Crowdlending and Credit Models

Pearpilai Jutasompakorn and Arif Perdana

Both are Assistant Professors in Accountancy Programme, Design and Specialised Business Cluster, Singapore Institute of Technology

A. Crowdlending Landscape in Singapore

Crowdfunding in the Southeast Asia region is still untapped. Only Malaysia, Thailand and Singapore have made a relatively significant impact through crowdfunding. Singapore managed just $7.5 million from 2103 campaigns (Statista, 2019).[1] This output is far from that of its US counterparts. In 2019, crowdlending in the US generated a total transaction value of $1.539 million from 217,400 campaigns (Statista, 2019).[2] They are geographically the second-largest nation (behind China) regarding transaction value raised through crowdlending projects.

Specifically, crowdlending can be divided into crowdinvesting or crowdlending. These two terms belong to the for-profit model, where investors act for profit rather than altruistically (e.g., donation-based crowdfunding). In crowdlending, borrowers seek funding for loans. Businesses can use these loans for personal and business purposes, which allows for the capture of long-tail business - as many of these capital seekers would not receive traditional funding because of their business/background. Crowdlending, therefore, matches a group of capital seekers (borrowers) with a group of capital providers (lenders). These capital providers can range from individual investors looking for alternative ways to diversify their portfolio to small and mid-sized enterprises (SMEs) with excess cash looking for higher returns. These loans are also typically unsecured, and intermediaries are often the only source of information between borrowers and lenders. Therefore, the due diligence of intermediaries is of paramount importance. However, intermediaries are usually unwilling to disclose their risk assessment procedures in the absence of regulatory due diligence requirements.

While the crowdfunding scene in Singapore is rather lackluster, they have cultivated a business climate for SMEs. The Basel Committee at Banking Supervision requires banks to quantify their capital requirements for their credit risk for traditional lending. For non-traditional lending platforms, credit risk is even more explicit. Crowdlending loans are unsecured and carry a high credit risk as the risk of loss for startups (the borrowers) can be significant. They are associated with a high default rate. Despite the high risk of crowdlending/P2P, there is no tailored regulatory legislation for crowdlending in Singapore. The regulator, the Monetary Authority of Singapore (MAS), applies existing rules restricting financial intermediation. At the time of this writing, protections remain unavailable for investors unless it is debt crowdfunding.

Automated Decision in Credit Scoring

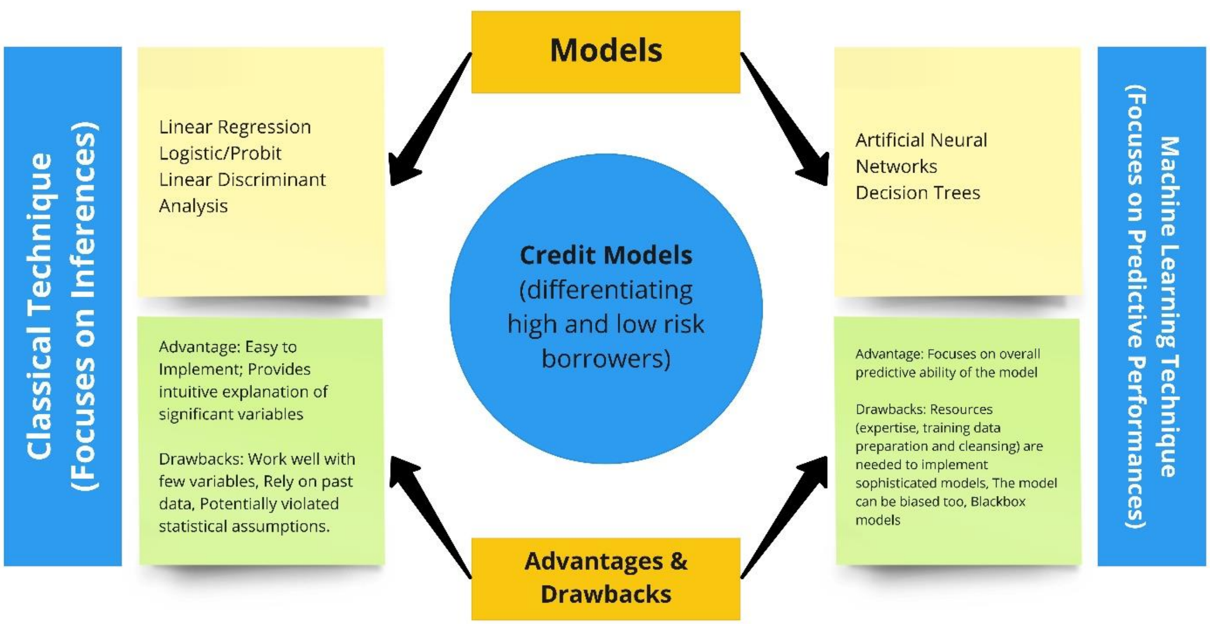
We can define the credit model as a systematic method for assessing credit risk that provides a consistent analysis of the factors identified as causing or influencing the level of risk. The accuracy of the credit model and approval decision can help the company mitigate the potential risk of loss from the default (i.e., approve a loan application from a high-risk borrower) as well as mitigate the potential forgone interest income (i.e.,

