

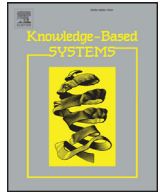
Central Lancashire Online Knowledge (CLOK)

Title	Predicting creditworthiness in retail banking with limited scoring data
Type	Article
URL	https://clock.uclan.ac.uk/id/eprint/24768/
DOI	https://doi.org/10.1016/j.knosys.2016.03.023
Date	2016
Citation	Abdou, Hussein, Tsafack, Marc D. Dongmo, Ntim, Collins G. and Baker, Rose D. (2016) Predicting creditworthiness in retail banking with limited scoring data. Knowledge-Based Systems, 103. pp. 89-103. ISSN 0950-7051
Creators	Abdou, Hussein, Tsafack, Marc D. Dongmo, Ntim, Collins G. and Baker, Rose D.

It is advisable to refer to the publisher's version if you intend to cite from the work.
<https://doi.org/10.1016/j.knosys.2016.03.023>

For information about Research at UCLan please go to <http://www.uclan.ac.uk/research/>

All outputs in CLOK are protected by Intellectual Property Rights law, including Copyright law. Copyright, IPR and Moral Rights for the works on this site are retained by the individual authors and/or other copyright owners. Terms and conditions for use of this material are defined in the <http://clock.uclan.ac.uk/policies/>



Predicting creditworthiness in retail banking with limited scoring data



Hussein A. Abdou^{a,b,*}, Marc D. Dongmo Tsafack^c, Collins G. Ntim^d, Rose D. Baker^e

^aHuddersfield Business School, University of Huddersfield, Huddersfield, West Yorkshire HD1 3DH, UK

^bManagement Department, Faculty of Commerce, University of Mansoura, Mansoura, Dakahlia, Egypt

^cSalford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK

^dHuddersfield Business School, University of Huddersfield, Huddersfield, West Yorkshire, HD1 3DH, UK

^eSalford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK

ARTICLE INFO

Article history:

Received 11 April 2015

Revised 21 March 2016

Accepted 25 March 2016

Available online 12 April 2016

JEL classification:

E50

G21

C45

Keywords:

Predicting creditworthiness

Credit scoring

Cascade correlation neural networks

CART

Limited data

ABSTRACT

The preoccupation with modelling credit scoring systems including their relevance to predicting and decision making in the financial sector has been with developed countries, whilst developing countries have been largely neglected. The focus of our investigation is on the Cameroonian banking sector with implications for fellow members of the Banque des Etats de L'Afrique Centrale (BEAC) family which apply the same system. We apply logistic regression (LR), Classification and Regression Tree (CART) and Cascade Correlation Neural Network (CCNN) in building our knowledge-based scoring models. To compare various models' performances, we use ROC curves and Gini coefficients as evaluation criteria and the Kolmogorov-Smirnov curve as a robustness test. The results demonstrate that an improvement in terms of predicting power from 15.69% default cases under the current system, to 7.68% based on the best scoring model, namely CCNN can be achieved. The predictive capabilities of all models are rated as at least very good using the Gini coefficient; and rated excellent using the ROC curve for CCNN. Our robustness test confirmed these results. It should be emphasised that in terms of prediction rate, CCNN is superior to the other techniques investigated in this paper. Also, a sensitivity analysis of the variables identifies previous occupation, borrower's account functioning, guarantees, other loans and monthly expenses as key variables in the forecasting and decision making processes which are at the heart of overall credit policy.

© 2016 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

The capability of statistical credit scoring systems to improve decision-making and time efficiencies in the financial sector has widely attracted researchers and practitioners particularly in recent years (see for example, [4,37,43–45,49,51,53,54]). Credit scoring systems are now regarded as virtually indispensable in developed countries. In developing countries statistical scoring models are needed not least to support judgemental techniques subject to each bank's individual policies. In building a scoring system a number of particular client's characteristics are used to assign a score. These scores can provide a firm basis for the lending and re-lending decision [9,17,23,48,49,52,53].

Background of the Cameroonian banking sector: Credit scoring is not popular in Africa at present. It appears neither to have been applied nor considered in the case of the Cameroonian banking

sector¹ and across the BEAC family. Cameroon is one of the developing countries in west and central Africa and is estimated to have a population just over 19 million people. The labour force was estimated in 2009 to be 7.3 million. Employment derives mainly from three sectors. Firstly, from industry: petroleum production and refining, aluminium production, food processing, light consumer goods, textiles, lumber, ship repair; secondly, from ser-

¹ The Bank of Issue for Cameroon is the "Bank of the Central African States" (Banque des Etats de L'Afrique Centrale, BEAC) which was created on November 22nd 1972. It was introduced to replace the "Central Bank of the State of Equatorial Africa and Cameroon" (Banque des Etats de l'Afrique Equatoriale et du Cameroun, BCEAC) which had been operating since April 14th 1959. BEAC is the central bank for the following six countries, in no particular order of priority: Cameroon, Central African Republic, Chad, Republic of the Congo, Equatorial Guinea and Gabon. Together these six countries also form the "Economic and Monetary Community of Central Africa" (Communauté Economique et Monétaire de l'Afrique Centrale, CEMAC). BEAC's headquarters are located in Yaounde, the capital of Cameroon. The issued currency is the "CFA Franc", which stands for "Financial Cooperation in Central Africa" (Coopération Financière en Afrique Centrale) and is pegged to the Euro at a rate of €1 = CFA665.957 [8].

* Correspondence author. Tel.: +44 1484473872; fax: +44 1484473148.

E-mail address: h.abdou@hud.ac.uk (H.A. Abdou).

vices; and finally, from the main sector which is agriculture, predominantly coffee, cocoa, cotton, rubber, bananas, oilseed, grains and root starches. The Gross Domestic Product (GDP) in 2008 was US\$20.65 billion. Total domestic lending was US\$1.3 billion which represented approximately 6.3% of its GDP. By contrast, in an advanced economy such as the Netherlands with a population only 2 million fewer than the Cameroon, domestic lending represented an estimated 219% of their GDP (CIA, 2009). Thus, there is at least a case for investigating the scope for the growth of the credit industry in the Cameroonian market (for details see [Appendix A](#)) including the selection of appropriate scoring techniques.

In Cameroon and across BEAC, a judgemental and traditional system called Tontines² remains very popular. Cameroonian banks are reluctant to take risks so most people rely on Tontines to overcome loss of income and, in the case of small entrepreneurs, to raise funds to finance their operations. Members' behaviour is to some extent guaranteed by the wish not to be excluded from help and solidarity which is important in the context of a background of great social and economic uncertainty. Tontines have some drawbacks as credit tools. They can only be used for the short-term as the debt will have to be repaid at the end of the Tontine's cycle; the interest on Tontine credit is relatively high (between 5–10% per month); a huge sum of money cannot be easily obtained to fund a large investment [31,35].

The aims of this paper are: firstly, to identify and investigate the currently used approaches to assessing consumer credit in the Cameroonian banking sector; secondly, to build appropriate and powerfully predictive scoring models to predict creditworthiness then to compare their performances with the currently used traditional system; and finally and freshly to discern which of the variables used in building the scoring models are most important to the decision making process.

Our practical contribution emerges from the foregoing. It would clearly be in the interests of both borrowers and banks to have decision making models which make credit available on terms which reflect the needs of borrowers and their ability to repay. Provision of such a service requires a sensitive and efficient credit scoring system. This is essential to establishing and monitoring the creditworthiness of borrowers in the joint interests of themselves and their lenders. The credit scoring system of choice needs to be tailored to the particular society and credit granter. The range of available models has to be compared and the preferred scoring systems should include direction of credit grantors' attention to the crucially relevant variables. However, in so far as Tontines are in use across six BEAC countries, a scoring system which potentially improves on these is likely to respond to the needs of more than one of the countries. Investors within and beyond the six stand to benefit from a more stable banking system which adopts a powerful scoring system to predict the soundness and profitability of

banks and their borrowers. The rest of our paper is organised as follows: section two reviews related studies; section three deals with the research methodology, section four explains the results and section five comprises the conclusion with policy recommendations and suggestions for future research.

2. Related studies

The purpose of credit scoring is to provide a concise and objective measure of a borrower's creditworthiness. Historically, Fisher [28] is the first to have used discriminant analysis to differentiate between two groups. Possibly the earliest application of applying multiple discriminant analysis is by Durand [24] who investigated car loans. Altman [62] introduced a corporate bankruptcy prediction scoring model based on five financial ratios.

Advances in information processing have fuelled progress in credit scoring techniques and applications. Conventional statistical techniques including logistic regression have been widely used and compared with non-parametric techniques such as classification and regression tree (CART) in building scoring models (e.g. [7,9,12,13,16,30,39,51,55,58,61]). Logistic regression deals with a dichotomous dependent variable which distinguishes it from a linear regression model, and makes the assumption that the probability of the dependent variable belonging to any of two different classes relies on the weight of the characteristics attached to it [1,4,5,37,41,48]. LR varies from other conventional techniques such as discriminant analysis in that it does not require the assumptions necessary for the discriminant problem [4,22]. Classification and regression tree is a tree-like decision model which is also used for classification of an object within two or more classes [18]. CART can be used to analyse either quantitative or categorical data and is widely used in building scoring models (e.g. [10,13,16,32,39,59,60]).

Advanced statistical techniques such as neural networks have been widely used in building scoring models ([1,4–6,9,18,29,38,42,55,56]). Also, by way of comparison between neural networks and other non-parametric techniques such as CART, Davis et al. [21] compared CART with Multilayer Perceptron Neural Network for credit card applications, and found comparable results for decision accuracy. Zurada and Kunene [63] found in their investigation of loan granting decisions comparable results for neural networks and decision trees across five different data-sets. A neural network is a system made of highly interconnected and interacting processing units that are based on neurobiological models mimicking the way the nervous system works. It usually consists of a three layered system comprising input, hidden, and output layers [1,4,5,33]. A Cascade Correlation Neural Network (CCNN) is a special type of neural network used for classification purposes. CCNN can avoid Multilayer Perceptrons Neural Network's drawbacks, such as the design and specification of the number of hidden layers and the number of units in these layers [19,27]. Various scoring models' evaluation criteria including receiver operating characteristic (ROC) curves and Gini coefficients are widely used and serve to assess the predictive capabilities of scoring models [2,4,11,18,20,46].

World-wide evolution of thought and practice in credit scoring can be substantially attributed to increasingly rigorous models of personal and corporate finance, increasingly powerful and discriminating statistical techniques and enormously more potent and economic processing capacity. This progress has been matched by a huge increase in the global demand for credit, not least in Africa including the BEAC family. All countries stand to benefit from wisely supervised credit's contribution to a healthy economy. Credit scoring already plays a key role in developed countries but our early investigation revealed that this is not the case for Cameroon and across BEAC, where judgemental approaches with

² A Tontine is a scheme in which members of a group combine resources to create a kitty [35]. Under a complex Tontine scheme the kitty is divided into lots and then auctioned. A small auction is held whereby a pre-set nominal fee is deducted from the kitty for every bid and the winner is the person ready to accept the least funds [31]. The difference between the original fund raised and the amount the member receives after the auction is a fee which is paid to the recipient of that lot at that session. The money usually has to be repaid within one or two months [35]. The fee paid by the 'beneficiary' at a particular session can be seen as interest paid on that money over the length of time before the loan is repaid. It also acts as an investment yielding a dividend for the other members since the sum of fees collected during the lending activities are then divided and distributed to the members of the Tontine at the end of each round of meetings. Despite relying solely on a tacit judgemental technique to select its members who do not even need to provide collaterals, Tontines are estimated to handle about 90 per cent of individuals' credit needs in Cameroon, and across BEAC, whereas the commercial and savings and loan banks realize a volume of about 10 per cent of all national loan business [35]. Tontines experience very high repayment rates relying on trust among members and most of all on their fear of being cast out of the Tontine.

their drawbacks still prevail. Judgemental techniques tend to encourage only very safe lending as successful borrowers will most likely have to be existing clients of the bank with a long and creditable financial history and/or powerful collateral. Statistical modelling techniques help to break these bounds by equipping any bank to expand lending activities within and beyond its existing clientele. The result is a growing credit industry with a concomitant boost to the economy. Our fresh contribution consists in the fact that, to the best of our knowledge, other authors do not distinguish the most important variables and none has investigated the potential benefits of scoring models in assessing Cameroonian personal loan credit.

