

Credit Risk Classification in Peer-to-Peer Marketplaces: The Nexus of Neural Network Approach

Baah Alexander (Corresponding author)

School of Finance and Economics, Jiangsu University, Zhenjiang City, 212013, P.R China

Tel: 86-186-5192-3738 E-mail: alexbaah25@gmail.com

Tan Zhongming

School of Finance and Economics, Jiangsu University, Zhenjiang City, 212013, P.R China

E-mail: tzmyx@ujs.edu.cn

Ding Guoping

School of Finance and Economics, Jiangsu University, Zhenjiang City, 212013, P.R China

E-mail: pillding@126.com

Albert Henry Ntarmah

School of Finance and Economics, Jiangsu University, Zhenjiang City, 212013, P.R China

E-mail: henritoalberto@gmail.com

Asare Evans Kwabena

School of Finance and Economics, Jiangsu University, Zhenjiang City, 212013, P.R China

E-mail: 5102170288@stmail.ujs.edu.cn

Received: December 23, 2019 Accepted: January 18, 2020 Published: February 3, 2020

doi:10.5296/ber.v10i1.16372

URL: <https://doi.org/10.5296/ber.v10i1.16372>

Abstract

Financial innovation in recent years have prominently contributed to the growth of Peer-to-Peer lending marketplaces allowing individual and businesses to secure loans on a common internet-based network. Similar to the ‘bricks and mortar’ banking system, online lending is coupled with the problem of information asymmetry. Borrower risk assessment has henceforth become the major concerns of various platforms that aim to reducing information imbalance towards mitigating credit risk. In this article, authors compared two learning algorithms – Logistic regression and Artificial Neural Network to classify borrowers based on loan repayment schedule. We revealed that both approaches were robust in classifying late borrowers with logistic regression being 0.02% more robust than Neural Network. Regarding variable relative importance, gender is considered the least important variable whereas terms-of-repayment is the most important variable affecting borrowers’ intention to pay off loans. Even though our study contributes to existing literature, it is however not limited to determining factors that may affect lenders’ investment decision in social lending.

Keywords: Artificial Neural Network, Financial innovation, Information asymmetry, Algorithm

1. Introduction

Electronic commerce, especially online lending platforms have become a renowned area of interest among financial experts, regulators and market participants in the financial sector. The arrival of Web 2.0, smartphones and enactment of strong financial innovation and reforms-financial liberalization in countries such as the United States, China, United Kingdom, Hong Kong, Malaysia, etc, however have stimulated the growth of the novel financing model. Besides, social lending credit is easy to access because of high speed internet upgrade and the ability to connect individual borrowers and lenders (Jagtiani & Lemieux, 2017). Platform operators rely solely on sophisticated networks (both online and offline) that involve institutional investors, direct loans (in terms of balance sheet) and transactions securitization. It has also been suggested that these alternative lenders may find it advantageous to partner with banking institutions to grant unsecured personal loans, small business loans and auto refinance loans. For example, Lending Club contracts part of its loans through Web-Bank (<https://www.webbank.com/our-brand-partners/lendingclub>).

Alternatively, online loan markets offer credits to individuals and SMEs, who for other reasons excluded from assessing traditional banking credit facilities. P2PL funds may be thought to be insecure and risky, but once provided can facilitate financial growth, manage working capital and manage the intermittent financial needs of debtors. On the other hand, P2P financing also helps divert significant capital from banks to Internet-based financing platforms. On the contrary, unsecured loans carry a higher risk and the lender assumes the sole risk of default but cannot overshadow the basic functionality of the P2P online loan platform (its ability to bring borrowers and investors together on a common internet). The platform and the investor generate income by lending to the borrower. Loans disbursed generally fulfill the financial purpose of the clients, who are required to pay a monthly payment of principal and interest in accordance with the contractual agreement between the

parties.

