Article

Use of Machine Learning Techniques to Create a Credit Score Model for Airtime Loans

Bernard Dushimimana [1], **Yvonne Wambui** [1], **Timothy Lubega** [1] and
Patrick E. McSharry [1,2,3,*]

1   College of Engineering, Carnegie Mellon University Africa, Kigali BP 6150, Rwanda;
     mathberd@gmail.com (B.D.); yvonne.wambui.gitau@gmail.com (Y.W.); lubzo23@gmail.com (T.L.)
2   African Center of Excellence in Data Science, University of Rwanda, Kigali BP 4285, Rwanda
3   Oxford Man Institute of Quantitative Finance, Oxford University, Oxford OX2 6ED, UK
*   Correspondence: patrick@mcsharry.net

check for
updates

Abstract: Airtime lending default rates are typically lower than those experienced by banks and microfinance institutions (MFIs) 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. Over three million loans belonging to more than 41 thousand customers with a repayment period of three months are analysed. Logistic Regression, Decision Trees and Random Forest are evaluated for their ability to classify defaulters using several cross-validation approaches and the latter model performed best. When the default rate is below 2%, it is better to offer everyone a loan. For higher default rates, the model substantially enhances profitability. The model quadruples the tolerable level of default rate for breaking even from 8% to 32%. Nonlinear classification models offer considerable potential for credit scoring, coping with higher levels of default and therefore allowing for larger volumes of customers.

---

1. Introduction

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 (Madis 2015). 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 of around 10% (Chen and Faz De Los Santos 2015). 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 analyzed 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.

Credit scoring is a tool used by financial institutions to facilitate the decision-making process of whether to accept or reject a loan (Deloitte 2017). The advantages include: enabling faster credit decisions; reducing the cost of credit analysis; and monitoring the portfolio of existing accounts (Mode 2017). Credit scoring models are typically estimated using a variety of historical personal and financial data obtained from customers. In developing countries, however, where there is a large population of unbanked adults, such data are not readily available (ComzAfrica 2017). Therefore, there is a need to search for alternative datasets in order to determine whether a customer

is creditworthy. MNOs can monitor the customer's calls and recharge history to determine their creditworthiness because they have access to such data.

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 is Safaricom, through its service known as 'okoa jahazi'. When a subscriber runs out of airtime, they are able to borrow money equivalent to the amount topped up in the last seven days with the expectation that it will be repaid within five days (Safaricom 2017).

The second mechanism involves a partnership between the MNO and a third party lender such that the MNO provides access to the customers and mobile network and the risk is transferred to the third party. An example of a lender which relies on such a partnership is ComzAfrica. Within such an ecosystem, MNOs tend to be protective of their customers' privacy which limits the amount of data shared with the lenders. Consequently, lenders have less information about their customers than financial institutions usually rely upon. In this mechanism, the third party lender pays the MNO the credit amount before the customer repays. Hence, in the case of a default, the third party loses the full amount. To remain financially viable, the lender must ensure that the profits from loans successfully repaid exceed the losses from non-performing loans. For example, for every USD 10 lost through a default, the lender needs ten customers to borrow and repay airtime amounts of USD 10.

This second airtime lending mechanism is the basis of study presented here in this paper. A partnership with ComzAfrica during a Carnegie Mellon University (CMU) practicum offered a unique opportunity to collect data and investigate the role of credit scoring based on machine learning (ML) approaches. The study compares a number of models and cross-validation techniques to determine whether credit scoring is required and its impact given the limited data available.

1.1. Financial Lending Industry

The main actors in this industry are financial institutions and money lenders who lend to those they deem creditworthy. These loans are usually monetary and the expectation is that the amount is repaid within a specific time frame with interest. This industry has evolved over time and extends beyond financial institutions to include players in other industries such as the telecommunication sector (Aker and Mbiti 2010).

