Forecasting creditworthiness in retail banking: a comparison of cascade correlation neural networks, CART and logistic regression scoring models

Hussein A. Abdou*
The University of Huddersfield, Huddersfield Business School, Huddersfield, West Yorkshire, UK, HD1 3DH

Marc D. Dongmo Tsafack
Salford Business School, University of Salford, Salford, Greater Manchester, M5 4WT, UK

ABSTRACT

The preoccupation with modelling credit scoring systems including their relevance to forecasting 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 the Cameroonian commercial banking sector with implications for fellow members of the Banque des Etats de L'Afrique Centrale (BEAC) family which apply the same system. We investigate their currently used approaches to assessing personal loans and we construct appropriate scoring models. Three statistical modelling scoring techniques are applied, namely Logistic Regression (LR), Classification and Regression Tree (CART) and Cascade Correlation Neural Network (CCNN). To compare various scoring models' performances we use Average Correct Classification (ACC) rates, error rates, ROC curve and GINI coefficient as evaluation criteria. The results demonstrate that a reduction in terms of forecasting power from 15.69% default cases under the current system, to 3.34% based on the best scoring model, namely CART can be achieved. The predictive capabilities of all three models are rated as at least very good using GINI coefficient; and rated excellent using the ROC curve for both CART and CCNN. 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 borrower's account functioning, previous occupation, guarantees, car ownership, and loan purpose as key variables in the forecasting and decision making process which are at the heart of overall credit policy.

Keywords: Forecasting creditworthiness; credit scoring; cascade correlation neural networks; CART; predictive capabilities.

JEL Classification: E50; G21; C45

1. Introduction

The capability of statistical credit scoring systems to improve forecasting decision-making and time efficiencies in the financial sector has widely attracted researchers and practitioners particularly in recent years (see for example, Abdou & Pointon, 2011; Šušteršic, *et al,* 2009; Ong, *et al,* 2005; Lee *et al,* 2002; Thomas *et al,* 2002; Thomas, 2000). Credit scoring systems are now regarded as virtually indispensible in developed countries. In developing countries the 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 (Crook & Banasik, 2012; Šušteršic, *et al,* 2009; Thomas, 2009; Dinh & Kleimeier, 2007; Thomas *et al,* 2002; Steenackers & Goovaerts, 1989).

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 sector[1]. 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 services; 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 2007 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[2] including the selection of appropriate scoring techniques.

In Cameroon and across BEAC, a judgemental and traditional system called Tontines remains very popular. A Tontine is a scheme in which members of a group combine resources to create a kitty (Kouassi *et al*, undated). 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 (Henry, 2003). 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 (Kouassi *et al*, undated). 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

[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 Financiere en Afrique Centrale) and is pegged to the Euro at a rate of €1= CFA665.957 (BEAC, 2010).

[2] 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 (COBAC, 2010). As of 2008, 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 (COBAC, annual report, 2008). 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 (BEAC, 2010). A list of Cameroon's banks, their acronyms, their capital distribution and number of branches is provided in the Appendix. Cameroon's banking system is also monitored by the Ministry of Finance and Economy.

provide collaterals, Tontines are estimated to handle about 90 per cent of individuals' credit needs in Cameroon, whereas the commercial and savings and loan banks realize a volume of about 10 per cent of all national loan business (Kouassi *et al*, undated). 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.

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 (Kouassi *et al*, undated; Henry, 2003).

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 forecast 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 forecast 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 (1936) 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 (1941) who investigated car loans. Altman (1968) introduced a corporate bankruptcy prediction scoring model based on five financial ratios.