The long-term growth trend of the industry and how it will affect the financial landscape is uncertain, hence rousing the need to probe in to the future. Unlike the traditional financing system, online P2PL use big data mining to evaluate borrowers information (Garson, 1991). It implies that a typical P2PL platform often filters loan applicants information with proprietary formula to ascertain the creditworthiness of the client before signaling potential investors to raise funds when necessary (Foster & Warren, 2016). The mathematical algorithms adopted by alternative online lenders and other globally accredited third-party organizations have led to faster and cheaper credit cost assessments, but could carry some degree of credit risk. These innovations could involve the risk of unequal treatment and fair lending crimes. According to Sufian (2009) any attempt to determine credit risk, economic activity, financial history and the macroeconomic environment of borrowers must be critically examined. Many scholars therefore have attempt to shed light on the need to evaluate the borrower's information as a criterion for classifying good and bad borrowers. In this article, we argue that the key information provided by the loan applicant can accurately predict the risk position of the claimant. To improve the methodology of credit risk assessment, we applied Artificial Neural Network and a logistic regression to the analyze P2P dataset in Australia. This study, as far as we know, is one of the few studies, but first to extract dataset from RateSetter.com in Australia to predict the possible factors that can influence the classification of borrower repayment decision. From a theoretical point of view, this study explains the relevance of machine learning algorithms in the field of finance and adds to the existing body of literature. Lenders and financial institutions can also instill the adoption of the methods applied in this academic paper to improve the loan selection process. Finally, it could also help investors to evaluate clients when making long-term investment decisions.

The structure of this paper is organized as follows. In section II, previous literature is summarized from the domains of P2P online lending, and empirical review of other comparative studies. Section III however comprise of ANN theory, data description and the methods adopted in the article. Section IV on the other hand presents the results and discussions whilst section V provides the summary and conclusion of the article.

2. Related Studies

2.1 P2P Online Lending

P2P lending platforms connect investors with excess funds to borrowers and SMEs via the Internet. P2P microfinance determine the trustworthiness of borrowers and monitor their credit risk once loans are granted. However, Liu, Brass, Lu, and Chen (2015) believe that P2P loans dissent from bank loans therein the investor, and not the P2P loan platform, is liable for deciding whether or not loans ought to be extended and assume the risk of default. This is often because of the fact that P2P lending platforms do not engage in business using their balance sheets as banks. To put in a broader perspective, the operations of P2P lending platforms are therefore not subject to banking regulations, notably with regard to capital requirements, leverage ratios and liquidity requirements.

P2P loans could also improve social well-being, as there is no systemic risk similar to that created by banks (Boyd & De Nicolo, 2005). First, P2P lending platforms do not have a vulnerable balance sheet composition. Alternatively, P2P lending platforms do not use balance sheets with the same approach as banks and hence there is no inherent risk on their balance sheets. Secondly, P2P lending companies are not structured so that the collapse of a platform results in the collapse of others in a chain reaction.

In other words,

- (i) P2P loan platforms are not required to repay lenders;
- (ii) P2P lending platforms do not assume the credit risk of counterparties; and
- (iii) Interbank settlement function is totally absent in online social lending.

It is therefore obvious that P2P is associated with various risks, for example default risk, operational risk and political risk are among the most common P2P risks highlighted by (Li et al., 2016) but notable among them is that P2P loans are not backed by collateral (Wang & Greiner, 2011).

2.2 Empirical Review

Big data mining techniques have been widely applied in different business applications. Financial institutions rely heavily on machine learning methods to rank borrowers based on certain characteristics. Popular credit rating techniques adopted over the years include the logistic regression by Tan, Baah, Ding, Owusu-Ansah, and Agyemang (2019), artificial neural network - ANN (Byanjankar, Heikkilä & Mezei, 2015), vector support machines (Bellotti & Crook, 2009), decision trees (Ince & Aktan, 2009), discriminant analysis (West, 2000) and nearest K-neighbors (Twala, 2010). In addition to the machine learning techniques mentioned above, random forests (Malekipirbazari & Aksakalli, 2015), linguistic interval analyzes (Mezei, Byanjankar, & Heikkilä, 2018) and survival analysis (Baesens, Setiono, Mues, & Vanthienen, 2003) are part of the recently applied credit risk assessment methods examined.

The most recent study by Tan et al. (2019) for instance extracted dataset from RateSetter P2P from 2014 to 2016 and empirically identified housing dummies: - prepayment, no mortgage renter; car, debt consolidation, investment, major events, professional services, 3-year loan term, 4-year loan term, interest rate and income to have a significant influence on the probability of borrowers' default. The estimated coefficients of the authors suggest a probabilistic default rate of 83.4% and therefore recommended a critical review of borrowers information presented to the platform.