3. Research methodology

In our research methodology, we adopt a two-stage approach. At the investigative stage we establish the currently applied approaches in the chosen environment for personal loans. At this stage, three informal interviews were conducted over the telephone with key credit lending officers from three major banks in Cameroon. Two out of the three lending officers provided a list of characteristics that are currently used in their evaluation process and this helped in deciding the list of variables included in our scoring models, details of which are given later. At the evaluative stage, we build the scoring models for personal loans in the chosen banking sector, and use three different statistical techniques, namely, Logistic Regression (LR), Classification and Regression Trees (CART) and Cascade Correlation Neural Network (CCNN). This is followed by an evaluation of the predictive capabilities of the scoring models using both Receiver Operating Characteristic (ROC) curve and Gini coefficients and then using the Kolmogorov-Smirnov curve as a robustness test. Here, different software is applied, including Scorto Credit Decisions and IBM SPSS 22. Finally, a sensitivity analysis is undertaken to determine the key variables under each technique, and to compare them with the variables currently used by the credit officers.

We submit that our work enables decision makers not only in the Cameroonian banking sector but throughout the BEAC family which applies the same system to go on to a third - implementation - stage of credit scoring. This facilitates progress beyond the present system with its shortcomings generating huge potential economic and social benefits. These benefits include externalities for the economy as a whole. Later, we discuss the data collection and the identification of variables used in building the scoring models.

3.1. Statistical techniques for constructing the proposed scoring models

3.1.1. Logistic regression

LR is one of the most widely used statistical models for deriving classification algorithms. It can simultaneously deal with both quantitative variables, such as age or number of dependants, and/or categorical variables, such as gender, marital status and purpose for the loan. In the case of LR it is assumed that the following model holds (see for example, [18], for a similar expression):

$$\log(P_{gi}/(1 - P_{gi})) = \alpha + \beta_1 K_{1i} + \beta_2 K_{2i} + \beta_3 K_{3i} + \dots$$

where, α , β_1 , β_2 , β_3 , ... are coefficients of the model and K_{ji} represents the respective characteristic variable j for applicant i under review, and P_{gi} represents the probability that applicant i is of good credit worthiness.

The probability that applicant i will be good is therefore given by:

$$P_{gi} = [\exp(\alpha + \beta_1 K_{1i} + \beta_2 K_{2i} + \beta_3 K_{3i} + \dots)] / [1 + \exp(\alpha + \beta_1 K_{1i} + \beta_2 K_{2i} + \beta_3 K_{3i} + \dots)]$$

The parameters in the equations are estimated using maximum likelihood. The value of P_{gi} can then either fall above the cut-off point and allow the application to be classified as 'good' or fall below it classifying it as 'bad'. The cut-off point represents a threshold of risks that the bank would be prepared to take on borrowers. Hence, the higher P_{gi} is above the cut-off point, the more credit-worthy the application will be regarded by the bank.

3.1.2. Classification and regression tree

CART is a popular classification model that can handle both quantitative and categorical data simultaneously. The construction of decision trees reflects the separation of attributes from each characteristic involved into 'good' and 'bad' risk classes. It is constructed using recursive partitioning, for which the separation produces the over fitted tree with a large number of branches and nodes. A pruning process is then necessary to obtain an optimal and practical model that will be effective in the field. Different algorithms exist to assess the quality of that separation between 'good' and 'bad'. A common algorithm is the $C_{4.5}$ which is the algorithm of the CART model used in this paper, and which uses the *GainRatio* criterion. Assuming T is a group formed in a certain node and T_i is the family of its sub-groups (see, for example, [7], p. 631), the *GainRatio* can be expressed as follows:

$$GainRatio_X = \frac{GainInfo_X}{I(X)}$$

where, $GainInfo_X$ is a criterion used by the $C_{4.5}$ algorithm to define further divisions into sub-groups for each of the original groups, when building the tree; $I(X) = SplitInfo$ is the entropy of group T , in which their formulae (see directly above for references) are given as follows:

$$GainInfo_X = H(T) - H_X(T)$$

$$I(X) = - \sum_{i=1}^m \frac{|T_i|}{|T|} \log_2 \left(\frac{|T_i|}{|T|} \right)$$

where, $H(T)$ is the entropy of the group T , and can be calculated as follows:

$$H(T) = [-p_1 \log_2(p_1) - p_0 \log_2(p_0)]$$

where, $p_1(p_0)$ is the proportion of examples of class 1 (0) in group T . This entropy is maximally = 1 when $p_1 = p_0 = 0.50$, and minimally 0 when $p_1 = 0$ or $p_0 = 0$. Whilst $H_X(T) = \sum_{i=1}^m \frac{|T_i|}{|T|} H(T_i)$, and $H(T_i)$ is the entropy of a sub-group of T .

In building a decision tree, the significance level of pruning requires the algorithm to monitor the increase in the number of errors after a node is replaced with a leaf or stronger sub-branch. If after such a replacement, the number of the errors does not exceed the number of the errors in the initial tree under an increase in the error frequency at the set significance level, the node is replaced with a leaf or the corresponding branch. The higher is the set significance value, the less the tree will be pruned.

3.1.3. Cascade correlation neural network

CCNN is a supervised learning architecture that builds a 'near-minimal multi-layer network topology' in the course of training. Primarily the network contains only inputs, output units, and the connections between them. This single layer of connections is trained, 'using the Quickprop algorithm [25] to minimize the error'. When no further improvement is seen in the level of error, the network's performance is evaluated. If the error is small enough, the network stops. Otherwise a new hidden unit is added to the network in an attempt to reduce the residual error [26].

CCNN refers to an architecture with a unique feature used in the discrimination between good and bad credit applications. It automatically trains nodes and increases its architecture size when

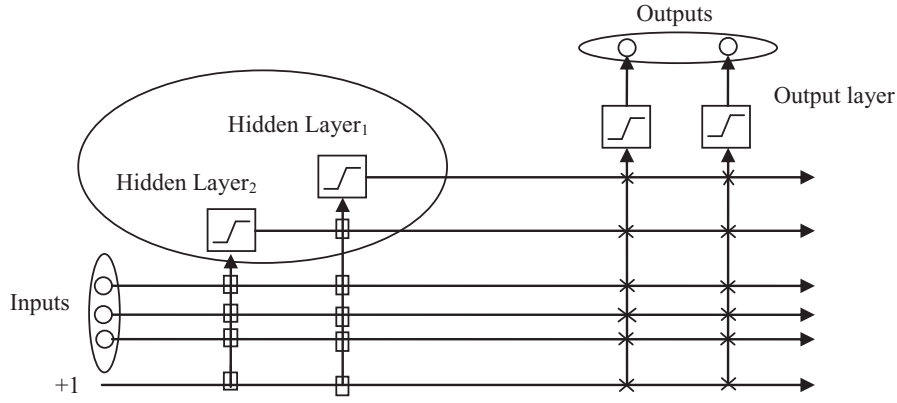


Fig. 1. Cascade Correlation Neural Network (CCNN) structure.

CCNN consists of one input layer, one hidden layer and one output layer. CCNN is based on two key principles. The first one is the cascade architecture of the network, in accordance with which the neurons of the hidden layer are added sequentially over time and then undergo no changes. According to the second principle the addition of each new component aims to maximize the value of the correlation between the output of the new component and the network error.

Source: Source: Fahlman and LeBiere [27] and Fahlman [26], modified.

analysing data until the analysis is complete or no further progress can be made. Thus, it allows avoiding one of the major problems in designing a neural network, which is obtaining the right size of the network by varying the number of hidden layers and connections between them as it is not possible to predetermine what would be suitable [19,26], as shown in Fig. 1.

CCNN is able to analyse a data-set comprising of both quantitative and categorical variables. The idea of CCNN is based on maximizing the correlation C , which can be calculated as follows (see, for example, [27], p.5; [19], p.2):

$$C = \sum_o \left| \sum_t (N_t - \bar{N})(E_{t,o} - \bar{E}_o) \right|$$

where, C is the sum from all output units and captures the magnitude of the correlation between the candidate units and the residual output error of the network. o is the output of the network at which the error is measured; t is the training pattern; N is the candidate neuron's output value; E_o is the residual output error sustained at output o ; $\bar{N}$ is the average of N over all patterns; $\bar{E}_o$ is the average of the E_o overall patterns; When C ceases to yield any improvement, a new unit is added to the architecture for the process to continue; this is the last until the result is found or further progress stagnates. C can be maximized through gradient ascent calculated through the computation of $\partial C / \partial w_i$, the partial derivative of C with respect to each of the candidates' weights, w_i , as follows (see, for example, [19], p.2, [27], p.5):

$$\frac{\partial C}{\partial w_i} = \sum_{t,o} \sigma_o (E_{t,o} - \bar{E}_o) d'_t l_{i,t}$$

where, σ_o is the sign of the correlation between the candidate's value and output o ; d'_t is the derivative for training pattern t of the candidate unit's activation function with regards to the sum of its inputs; $l_{i,t}$ is the input received by the candidate's unit from unit i for pattern t .

In building CCNN models the network algorithm presupposes conditions for the cessation of the network's training. These comprise three model parameters, the maximum iterations number where the parameter sets the number of iterations upon the completion of which the network training will be stopped; the correct classification rate where the parameter sets the condition for the stopping of the network's training when the value has reached the level of the set value's correct classification, and the network error improvement where the parameter sets the condition for the stopping of the network's training. The process stops when the network error value between the iterations has reached the set value.

3.2. Proposed performance evaluation criteria for scoring models

The Average Correct Classification (ACC) rate can be used to analyse the predictability of binary classifiers. The ACC rate = [observed good predicted good + observed bad predicted bad] / [total number of observations], and total error rate = [observed good predicted bad + observed bad predicted good] / [total number of observations]. Thus the ACC rate summarizes the accuracy of the predictions for a particular model. By contrast, the error rate refers to any misclassification performed by a predictive classifier and can be derived from the classification matrix. Those actually good but incorrectly classified as bad form the basis of the Type I error, and those actually bad but incorrectly classified as good represent the Type II error. For further discussion of the ACC rate and error rates, the reader is referred to Abdou [2].

3.2.1. Area under the ROC curve (AUC) and Gini coefficient

The ROC curve plots the relationship between sensitivity and (1 – specificity) for all cut-off values. Sensitivity refers to those cases which are both actually bad and predicted to be bad as a proportion of total bad cases. Specificity refers to cases which are both actually good and predicted to be good as a proportion of total good cases. The Area under the Curve (AUC) is used for the comparison of different classification models in order to assess their effectiveness. ROC is very powerful when dealing with a narrow cut-off range [18]. It does not require any adjustment for misclassification cost on its simplest form used for two classes' classifiers.

When comparing models for a given level of specificity the model with the higher sensitivity is preferred. Additionally, for a given level of sensitivity, the model with a higher level of specificity is also preferred. As we change the cut-off point, the ratio of type I to type II errors changes. Thus, there is a trade-off between the error types. AUC values, (see, for example, [36,40,50]), can be interpreted as: $0 \leq \text{AUC} < 0.6$ = fail; $0.6 \leq \text{AUC} < 0.7$ = poor; $0.7 \leq \text{AUC} < 0.8$ = fair; $0.8 \leq \text{AUC} < 0.9$ = good; and $0.9 \leq \text{AUC}$ = excellent.