In most emerging economies, a significant number of MNOs are also competing to offer credit services in the form of mobile money and airtime (Aamo et al. 2017). Subscribers can apply for airtime loans and it is at the discretion of the MNO to offer the airtime loan at a specified interest rate to be repaid within a particular time period (Chen and Faz De Los Santos 2015). While mobile lending is innovative and has the potential to improve financial inclusion, it also poses challenges for regulators.

1.2. Credit Scoring in Related Industries

The concept of credit scores dates back to the 1950s (Pell 2017), when the lending decisions were made by loan approval officers. This method was not effective as it relies on the subjective judgment of the individual loan officer. Furthermore, there was no accurate way of determining and monitoring the defaulters and non-defaulters.

In the 1950s, Bill Fair and Earl Isaac introduced a statistical number designed to represent the creditworthiness of an individual and most of the time their predictions were accurate. It was not until the 1970s that these credit scores became an important component of the lending industry. Many lenders, from banks to microfinance institutions, currently use credit scoring to measure a potential borrower's creditworthiness (Abdulrahman et al. 2014).

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.

1.3. ComzAfrica

ComzAfrica is a micro-lending company operating in 16 countries across Africa and Asia. ComzAfrica has built an Airtime Credit Service (ACS) which allows users to access airtime on a credit basis. Given that the users do not always have access to a retailer or direct funds, this service allows them to access airtime on a credit basis and make calls or send messages (ComzAfrica 2017).

The service is offered through two main interfaces: Short Message Service (SMS) and Unstructured Supplementary Service Data (USSD). When a customer makes an airtime on credit request using one of the interfaces mentioned above, the system will check if they meet certain eligibility criteria. If the criteria are satisfied, the airtime requested is loaded onto their account. Repayment occurs when the customer performs their next recharge (ComzAfrica 2017).

1.4. Theoretical Framework

This study uses an empirical approach to determine the business value of using machine learning to predict the behaviour of customers that accept airtime loans. The theoretical framework develops a logical relationship between explanatory variables and loan repayment outcomes and provides a mechanism for exploring the required information through quantitative analysis, guided by a meta-analysis of published studies on credit scoring. The variables include financial information pertaining to the loan and customer behavior on the mobile network. The outcomes are binary in that a particular loan is either repaid or results in a default. Three classification models are evaluated and compared: logistic regression; a decision tree; and an ensemble approach known as random forest. Predictability is measured by calculating classification accuracy metrics. In order to avoid over-fitting and to ensure that the results generalise to new loans and customers, careful attention is given to designing an appropriate cross-validation approach.

1.5. Motivation and Contribution

Motivation for this study arises from the fact that, according to the GSMA, there are over a billion registered mobile money accounts and close to USD 2 billion in daily transactions. The study contributes to a better understanding of how mobile money is enhancing financial inclusion and offering access to a range of financial services. First, it investigates an innovative form of lending known as an airtime credit service (ACS), which is a cashless microloan that allows users to easily access airtime on a credit basis. Second, airtime lending is growing rapidly, available in 16 countries in Africa and Asia and accounting for one fifth of airtime, thereby increasing the need for a systematic approach to credit scoring as requested by ComzAfrica. Third, a comparison of different models provides information about the benefits of different model structures and the relevance of nonlinearity. Fourth, a rigorous cross-validation approach is designed to ensure that no overfitting takes place and that the estimated performance can be generalized. Fifth, the success of the credit scoring models developed here demonstrates that it is possible to predict lending outcomes without customer demographics. Sixth, the relevance of the results to the business application is assessed by considering the costs and benefits associated with all four classification outcomes in the confusion matrix (true positives, true negatives, false positives and false negatives), thereby utilizing a detailed evaluation of profitability to identify when a classifier has the potential to offer a business case for its implementation. This approach is used to understand whether a credit scoring model is needed and to assess its economic value.

The paper is organized as follows. Section 2 provides the literature review for the study by gleaning information using a meta-analysis of relevant published papers that explore the utility and accuracy of credit scoring. Section 3 presents the methodology which describes the data collection, analysis, features, evaluation techniques and the role of cross-validation. Section 4 provides the results for all the machine learning models and different cross-validation techniques and discusses the implications for business in terms of economic costs and benefits. Section 5 concludes by presenting the key findings of the study.