Earlier authors have also attempted to use two or more learning techniques to model the classification problem. Notable among them is Ince and Aktan (2009) who devised four different machine learning techniques to evaluate the credit rating of the credit card policy in banks. Out of the models applied, ANN was superior because it produced the lowest Type II error compared to the other three methods - discriminant analysis, LR, and decision tree. A positive measure affirmed by ANN confirms a minimal risk of misclassification of marginal

costs related to Type II errors. Bahrammirzaee (2010) also compared three artificial intelligence techniques in his research. The techniques adopted were ANN, expert systems and a hybrid intelligence system and concluded that artificial intelligence is robust to solving financial problems. The study conducted by Baesens et al. (2003) also examined the ability of the artificial neural network to implement advanced and user-friendly decision support systems for assessing credit risk.

Malhotra and Malhotra (2003) generally took data from 12 different savings and credit cooperatives to create borrower learning classifications. The study used a neural network and multiple discriminant analysis to classify potential borrowers. It was evident that the neural network was more efficient than multiple discriminant analysis in fault identification.

With regard to Mezei et al. (2018), the use of interval-valued linguistic labels and entropy-based discretization can enhance the classification performance of traditional supervised learning. This study proposed the use of traditional machine learning methods enhanced by data transformation based on fuzzy set theory to improve the quality of loan identification with high probability of failure.

The empirical results of Guo (2017) indicate that the Skewed t distribution in Hansen (1994) has more predictive power than the three other distributions applied (namely Student's t distribution, the normal reciprocal inverse Gaussian distribution and the generalized hyperbolic distribution). It further estimated the Skewed t distribution predicts VaR value closest to the non-parametric historical VaR estimates relative to other heavy-tailed distributions.

According to the literature, the use of the neural network is considered superior to other machine learning algorithms and has therefore been widely applied not only in the field of finance, but also in that of biology and genetics.

3. Methodology

3.1 Artificial Neural Network

ANN is a mathematical model that attempts to simulate the structure and functionality of biological neural networks (Bourquin, Schmidli, van Hoogevest, & Leuenberger, 1997). The base of each network of artificial neurons is an artificial neuron, that is a simple mathematical function. There are three simple sets of rules in the artificial modelling technique: multiplication, summation, and activation. At the beginning of the artificial nodes, the inputs are weighed, then each input value is multiplied by the individual weight. This is seen in the equation 1.0 below,

$$f(k, w) = \rho(k \cdot w) = \rho(\sum_{i=1}^n (k_i \cdot w_i) + b) \quad (1)$$

In the equation, the variables k and w represent the independent variables (input vectors) and the weight vector of the neuron when there are entries in the nodes while ρ designates an activation function and b represents the angle.

ANN is a typical complex system consisting of highly interconnected structures called

neurons, nodes or units (Eletter & Yaseen, 2017; Malhotra & Malhotra, 2003; Oreski, Oreski, & Oreski, 2012). The neural network consists essentially of three forms: the input layer, the hidden layer and the output layer. The input layer is simply weighed and summed, and the size of the input and output vectors will be the same if the neural network has nodes both at the input and at the output. There are two important features of nodes: receiving inputs from other nodes, such as hidden or hidden nodes, and sending to other nodes, whether or not they are output, following the hidden nodes. Hidden nodes are not added straightaway to the incoming data but are often grouped into a compact and connected hidden layers which transmit information out of the network as a target output, then corresponds to a dependent variable in the data (Lee & Chen, 2005; Tsai & Wu, 2008). A simplified three-layer output is illustrated in Figure 1 below;

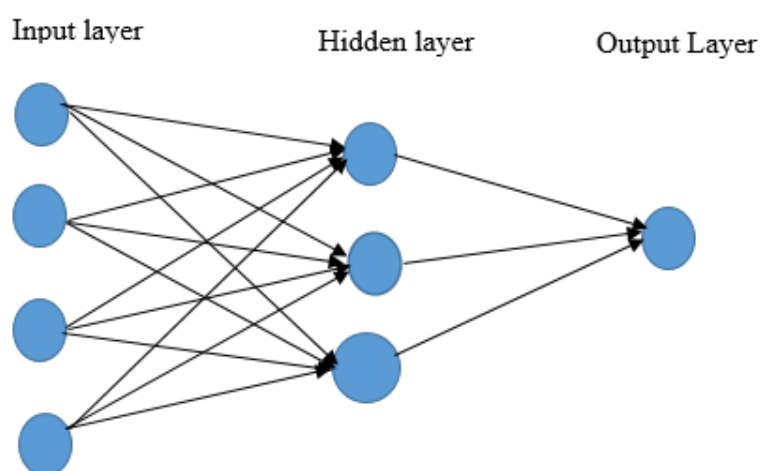


Figure 1. Schematic of 3-Layer ANN

ANN machine learning is a two-step integrated process, namely the training step and the testing step. The first stage is formed with a set of training dataset which is either supervised or unsupervised. In supervised training, the artificial network is trained to execute both input and output, but the unsupervised network provides only one input without output. In the other words, the test phase is analyzed with the test data.

