Loan Risk Prediction Using Transaction Information

1. Problem Statement

The objective of proposed work is to predict loan credit risk and determine the probability of non-payment of bank financial services e.g. whether a person will pay back a loan or not. The other objective of the project is to study the ability of neural network algorithms to handle the problem of predicting credit default that measures the creditworthiness of the loan application over a time period. However, there are many risks related to bank loans, for the bank and for those who get the loans. Risk prediction and monitoring is critical for the success of the business model. Credit risk is the probability that a customer won't be able to make a required payment, causing a loss for the bank or financial institute that provided the loan.

2. Background

To predict the credit default, several methods have been proposed. The use of method depends on the complexity of banks and financial institutions, size and type of the loan [1]. The commonly used method has been discrimination analysis. This method uses a score function that helps in decision making whereas some researchers have stated doubts on the validity of discriminates analysis because of its restrictive assumptions; normality and independence among variables. Artificial neural network models have created to overcome the shortcomings of other inefficient credit default models [1-3].

By using scoring models that are AI-based and use deep learning, banks and financial institutions can access more realistic predictions on credit risk, using customers' credit history and the power of big data. This way credit can be approved to the right people and better pricing options offered to people who deserve it [2].

Feed-forward neural network algorithm is applied to a small dataset of residential mortgages applications of a bank to predict the credit default. The output of the model will generate a binary value that can be used as a classifier that will help banks to identify whether the borrower will default or not default. This paper will follow an empirical approach which will discuss neural network-based models and experimental results will be reported by training and validating the

models on residential mortgage loan applications. As the final step in the direction, linear regression method is also performed on the dataset.

3. Methodology

Step1: Collect the data from kaggle.com (url: https://www.kaggle.com/wordsforthewise/lending-club). The dataset consists of more than 8.5 million records.

The feedforward network consists of an input layer with x input variables, y hidden layers and an output layer with one neuron that represents a classifier.

Step-2: The network is trained by using a supervised learning algorithm (i.e. backpropagation algorithm). The algorithm optimizes the neuron weights which minimizing the error between actual and desired output

Step 3: The output of the model will generate a binary value that can be used as a classifier that will help banks to identify whether the borrower will defaulter or not. Once the dataset will be trained, test the performance of the model on the test dataset.

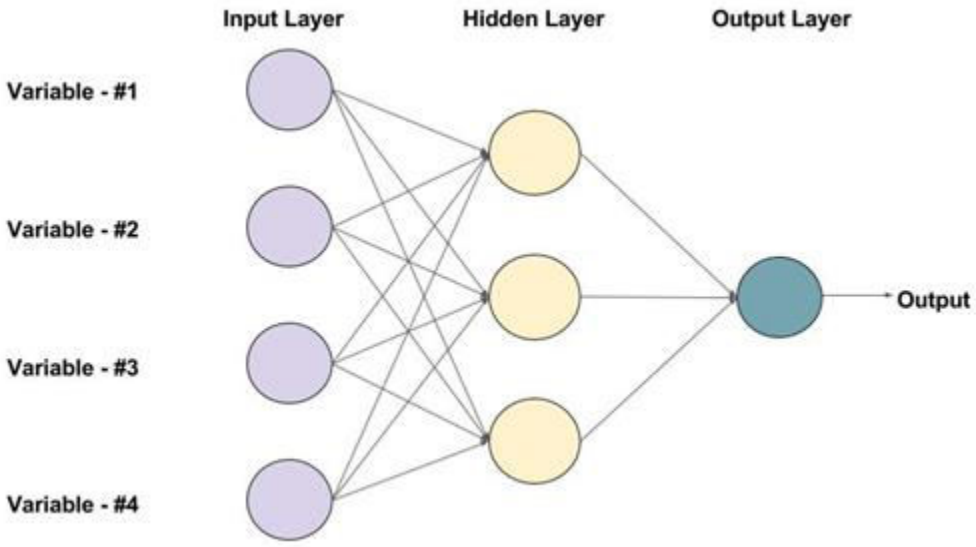

Figure 1: An example of a Feed-forward Neural Network with one hidden layer (3 neurons)

4. Experimental Design

4.1 *Evaluation Measures:*

The downloaded dataset will be divided in two parts 1) training dataset 2) testing dataset. Training dataset will be used to train the model and to validate the performance of the model, test dataset will be used. For measuring the accuracy any regression method will be applied in the same dataset.

4.2 *Software & Hardware Requirements:*

Python based Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Navigator, Python, and libraries such as Tensorflow, and Keras will be utilized for this process.

5. References

[1] Chen, N., Ribeiro, B., & Chen, A. (2016). Financial credit risk assessment: a recent review. Artificial Intelligence Review, 45(1), 1-23.

[2] Trustorff, J., Konrad, P., & Leker, J. (2011). Credit risk prediction using support vector machines. Review of Quantitative Finance and Accounting, 36(4), 565-581.

[3] Zięba, M., Tomczakb, S., & Tomczaka, J. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. Expert Systems with Applications, 58, 93-101.