**Default Loan Prediction Model**

**Sharlin Kahlon**

**Ryerson University**

**501124232**

**Abstract**

In finance industry, retail & commercial banks have set business goals to extend credit to their prospects and target high outstanding balances. However, bank has low risk appetite to factors like provision for credit loss and loans closed to collection. Despite modern era, commercial banks’ operations are cumbersome, lack automation and models to predict credit risk in short period of time. Lending credit to any individual or businesses include but not limited to adjudication based on their personal bureau report, previous paid or unpaid bankruptcy, score, income, net worth, financial ratios and retained earnings.

The main problem in extending credit is financial institute’s low risk tolerance and fail to determine probability of default. Secondly, loan officers face numerous challenges working manually with paper forms and dive deeper to analyse risks using financial statements which is time consuming and less efficient. Dataset used to train model is available publicly through google dataset search -url: *https://data.world/jaypeedevlin/lending-club-loan-data-2007-11/workspace/file?filename=lending\_club\_loans.csv*. The theme chosen to complete our research is classification using various methods like Decision Tree, Random Forest, Naïve Bayes emphasizing supervised learning. This paper focuses on three research questions: a.) evaluate performance of different classification models b.) which attribute plays major role in our decision? and c.) which model exhibit higher accuracy?

The tool used to train model is Python programming language and its libraries, Jupyter as an interface for coding, GitHub repository to store project deliverables and MS PowerPoint for final presentation. Trained model anticipates determining bad loans in initial stage of application request and would help bank to pre-approve or decline client’s request efficiently.