Advances in information processing have fueled progress in credit scoring techniques and applications. Conventional statistical techniques including logistic regression (LR) have been widely used and compared with non-parametric techniques such as classification and regression tree (CART) in building scoring models (e.g. Hand & Jacka, 1998; Thomas, 2000; Baesens *et al*., 2003; Zekic-Susac, *et al*. 2004; Lee *et al.*, 2006; Chuang & Lin, 2009; Crone & Finlay, 2012). Logistic regression deals with a dichotomous dependent variable which distinguishes it from a linear regression model. Logistic regression 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 (Steenackers & Goovaerts, 1989; Lee *et al*, 2002; Abdou & Pointon, 2011). LR varies from other conventional techniques such as discriminant analysis in that it does not require the assumptions necessary for the discriminant problem (Desai *et al*, 1996; Abdou & Pointon, 2011). Classification and regression tree is a tree-like decision model which is also used for classification of an object within two or more classes (Crook *et al,* 2007). CART can be used to analyse either quantitative or categorical data and is widely used in building scoring models (e.g. Lee *et al*, 2006; Hsieh & Hung, 2010; Chuang & Lin, 2009; Zhang *et al*, 2010; Bellotti & Crook, 2012; Crone & Finlay, 2012; Zhang & Thomas, 2012).

Advanced statistical techniques such as neural networks have been widely used in building scoring models (Glorfeld and Hardgrave, 1996; West, 2000; Malhotra & Malhotra, 2003; Lee & Chen, 2005; Crook *et al.* 2007; Abdou & Pointon, 2011; Brentnall *et al.* 2010; Loterman *et al.* 2012). Also, by way of comparison between neural networks and other non-parametric techniques such as CART, Davis *et al.* (1992) compared CART with Multilayer Perceptron Neural Network for credit card applications, and found comparable results for decision accuracy. Zurada and Kunene (2011) 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. A neural network usually consists of a three layered system comprising input, hidden, and output layers (Huang *et al*, 2006; Abdou & Pointon, 2011). 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 (Fahlman & Lebiere, 1991; Da Silva, undated). Various scoring models' evaluation criteria including average correct classification rates, error rates, receiver operating characteristic (ROC) curve and Gini coefficient are widely used and serve to assess the predictive capabilities of scoring models (Damgaard & Weiner, 2000; Crook *et al,* 2007; Abdou, 2009; Chandra & Varghese, 2009; Sarlija *et al*, 2009; Abdou & Pointon, 2011).

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 Cameroon. 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, where judgemental approaches with 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 Cameroonian banking sector for personal loans. At this stage, a pilot study comprising three informal interviews was 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 Cameroonian banking sector, and use three different statistical techniques, namely, LR, CART and CCNN. This is followed by an evaluation of the predictive capabilities of the scoring models using ACC rates, error rates, ROC curve and GINI coefficients. Here, different software is applied, including Scorto Credit Decisions. 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 BEAC family which apply 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, Crook *et al*, 2007, for a similar expression):

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

where,

$\boldsymbol{\alpha}$, 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 an applicant under case i will be good is given by:

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

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 above the cut-off point, the more creditworthy the application will 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' class risk. 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, 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, Baesens *et al*., 2003, p. 631; Scorto, 2007, p. 53), the *GainRatio* can be expressed as follows:

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

where,

$GainInfo_x$ is a criterion used by the C4.5 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:

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

whereby,

$p_1(p_0)$ 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*.

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 (Fahlman, 1988) 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 to the network in an attempt is added to reduce the residual error (Fahlman, 1991, p. 1).

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 net work error (Fahlman & Lebiere, 1991). 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 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 (Fahlman, 1991; Da Silva, no date), as shown in Figure 1.

FIGURE (1) HERE

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*, in which it can be calculated as follows (see, for example, Fahlman & Lebiere, 1991, p.5; Da Silva, no date, p.2):

$$C = \sum_{o} \left| \sum_{t} (N_t - \overline{N})(E_{t,o} - \overline{E_o}) \right|$$

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*; $\overline{N}$ is the average of *N* over all patterns; $\overline{E_o}$ 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, Da Silva, undated, p.2; Fahlman & Lebiere, 1991, p.5):

$$\frac{\partial C}{\partial w_i} = \sum_{t,o} \sigma_o \left(E_{t,o} - \overline{E_o}\right) d_t' I_{i,t}$$

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; $I_{i,t}$ is the input received by the candidate's unit from unit *i* for pattern *t*.

3.2. Proposed performance evaluation criteria for scoring models

3.2.1. Classification matrix and error rates

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 criterion, the reader is referred to Abdou (2009).