2. Literature Review

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.

The studies use several models such as linear and quadratic discriminant analysis (LDA, QDA) (Blanco et al. 2013), logistic regression (Baklouti 2013), binary logit model (Van Gool et al. 2009), classification trees, multilayer perceptron (MLP), support vector machine (SVM), random forest and boosting (Cubiles-de-la-Vega et al. 2013). The best performing models were selected and described in Table 1 in terms of the amount of data and number of variables. Evaluation criteria presented in the papers included area under the receiver operating curve (AUC) (Blanco et al. 2013), Kolmogorov–Smirnov Statistic (KS) (Van Gool et al. 2009) and expected misclassification cost (EMC) (Cubiles-de-la-Vega et al. 2013).

Table 1. Summary of best performing models.

Papers	#Variables	Model	Defaulters	Non-Defaulters	ACC%	N
Cubiles-de-la-Vega et al. (2013)	39	MLP	2673	2778	88.33	5451
Van Gool et al. (2009)	16	Binary Logit Model	1661	5061	76.8	6722
Blanco et al. (2013)	39	MLP	2673	2778	93.22	5451
Baklouti (2013)	10	Logistic Regression	1994	3028	–	5022

None of the papers above explicitly address the issue that a given customer may have taken multiple loans. The behaviour of a customer is often summarised by defining a variable that relates to their historical repayments and the outcome of different loans. It is therefore necessary to design an evaluation scheme that can effectively deal with the temporal dynamics concerning different loans. This temporal dimension has important implications for the type of cross-validation technique that is appropriate for this particular application. Another difference is the fact that loan disbursement is relatively frequent in airtime lending unlike bank loans and microfinance loans. The amount being borrowed is also generally quite small compared to the loans approved by financial institutions.

The meta analysis selection criteria endeavors to find statistically robust studies whereby the number of customers and loans are large enough to obtain statistically significant results. The criteria also focused on analysing papers that perform credit scoring research in developing countries where ComzAfrica operates. The objective is to find research relating to countries facing similar challenges in terms of data availability on prospective applicants and countries with a similar economic and development context.

Variables used in previous studies were also reviewed in order to identify which variables are most likely to offer predictive information for the present case study. The similarity test involves checking how the variables describe a prospective loan applicant and its relevance to the model measured via statistical significance. Finally, these variables employed in the models developed in the papers that are statistically significant and related to the mobile industry are listed in Table 2.

The majority of the variables listed above in Table 2 are customer details. The case study company, ComzAfrica, is a third party company that does not have access to extensive customer details such as demographic information (for example, age and gender). There were no papers that describe the application of building a credit score model without substantial customer details. This paper seeks to demonstrate that it is still possible to create accurate credit scoring models for airtime lending without demographic details of customers.

Table 2. Variables highlighted by the meta analysis.

Variables	(Cubiles-de-la-Vega et al. 2013)	(Van Gool et al. 2009)	(Baklouti 2013)	(Blanco et al. 2013)
Previous Loan Grant	Yes	–	–	Yes
Loan Grant	Yes	Yes	Yes	Yes
Loan Denied	Yes	–	–	Yes
Gender	Yes	–	Yes	Yes
Age	Yes	Yes	Yes	Yes
Interest Rate	Yes	–	–	Yes
Cycles	–	Yes	–	–
Beginning Month	–	Yes	–	–
Number of previous loans	–	–	Yes	Yes
previous loan default	–	–	Yes	–
Marital Status	Yes	Yes	–	Yes

3. Methodology

In this section, the methodology is outlined in terms of the data analysis, feature selection, evaluation and cross-validation.

