

Literature Survey

As the data is increasing daily due to digitization in the banking sector, people want to apply for loans through the internet. Machine Learning (ML), as a typical method for information investigation, has gotten more consideration increasingly. Individuals of various businesses are utilising ML calculations to take care of the issues dependent on their industry information. Banks are facing a significant problem in the approval of the loan. Daily there are so many applications that are challenging to manage by the bank employees, and also the chances of some mistakes are high. Most banks earn profit from the loan, but it is risky to choose deserving customers from the number of applications. There are various algorithms that have been used with varying levels of success. Logistic regression, decision tree, random forest, and neural networks have all been used and have been able to accurately predict loan defaults. Commonly used features in these studies include credit score, income, and employment history, sometimes also other features like age, occupation, and education level.

A recent development of machine learning techniques and data mining has led to an interest of implementing these

techniques in various fields [17]. The banking sector is no exclusion and the increasing requirements towards financial

institutions to have robust risk management has led to an interest of developing current methods of risk estimation.

Potentially, the implementation of machine learning techniques could lead to better quantification of the financial risks

that banks are exposed to. Within the credit risk area, there has been a continuous development of the Basel accords,

which provides frameworks for supervisory standards and risk management techniques as a guideline for banks to

manage and quantify their risks. From Basel II, two approaches are presented for quantifying the minimum capital

requirement such as the standardized approach and the internal ratings based approach (IRB) [16]. There are different

risk measures banks consider in order to estimate the potential loss they may carry in future. One of these measures is

the expected loss (EL) a bank would carry in case of a defaulted customer. One of the components involved in EL-

estimation is the probability if a certain customer will default or not. Customers in default means that they did not meet

their contractual obligations and potentially might not be able to repay their loans [18]. Thus, there is an interest of

acquiring a model that can predict defaulted customers. A technique that is widely used for estimating the probability of

client default is Logistic Regression [19]. In this thesis, a set of machine learning methods will be investigated and

studied in order to test if they can challenge the traditionally applied techniques.

A prediction is a statement about what someone thinks will happen in the future. People make predictions all the time.

Some are very serious and are based on scientific calculations, but many are just guesses. Prediction helps us in many

things to guess what will happen after some time or after a year or after ten years. Predictive analytics is a branch of

advanced analytics that uses many techniques from data mining, statistics, modeling, machine learning, and artificial

intelligence to analyze current data to make predictions. “Adyan Nur Alfiyatin, Hilman Taufiq [14] and their friends

work on the house price prediction. They use regression analysis and Particle Swarm Optimization (PSO) to predict

house price”. One other similar work on the Mohamed El Mohadab, Belaid Bouikhalene [15] and Said Safi to predict

the rank for scientific research paper using supervised learning. Kumar Arun, Garg Ishan and Kaur Sanmeet [13] work

on bank loan prediction on how to bank approve a loan. They proposed a model with the help of SVM and Neural

networks like machine learning algorithms. This literature review helps us carry out our work and propose a reliable

bank loan prediction model