3.2.2. 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 other to assess their effectiveness. ROC is very powerful when dealing with a narrow cut-off range (Crook *et al*, 2007). 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 (1– specificity) the model with the higher sensitivity is preferred. Additionally, for a given level of sensitivity, the model with a lower level of (1 – specificity) is also preferred. These criteria are simple to apply. 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, Larivière, & Poel, 2005; Lin, 2009; Tape, 2010), can be interpreted as: $0 \leq AUC < 0.6$ = fail; $0.6 \leq AUC < 0.7$ = poor; $0.7 \leq AUC < 0.8$ = fair; $0.8 \leq AUC < 0.9$ = good; and $0.9 \leq AUC$ = 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' class risks have been separated.

The relationship between the GINI coefficient and the AUC value is given by AUC = $\frac{\text{GINI}+1}{2}$ (see, for example, Scorto, 2007, p.77). The following are some interpretations of the GINI values for assigning levels of quality to classifiers (Scorto, 2007, p.77):

$0 \leq$ GINI < 0.25 = low quality classifier
$0.25 \leq$ GINI < 0.45 = Average quality classifier
$0.45 \leq$ GINI < 0.60 = Good quality classifier, and
$0.60 \leq$ GINI = very good quality classifier.

3.3. Data collection and sampling

The data-set for the construction of the different models comprises 599 historical blind consumer loans provided by a Cameroonian bank. This data-set consists of 505 good and 94 bad credit cases. To test the predictive capabilities of the scoring models, this data-set has been divided into a training set of 480 cases and a testing set of 119 cases selected randomly. 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 independent/predictor variables and 1 dependent variable, namely, loan status. For all 599 cases there were no missing attributes from the data-set. Some variables attracted the same values for all cases in this data-set and so these variables were excluded. Table 1 portrays information about the nature of the loan, the personal characteristics of the borrower and the borrower's history.

TABLE (1) HERE

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.

3.1. 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, 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 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 and their encoding. Finally, 18 predictor variables are used to build the scoring models. In order to construct the proposed models, we use SPSS 17.0, STATGRAPHICS 5.1 and Scorto Credit Decision. The detailed results from all three statistical modelling techniques, namely, LR, CART and CCNN are summarised next. The respective predictive capability of the classification models is also investigated.

4.2.1. Analysis of the scoring models

4.2.1.1. Logistic regression

It can be observed from Table 2 that for the LR the correct classification of 'good' within a good risk-class is 95.64%, its correct classification of 'bad' within a bad risk-class is 62.76%, and its ACC rate is 90.48% amongst the overall set using a cut-off point of 0.5. The overall ACC rate of training and testing samples are 93.75% and 77.31%, respectively. As a result of conducting a sensitivity analysis of the 18 predictor variables used in building the LR scoring model, Table 4 shows that POC, GRT, BAF, LOB and LPE are the most important variables with contribution weightings of 0.289, 0.181, 0.119, 0.115 and 0.073, respectively. 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 six predictor variables are the least important, namely: HST, EDN, NDP, AGE, LDN and LAT.

4.2.1.2. Classification and Regression Tree

Using a tree[3] depth of 8 and 44 nodes, Table 2 also presents the CART classification matrix, where it can be noted that 100% of 'good' have been correctly classified as good risk-class, 78.72% of 'bad' have been correctly

[3] In building the CART model, the working mode selected decision tree over decision rules. Also, the significant level of tree pruning was 0.25, selected by default, with iterative building of trees and use of the Gain Ratio criterion. It should be emphasised that without the use of these options as part of the software design, different

classified as bad risk-class with an overall ACC rate of 96.66%. A 99.58% and an 84.03% are the ACC rates for the training and testing samples, respectively. In Table 4, conducting a sensitivity analysis, it can be noted that for this model the most important variables are BAF, POC, CON, GRT and LPE with contribution weightings in turn of 0.087, 0.086, 0.066, 0.063 and 0.063, respectively. Our investigative stage identifies POC, GRT and BAF as the most important variables based on the currently used system; this is consonant with our findings applying CART, but with a much lower default rate than in the case of the current system. The least important variables are TPN, HST, LDN, NDP and LOB.