Bias nodes are sometimes added to anticipatory neural networks to facilitate model explanation. The skew nodes function as an input node that always yield one (1) – a constant value or another constant. Because of this property, they are not connected to the previous layer. Constant 1 is called bias activation. All neural networks do not have a polarization node. Bias neurons are used to shift the output of an activation function. Regardless of the type of neuron, node, or processing unit, neural networks are always constructed from weighted connections between these units (Montana & Davis, 1989).

3.2 Data Description

In order to successfully categorize credit risk, we used data from RateSetter lending platform based in Australia. The data was obtained on www.ratesetter.com.au from October 2014 to March 2019 and contained 32,691 observations of matured loans. The data consists of 15

variables but excluded the following variables from the final analysis; date, identifier, outstanding principal and refund status. These variables however have no effect on the results obtained from the analysis. The remaining 11 variables were considered. Our data sample included 42.2% and 57.8% respectively of prepaid and delayed loans. The dependent variable in this classification modeling, also referred to as the output variable, is early repayment, which is not much different from defaulted loans, as has been used frequently in most studies. DV is a dichotomous with the values 0 and 1. The value 1 indicates that the borrower could not meet the repayment schedule and therefore made a late repayment and that 0 provides information on the borrower who was able to repay loans on time. The variables are described in Table 1 below;

Table 1. Variable Description

Variables	Description	Type
Contract length	Length of the finance contract in months	<i>Scale</i>
Annual rate	Annual interest rate paid by the borrower on the finance amount, where contracts are variable rate the rate disclosed is the interest rate originally specified in the finance contract	<i>Scale</i>
Finance amount	Total principal borrowed inclusive of capitalized fees and charges.	<i>Scale</i>
Loan Purpose	Use of finance proceeds declared by the borrower. Dummies were created for this variable.	<i>Categorical</i>
Borrower state	State in which the borrower lives also a dummy	<i>Categorical</i>
Borrower age	Age of the individual borrower	<i>Scale</i>
Employment status	Employment status of borrower at time of entering the finance agreement. The employment was coded as dummy variable	<i>Categorical</i>
Borrower income	Gross annual income of the borrower at the time of entering the finance agreement	<i>Scale</i>
Housing status	Describes the housing status of the borrower at the time of entering the finance agreement	<i>Categorical</i>
Gender	Gender of borrower	<i>Binary</i>
Dependent variable	Borrower has made additional payments over and above those set out in their repayment schedule	<i>Binary</i>

NB: dependent variable = Early repayment made

3.3 Research Design

An artificial neural network for classification, integrated into SPSS, was applied to analyze the dataset. The data was classified into two subsets: 70% for training and 30% was considered as handout models. The query was automatically executed so that the test data set of the training data includes default cases of about the same percentage. This is shown in Table 2 below;

Table 2. Data partition to build the model

Samples	No. Of Observations	Percent (%)
Training	23,008	70%
Testing	9,683	30%
<i>Total</i>	32,691	100%

The Step function of NN specifies binary function that has only two possible output values. This means that if the input value reaches a specific threshold, the output value gives a value and if a specific threshold is not reached, the output value is different. The dependent variable is therefore described in the equation below.

$$y = \begin{cases} 1 & \text{if } w_i k_i \geq \text{non paid threshold} \\ 0 & \text{if } w_i k_i \leq \text{repaid threshold} \end{cases} \quad (2)$$

This type of transfer function is generally used in an artificial neuron and known in theory as artificial perceptron neuron. Perceptron is used to solve classification problems and, as such, is most frequently found in the last layer of artificial neural networks. In the case of a linear transfer function, an artificial neuron performs a simple linear transformation in the sum of weighted inputs and bias. This artificial neuron contrasts with the perceptron but more commonly used in the input layer of artificial neural networks (Suzuki, 2011)

4. Results and Discussion

According to our data set, we use ten (10) neurons in the input layer that correspond to the independent variables. The output of SPSS further revealed seven (7) hidden layers with two (2) layers in the output. To complete the process of selecting the required neurons in the hidden layer, AUC approach was used. The AUC is calculated based on the area under the ROC curve. ROC measures the degree of association between TP and FP in the classification problem for different threshold values used. According to Byanjankar et al. (2015) the rule for AUC is based on its probabilistic value. All things being equal, the closer the AUC is to 1, the greater the probability assigned to the classification class for correct accuracy. However, the model predicted 77.0% of the classification algorithm with a focus on the correct classification of delinquent borrowers. We observed modelling biasedness from the middle (hidden) layer suggesting a meaningful explanation to the output layer. Figure 2 below illustrates the results of the credit rating of the neural network in the studies.