3.1. 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 ComzAfrica. After negotiating legal and confidentiality agreements with ComzAfrica as part of the practicum organized by CMU-Africa, data for the period 1 January 2016 to 30 June 2017 were accessed. 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 considered from 1 January 2016 to 30 April 2017 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 ComzAfrica 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 behavior for loan duration but defaulters tend to have lower loan counts and lower usage amounts.

Table 3. Summary of the ComzAfrica dataset.

Summary Description	Value
Total loans	3,002,667
Total customers	41,391
Number of customers who repaid loans	21,608
Number of defaulters customers	19,783
Total months of data	16
Avg. loans per customer	72.5
Avg. amount borrowed per customer	35.05 USD
Avg. loans per month	187,666
Avg. loans per month per customer	4.5
Number of defaulted loans	30,809
Loan default rate	0.01%

Table 4. Summary of feature distribution.

Variable	Status	Min	Max	Mean	Std Deviation
Average Loan duration (days)	Defaulters	0	290	9.3	14.14
	Non-Defaulters	0.27	505	9.58	16.11
Average Loan count	Defaulters	1	505	44.36	45.21
	Non-Defaulters	1	713	81.38	65.62
Average Usage amount (USD)	Defaulters	0	183.43	5.94	8.14
	Non-Defaulters	0	295.46	6.86	8.95

3.2. Feature Engineering and Selection

The meta analysis suggests that features can be grouped into three categories: loan details, customer behavior and customer details (Age, Gender). The features that are available for this study are: loan amount, number of recharges for each month (how many times the customer recharged), usage amount (amount used the previous month), activation date (when the customer's account was activated), the date the loan was taken, the date of loan payment, and total amount used every month. From the list of features above, this study is limited to the loan details and customer behavior. The customer details (Age, Gender) are available only to the MNO and not provided to ComzAfrica. Other features are also constructed based on the available variables and included: loan count (how many loans the customer has at any time), loan duration (how long the customer took to repay the loan), age on network (how long the customer has been with the MNO) and the loan month (the month that the loan was taken).

3.3. Machine Learning Models

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 (Cox 1958). 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 (William 1959). 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 (Breiman 2011). By including the RF model, it is possible to ascertain whether ensemble techniques can offer any benefits beyond the LR and DT techniques.

3.4. Evaluation

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. The accuracy formula is given by:

$$ACC = (TP + TN)/(TP + TN + FN + FP) \tag{1}$$

Table 5 shows the confusion matrix for a predictive classifier and describes the impacts associated with the four potential outcomes. The matrix gives the cost or benefit in terms of profit for each prediction and actual outcome where the objective is to identify customers that will repay. Lending is a

cost-sensitive business, 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:

$$Specificity = TN/(TN + FP) \tag{2}$$

Specificity measures the probability that a classifier correctly predicts default when considering all those that actually default. The priority for profitable lending is to avoid customers that are likely to default, which is achieved by maximizing specificity.

Table 5. Confusion matrix and associated profits.

Actual\Prediction	Predict Repaid	Predict Default
Actual Repaid	TP profit = +10%	FN profit = −10%
Actual Default	FP profit = −100%	TN profit = 0%

3.5. Cross-Validation Scenarios

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.

Three different cross-validation scenarios are introduced for evaluating the predictive accuracy of the models and their properties are described in Table 6. The first scenario referred to as CV1 is where the loans are divided in a ratio of 70:30 randomly without considering any variable. In this case, the focus is on the loans as individual entities. The repercussion is that a loan in the future can be used to predict a loan in the past. Furthermore, this approach does not consider the particular customer and implies that some of the loans belonging to a customer can be used for training and some for testing. To assess the problems associated with scenario CV1, it is necessary to consider a second scenario, referred to as CV2 where the loans are split in a ratio of 70:30 based on the customer. This means that loans for a particular customer can either be included in the training or testing but not both. This avoids situations where loans for an individual customer are being used for both training and testing. However, CV2 does not address the time issue although the design is not as flawed as the CV1 scenario.

