

**Micro-Credit Defaulter Model**

**Submitted by:**

**Shivakumar H R**

**ACKNOWLEDGMENT**

Micro credit defaulter lending default rates are typically lower than those experienced by banks but are likely to grow as the service is offered more widely. In this paper, credit scoring techniques are reviewed, and that knowledge is built upon to create an appropriate machine learning model for airtime lending.

**INTRODUCTION**

* **Business Problem Framing**

Airtime in developing countries is quickly becoming a basic commodity among the rapidly growing middle class. Failure to have sufficient airtime in order to communicate or load data bundles is proving a challenge for many prepay customers . Many mobile network operators (MNOs) active in emerging markets have spotted this as an opportunity to offer their subscribers short-term airtime loans at a moderate interest rate. This new service has the potential to increase their average revenue per user. Repayment of the loan is achieved once the subscriber’s account is credited. However, the risks of defaulting need to be analysed to obtain a deeper understanding of this new product. In this paper, we define default as a subscriber failing to repay the loan within a specified time frame. This risk transcends most institutions that lend money to their customers. To mitigate this risk, credit scoring models are required to assess the capability of the customer to pay a certain amount within the specified period Many microfinance institutions (MFI), experts and donors are supporting the idea of using mobile financial services (MFS) which they feel are more convenient and efficient, and cost saving, than the traditional high-touch model used since long for the purpose of delivering microfinance services. Though, the MFI industry is primarily focusing on low income families and are very useful in such areas, the implementation of MFS has been uneven with both significant challenges and successes.

* **Conceptual Background of the Domain Problem**

We have identified two distinct mechanisms in the airtime lending industry. In the first, the MNOs offer airtime loans to customers and bear the risk of non-performing loans. An example of a company that uses such a mechanism. When a subscriber runs out of airtime, they are able to borrow money equivalent to the amount topped up. Airtime lending is unique because loan repayment is encouraged by the customer’s need to use services such as calling, messaging, access to the Internet and mobile applications. A customer cannot use these services if they have no airtime. Buying and loading the airtime onto the customer’s account provides a direct loan repayment process. This incentive structure distinguishes airtime lending from other credit services. They understand the importance of communication and how it affects a person’s life, thus, focusing on providing their services and products to low income families and poor customers that can help them in the need of hour.

* **Review of Literature**

The main purpose of this part of the study is to understand the application of credit scoring in the financial services sector. Based on previously published research, it is possible to identify commonly used datasets, relevant features and the types of models constructed. This valuable information then provides a basis for developing a credit scoring system for mobile airtime lending. While the literature review does not offer any papers specifically addressing airtime lending, a number of papers are identified that undertake research in related fields, in particular microfinance in developing countries, and these offer useful insights. The following paragraphs summarize the different model structures that have been deployed and describe their performance and the variables commonly used.

* **Motivation for the Problem Undertaken**

In order to understand the business implications of credit scoring, it is necessary to consider the profits in relation to varying levels of default that will likely result as the volume of customers . This is calculated by adding the total value of the loan to the interest that is incurred and then subtracting the loans that defaulted. When the default rate is as low as 0.01%, it is better to accept all the loan requests. The model cannot beat this simple approach. However, as the default rate increases to >2%, the company will generate more profits by using the model rather than approving all loans. When there are no defaults, the company makes a profit equal to the interest rate. The company breaks even with zero profits at a default rate of 8%. In contrast, by using the model, the maximum default rate that the company can tolerate before making losses is 32%. This means that the model effectively quadruples the level of default risk that can be tolerated. They are collaborating with an MFI to provide micro-credit on mobile balances to be paid back in 5 days. The Consumer is believed to be defaulter if he deviates from the path of paying back the loaned amount within the time duration of 5 days. For the loan amount of 5 (in Indonesian Rupiah), payback amount should be 6 (in Indonesian Rupiah), while, for the loan amount of 10 (in Indonesian Rupiah), the payback amount should be 12 (in Indonesian Rupiah).

**Model/s Development and Evaluation**

* **Identification of possible problem-solving approaches (methods)**

**Data Analysis:**

Building on this meta-analysis, the next step towards building a credit scoring model is to perform an exploratory analysis of the data and provide summary statistics about the variables. The study period is conditional on the data made available by Indonesian Rupiah. After negotiating legal and confidentiality agreements with Indonesian Rupiah as part of the practicum organized, These data were associated with three million loans from a total of 46 thousand customers. Additional time was required to provide clearance for publication of the results given the commercial sensitivity of the study. Loans are consider in order to evaluate each individual customer’s performance in the next three months (performance window) following the previous loan. Those who did not pay within the performance window are classified as defaulters. Table 3 describes the characteristics of the Indonesian Rupiah dataset used in the study and Table 4 provides summary statistics for the average loan duration, loan count and average usage amount for both non-defaulting and defaulting customers. The average is calculated because each customer could have accessed multiple loans in the past. All distributions are non-negative and right skewed with the majority of customers associated with smaller values. An inspection of non-defaulting and defaulting customers suggests remarkably similar behaviour for loan duration but defaulters tend to have lower loan counts and lower usage amounts.