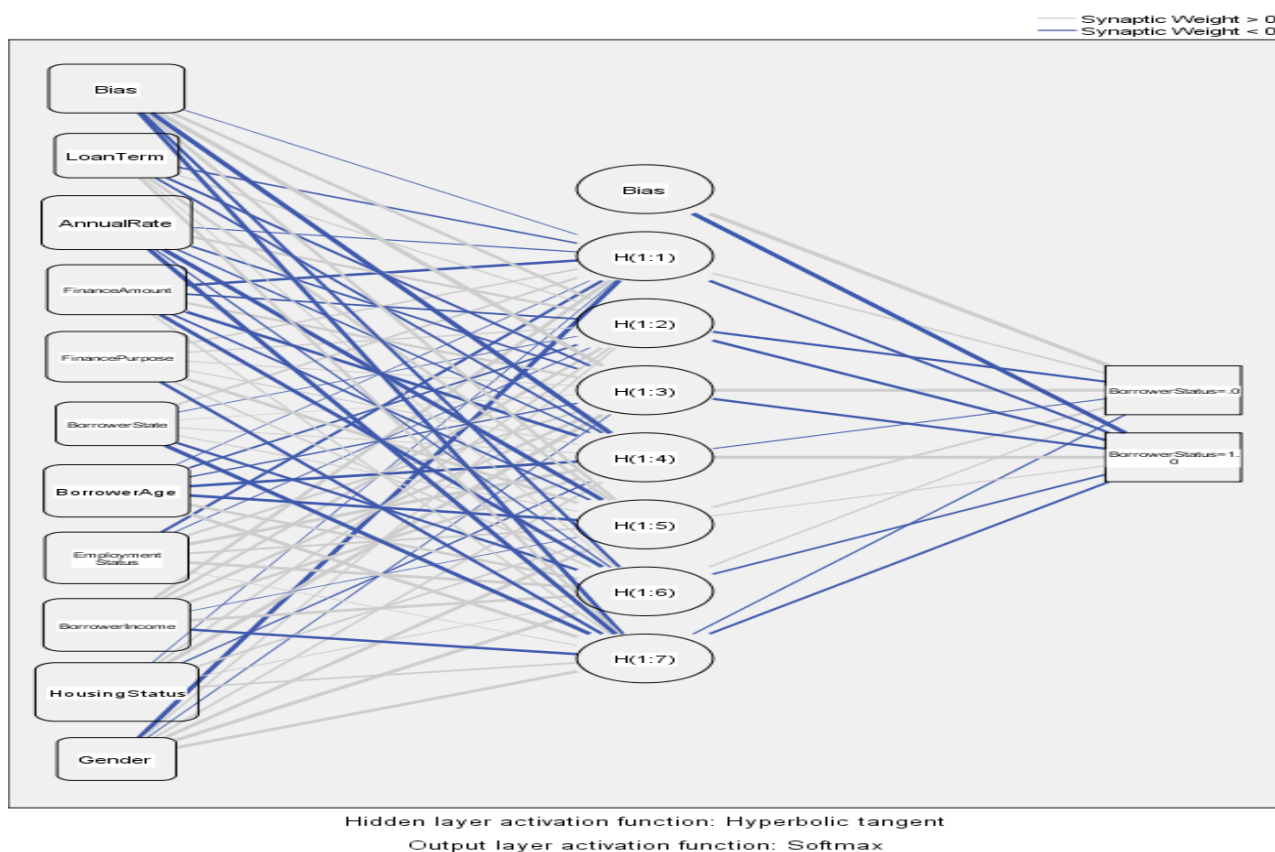


Figure 2. Neural network Credit rating model; Loan payment Classification of borrowers

Based on the results obtained, our model accurately classified 84.4% of the repayment of late loans among borrowers, while 54.6% was classified as prepaid loans for training data. At the same time, the accuracy classification observed for late repayment recorded 84.6% and 54.6% of good borrowers, also known as prepaid loans, from the test data set.