In the two scenarios CV1 and CV2, the default rate is extremely small. This is a problem since, if one guessed that the loan will be paid, there is a very high probability that this prediction will be correct. This insight about the existence of a simple benchmark classifier motivates the third scenario, CV3. The loan details for each customer are first summarized and their last loan status is selected for entry. This reduces the size of the dataset to the number of customers available. The ratio of defaulting customers to non-defaulting customers is approximately 50:50 ensuring a balanced dataset with a default rate of 50% by design. The data are randomly split based on customers into 70:30 for training

and testing. A single customer can only be in one of the two groups. This CV3 scenario therefore has no problem with repeat customers or time continuity.

Table 6. Cross validation scenarios and their properties.

Property	CV1	CV2	CV3
Selection	Random loans	Random customer	Random customer and one loan
Evaluation	Evaluate loans	Evaluate loans	Evaluate customer
Time Issue	Yes	Yes	Yes
Customer Issue	Yes	No	No
Default Rate	0.01%	0.01%	50%

4. Results and Discussion

Table 7 shows the performance of all models for evaluations undertaken in-sample and using each of the three cross-validation scenarios. 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.

Nonlinearity is relevant with both the decision tree and Random Forest beating logistic regression in terms of specificity. Random Forest is superior with an accuracy of 82.3% and specificity of 84.7% for CV3, confirming the advantages of the ensemble approach. Using the RF and cross-validation scenario CV3 (Table 8), the predictions for loans repaid is correct for 85% and incorrect for 15%. The predictions for loans defaulted are correct for 80% and incorrect for 20%.

Table 7. Model evaluation results.

Model	Evaluation Metric	In-Sample	CV1	CV2	CV3
LR	ClassAcc	99.1%	99.1%	99.1%	80.0%
	Specificity	0.2%	0.2%	0.2%	78.8%
DT	ClassAcc	99.1%	99.1%	99.0%	79.7%
	Specificity	0.0%	0.0%	0.0%	79.8%
RF	ClassAcc	99.1%	99.1%	99.9%	82.3%
	Specificity	0.8%	0.8%	0.0%	84.7%

Table 8. Confusion matrix of random forest model for CV3.

Actual/Prediction	Predict Repaid	Predict Default
Actual Repaid	85%	15%
Actual Default	20%	80%

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

increases. 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.
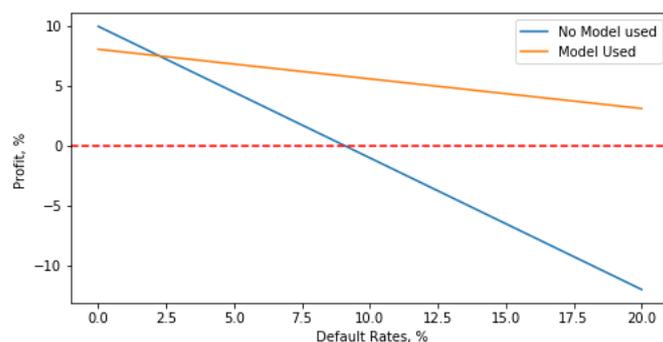
Figure 1. Profit generated by a company charging 10% interest on each loan for a range of default rates.

5. Conclusions

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 behavior 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.

The study yielded the following key findings:

- 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.
- Random forest was the best classifier with an accuracy of 82.3% which showed that nonlinearity and an ensemble approach was superior.
- When the default rate is low, it is better to offer the loans to every customer. With increasing default rates, a point is eventually reached whereby the model will outperform this simple approach of offering loans to everyone.

- The maximum tolerable default rate is increased by the optimal model to 32% compared with 8% when the company does not use a model.

ComzAfrica is the primary beneficiary of this study in that the machine learning algorithms and empirical study is based on data obtained from its customers. Furthermore, the results demonstrate that there is a strong business case for using credit scoring to increase the profitability of the ACS product. The methodology and approach studied in this paper are also relevant for a wide range of pay-as-you-go mobile products where credit is offered for basic services: electricity tokens, smart water meters; smart cooking devices and solar energy.