---

[1] Statista. (2019). Statista, Crowdfunding/Crowdinvesting. Retrieved from Statista: https://www.statista.com/outlook/377/124/crowdinvesting/singapore#market-revenue
[2] Statista. (2019). Statista, United States Crowdfunding Outlook. Retrieved from Statista: https://www.statista.com/outlook/335/109/crowdfunding/united-states

reject a loan application from a low-risk borrower). Traditionally, credit scoring models differ based on loans, such as real estate, consumer, and small business loans. Consumer and small business lending is scored by determining the credit quality of the borrower (e.g., personal characteristics and behavioural analysis of the business owner) combined with computerized financial analysis of the borrower (e.g., financial statements). Mid-market commercial and industrial businesses is based on the 5Cs (character, capital, collateral, capacity, and economic conditions) and financial statement analysis. Whereas credit analysis for Large commercial and industrial corporations lending relies on rating agencies, market analysts and sophisticated credit scoring models (e.g. Altman Z-Score).

A trend emerges in crowdlending to move away from traditional approaches to identifying potentially delinquent borrowers who are not delinquent. Crowdlending are leveraging on their ample data and using both classical and machine learning. While the classical model focuses on making inferences about the critical variables in credit models, the machine learning model focuses on the predictive ability of all variables in the credit model. The machine learning model for credit scoring is supervised learning, which means that the target variable in the training data is labeled with known outcomes (e.g., default or non-default). The model uses the training data for model development and then applies this model to new data with unknown outcomes. The machine learning model focuses more on the overall predictive performance than looking at the individual contributing variables in the model. Therefore, the goal of machine learning is to improve efficiency and effectiveness in credit risk management. The figure below illustrates the two techniques. We also briefly describe the most used models among the classical and machine learning techniques. Our explanation is intended for non-technical readers. We provide simple illustrations to complement our description.

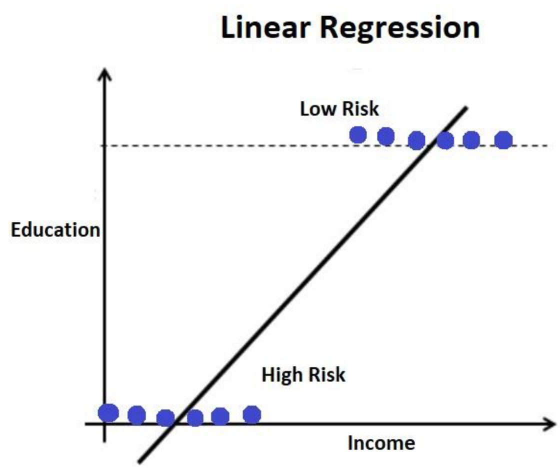


Classical Credit Scoring Models

There are four widely accepted classical credit scoring models:

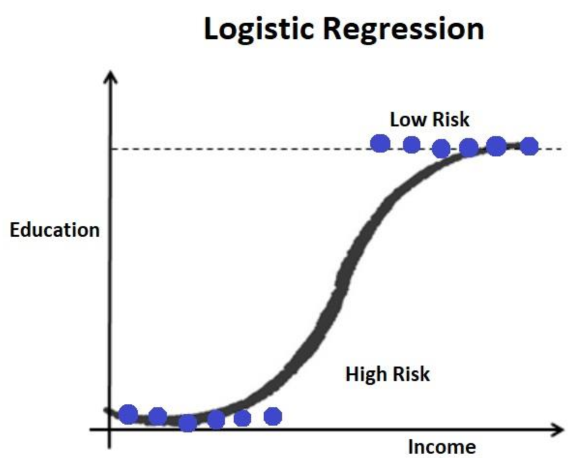
1. Linear Regression Model

Also known as a linear regression model, this is the most used model in practice; it provides a relatively robust estimate based on Bernoulli assumptions given the available information. With the probabilistic distributional assumption, the probability of default can be estimated. These models are developed by regressing a selection of quantitative or qualitative variables reflecting the characteristics of the (loan) applicant against a dependent variable that takes the value of 1 (if the loan is in default) and 0 (if the loan is not in default). In other words, the probability of default is assumed to be explained by characteristics of the borrower. The predicted value may fall outside the range of 0 to 1 under the linear probability model.