A related measure is the Gini coefficient. This coefficient is another good tool to evaluate the performance of different credit scoring models. It will suggest how well the 'good' and 'bad' risk classes have been separated. The relationship between the Gini coefficient and the AUC value is given by $\text{AUC} = \frac{\text{Gini} + 1}{2}$. The following are some interpretations of the Gini values for assigning levels of quality to classifiers [47]:

Table 1
Variables used in building the scoring models.

Predictive variable	Encoding	Attribute's encoding	Comments
Loan amount*	LAT	Quantitative	–
Loan duration*	LDN	Quantitative	Initial duration of loan
Loan purpose*	LPE	Construction materials, auto parts = 0; edibles = 1; clothing, jewellery = 2; electrical items = 3; other purchases = 4	–
Age*	AGE	Quantitative	Borrower's age at time of lending
Marital status*	MST	Married = 0; Single = 1; Polygamy = 2; Engaged = 3	–
Gender*	GNR	Male = 0; Female = 1	–
No. of dependants*	NDP	Quantitative	Number of individuals, relying on the borrower for financial support
Current Job*	JOB	Public sector = 0; Private sector = 1	–
Education*	EDN	High school = 0; Undergraduate = 1; Postgraduate = 2	Highest level of academic instruction of the borrower
Housing*	HST	Not renting (e.g. living with relatives and no rental charge) = 0; Renting = 1	Establishes if the borrower pays rent
Telephone*	TPN	No = 0; Yes = 1	–
Monthly income*	MNC	Quantitative	Includes salary and other sources of income
Monthly expenses*	MCR	Quantitative	Includes other loan repayments and utility bills
Guarantees*	GRT	No = 0; Yes = 1	This includes support by a guarantor
Car ownership*	CON	No = 0; Yes = 1	–
Borrower's account functioning*	BAF	Account mostly in debit = 0; Account mostly in credit = 1; Alternately debit/credit = 2	How well the borrower manages his/her bank account
Other loans *	LOB	No = 0; Yes = 1	Loans from other banks
Previous employment*	POC	No = 0; Yes = 1	Exceeding one year
Feasibility study	N/A	–	Not required by the bank
Identification	N/A	–	All applicants had provided valid identification documents
Personal reputation	N/A	–	All applicants had a good reputation according to the bank
Field investigation	N/A	–	Not required by the bank
Central bank enquiries	N/A	–	Not required by the bank
Loan status*	LST	Bad = 0; Good = 1	Quality of the loan

* Variables are finally selected in building the scoring models.

$0 \leq \text{Gini} < 0.25$ = low quality classifier
 $0.25 \leq \text{Gini} < 0.45$ = Average quality classifier
 $0.45 \leq \text{Gini} < 0.60$ = Good quality classifier, and
 $0.60 \leq \text{Gini}$ = very good quality classifier.

3.3. Data collection and sampling

The data-set for the construction of the different models comprises 599³ historical blind consumer loans provided by one of the largest Cameroonian banks in 2011. This data-set consists of 505 good and 94 bad credit cases. To test the predictive capabilities of the scoring models, we use a stratified 5-fold cross-validation technique. We randomise the data so that the percentage of bad customers in each group is the same. This is done by separating the two groups of customers, randomly permuting each group, and taking 1/5 of each group for each of the 5-folds. This procedure gives a constant ratio between the number of good and bad cases, leading to have 101 good credit and 19 bad credit in each fold (except for one group which is short by one defaulter). This was done using a purpose-written program. The training set consists of 479 cases⁴ and the hold-out set consists of 120 cases. Each applicant is linked to 24 variables, mostly describing his/her demographic and financial information as presented in Table 1.

For each customer there are 23 explanatory variables and 1 dependent variable, namely, loan status. For all the 599 cases there were no missing attributes from the data-set. Some variables took the same values for all cases inclusive in this data-set and so these variables were excluded. We also investigate the correlation be-

tween the final 18 predictor variables and find no large correlation (i.e. >0.50) amongst them, as shown later in our results section. Table 1 portrays information about the nature of the loan, the personal characteristics of the borrower and the borrower's history⁵.

4. Results and discussions

In this section, a summary of the pilot study (in terms of telephone interviews) is discussed. Next, credit scoring models are built using statistical techniques, namely, LR, CART and CCNN. It should be emphasised that the data-set consists of 84.3% (505/599) good loans and 15.7% (94/599) bad loans. Statistically a data-set with 50% of defaulters would give the best discrimination between the two groups. However, our observed 15.7% of defaulters is still enough to allow firm conclusions to be drawn (for further discussion of class imbalance the reader is referred to [34]).

4.1. Investigative stage

From the pilot study it was understood that all applications have to be submitted to branches by existing customers as non-existing customers' applications are invariably not welcomed and it is not possible to make online applications. The criteria that they use in their analysis of credit applications are mainly selected according to the information from BEAC (Central Bank) and COBAC (banking supervisory agency). The requirements for each application are: to compute a financial ratio of the prospective borrower's current income in relation to current indebtedness; to establish as accurately as possible their current monthly expenditures; to conduct an identity check; and to establish clearly where they reside,

³ Although our scoring data-set is limited, however it does reflect the overall bank's default rate.

⁴ This consists of 404 good-risk class and 75 bad-risk class.

⁵ Prior to the processing of the original data, we checked for any typos and we coded the data as shown in Table 1.

their job status and the number of dependants. Personal reputation is considered too, as well as guarantees and/or guarantors. It should be emphasised that 'Previous Occupation' 'Guarantees' and 'Borrower's Account Functioning' are considered by the credit officers to be the most important attributes in their current evaluation process.

Once all the requested documents in support of the application have been received and validated by the bank, at least two lending officers will then analyse the application, and make appropriate comments. Next, a senior bank officer (such as branch manager, or head credit analyst) conducts a review and makes the final decision either to grant or refuse the credit. Validating the customer's documents involves actual field checks where applicable. Then, they use judgemental techniques to analyse applications. It is a long, difficult process involving many people and much unspoken informality. Credit card facilities are not offered by the BEAC family including Cameroonian banking sector at present. The banks provide a small proportion of total consumer credit, consumers relying instead on informal, typically Tontine-based lending for an estimated 90% of total consumer credit. Such a profile is arguably attributable, firstly to the absence of small lines of credit otherwise conveniently offered by credit cards and secondly to the lengthy, laborious and restrictive process undergone to obtain credit from the banks. These inhibitions underscore the case for building appropriate credit scoring models as a decision support tool.

4.2. Evaluative stage

At this stage some variables, such as 'central bank enquiries', 'personal reputation', 'field visit' and 'identifying documents' had to be excluded as they had identical values in each case. Table 1 presents the variables that are used in building various scoring models and their encoding. Finally, 18 predictor variables are used to build the scoring models. In order to construct the proposed models, we use Scort Credit Decision and IBM SPSS Statistics 22. Table 2 presents correlation results between the final 18 predictor variables including the dependent variable (loan quality). As shown in Table 2, all correlations between predictor variables are within acceptable range i.e. <0.50.

Table 3 shows the descriptive statistics for 12 categorical variables. It is obvious that previous employment (POC) is the most important variable based on the highest information value⁶ score of 1.361. This is followed by three variables, namely, guarantees (GRT), borrower's account functioning (BAF) and other loans (LOB), but of less importance compared to POC. However, the least important variables are telephone (TPN), housing (HST) and JOB, as shown in Table 3. In addition, six numerical variables are used in building the scoring models. As to the later, credit limit is up to 15,000,000 CFA; term ranges from 3 to 13 years; age ranges from 21 to 72 years old; income ranges from 50,000 CFA to 13,800,000 CFA; expenses range from 15,000 CFA to 15,000,000 CFA and finally number of dependants ranges from 0 up to 14.

The detailed results from our statistical modelling techniques, namely, LR, CART and CCNN are summarised next. The respective predictive capability of the classification models is also investigated.

⁶ Information Value, or total strength of the characteristics, which relates directly to the Weight of Evidence (WOE), is an alternative to chi-square which may be used to identify the strength of different variables. On the one hand, the effect of the information value as a measure is to provide the greatest contribution to the attributes that have the greatest impact on the score. On the other hand, chi-square value may identify attributes with a large difference between the expected and actual, but has little impact on the final decision.

Table 2
Correlation matrix.

LAT	LDN	LPE	AGE	MST	GNR	NDP	JOB	EDN	HST	TPN	MNC	MCR	GRT	CON	BAF	LOB	POC	LNS
1																		
LDN	1																	
LPE	.317**	1																
AGE	-.005	-.015	1															
MST	.109**	.033	-.019	1														
GNR	-.050	.017	.053	.114**	1													
NDP	-.045	.108**	-.043	.066	.023	1												
JOB	.073	.061	-.156**	.313**	.027	.020	1											
EDN	.084*	.038	-.049	.090*	.026	.093	.044**	1										
HST	.071	-.022	.152**	.090*	.024	.093	.044**	.061	1									
TPN	.026	-.040	-.034	.142**	.026	.093	.044**	.061	-.035	1								
MNC	.469**	.179**	.054	.025**	.054	.038**	.049	.019	-.023	-.047**	1							
MCR	.139	.068**	.106	.025**	.036	.041	.013	.097**	-.004	.014	.069**	1						
GRT	.010**	.079	-.085	.063	.045	.024	-.007	.043	-.046	-.019**	.044	.028	1					
CON	.320	.034	.030	.019	-.085	-.012	.018	-.024	.008	-.018	.292	.089	-.139	1				
BAF	-.017**	-.030	.094	.018	.063*	-.068	-.026	.061	.033*	.058**	.014	-.165	.124	.075*	1			
LOB	.035	-.023	.123	.021	.050	-.006	.066	-.012	-.042	-.012	.005	.021	-.103*	.098	.020	1		
POC	.047	.059*	-.005**	.136	.096	.049**	.049**	.048	-.035	-.069	.042**	.024**	.004	-.062	-.044	-.042**	1	
LNS	.003	.052	-.036**	.111	.041	.024	.015**	.006**	-.013	.003	-.037	-.109**	.307**	-.098	.073	-.215	.517**	1

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed). LAT = Loan Amount; LDN = Loan Duration; LPE = Loan Purpose; AGE = Borrower's Age at Time of Lending; MST = Marital Status; GNR = Gender; NDP = Number of Dependents; JOB = Current Job; EDN = Education; HST = Housing Status; TPN = Telephone; MNC = Monthly Income; MCR = Monthly Expenses; GRT = Guarantees; CON = Car Ownership; BAF = Borrower's Account Functioning; LOB = Other Loans; POC = Previous Employment; LNS = Loan Status (dependent variable).

Table 3
Descriptive statistics for the categorical variables.