Limitations of the study arise from the use of a single dataset obtained from one company. However, the large number of loans and customers considered and generic applicability of credit scoring for mobile credit suggests that the variables and models investigated are relevant for other business applications. Further work may include the use of demographic information which could be obtained from mobile network operators and consideration of additional pay-as-you-go mobile products.

References

Aamo, Iorliam, Atu Myom, and Yahaya I. Shehu. 2017. Airtime Credit Banking: From Two Applications to One Application. *Journal of Computer and Communications* 5: 10–15. [CrossRef]

Abdulrahman Umar Farouk, Joseph Kobina Panford, and James Ben Hayfron-Acquah. 2014. Fuzzy Algorithm to Credit Scoring for Microfinance in Ghana. *International Journal of Computer Applications* 94: 11–18. [CrossRef]

Aker, Jenny C., and Isaac M. Mbiti. 2010. Mobile Phones and Economic Development in Africa. *Journal of Economic Perspectives* 24: 207–32. [CrossRef]

Baklouti, Ibtissem. 2013. Determinants of Microcredit Repayment: The Case of Tunisian Microfinance Bank. *In African Development Review* 25: 370–82. [CrossRef]

Belson, William A. 1959. Matching and Prediction on the Principle of Biological Classification. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 8: 65–75. [CrossRef]

Blanco, Antonio, Rafael Pino-Mejías, Juan Lara, and Salvador Rayo. 2013. Credit scoring models for the microfinance industry using neural networks: Evidence from Peru. *Expert Systems with Applications* 40: 356–64. [CrossRef]

Breiman, Leo. 2001. Random Forests. *Machine Learning* 5: 5–32. [CrossRef]

Chen, Gregory C., and Pedro Xavier Faz De Los Santos. 2015. *The Potential of Digital Data: How Far Can It Advance Financial Inclusion?* Focus Note 100. Washington, DC: CGAP.

ComzAfrica. 2017. Products|ComzAfrica. Available online: http://www.ComzAfricafrica.com/products (accessed on 11 October 2017).

Cox, David R. 1958. The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society: Series B (Methodological)* 20: 215–42.

Cubiles-De-La-Vega, María-Dolores, Antonio Blanco-Oliver, Rafael Pino-Mejías, and Juan Lara-Rubio. 2013. Improving the management of microfinance institutions by using credit scoring models based on Statistical Learning techniques. *Expert Systems with Applications* 40: 6910–17. [CrossRef]

Deloitte. 2017. Credit scoring Case study in Data Analytics. Available online: https://www2.deloitte.com/content/dam/Deloitte/global/Documents/Financial-Services/gx-be-aers-fsi-creditscoring.pdf (accessed on 12 October 2017).

Madise, Sunduzwayo. 2015. Mobile Money and Airtime: Emerging Forms of Money. *SSRN Electronic Journal*. [CrossRef]

Mode. 2017. Products. Available online: http://www.mo-de.com/products/ (accessed on 11 October 2017).

Pell, Nicholas. 2017. A Secret History of Credit Scores: Who Determined What Matters and Why. Available online: https://www.thestreet.com/personal-finance/credit-cards/a-secret-history-of-credit-scores-who-determined-what-matters-and-why-13097739 (accessed on 12 October 2017).

Safaricom. 2017. Okoa Jahazi. Available online: https://www.safaricom.co.ke/personal/plans/gettingstarted/okoa-jahazi (accessed on 11 October 2017).

Tang, Yuchun, Yan-Qing Zhang, Nitesh V. Chawla, and Sven Krasser. 2009. SVMs Modeling for Highly Imbalanced Classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 39: 281–88. [CrossRef] [PubMed]

Van Gool, Joris, Bart Baesens, Piet Sercu, and Wouter Verbeke. 2009. An Analysis of the Applicability of Credit Scoring for Microfinance. Paper presented at the Academic and Business Research Institute Conference, Orlando, FL, USA, September 24–26.