteristics of the project or business and the characteristics of the loan. The study discovers that the probability for default is influenced by the gender of the borrower, type of business activity, amount of loan, training and the repayment period (A.H. Roslan and Mohd Zaini Abd Karim, 2009).

This is relevant to my thesis because they are also trying to answer the same question as I am. This gives me a good reference and comparison once I have some result from my analysis.

1. **“Factors Affecting Loan Repayment Performance Among Yam Farmers in the Sene District, Ghana” by C. A. Wongnaa, D. Awunyo-Vitor (Department of Agropreneurship, Institute of Entrepreneurship and Enterprise Development, Kumasi Polytechnic, Ghana | Department of Agricultural Economics, Agribusiness and Extension, Faculty of Agriculture, Kwame Nkrumah University of Science and Technology, Ghana)**

This peer-reviewed article discusses different factors affecting loan repayment performance among farmers in the Sene District, Ghana. The study found that educational level, number of years of farming experience, profit gained from loan, age of farmer, supervisory visits to farmers and access to off-farm income have positive effects on yam farmers’ ability to repay the loans given to them by financial institutions. A rise in each of these factors will enhance yam farmers’ loan repayment abilities. On the other hand, gender and marriage have negative effects on yam farmers’ loan repayment abilities (C.A. Wongnaa and D. Awunyo-Vitor, 2013). This is relevant to my thesis for the same reason as the second paper.

1. **“Credit Default Risk and its Determinants of Microfinance Industry in Ethiopia” by Samuel Setargie (Lecturer of Aksum University, College of Business and Economics)**

This paper also focuses on credit default risk and different factors leading to this risk. Factors that are found to be significant determinants of credit default performance were education, credit/loan size, and credit diversion, availability of other credit sources, credit/loan supervision, and suitability of credit repayment period and income (Setargie, 2013). This is relevant to my thesis for the same reason as the second and the third paper.

1. **“Neural network approach to loan default prediction” by Manjeet Kumar1, Vishesh Goel2, Tarun Jain3, Sahil Singhal4, Dr. Lalit Mohan Goel5 (1,2,3,4 Students, CSE, Bharati Vidyapeeth's College of Engineering, New Delhi | 5Prof, CSE Department, Bharati Vidyapeeth's College of Engineering, New Delhi)**

In this paper, the Multilayer Perceptron Model with Adam Optimizer as the optimization function is used to predict the loan default rate. Their proposed neural network model is tested for effectiveness using dataset provided by the Lending club bank. Then, Principal Component Analysis (PCA) was done on this dataset to reduce the number of dimensions present in the dataset.

I choose this paper because the authors also utilize Artificial Intelligent technique, or Deep Learning to be more precise. This will benefit my research a lot as I can see how other people are applying AI in this topic.

1. **Kaggle Competition - Home Credit Default Risk: Can you predict how capable each applicant is of repaying a loan by Home Credit Group**

This is a dataset about Default Risk provided by Home Credit two years ago. This will be the main dataset I will use for my honor thesis. The main train dataset (300,000 by 127) has more than 300,000 observations, each represents one loan application. Each observation has more than 100 features, including useful information about the customer to predict their default likelihood. These features range from specific information of the application such as income, gender, number of house/children, etc. to general information such as details of the building they live, did they submit this form, etc. Clearly, some features will be more important than the other, and my job will be finding out those features to build the best model possible. There will be a target column with only 2 possible values: 0 and 1. 1 indicates the person has some difficulty in paying back their loan, 0 indicates otherwise.

# References

A.H. Roslan and Mohd Zaini Abd Karim. (2009). Determinants of Microcredit Repayment in Malaysia: The Case of Agrobank. *Humanity & Social Sciences Journal*, 45 - 52.

Balogun, E.D. & Alimi, A. (1988). Loan Delinquency Among Small Farmers in developing countries: A case study of small-farmers credit programme in Lagos State of Nigeria.

C.A. Wongnaa and D. Awunyo-Vitor. (2013). Factors Affecting Loan Repayment Performance Among Yam Farmers in the Sene District, Ghana. *Agris on-line Papers in Economics and Informatics, V*, 111 - 122.

Home Credit. (2018). Retrieved from Kaggle: https://www.kaggle.com/c/home-credit-default-risk

Manjeet Kumar, Vishesh Goel, Tarun Jain, Sahil Singhal, Dr. Lalit Mohan Goel. (2018, April). Neural Network approach to loan default prediction. *International Research Journal of Engineering and Technology (IRJET)*, 4231 - 4234.

Merritt, E. (2009). Mortgage Default: Causes and Cures, Mortgage Loan Compliance, the Forensic Loan Audit Company.

Ntiamoah, B.E. , Oteng, E., Opoku, B., Siaw, A. (2014). Loan Default Rate and its Impact on Profitability in Financial Institutions. *Research Journal of Finance and Accounting*, 67 - 72.

Setargie, S. (2013). Credit Default Risk and its Determinants of Microfinance Industry in Ethiopia.