Characteristic	Code	Count	Total %	Goods	Goods %	Bads	Bads %	Bad Rate	WOE*
Loan Purpose (LPE)									
Construction materials, auto parts	0	54	9.02%	46	9.11%	8	8.51%	14.81%	6.794
Edibles	1	287	47.91%	244	48.32%	43	45.74%	14.98%	5.47
Closing, jewellery	2	161	26.88%	138	27.33%	23	24.47%	14.29%	11.05
Electrical items	3	48	8.01%	36	7.13%	12	12.77%	25.00%	−58.265
Other purchases	4	49	8.18%	41	8.12%	8	8.51%	16.33%	−4.713
Information value: 0.038									
Marital Status (MST)									
Married	0	320	53.42%	259	51.29%	61	64.89%	19.06%	−23.531
Single	1	192	32.05%	166	32.87%	26	27.66%	13.54%	17.263
Polygamy	2	84	14.02%	77	15.25%	7	7.45%	8.33%	71.663
Engaged	3	3	0.50%	3	0.59%	0	0.00%	0.00%	11.05
Information value: 0.098									
Gender (GNR)									
Male	0	290	48.41%	240	47.52%	50	53.19%	17.24%	−11.265
Female	1	309	51.59%	265	52.48%	44	46.81%	14.24%	11.428
Information value: 0.013									
Current Job (JOB)									
Public sector	0	372	62.10%	312	61.78%	60	63.83%	16.13%	−3.261
Private sector	1	227	37.90%	193	38.22%	34	36.17%	14.98%	5.507
Information value: 0.002									
Education (EDN)									
High school	0	393	65.61%	333	65.94%	60	63.83%	15.27%	3.253
Undergraduate	1	178	29.72%	146	28.91%	32	34.04%	17.98%	−16.339
Postgraduate	2	28	4.67%	26	5.15%	2	2.13%	7.14%	88.369
Information value: 0.036									
Housing (HST)									
Not renting	0	334	55.76%	283	56.04%	51	54.26%	15.27%	3.236
Renting	1	265	44.24%	222	43.96%	43	45.74%	16.23%	−3.979
Information value: 0.001									
Telephone (TPN)									
No	0	50	8.35%	42	8.32%	8	8.51%	16.00%	−2.304
Yes	1	549	91.65%	463	91.68%	86	91.49%	15.66%	0.212
Information value: 0.000									
Guarantees (GRT)									
No	0	46	7.68%	21	4.16%	25	26.60%	54.35%	−185.562
Yes	1	553	92.32%	484	95.84%	69	73.40%	12.48%	26.671
Information value: 0.476									
Car Ownership (CON)									
No	0	470	78.46%	405	80.20%	65	69.15%	13.83%	14.824
Yes	1	129	21.54%	100	19.80%	29	30.85%	22.48%	−44.339
Information value: 0.065									
Borrower's Account Functioning (BAF)									
Account mostly in debit	0	27	4.51%	12	2.38%	15	15.96%	55.56%	−190.441
Account mostly in credit	1	547	91.32%	478	94.65%	69	73.40%	12.61%	25.424
Alternately debit/credit	2	25	4.17%	15	2.97%	10	10.64%	40.00%	−127.58
Information value: 0.410									
Other Loans (LOB)									
Other Loans	0	477	79.63%	421	83.37%	56	59.57%	11.74%	33.602
Other Loans	1	122	20.37%	84	16.63%	38	40.43%	31.15%	−88.803
Information value: 0.291									
Previous Employment (POC)									
No	0	50	8.35%	11	2.18%	39	41.49%	78.00%	−294.693
Yes	1	549	91.65%	494	97.82%	55	58.51%	10.02%	51.394
Information value: 1.361									

* Refers to the weight of evidence; one of the earliest measures used in credit scoring models, and it depends on the odds ratio of good scores expressed as a proportion of bad scores.

4.2.1. Analysis of the scoring models

4.2.1.1. Logistic regression. Five Logistic Regression (LR) credit scoring models are built and their classification results of the corresponding hold-out samples are shown in Table 4. It can be observed from Table 4 that the average correct classification rate for the 5-folds is 88.65% with 95.05% and 54.26% for good risk-class and bad risk-class, respectively, using a cut-off point of 0.5. Also, the average Type I error is 4.95% and the average Type II error is 45.74% resulting a total error rate of 11.35%, as shown in Table 4.

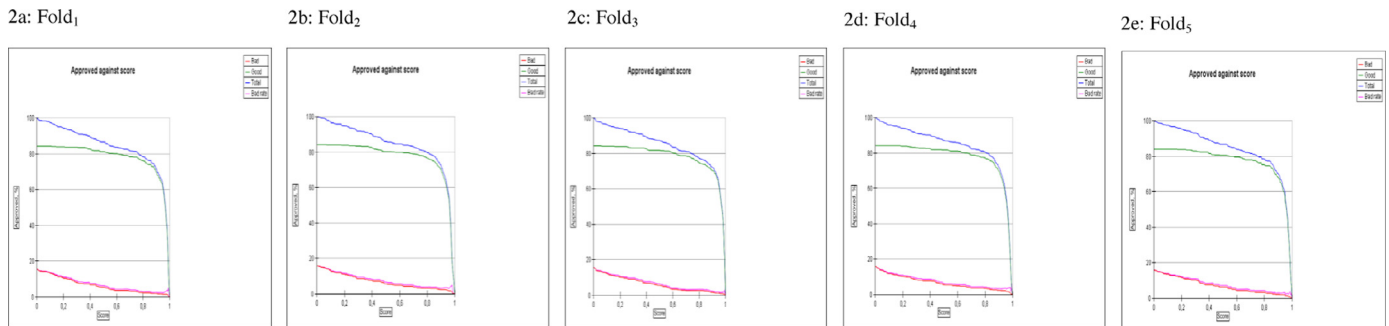
The approved against score chart can provide accurate graphical information to the decision makers. Five sub-figures for the 5 logistic regression scoring models are shown in Fig. 2. For example, for the first LR scoring model (Fold₁), the far right-hand side, the total number of accepted cases is below 5 cases (approximately 4 cases), as shown in Fig. 2.a. Therefore, the final decision depends on the decision makers' point of view. For instance, a cut-off score of 50% gives a chance to approximately accept a total num-

Table 4

Cross-validation results for the 5 logistic regression (LR) scoring models.

LR	Classification results			Error results		
	GG	BB	ACCR%	Type I	Type II	TE
Fold ₁	94.06(95/101)	63.16 (12/19)	89.17 (107/120)	5.94 (6/101)	36.84 (7/19)	10.83 (13/120)
Fold ₂	96.04 (97/101)	47.37 (9/19)	88.33 (106/120)	3.96 (4/101)	52.63 (10/19)	11.67 (14/120)
Fold ₃	96.04 (97/101)	47.37 (9/19)	88.33 (106/120)	3.96 (4/101)	52.63 (10/19)	11.67 (14/120)
Fold ₄	91.09 (92/101)	68.42 (13/19)	87.50 (105/120)	8.91 (9/101)	31.58 (6/19)	12.50 (15/120)
Fold ₅	98.02 (99/101)	44.44 (8/18)	89.92 (107/119)	1.98 (2/101)	55.56 (10/18)	10.08 (12/119)
Mean	95.05 (480/505)	54.26 (51/94)	88.65 (531/599)	4.95 (25/505)	45.74 (43/94)	11.35 (68/599)

Notation: LR=Logistic Regression Model; GG=Good credit correctly classified as good; BB=Bad credit correctly classified as bad; ACCR=Average correct classification rate; Type I=good credit misclassified as bad; Type II=bad credit misclassified as good and TE=Total errors (Type I+Type II).

**Fig. 2.** Approved against score (%) for the 5-folds Logistic Regression (LR) models.**Table 5**

Cross-validation results for the 5 decision tree (CART) scoring models.

CART	Classification results			Error results		
	GG	BB	ACCR%	Type I	Type II	TE
Fold ₁	95.05 (96/101)	63.16 (12/19)	90.00 (108/120)	4.95 (5/101)	36.84 (7/19)	10.00 (12/120)
Fold ₂	97.03 (98/101)	57.89 (11/19)	90.83 (109/120)	2.97 (3/101)	42.11 (8/19)	9.17 (11/120)
Fold ₃	95.05 (96/101)	47.37 (9/19)	87.50 (105/120)	4.95 (5/101)	52.63 (10/19)	12.50 (15/120)
Fold ₄	96.04 (97/101)	57.89 (11/19)	90.00 (108/120)	3.96 (4/101)	42.11 (8/19)	10.00 (12/120)
Fold ₅	97.03 (98/101)	61.11 (11/18)	91.60 (109/119)	2.97 (3/101)	38.89 (7/18)	8.40 (10/119)
Mean	96.04 (485/505)	57.45 (54/94)	89.98 (539/599)	3.96 (20/505)	42.55 (40/94)	10.02 (60/599)

Notation: CART=Classification and Regression Tree Model; GG=Good credit correctly classified as good; BB=Bad credit correctly classified as bad; ACCR=Average correct classification rate; Type I=good credit misclassified as bad; Type II=bad credit misclassified as good and TE=Total errors (Type I+Type II).

ber of 102 cases; this consists of 95 good credit and 7⁷ bad credit (with a bad rate of 6.86%), based on LR credit scoring model. These graphical results confirm our numerical modelling results shown in Table 4.

As a result of conducting a sensitivity analysis of the 18 explanatory variables used in building different LR scoring models, we calculate the average of the ranking of the contribution weights for the 5 LR models which allows us to establish the five most importantly ranked variables, as follows: POC, GRT, BAF, LOB and LPE are the most important variables with average contribution weightings in turn of 0.289, 0.182, 0.121, 0.116 and 0.059, respectively, as shown in Table 8. The prominence of POC, GRT and BAF accords with our findings from the investigative stage, but with a notably lower default rate. Conversely, the following five predictor variables are the least important, namely: LAT, LDN, AGE, NDP and HST, as shown in Table 8.

4.2.1.2. Classification and regression tree. Table 5 presents classification results for the 5 CART models and their corresponding hold-out samples. In building the decision tree the following criteria

are used: significance level of tree pruning is 0.25; the significance level for the pruning of the rules is 0.25; and significance level for the Fisher test is 0.10; selected by default as part of the software design, with iterative building of trees and use of the Gain-ratio criterion. It can be noted from Table 5 that the average correct classification rate for the 5-folds CART scoring models is 89.98% with 96.04% and 57.45% for good risk-class and bad risk-class, respectively. The average Type I and Type II error are 3.96% and 42.55%, respectively, resulting a total error rate of 10.02%.

Fig. 3 shows the approved against score for the five decision tree models. For example, for the first CART scoring model (Fold₁), the far right-hand side, the total number of accepted cases⁸ is below 50 cases (approximately 43 cases), as shown in Fig. 3.a. As the final decision depends on the policy makers' point of view, various cut-off scores surely provide different combinations of accepted and rejected cases. A cut-off score of 50%, for example, gives a chance to approximately accept a total number of 103 cases (this consists of 96 good credit and 7 bad credit -with a bad rate of 6.80%), based on the CART scoring model which confirms our results shown in Table 5.

⁷ It should be emphasised, as part of the currently used software design, that these numbers can accurately be identified.

⁸ This presupposes a 100% cut-off score or a bank with a strict/conservative credit policy.

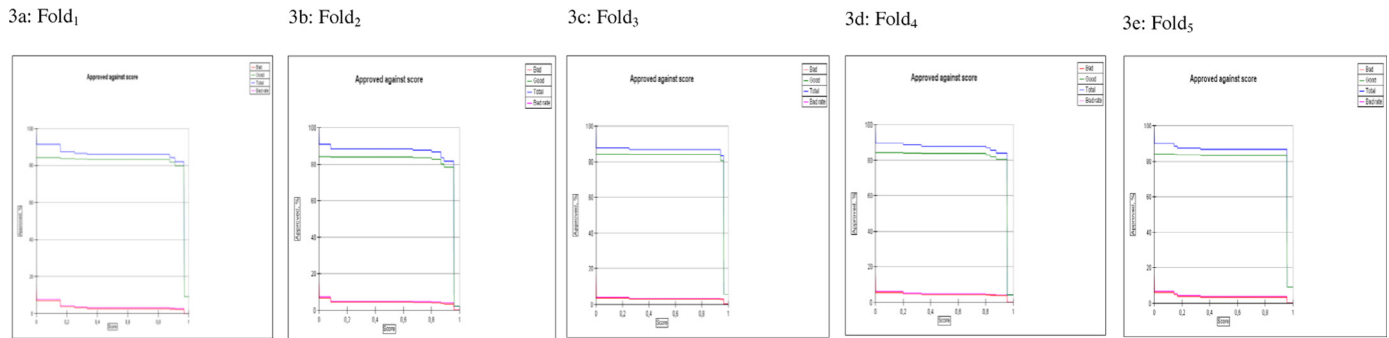


Fig. 3. Approved against score (%) for the 5-folds decision tree (CART) models.