* **Testing of Identified Approaches (Algorithms)**

Three machine learning models are considered in order to study the relevance of nonlinearity and the potential benefits of using an ensemble technique. By utilising the three models, it is therefore possible to determine the most appropriate model structure for describing the relationships between the explanatory variables and default for airtime lending. First, logistic regression (LR) provides binary classifications using linear relationships . LR is traditionally used as a relatively simple model structure and sets a benchmark for comparing the performance of the classifiers. Second, a decision tree (DT) is constructed to assess the potential improvement using a nonlinear model. While offering the possibility of a nonlinear model structure, DT has the added benefit of resulting in a set of rules that are relatively easy to implement. Third, an ensemble approach known as Random Forest (RF) is deployed by averaging over a collection of decision trees . By including the RF model, it is possible to ascertain whether ensemble techniques can offer any benefits beyond the LR and DT techniques.

* **Run and Evaluate selected models**

In the financial services sector, it is more important to predict those who will default than those who will repay. This is because the financial risk associated with defaulters is high. A confusion matrix is used to evaluate the classification models with positive (negative) outcomes denoting repayment (default), respectively. This 2x2 matrix measures the number of predicted/actual cases that are True Positive (TP), True Negative (TN), False Positive (FP) and False Negative (FN). From this matrix, it is possible to calculate the classification accuracy. where the cost of one non-performing loan from a customer defaulting is much greater than the benefit of a customer repaying, a factor of ten in this case. The loss from rejecting a customer that will repay is much less than that suffered from approving a loan for a customer that will default. Accuracy is a useful summary statistic but is not the most relevant performance metric for this particular business application. The greatest threat to financial sustainability arises when the classifier predicts that a customer will repay a loan and they actually default (FP). Therefore, it is most important to correctly predict the customers who will not repay. For applications that require highly effective detection ability for only one class, it is recommended to consider an alternative metric to accuracy (Tang et al. 2009). The loan approval application is best assessed using the classification metric known as specificity defined as: Specificity = TN/(TN + FP)

* **Key Metrics for success in solving problem under consideration**

When evaluating the performance of a predictive model, it is necessary to create a distinct training and testing dataset in order to avoid over-fitting and ensure the model will generalise. Performance evaluation metrics can either be in-sample (data used for training is also used for testing) or out of sample (the data used to train the model is different from what is used for testing). An out-of-sample (OS) evaluation approach is required to ensure that the results are likely to generalize to new datasets. When dealing with data where the distribution of two prediction classes are highly imbalanced (low percentage of defaulters), a lot of care should be taken when creating the training and testing datasets.

* **Visualizations**

This is calculated by adding the total value of the loan to the interest that is incurred and then subtracting the loans that defaulted. Figure 1 shows the profit as the default rate increases. When the default rate is as low as 0.01%, it is better to accept all the loan requests. The model cannot beat this simple approach. However, as the default rate increases to >2%, the company will generate more profits by using the model rather than approving all loans. When there are no defaults, the company makes a profit equal to the interest rate. The company breaks even with zero profits at a default rate of 8%. In contrast, by using the model, the maximum default rate that the company can tolerate before making losses is 32%. This means that the model effectively quadruples the level of default risk that can be tolerated.

* **Interpretation of the Results**

For the in-sample and first two cross-validation evaluations, CV1 and CV2, all models have high accuracy but struggled to achieve a reasonable specificity. Any simple benchmark classifier that predicts that all the loans will be repaid can easily achieve a high accuracy. However, the low specificity reveals its inability to predict the customers who will default. CV3, which has a balanced distribution of customers who repaid and defaulted based on accuracy, performed worse than the other two cross-validation scenarios. This is because, when default rate is close to 50%, it becomes more difficult for the model to generate accurate predictions although the specificity is high. In order to adequately evaluate the potential of predicting the customers who default, more data about loans and customers who defaulted are required. This is why evaluation scenario CV3 specifically operates on a dataset with one loan per customer and approximately half of these resulted in default by design to obtain a balanced set of categories. The variables collected provided information about previous loans. Adding the month that the loan was borrowed as dummy variables increased the accuracy of the model. This is likely due to annual seasonality in customer’s incomes. The number of recharges is not selected as a relevant variable.

**CONCLUSION**

An innovative financial product known as an airtime credit service (ACS), which is a cashless microloan that allows users to easily access airtime on a credit basis was investigated using an empirical analysis of over three million loans belonging to more than 41 thousand customers. The study started with a meta-analysis of previous publications that consider the construction of credit scoring algorithms and helped to select the relevant features and model structures that have been effective. The aim was to determine an appropriate quantitative model for using financial information pertaining to the loan and customer behaviour on the mobile network to predict the outcome of the loan. Binary classifiers are appropriate for dealing with the two discrete outcomes for customer behaviour: repayment or default. Three different machine learning model structures were considered: logistic regression; a decision tree; and random forests.

* It would be beneficial to have access to more variables. The literature review suggests that obtaining customer details from the MNOs would improve performance.
* For a classification problem with an imbalanced number of categories, classification accuracy is not an appropriate performance metric. One solution is to use specificity instead in order to focus on the classifier’s ability to predict the occurrence of a particular category. In the case of credit scoring, predicting defaulters is of utmost importance.
* Great care is required when selecting the appropriate cross-validation technique. There are many factors that affect the data structure and the specific application should be taken into consideration. For credit scoring, correct handling of the time of loan disbursement and customer identity are crucial to avoid over-fitting and unrealistically high estimates of accuracy.
* Both nonlinear classification models outperformed logistic regression demonstrating the added value of using a nonlinear model structure