The above statistics are shown in Table 3 below;

Table 3. Prediction table

Samples	Actual	Prediction		
		Early Repayment	Late Repayment	Correct Percent
Training	Early Repayment	5303	4405	54.6%
	Late repayment	2077	11223	84.4%
Testing	Early Repayment	2225	1851	54.6%
	Late repayment	864	4743	84.6%

The behaviour of the individual variables in the model was examined by weighing their relative importance in the analysis. In the model, relative importance is described in relation to the steps described in (Byanjankar et al., 2015; Garson, 1991). The importance of the

variable indicates the extent of which independent variables are different based on ranking. The relative importance is illustrated in Figure 3 below. It turns out that the explanatory variable "payment time" is the primary determinant that influences the borrower's decision to pay acquired loan from RateSetter.com in Australia. The second most important variable was the "interest rate" followed by "the purpose of the loan." The repayment term of the loans has a strong influence on the credit decision of the borrowers and principally in contrast to the conclusions of Byanjankar et al. (2015). Sex (gender) is however the least important variable, suggesting that it has minimal influence on the borrower's repayment plan in our model.

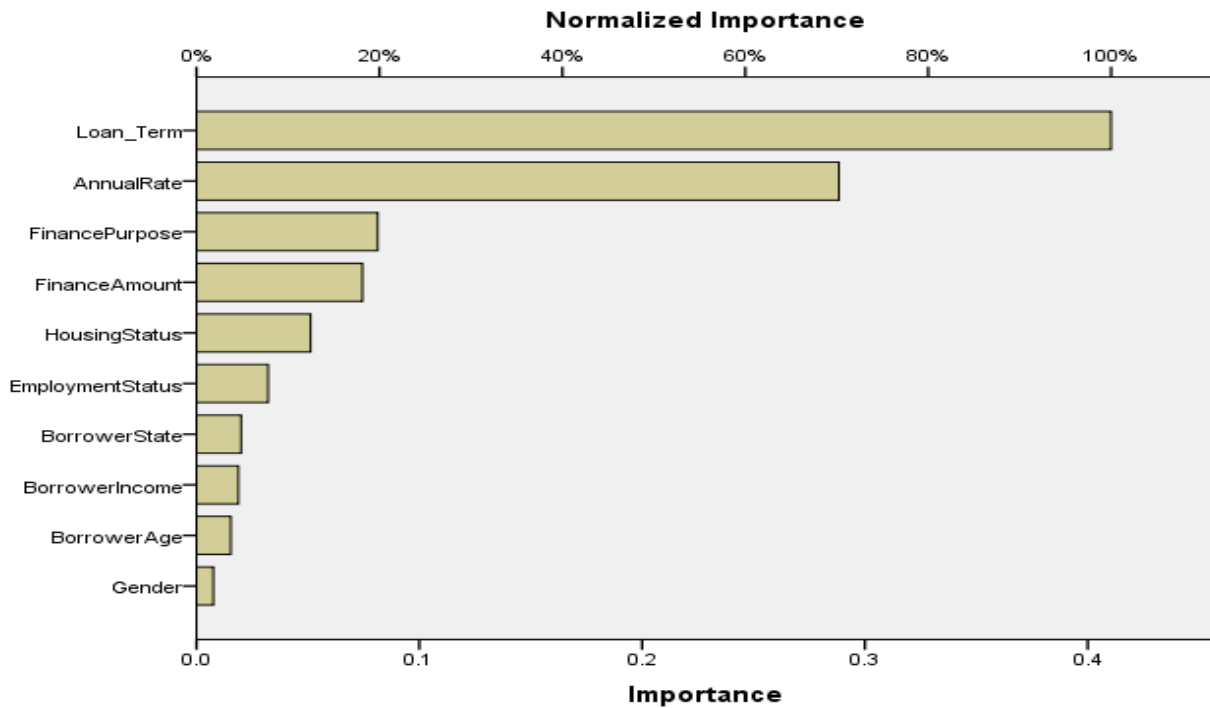


Figure 3. Relative importance of the variables as reported by SPSS

4.1 Model Comparison

The applied NN and logistics regression model successfully demonstrates the applicability of the credit rating for the classification and projection of loans repayment in P2P loans online. We have seen the application of big data mining in the studies by (Zhou, Chawla, Jin, & Williams, 2014) five years ago that analysed credit risk assessment. In the same approach, we used different data set to classify borrowers by examining the predictive power of each learning technique (i.e. binary logistic regression model and neural network). The results of both models are shown in Table 4 below.