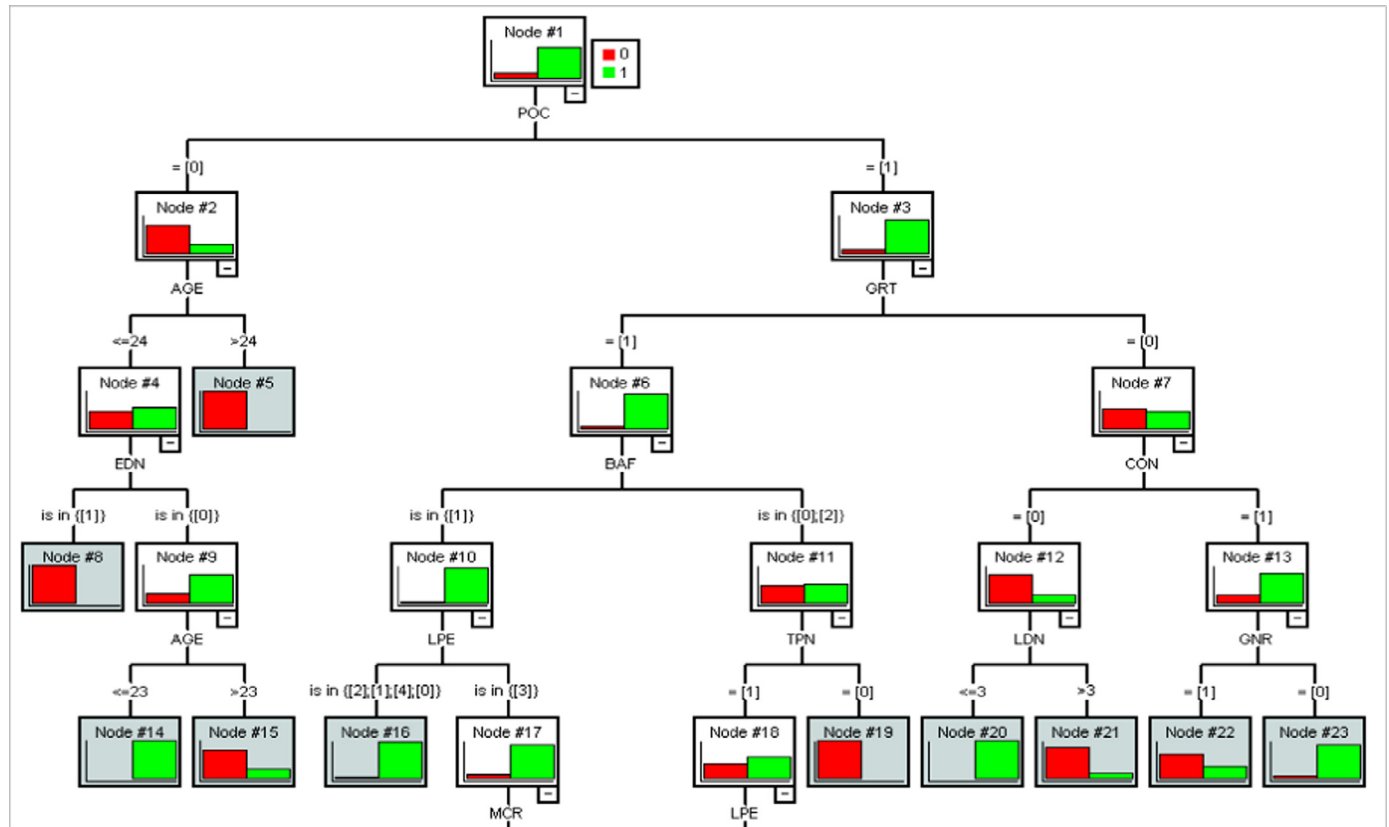


Fig. 4. Decision tree for the first fold.

Note: This tree shows 5 out of 10 rules (total tree depth is 10) and 23 out of 37 nodes. Significance level of tree pruning is 0.25 and using the Gain-ratio criterion; the significance level for the pruning of the rules is 0.25; and significance level for Fisher test is 0.10. LDN=Loan Duration; LPE=Loan Purpose; AGE=Borrower's Age at Time of Lending; GNR=Gender; EDN=Education; TPN=Telephone; MCR=Monthly Expenses; GRT=Guarantees; CON=Car Ownership; BAF=Borrower's Account Functioning; POC=Previous Employment.

Furthermore, in decision analysis the decision tree is an essential tool to visualize any analytical decision. For example, Fig. 4 shows the decision tree for the first fold (total number of rules is 10 and the total number of nodes is 37). As shown in the tree the first rule splits the data by presence of POC which considered the most important predictor. When POC is given the value of (1) subsequent splitting is based on GRT, when POC is given the value of (0) subsequent splitting is based on AGE (for example: Rule #1 If AGE > 24 and POC=[0], then 0; and Rule #2 If EDN is in (1) and POC=[0], then 0). When GRT is given the value of (1) subsequent splitting is based on BAF. When BAF is given the value of (1) subsequent splitting is based on LPE; and when LPE is given the value of (3) subsequent splitting is based on MCR and so on.

In Table 8, conducting a sensitivity analysis for the five CART scoring models we calculate the average of the ranking of their

contribution weights. As a result, the most important predictors are POC, BAF, GRT, LPE and MCR with contribution weightings in turn of 0.211, 0.114, 0.099, 0.061 and 0.057; whilst the least important predictors are HST, LOB, GNR, JOB and MNC, respectively. Our investigative stage identifies POC, GRT and BAF as the key variables based on the currently used system; this is consonant with our findings applying the CART scoring model but with a much lower default rate than in the case of the current system. It should be emphasised that these results do agree with the decision tree rules shown in Fig. 4.

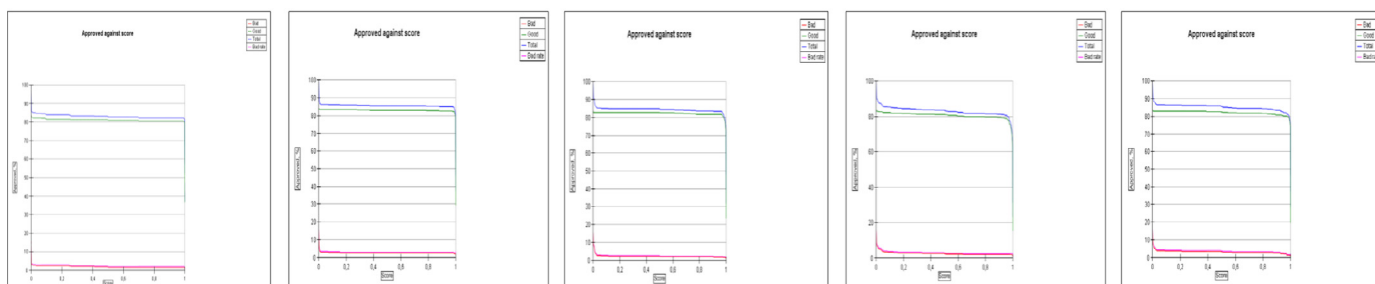
4.2.1.3. Cascade correlation neural networks. Five Cascade Correlation Neural Networks (CCNN) credit scoring models are built and their classification results of the corresponding hold-out samples are shown in Table 6. In building the CCNN scoring models, the

Table 6

Cross-validation results for the 5 Cascade Correlation Neural Network (CCNN) scoring models.

CCNN	Classification results			Error results		
	GG	BB	ACCR%	Type I	Type II	TE
Fold ₁	96.04 (97/101)	68.42 (13/19)	91.67 (110/120)	3.96 (4/101)	31.58 (6/19)	8.33 (10/120)
Fold ₂	96.04 (97/101)	73.68 (14/19)	92.50 (111/120)	3.96 (4/101)	26.32 (5/19)	7.50 (9/120)
Fold ₃	99.01 (100/101)	47.37 (9/19)	90.83 (109/120)	0.99 (1/101)	52.63 (10/19)	9.17 (11/120)
Fold ₄	96.04 (97/101)	68.42 (13/19)	91.67 (110/120)	3.96 (4/101)	31.58 (6/19)	8.33 (10/120)
Fold ₅	100 (101/101)	66.67 (12/18)	94.96 (113/119)	0.00 (0/101)	33.33 (6/18)	5.04 (6/119)
Mean	97.43 (492/505)	64.89 (61/94)	92.32 (553/599)	2.57 (13/505)	35.11 (33/94)	7.68 (46/599)

Notation: CCNN=Cascade Correlation Neural Network Model; GG=Good credit correctly classified as good; BB=Bad credit correctly classified as bad; ACCR=Average correct classification rate; Type I=good credit misclassified as bad; Type II=bad credit misclassified as good and TE=Total errors (Type I+Type II).

5a: Fold₁5b: Fold₂5c: Fold₃5d: Fold₄5e: Fold₅**Fig. 5.** Approved against score (%) for the 5-folds Cascade Correlation Neural Network (CCNN) models.

following criteria are used: an iteration limit value of 5000, correct classification rate limit value of 95%, an error improvement value of 3, and an error improvement iterations number of 5, selected by default as part of the software design. The maximum iteration number is used over the other two model parameters (i.e. the correct classification rate and the network error improvement), as chosen automatically by the software. It can be noted from Table 6 that the classification results for the 5-folds CCNN are as follows: the correct classification rates of 'good' into good risk-class is 97.43% and the correct classification rates of 'bad' into bad risk-class is 64.89% with an overall average correct classification rate of 92.32%. The average of total errors is 7.68% with an average Type I error of 2.57% and an average Type II error of 35.11%.

Fig. 5 shows the approved against score for the five CCNN scoring models. For example, for the first CCNN scoring model (Fold₁), the far right-hand side, the total number of accepted cases is below 200 cases (approximately 176 cases), as shown in Fig. 5.a. As different cut-off scores can provide different combinations of accepted and rejected cases, therefore the choice of a particular cut-off points depends on decision and policy makers' view points and how they may be optimistic (or pessimistic) in relation to their credit policy expectations. For instance, a cut-off score of 50% gives a chance to approximately accept a total number of 103 cases (this consists of 97 good credit and 6 bad credit - with a bad rate of 5.83%) based on CCNN scoring model. These graphical results confirm our numerical modelling results shown in Table 6.

It can also be observed from Table 8 that we conduct a sensitivity analysis for the five CCNN scoring models and we calculate the average of the ranking of their contribution weights. Out of the 18 predictor variables, POC, BAF, LOB, CON, GRT, and MCR are the most important variables with contribution weightings in turn of 0.090, 0.087, 0.087, 0.086, 0.078, and 0.078, respectively. On the other hand, the least important variables are LPE, LDN, LAT, AGE, and MST. Again, this is consonant with our findings from the investigative stage, but with much lower default rates compared to the rates in the current system.

4.2.2. Comparison of different scoring models

It can be observed that, when comparing various scoring techniques, CCNN has the highest ACC rate of 92.32% for the five CCNN scoring models compared to 88.65% and 89.98% for LR and CART scoring models, respectively, as shown in Table 7. Our scoring models are evaluated in this paper also using other criteria, namely, AUC and the Gini coefficients. Table 7 summarises the different values under each criterion for each of the scoring models. By inspecting the ACC rate, it can be noted that the accuracy across all different models varies from 88.65% for LR to 92.32% for CCNN. From the judgemental techniques currently being practised in Cameroon and the BEAC family, the default cases are 15.69% (94/599) signifying that, those default cases could potentially be reduced by at least 4.34% (15.69%–11.35%) through utilisation of LR and at most by 8.01% (15.69%–7.68%) through CCNN.