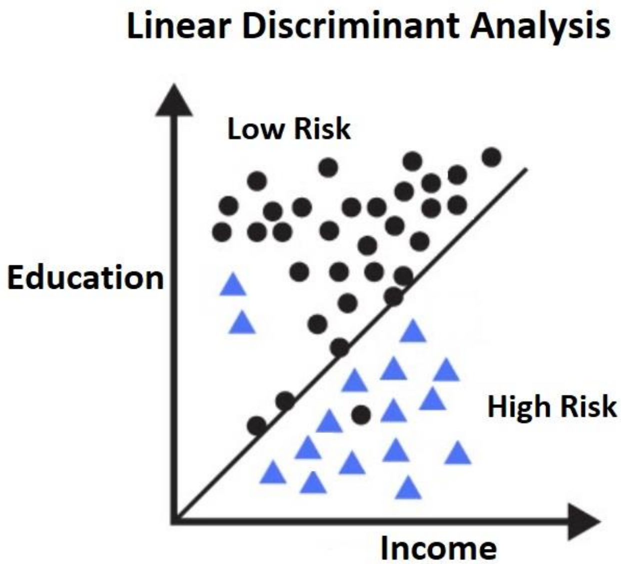
2. Logistic Regression/Probit Model

Both logit and probit models represent a refined form of regression in which we restrict the default probabilities to a range of 0 to 1. Both models assume Bernoulli distribution on the dependent variable. Logistic and probit are link function meaning that they describe the relationship between the predictor and the mean of the distribution function. The variables can be discrete in both models, which is not possible in discriminant analysis and linear regression (probability) models. Thus, the logistic model in credit scoring aims to determine the conditional probability that a given observation belongs to a class, given the values of the independent variables of the credit applications. For example, if the high risk is equal to 0 and low risk equal to 1, any probability that is less than 0.5 will be classified as high risk. In contrast, the probability above 0.5 will be classified as low risk.



3. Linear Discriminant Analysis Models.

The linear discriminant analysis (LDA) model attempts to establish a linear classification rule that distinguishes between particular groups of borrowers (between delinquent and non-delinquent borrowers). LDA assumes that the data is normally distributed. In other words, the discriminant analysis model divides borrowers into high and low default risk classes. We can build a discriminant analysis model in a stepwise fashion, with all variables reviewed and evaluated at each step to determine which contributes most to the distinction between groups.



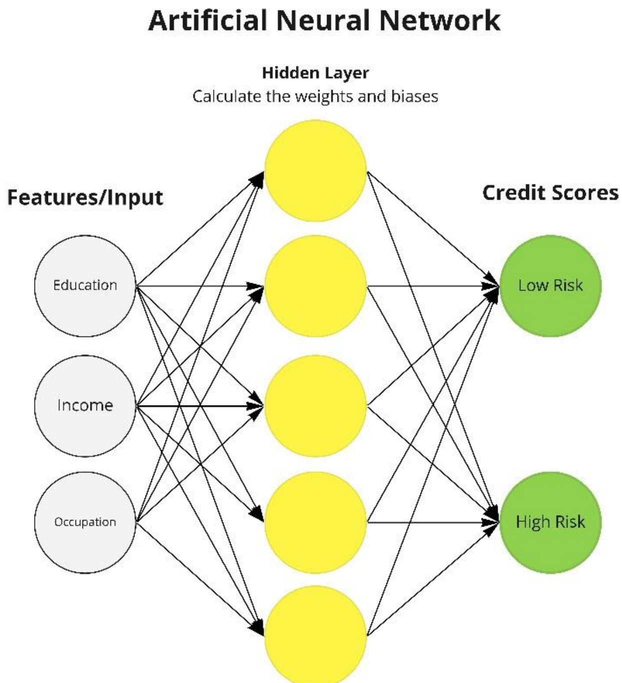
Classical models are relatively easy to implement and can work with small data. However, there are three disadvantages that can occur when using classical models. First, classical models are not optimal when the analysis involves numerous influencing variables. Second, the models rely heavily on the applicant's past credit history to approve applications. This reliance leads to a generalization of rejected applicants because they have no credit history. Second, these classical models make statistical assumptions that are frequently violated in the practice of credit modeling, making these techniques theoretically invalid for finite samples. An example of an assumption that is frequently violated in statistical credit models is multivariate normality for independent variables.

Artificial Intelligence and Machine Learning (AI/ML) Models

AI/ML models can be used as alternatives to traditional statistical models. Beneath are the four commonly used model for credit evaluation.