Table 4. Comparison between NN and LR with test data

Samples	Actual	Prediction		
		<i>Early Repayment</i>	<i>Late Repayment</i>	<i>Correct Percent</i>
Neural Network	<i>Early Repayment</i>	2225	1851	54.6%
	<i>Late repayment</i>	864	4743	84.6%
Logistic Regression	<i>Early Repayment</i>	7772	6012	56.4%
	<i>Late repayment</i>	287	16030	84.8%

Table 4 above clearly shows that both models performed relatively well to accurately predict the repayment behaviour of borrowers. The results posit that logistic regression is more accurate in predicting borrowers who are more likely to pay their loans as determined in the loan contract application (56.4% against 54.6%) than the neural network. However, logistic regression outperformed the neural network by classifying customers based on default (84.8% vs. 84.6%). The empirical estimations show that LR has the highest filtering power of delinquent borrowers which is contrary to the findings of Byanjankar et al. (2015).

The task of the credit agencies goes beyond the determination of the credit rating of the borrowers, but it is not limited to the verification of the credit report of individuals and companies. However, financial institutions, microfinance and individual lenders can rely on reports such as current credit information, banks' balance sheets, overdue payments, bankruptcies, personal identification information and inquiries to rate debtors when applying for loans. Once big data algorithms correctly predicts the repayment behaviour otherwise refer in literature as the probability of default, potential and creditworthy applicants are allowed to invest and borrow. Interest rates on the other hand, are calculated based on the credit grade assigned to the borrower whereas the later rank customers based on soft and hard information. It can be argued that the higher the interest rate, the more likely the borrower will default on the acquired loan.

5. Summary and Conclusion

The new digital financial intermediation and its inherent risk of non-compliance demand rigorous data extraction among risk managers and researchers. In this study, authors analyzed borrowers' loan repayment structure to measure the risk of default in Australia. The study adopted logistic regression and the neural network to analyze this classification problem. These approaches offered meaningful explanation following this underscore and under-researched field of study. Both methods logistic regression and the neural network are effective in identifying the risk of borrowers based on the lifespan of the loan contracted but settled on the superiority of former compared with the later. We also explore the sequential importance of the I.Vs that online P2P markets must take into account. Compared to recent articles, our empirical results have drawn significant conclusions and contribute to the existing literature. However, we cannot generalize the results of our studies because the lending and credit filtering strategies of P2P marketplaces differ and studies utilized data set

from a lending platform. We therefore suggest the use of dataset from different lending platforms. In summary, future studies may also consider other machine learning algorithms to classify borrowers.

Reference

- Baesens, B., Setiono, R., Mues, C., & Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. *Management science*, *49*(3), 312-329. <https://doi.org/10.1287/mnsc.49.3.312.12739>
- Bahrammirzaee, A. (2010). A comparative survey of artificial intelligence applications in finance: artificial neural networks, expert system and hybrid intelligent systems. *Neural Computing and Applications*, *19*(8), 1165-1195. <https://doi.org/10.1007/s00521-010-0362-z>
- Bellotti, T., & Crook, J. (2009). Support vector machines for credit scoring and discovery of significant features. *Expert Systems with Applications*, *36*(2), 3302-3308. <https://doi.org/10.1016/j.eswa.2008.01.005>
- Bourquin, J., Schmidli, H., van Hoogevest, P., & Leuenberger, H. (1997). Application of artificial neural networks (ANN) in the development of solid dosage forms. *Pharmaceutical development and technology*, *2*(2), 111-121. <https://doi.org/10.3109/10837459709022616>
- Boyd, J. H., & De Nicolo, G. (2005). The theory of bank risk taking and competition revisited. *The Journal of finance*, *60*(3), 1329-1343. <https://doi.org/10.1111/j.1540-6261.2005.00763.x>
- Byanjankar, A., Heikkilä M., & Mezei, J. (2015). *Predicting credit risk in peer-to-peer lending: A neural network approach*. Paper presented at the 2015 IEEE Symposium Series on Computational Intelligence. <https://doi.org/10.1109/SSCI.2015.109>
- Eletter, S. F., & Yaseen, S. G. (2017). Loan decision models for the Jordanian commercial banks. *Global Business and Economics Review*, *19*(3), 323-338. <https://doi.org/10.1504/GBER.2017.083960>
- Foster, F. D., & Warren, G. J. (2016). Interviews with institutional investors: The how and why of active investing. *Journal of Behavioral Finance*, *17*(1), 60-84. <https://doi.org/10.1080/15427560.2015.1095754>
- Garson, G. D. (1991). Interpreting neural-network connection weights. *AI expert*, *6*(4), 46-51.
- Guo, Z.-Y. (2017). Heavy-tailed Distributions and Risk Management of Equity Market Tail Events. <https://doi.org/10.2139/ssrn.3013749>
- Ince, H., & Aktan, B. (2009). A comparison of data mining techniques for credit scoring in banking: A managerial perspective. *Journal of Business Economics and Management*, *10*(3), 233-240. <https://doi.org/10.3846/1611-1699.2009.10.233-240>
- Jagtiani, J., & Lemieux, C. (2017). Fintech lending: Financial inclusion, risk pricing, and alternative information. <https://doi.org/10.21799/frbp.wp.2017.17>