The error results in Table 7 also show that the Type I errors are very low compared with the Type II errors for all models. CCNN also has the lowest average Type I error of 2.57% compared to 4.95% and 3.96% for LR and CART, respectively. The average Type II error is much lower for CCNN (35.11%) compared to both LR and CART (45.74% and 42.52%, respectively) scoring models. Decision-makers should be careful which model they choose to apply because Type II errors are much more important, due to the fact that a Type II error necessarily involves default with its consequentially much higher cost. It is potentially more costly for a bank to misclassify a bad loan as good (Type II) than a good loan as bad (Type I) since in the latter case at worst opportunity cost is involved. These results are consonant with the literature where it has been found that advanced scoring models have lower error rates compared to conventional scoring models (see for example, [1,3,37]). Our results show the superiority of neural networks in predicting default rate in a stronger and more revealing manner – clearly of considerable economic value in a community where borrowers are all too frequently prone to default.

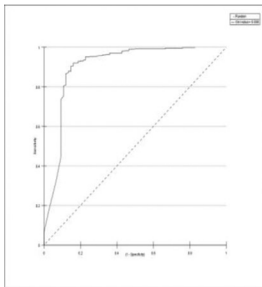
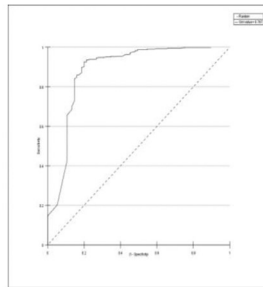
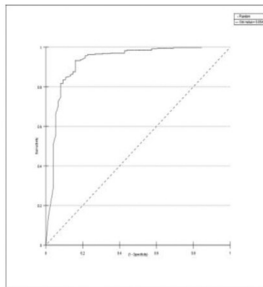
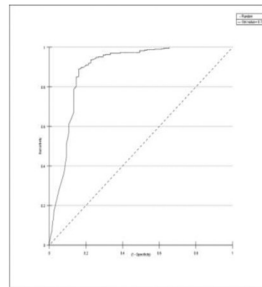
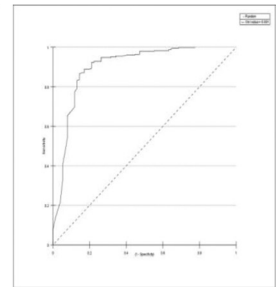
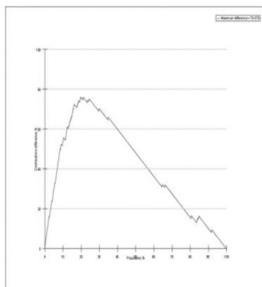
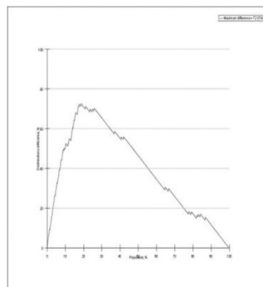
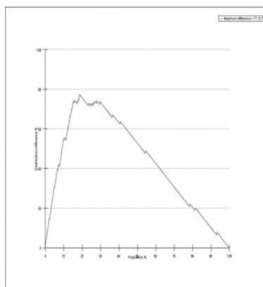
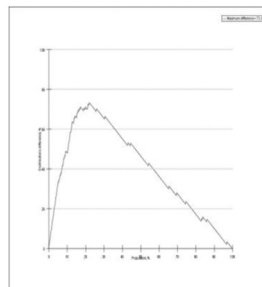
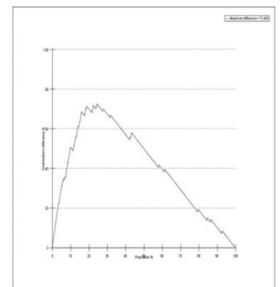
Figs. 6–8 present the ROC curves for all scoring models. The computations of the average AUC show that their values are

Table 7

Comparing cross-validation results, error rates, AUC values, Gini coefficients and K-S values.

CSMs	Classifications results			Error results		Evaluation criteria		Robustness test
	GG	BB	ACCR%	Type I	Type II	AUC	Gini	
LR								
Fold ₁	94.06	63.16	89.17	5.94	36.84	0.904	0.808	76.079
Fold ₂	96.04	47.37	88.33	3.96	52.63	0.884	0.767	72.574
Fold ₃	96.04	47.37	88.33	3.96	52.63	0.927	0.854	77.317
Fold ₄	91.09	68.42	87.50	8.91	31.58	0.891	0.781	73.356
Fold ₅	98.02	44.44	89.92	1.98	55.56	0.901	0.801	72.408
Mean	95.05	54.26	88.65	4.95	45.74	0.901	0.802	74.347
CART								
Fold ₁	95.05	63.16	90.00	4.95	36.84	0.929	0.857	81.525
Fold ₂	97.03	57.89	90.83	2.97	42.11	0.887	0.773	73.772
Fold ₃	95.05	47.37	87.50	4.95	52.63	0.915	0.830	81.333
Fold ₄	96.04	57.89	90.00	3.96	42.11	0.886	0.772	74.267
Fold ₅	97.03	61.11	91.60	2.97	38.89	0.905	0.809	78.205
Mean	96.04	57.45	89.98	3.96	42.52	0.904	0.808	77.820
CCNN								
Fold ₁	96.04	68.42	91.67	3.96	31.58	0.933	0.865	85.373
Fold ₂	96.04	73.68	92.50	3.96	26.32	0.926	0.852	84.439
Fold ₃	99.01	47.37	90.83	0.99	52.63	0.943	0.886	86.459
Fold ₄	96.04	68.42	91.67	3.96	31.58	0.923	0.846	83.297
Fold ₅	100	66.67	94.96	0.00	33.33	0.951	0.901	87.402
Mean	97.43	64.89	92.32	2.57	35.11	0.935	0.870	85.394

Notation: LR=Logistic Regression Model; CART=Decision Tree Model; CCNN=Cascade Correlation Neural Network Model; CSMs=Credit Scoring Models; GG=% of good correctly classified as good; BB=% of bad correctly classified as bad; Type I=% of good misclassified as bad; Type II=% of bad misclassified as good.

6a: Fold₁6b: Fold₂6c: Fold₃6d: Fold₄6e: Fold₅6f: Fold₁6g: Fold₂6h: Fold₃6i: Fold₄6j: Fold₅**Fig. 6.** The ROC curves (in the top) and The K-S Curves (in the bottom) for the 5-folds Logistic Regression (LR) scoring models.

superior to 0.90 and vary from 0.901 for LR to 0.935 for CCNN (compared to 0.904 for CART model). The average value of AUC for the scoring models represents a classifier of excellent quality (as explained earlier in the methodology section). Clearly, CCNN has the most superior quality by the AUC criterion. In addition, the average Gini coefficient for the different models varies between 0.802 for LR to 0.870 for CCNN (compared to 0.808 for CART model). All coefficients are greater than 0.6 so, as discussed in the methodology section, it demonstrates that all models are of very good

quality. It should also be emphasised that our results are consistent and based on ACC rates' results CCNN is considered the best classifier above other techniques with 92.32% correct classification rate for the five hold-out (testing) sub-sample. In line with this, error rates' results show that CCNN is superior to other techniques as explained above. Clearly CCNN appears to be superior to the other techniques using our evaluation criteria in forecasting default. These predictive capabilities should carry over into practice in classifying future credit applications into good and bad risk-

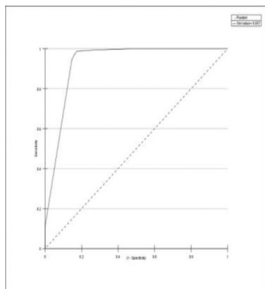
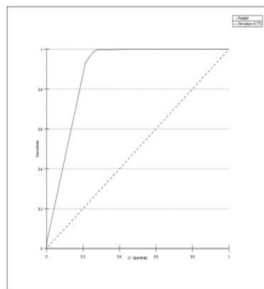
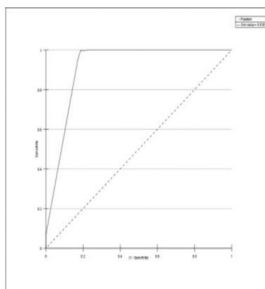
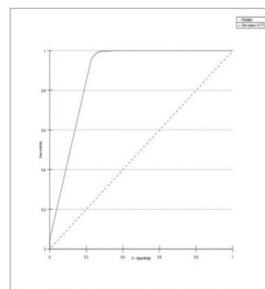
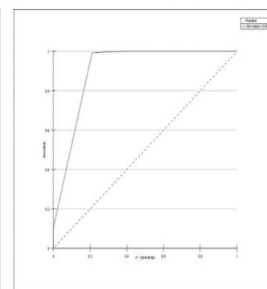
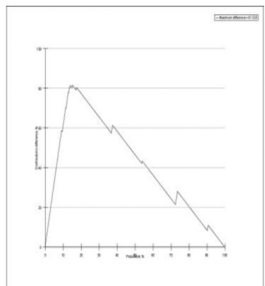
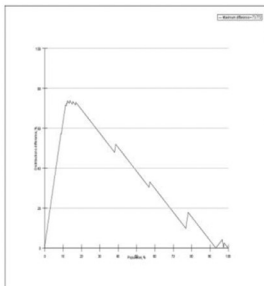
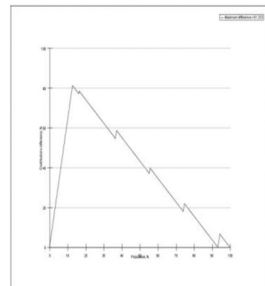
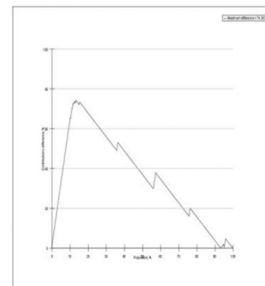
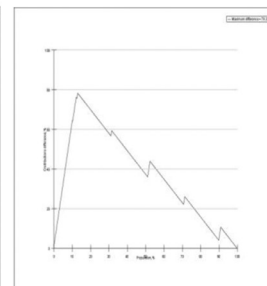
7a: Fold₁7b: Fold₂7c: Fold₃7d: Fold₄7e: Fold₅7f: Fold₁7g: Fold₂7h: Fold₃7i: Fold₄7j: Fold₅

Fig. 7. The ROC curves (in the top) and The K-S Curves (in the bottom) for the 5-folds decision tree (CART) scoring models.