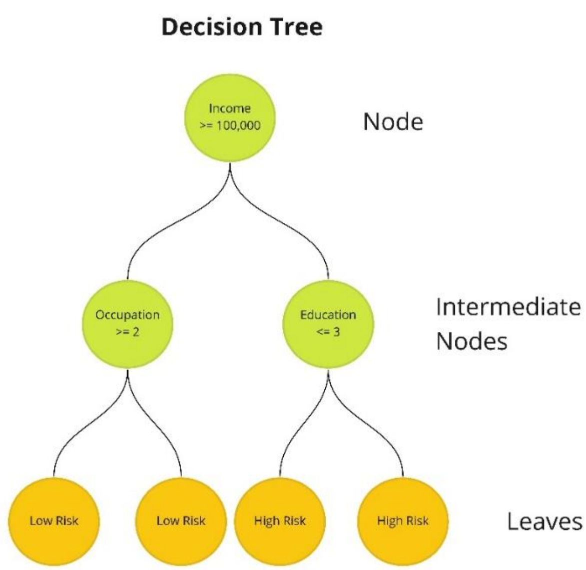
1. Artificial Neural Networks

An artificial neural network (ANN) is a loosely modeled on the human brain to recognize patterns. The idea of ANN is a special case of logistic regressions where numbers of models are stacked together. This allows the interconnections between models which have the similar structure of brain function via neurons. For example, a credit score is classified as high and low risks. Neural networks require labeled training data to develop the model. The features from the customer data form the input which is processed in multiple layers (neural networks). This neural network calculates the weight and bias for each feature and determine its probability. The features that have the highest probability are applied to the target variable for credit scoring, resulting in whether borrowers can be classified as high or low risks.



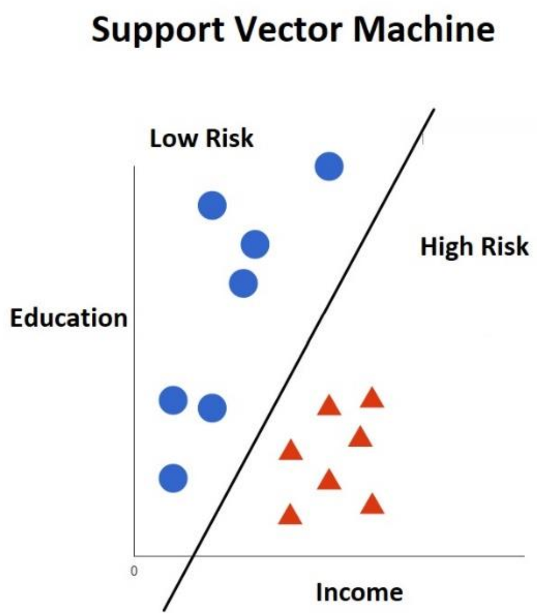
2. Decision Trees

A decision tree is a tool in decision analysis that can be used to visually represent decisions. It is a classification model provided yes/no structure for each node. then the multiple nodes are combined to a tree strucures. In a decision tree, the decision path is represented in nodes, intermediate nodes, and leaves. Following the previous example of credit score, there could be two leaves (i.e., high, and low risks) in the decision tree. The nodes and intermediate nodes appear according to the credit data features. Each leaf indicates the probability. The highest probability in a particular leaf is the selected credit score.



3. Support Vector Machines

Support Vector Machine (SVM) technique is a discriminative classifier that divide the borrowers' data. SVM can then draw a line to provide a decision boundary, for example, for potentially default and not default borrowers. The technique can work with small datasets. While LDA required normal distribution assumption, SVM does not require such assumption.



Although machine learning models are better than classical models in their predictive ability, no "one size fits all" machine learning for building credit scoring models. However, ensemble learning can optimise machine learning models development. Ensemble learning is a machine learning paradigm in which multiple models (commonly referred to as "weak learners") are trained to solve the same problem. There are three potential drawbacks of machine learning models. First, expertise and a large amount of relevant data are required to build an appropriate model. Developers needs to perform training data curation appropriately. Second, a machine learning model can also be biased. Even if the collected data does not contain any visible biased variables, other existing variables may have hidden bias. When this hidden bias is present, model developers need to be aware of it. They can preprocess the data and optimise the model algorithm to minimise the possible bias. Third, there is no intuitive explanation for significant variables in machine learning models such as Neural Networks. Machine learning models can potentially form a "black box" where it is difficult to understand how the model generates the predictions.

Conclusion

Machine learning models allows crowdlending companies to increases effectiveness and efficiency in credit approval process compared to traditional approaches; companies can leverage on existing data, recalibrate model in real-time, quickly evaluate several potentially influencing variables. However, companies need to be aware of the benefits and limitations of models that use either classical or machine learning models.