- Lee, T.-S., & Chen, I.-F. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Systems with Applications*, 28(4), 743-752. <https://doi.org/10.1016/j.eswa.2004.12.031>
- Li, J., Galley, M., Brockett, C., Spithourakis, G. P., Gao, J., & Dolan, B. (2016). A persona-based neural conversation model. *arXiv preprint arXiv:1603.06155*. <https://doi.org/10.18653/v1/P16-1094>
- Liu, D., Brass, D., Lu, Y., & Chen, D. (2015). Friendships in online peer-to-peer lending: Pipes, prisms, and relational herding. *Mis Quarterly*, 39(3), 729-742. <https://doi.org/10.25300/MISQ/2015/39.3.11>
- Malekipirbazari, M., & Aksakalli, V. (2015). Risk assessment in social lending via random forests. *Expert Systems with Applications*, 42(10), 4621-4631. <https://doi.org/10.1016/j.eswa.2015.02.001>
- Malhotra, R., & Malhotra, D. K. (2003). Evaluating consumer loans using neural networks. *Omega*, 31(2), 83-96. [https://doi.org/10.1016/S0305-0483\(03\)00016-1](https://doi.org/10.1016/S0305-0483(03)00016-1)
- Mezei, J., Byanjankar, A., & Heikkilä, M. (2018). Credit risk evaluation in peer-to-peer lending with linguistic data transformation and supervised learning. <https://doi.org/10.24251/HICSS.2018.169>
- Montana, D. J., & Davis, L. (1989). *Training Feedforward Neural Networks Using Genetic Algorithms*. Paper presented at the IJCAI.
- Oreski, S., Oreski, D., & Oreski, G. (2012). Hybrid system with genetic algorithm and artificial neural networks and its application to retail credit risk assessment. *Expert Systems with Applications*, 39(16), 12605-12617. <https://doi.org/10.1016/j.eswa.2012.05.023>
- Sufian, F. (2009). Determinants of bank efficiency during unstable macroeconomic environment: Empirical evidence from Malaysia. *Research in International Business and Finance*, 23(1), 54-77. <https://doi.org/10.1016/j.ribaf.2008.07.002>
- Tan, Z., Baah, A., Ding, G., Owusu-Ansah, P., & Agyemang, K. (2019). Modelling Default Risk of Borrowers: Evidence from Online Peer to Peer Lending Platforms in Australia. *Research Journal of Finance and Accounting*, 10, 10. <https://doi.org/10.7176/RJFA>
- Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 34(4), 2639-2649. <https://doi.org/10.1016/j.eswa.2007.05.019>
- Twala, B. (2010). Multiple classifier application to credit risk assessment. *Expert Systems with Applications*, 37(4), 3326-3336. <https://doi.org/10.1016/j.eswa.2009.10.018>
- Wang, H., & Greiner, M. E. (2011). Prosper-The eBay for Money in Lending 2.0. *CAIS*, 29, 13. <https://doi.org/10.17705/1CAIS.02913>
- West, D. (2000). Neural network credit scoring models. *Computers & Operations Research*, 27(11-12), 1131-1152. [https://doi.org/10.1016/S0305-0548\(99\)00149-5](https://doi.org/10.1016/S0305-0548(99)00149-5)

Zhou, Z.-H., Chawla, N. V., Jin, Y., & Williams, G. J. (2014). Big data opportunities and challenges: Discussions from data analytics perspectives. *IEEE Computational Intelligence Magazine*, 9(4), 62-74. <https://doi.org/10.1109/MCI.2014.2350953>

Copyright Disclaimer

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).