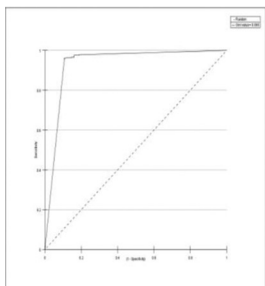
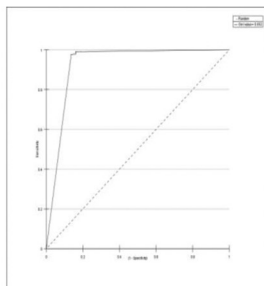
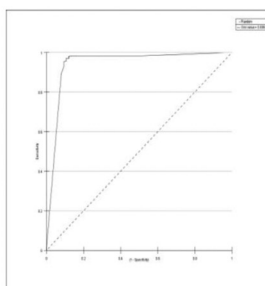
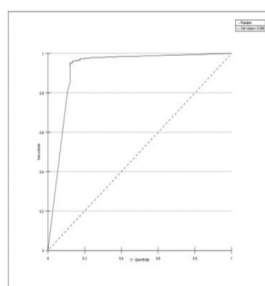
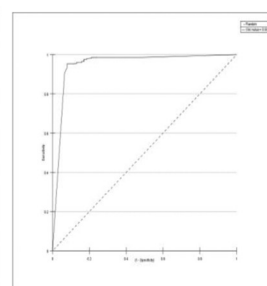
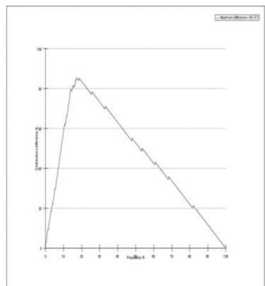
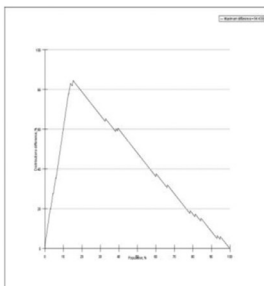
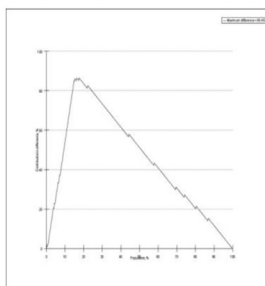
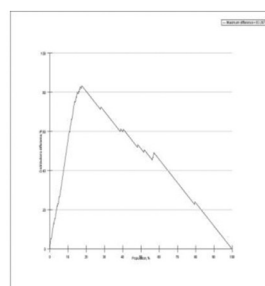
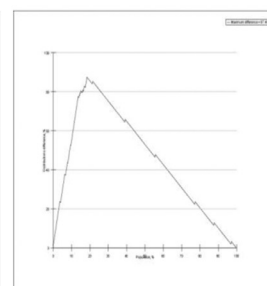
8a: Fold₁8b: Fold₂8c: Fold₃8d: Fold₄8e: Fold₅8f: Fold₁8g: Fold₂8h: Fold₃8i: Fold₄8j: Fold₅

Fig. 8. The ROC curves (in the top) and The K-S Curves (in the bottom) for the 5-folds Cascade Correlation Neural Network (CCNN) scoring models.

classes. These results are consonant with other authors such as Crook et al. [18] who came to a similar conclusion that advanced scoring models have higher ROC and Gini values compared to conventional techniques.

For the purpose of comparing the ROC curves results and in order to evaluate the overall scoring predictability and effectiveness, we consider Kolmogorov-Smirnov (K-S) curves as a robustness test. The K-S curve is one of a number of measures used throughout

Table 8
Importance of the variables under each model and their averages.

LR Models: Contribution weight												
Pre.	Fold ₁	Fold ₂	Fold ₃	Fold ₄	Fold ₅	CART Models: Contribution weight						
Pre.	Fold ₁	Fold ₂	Fold ₃	Fold ₄	Fold ₅	Pre.	Fold ₁	Fold ₂	Fold ₃	Fold ₄	Fold ₅	Mean
POC	0.267	0.324	0.241	0.329	0.283	POC	0.183	0.202	0.226	0.248	0.194	0.211
GRT	0.202	0.175	0.169	0.202	0.161	BAF	0.117	0.156	0.098	0.086	0.112	0.114
BAF	0.122	0.158	0.092	0.108	0.123	GRT	0.116	0.090	0.102	0.107	0.081	0.099
LOB	0.115	0.103	0.109	0.138	0.113	LPE	0.037	0.042	0.106	0.042	0.076	0.061
LPE	0.097	0.051	0.063	0.048	0.038	MCR	0.111	0.000	0.070	0.034	0.072	0.057
TPN	0.044	0.055	0.062	0.008	0.058	AGE	0.047	0.077	0.052	0.052	0.058	0.057
MNC	0.000	0.050	0.058	0.058	0.051	MST	0.030	0.034	0.085	0.042	0.050	0.048
MCR	0.000	0.051	0.000	0.048	0.036	LAT	0.051	0.056	0.020	0.063	0.050	0.048
MST	0.000	0.000	0.065	0.000	0.068	EDN	0.051	0.065	0.058	0.056	0.004	0.047
JOB	0.025	0.007	0.011	0.033	0.037	TPN	0.040	0.040	0.048	0.040	0.042	0.042
GNR	0.015	0.011	0.05	0.017	0.014	LDN	0.022	0.059	0.020	0.047	0.054	0.040
EDN	0.071	0.000	0.011	0.000	0.000	NDP	0.000	0.041	0.015	0.023	0.092	0.034
CON	0.027	0.003	0.041	0.005	0.003	CON	0.027	0.005	0.056	0.057	0.022	0.033
HST	0.015	0.012	0.016	0.006	0.015	HST	0.089	0.016	0.000	0.004	0.049	0.032
NDP	0.000	0.000	0.012	0.000	0.000	LOB	0.015	0.055	0.010	0.045	0.013	0.028
AGE	0.000	0.000	0.000	0.000	0.000	GNR	0.042	0.021	0.011	0.019	0.029	0.024
LDN	0.000	0.000	0.000	0.000	0.000	JOB	0.022	0.018	0.010	0.013	0.002	0.013
LAT	0.000	0.000	0.000	0.000	0.000	MNC	0.000	0.023	0.013	0.022	0.000	0.012
Σ	1.000	1.000	1.000	1.000	1.000	Σ	1.000	1.000	1.000	1.000	1.000	1.000

Notation: LR = Logistic Regression Model; CART = Cascade Correlation Neural Network Model; LAT = Loan Amount; LDN = Loan Duration; LPE = Loan Purpose; AGE = Borrower's Age at Time of Lending; MST = Marital Status; GNR = Gender; NDP = Number of Dependents; JOB = Current Job; EDN = Education; HST = Housing Status; TPN = Monthly Income; MCR = Monthly Expenses; GRT = Guarantees; CON = Car Ownership; BAF = Borrower's Account Functioning; LOB = Other Loans; POC = Previous Employment.

statistics to describe how far apart the distribution functions of two populations (i.e. the scores of the good credit and the bad credit) are. It can describe the general properties of the scorecard and does not depend on which cut-off score is used. This measure can give a feel for the robustness of the scorecard if the cut-off score is changed, also can be useful to determine what the cut-off score should be. This measure can be used as an indicator of the relative effectiveness of different scorecards (see for example, [57]).

The general formula for K-S statistics can be presented as follows (see for example, [57], p.905):

$$K - S = \max_s |P_G(s) - P_B(s)|$$

where, $P_G(s)$ and $P_B(s)$ are the 'good' and the 'bad' distribution functions with score s where it covers the whole the score range.

Figs. 6–8 show different models K-S curves, and the top point on each of these curves refers to the maximum difference between the distribution of 'good' and 'bad' credit. The K-S measure is often used together with the Gini coefficient to assess scorecards quality. The average K-S curve values vary between 74.347 for LR model and 85.394 for CCNN scoring models (compared to an average value of 77.820 for CART scoring models). Clearly CCNN considering maximum iteration number as a model parameter is superior to the other scoring models and The K-S curves results do confirm the ROC curves results for all scoring models, as shown in Figs. 6–8.

4.2.3. Sensitivity analysis of variables

From Table 8, it can be observed that different scoring models treat the variables differently as they respectively attribute to them different levels of importance. However, there is an agreement about three variables amongst them namely POC, BAF and GRT. Aggregating the ranking of the average contribution weights of the three scoring models allows us to establish the five most importantly ranked variables, as follows: POC, BAF, GRT, LOB and MCR. By contrast, the least important variables for these modelling techniques are as follows: LDN, LAT, AGE, NDP and GNR. Of these five most important variables three namely BAF, POC and GRT are identified in the investigative stage as being currently used in the present traditional system for evaluating consumer loans within the chosen banking sector. The other two variables namely LOB and MCR are not given due prominence in current practice in Cameroon and the BEAC family (in addition to TPN, which is very close in its ranking to MCR), yet we find that they are very important. Thus we submit a case for the Cameroonian banking sector, and the BEAC family, to pay more attention to the variables which we find to be important, even while they are not yet using scoring models. It is expected that, if implemented, credit scoring models could help the BEAC family banking industry to provide credit not only at lower cost to themselves but also more expeditiously and to a much larger population.

5. Conclusions

We have shown that there is clearly a powerful role for credit scoring models in emerging economies as exemplified by the Cameroonian banking sector, and the BEAC family which apply the same system, over the traditional approaches to credit prediction. We explore the case for the more sophisticated scoring techniques through two stages. At the investigative stage, we find that judgemental methods are used in Cameroon to meet the demand for credit, with statistical models playing no role. Local assessment practices are slow, costly, and laborious, and constrain the

banks into providing credit very largely to existing customers. Previous Occupation, Guarantees, and Borrower's Account Functioning are identified as the most important criteria preferred by credit officers.

At the evaluative stage, we demonstrate that statistical scoring models for credit decision making are a more effective means of forecasting than the currently applied judgemental approaches. Within the statistical models the advanced scoring techniques are found in this study to be superior to conventional scoring techniques. Our results show that CCNN is the best scoring model based on the hold-out samples achieving the lowest Type II error of 35.11% and the highest AUC value of 93.50%. Therefore, it can be concluded that neural networks, in terms of predictive accuracy, are superior to other scoring models as a classifier. Our results suggest that the default rate from 15.69% under the current approach would drop to 7.68% (100% – 92.32%) under CCNN (see Table 6). In addition ROC curves and Gini coefficients show that CCNN is more powerfully predictive than the other scoring models applied in this paper, which is also confirmed by our robustness test applying Kolmogorov-Smirnov Curves. From our sensitivity analysis, we find that the five key variables, based upon all different scoring modelling techniques are POC, BAF, GRT, LOB and MCR. Of these, Previous Occupation, Borrower's Account Functioning and Guarantees in particular are highlighted for their importance in the cultural and economic environment of BEAC banking industry. We consider this to be of critical interest to bankers.

Future research could be conducted again on a larger sample. We could also investigate whether different results can be achieved if different model parameters (i.e. the maximum iterations number, the correct classification rate and the correct classification rate) are applied using CCNNs. Additionally, other statistical techniques could be applied, such as fuzzy algorithms, genetic programming, hybrid techniques, and expert systems. Furthermore, real field studies could be undertaken into misclassification costs of forgone profit on good customers rejected and lost revenues from bad debts arising from bad customers misclassified as good. The scope of the present study could be extended to business loans and other products. Further research could investigate the socio-economic benefits of shifting the risk from the current Tontine system to formal banking.

Appendix. Cameroonian market

The Cameroonian banking sector and all activities relating to savings and/or credit in Cameroon are supervised by the "Banking Commission of Central Africa" (Commission Bancaire de l'Afrique Centrale, COBAC). COBAC was created by the BEAC member states in 1993 to secure the region's banking system. COBAC ensures that the banking rules are respected in the six BEAC countries and it can apply sanctions to banks that do not follow them scrupulously [14]. As of 2010, COBAC had twelve banks under its supervision in Cameroon. These are private banks, with important foreign and local participation and moderate state involvement without a majority stake. The twelve banks have a total of 128 branches across Cameroon with about CFA87.65 billion (€131.67 million) in assets [15]. CEMAC as a whole has a total of 39 banks with 245 branches and combined capital of CFA271.68 billion (€407.97 million). Hence, Cameroon holds about one third of the banking power of the six countries in the CEMAC zone and about half of all branches are situated in Cameroon [8]. A list of Cameroon's banks, their acronyms, their capital distribution and number of branches is provided below. Cameroon's banking system is also monitored by the Ministry of Finance and Economy.

List of Bank in Cameroon as per COBAC annual report 2010 [14]

Bank name	Short name	Capital (million CFA)	Capital distribution (%)	Number of branches
Afriland First Bank	First Bank	9000	Foreign 56.45 Private 43.55	14
Amity Bank Cameroon PLC	Amity	7400	Foreign 6.75	9
Banque Internationale du Cameroun pour l'Epargne et le Cr�dit	BICEC	6000	Private 93.25 Foreign 82.5	27
Commercial Bank of Cameroon	CBC Bank	7000	Public 17.5 Foreign 33.66	9
Citibank N.A. Cameroon	Citibank	5684	Private 66.44 Foreign 100	2
Ecobank Cameroun	Ecobank	5000	Foreign 86.05 Private 13.95	15
CA SCB Cameroun	CLC	6000	Foreign 65.00 Public 35.00	15
Soci�t� G�n�rale de Banques au Cameroun	SGBC	6250	Foreign 74.40	21
Standard Chartered Bank Cameroon	SCBC	7000	Public 25.60 Foreign 99.99	2
Union Bank of Cameroon PLC	UBC Plc	20,000	Private 00.01 Foreign 54.00	5
National Financial Credit Bank	NFC Bank	3317	Private 11.45 Public 34.55 Private 100	8
Union Bank of Africa	UBA	5000	2 Foreign 99.99 Private 00.01	
Total = 12 Banks		87,651		128 branches

References

- [1] H. Abdou, J. Pointon, A. Elmasry, Neural nets versus conventional techniques in credit scoring in Egyptian banking, *Expert Syst. Appl.* 35 (3) (2008) 1275–1292.
- [2] H. Abdou, Genetic programming for credit scoring: The case of the Egyptian public sector banks, *Expert Syst. Appl.* 36 (9) (2009) 11402–11417.
- [3] H. Abdou, An evaluation of alternative scoring models in private banking, *J. Risk Finance* 10 (1) (2009) 38–53.
- [4] H. Abdou, J. Pointon, Credit scoring, statistical techniques and evaluation criteria: a review of the literature, *Intell. Syst. Account., Finance Manage.* 18 (2–3) (2011) 59–88.
- [5] H. Abdou, S. Alam, J. Mulkeen, Would credit scoring work for Islamic finance? A neural network approach, *Int. J. Islamic Middle Eastern Finance Manage.* 7 (1) (2014) 112–125.
- [6] S. Akkoc, An empirical comparison of conventional techniques, neural networks and the three stage hybrid Adaptive Neuro Fuzzy Inference System (ANFIS) model for credit scoring analysis: the case of Turkish credit card data, *Eur. J. Oper. Res.* 222 (1) (2012) 168–178.
- [7] B. Baesens, T.V. Gestel, S. Viaene, M. Stepanova, J. Suykens, J. Vanthienen, Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring, *J. Oper. Res. Soc.* 54 (6) (2003) 627–635.
- [8] BEAC, Banque des Etats de l'Afrique Centrale, l'institut d'emission de l'Afrique centrale a travers le xxe siecle.
- [9] H. Bekhet, S. Eletter, Credit risk assessment model for Jordanian commercial banks: neural scoring approach, *Rev. Dev. Finance* 4 (1) (2014) 20–28.
- [10] T. Bellotti, J. Crook, Loss given default models incorporating macroeconomic variables for credit cards, *Int. J. Forecasting* 28 (1) (2012) 171–182.
- [11] F. Chandra, P. Varghese, Fuzzifying Gini Index based decision trees, *Expert Syst. Appl.* 36 (4) (2009) 8549–8559.
- [12] Y.-S. Chen, C.-H. Cheng, Hybrid models based on rough set classifiers for setting credit rating decision rules in the global banking industry, *Knowledge-Based Systems* 39 (2013) 224–239.
- [13] C.-L. Chuang, R.-H. Lin, Constructing a reassigning credit scoring model, *Expert Syst. Appl.* 36 (2, 1) (2009) 1685–1694.
- [14] COBAC, La Commission Bancaire de l'Afrique Centrale (COBAC).
- [15] COBAC, Annual Report.
- [16] S. Crone, S. Finlay, Instance sampling in credit scoring: an empirical study of sample size and balancing, *Int. J. Forecasting* 28 (1) (2012) 224–238.
- [17] J. Crook, J. Banasik, Forecasting and explaining aggregate consumer credit delinquency behaviour, *Int. J. Forecasting* 28 (1) (2012) 145–160.

- [18] J. Crook, D. Edelman, L. Thomas, Recent developments in consumer credit risk assessment, *Eur. J. Oper. Res.* 183 (3) (2007) 1447–1465.
- [19] Da Silva, J.D.S. (undated). The Cascade-Correlated Neural Network Growing Algorithm using the Matlab Environment. Available at: <http://www.lac.inpe.br/~demisio/cap351/m11-2slidep.pdf> (Accessed April, 2012).
- [20] C. Damgaard, J. Weiner, Describing inequality in plant size or fecundity, *Ecology* 81 (4) (2000) 1139–1142.
- [21] R.H. Davis, D.B. Delman, A.J. Gammerman, Machine learning algorithms for credit-card applications, *IMA J. Math. Appl. Bus. Ind.* 4 (4) (1992) 43–51.
- [22] V.S. Desai, J.N. Crook, G.A. Overstreet, A comparison of neural networks and linear scoring models in the credit union environment, *Eur. J. Oper. Res.* 95 (1) (1996) 24–37.
- [23] T.H.T. Dinh, S. Kleimeier, A credit scoring model for Vietnam's retail banking market, *Int. Rev. Financ. Anal.* 16 (5) (2007) 471–495.
- [24] D. Durand, Risk Elements in Consumer Instalment Financing, *Studies in Consumer Instalment Financing*, National Bureau of Economic Research, New York, 1941.
- [25] S.E. Fahlman, Faster-learning variations on back-propagation: an empirical study, in: *Proceedings of the 1988 Connectionist Models Summer School*, Morgan Kaufmann, 1988.
- [26] Fahlman, S. (1991). The Recurrent Cascade-Correlation Architecture. Available at: <http://pi.314159.ru/fahlman1.pdf> (Accessed April, 2012).
- [27] Fahlman, S. & Lebiere, C. (1991). The Cascade-Correlation Learning Architecture. Available at: <http://www.cs.iastate.edu/~honor/fahlman.pdf> (Accessed April, 2012).
- [28] R.A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugen.* 7 (2) (1936) 179–188.
- [29] L.W. Glorfeld, B.C. Hardgrave, An improved method for developing neural networks: the case of evaluating commercial loan creditworthiness, *Comput. Oper. Res.* 23 (10) (1996) 933–944.
- [30] D.J. Hand, S.D. Jacka, *Statistics in Finance*, Arnold Applications of Statistics, London, 1998.
- [31] Henry, A. (2003). Using tontines to run the economy. Available at: <http://ecole.org/seminaires/FS3/SEM105/VC190603-ENG.pdf/view> (Accessed March, 2012).
- [32] N.-C. Hsieh, L.-P. Hung, A data driven ensemble classifier for credit scoring analysis, *Expert Syst. Appl.* 37 (1) (2010) 534–545.
- [33] J. Huang, G. Tzeng, C. Ong, Two-stage genetic programming (2SGP) for the credit scoring model, *Appl. Math. Comput.* 174 (2) (2006) 1039–1053.
- [34] M.-J. Kim, D.-K. Kang, H.B. Kim, Geometric mean based boosting algorithm with over-sampling to resolve data imbalance problem for bankruptcy prediction, *Expert Syst. Appl.* 42 (3) (2015) 1074–1082.
- [35] Kouassi, A., Akpapunna, J. & Soedjedje, H. (undated). Cameroon. Available at: <http://fic.wharton.upenn.edu/fic/africa/Cameroon%20Final.pdf> (Accessed March, 2012).
- [36] B. Larivière, V.-D. Poel, Predicting customer retention and profitability by using random forests and regression forests techniques, *Expert Syst. Appl.* 29 (2) (2005) 472–484.
- [37] T. Lee, C. Chiu, C. Lu, I. Chen, Credit scoring using the hybrid neural discriminant technique, *Expert Syst. Appl.* 23 (3) (2002) 245–254.
- [38] T. Lee, I. Chen, A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression spines, *Expert Syst. Appl.* 28 (4) (2005) 743–752.
- [39] T. Lee, C. Chiu, Y. Chou, C. Lu, Mining the customer credit using classification and regression tree and multivariate adaptive regression spines, *Comput. Stat. Data Anal.* 50 (4) (2006) 1113–1130.
- [40] S.L. Lin, A new two-stage hybrid approach of credit risk in banking industry, *Expert Syst. Appl.* 36 (4) (2009) 8333–8341.
- [41] F. Louzada, P. Ferreira-Silva, C. Diniz, On the impact of disproportional samples in credit scoring models: an application to a Brazilian bank data, *Expert Syst. Appl.* 39 (9) (2012) 8074–8078.
- [42] R. Malhotra, D.K. Malhotra, Evaluating consumer loans using neural networks, *Omega Int. J. Manage. Sci.* 31 (2) (2003) 83–96.
- [43] K. Majeske, T. Lauer, The bank loan approval decision from multiple perspectives, *Expert Syst. Appl.* 40 (5) (2013) 1591–1598.
- [44] A. Ono, R. Hasumi, H. Hirata, Differentiated use of small business credit scoring by relationship lenders and transactional lenders: evidence from firm-bank matched data in Japan, *J. Bank. Finance* 42 (2014) 371–380.
- [45] C. Ong, J. Huang, G. Tzeng, Building credit scoring models using genetic programming, *Expert Syst. Appl.* 29 (1) (2005) 41–47.
- [46] N. Sarlija, M. Bensic, M. Zekic-Susac, Comparison procedure of predicting the time to default in behavioural scoring, *Expert Syst. Appl.* 36 (5) (2009) 8778–8788.
- [47] Scortto, Scortto Credit Decision – User Manual, Scortto™ Cooperation, 2009.
- [48] A. Steenackers, M.J. Goovaerts, A credit scoring model for personal loans, *Insur.: Math. Econ.* 8 (8) (1989) 31–34.
- [49] M. Šušteršič, D. Mramor, J. Zupan, Consumer credit scoring models with limited data, *Expert Syst. Appl.* 36 (3) (2009) 4736–4744.
- [50] Tape, T.G. (2010). Interpreting diagnostic tests. Available at: <http://gim.unmc.edu/dxtests/roc3.htm> (Accessed April, 2012).
- [51] L.C. Thomas, A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers, *Int. J. Forecasting* 16 (2) (2000) 149–172.
- [52] L.C. Thomas, Modelling the credit risk for portfolios of consumer loans: analogies with corporate loan models, *Math. Comput. Simulat.* 79 (8) (2009) 2525–2534.
- [53] L.C. Thomas, D.B. Edelman, L.N. Crook, *Credit Scoring and Its Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 2002.
- [54] E. Tong, C. Mues, L. Thomas, Mixture cure model in credit scoring: if and when borrowers default, *Eur. J. Oper. Res.* 218 (1) (2012) 132–139.
- [55] G. Wang, J. Ma, L. Huang, K. Xu, Two credit scoring models based on dual strategy ensemble trees, *Knowledge-Based Syst.* 26 (2012) 61–68.
- [56] D. West, Neural network credit scoring models, *Comput. & Oper. Res.* 27 (11–12) (2000) 1131–1152.
- [57] Z. Yang, Y. Wang, Y. Bai, X. Zhang, Measuring scorecard performance, *Comput. Sci.-ICCS* (2004) 900–906.
- [58] M. Zekic-Susac, N. Sarlija, M. Bensic, Small business credit scoring: a comparison of logistic regression, neural networks, and decision tree models, 26th International Conference on Information Technology Interfaces, 2004.
- [59] J. Zhang, L. Thomas, Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD, *Int. J. Forecasting* 28 (1) (2012) 204–215.
- [60] D. Zhang, X. Zhou, S.C.H. Leung, J. Zheng, Vertical bagging decision trees model for credit scoring, *Expert Syst. Appl.* 37 (12) (2010) 7838–7843.
- [61] X. Zhu, J. Li, D. Wu, H. Wang, C. Liang, Balancing accuracy, complexity and interpretability in consumer credit decision making: a C-TOPSIS classification approach, *Knowledge-Based Syst.* 52 (2013) 258–267.
- [62] E.I. Altman, Financial ratios, discriminant analysis and the prediction of corporate bankruptcy, *The Journal of Finance* XXIII (4) (1968) 589–609.
- [63] J. Zurada, N. Kunene, Comparisons of the Performance of Computational Intelligence Methods for Loan Granting Decisions, 2011.