TABLE (2) HERE

4.2.1.3. Cascade Correlation Neural Network

Table 2 above presents its correct classification of 'good' into good risk-class at 96.03%; its correct classification of 'bad' into bad risk-class at 89.36%; and an overall ACC rate at 94.99%. CCNN[4] has the best classification of 'bad' into bad risk-class out of the three models. The ACC rates for training and testing samples are 97.08% and 86.56%, respectively. Also, for CCNN it can be observed from Table 4 that, out of the 18 predictor variables, BAF, LOB, POC, GRT and MCR are the most important variables with contribution weightings of in turn 0.109, 0.109, 0.108, 0.093 and 0.093, respectively. This is consonant with our findings from the investigative stage, but with much lower default rate in the case of the current system. By contrast, JOB, GNR, AGE, LDN and MST are the least important variables.

4.2.2. Comparison of different scoring models

It can be observed that, when comparing all techniques, CART has the highest Average Correct Classification (ACC) rate of 99.58% for the training set, and 96.66% for the overall set, whilst CCNN has the highest ACC rate of 86.56% for the testing set, which shows the superiority of neural networks in forecasting 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. These scoring models are evaluated in this paper also using other criteria, namely, Error rates, AUC and the GINI coefficients. Table 3 summarises the different values under each criterion for each of the models. By inspecting the ACC rate, it can be noted that the accuracy across the three models varies from 90.48% for LR, 94.99% for CCNN to 96.66% for CART. From the judgemental techniques currently being practised in Cameroon, the default cases are 15.7% (94/599) signifying that, those default cases could potentially be reduced by 6.18% through utilisation of LR, 10.69% through CCNN and 12.36% through CART.

---

results are reported as follows: 98.75% and 95.83% correct classification rates for the training and overall samples, respectively. The same correct classification rate of 84.03% for the hold-out sample is recorded. But, a lower GINI coefficient of 81.10% is achieved under this model.

[4] It should be emphasised that in building the CCNN model a Maximum Iteration Number (MIN) is considered as a model parameter over both Correct Classification Rate (CCR) and Network Error Improvement (NEI). Also, an iteration limit value of 5,000 and an error improvement value of 3 are applied. However, applying NEI, as a model parameter, different results were found, as follows: an overall ACC rate of 95.20% is achieved; with 96.50% and 89.90% as the correctly classified rates for training and testing samples, respectively, but with a GINI coefficient value of 82.60%.

TABLE (3) HERE

The error results in Table 3 also show that the Type I errors are very low compared with the Type II errors for all models. However, CART has the lowest Type I error of 0.00%, whilst CCNN has the lowest Type II error of 10.64%. 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. In this respect also CCNN shows its particular power to discriminate between good and bad.

FIGURE (2) HERE

Figure 2 presents the ROC curves for the three models. The computations of the AUC show that its value varies from 0.8940 for LR, 0.9210 for CART, to 0.9475 for CCNN. The value of AUC for LR represents a classifier of good quality (between 0.8 and 0.9), whereas, the CART and CCNN based classifiers with AUC values superior to 0.9 translate into excellent quality (as explained earlier in the methodology section). Clearly, CCNN has the most superior quality by the AUC criterion. Finally, the GINI coefficient for the different models varies between 0.788 for LR, 0.842 for CART to 0.895 for CCNN. All three coefficients are greater than 0.6 so, as discussed in the methodology section, it demonstrates that all three models are of very good quality. Clearly CCNN appears to be superior to the other techniques under this criterion also in forecasting default. These predictive capabilities should carry over into practice in classifying future credit applications into good and bad risk-classes.

4.2.3. Sensitivity analysis of variables

From Table 4, it can be observed that the three models treat the variables differently as they respectively attribute to them different levels of importance. Aggregating the ranking of the contribution weights of the three models allows us to establish the five most importantly ranked variables, as follows: BAF, POC, GRT, CON and LPE. By contrast, the least important variables for these three modelling techniques are as follows: LDN, NDP, AGE, JOB 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 Cameroonian banking sector. The other two variables namely CON and LPE are not given due prominence in current practice in Cameroon (in addition to LOB and MCR, which are very close in their ranking to LPE), yet we find that they are very important. Thus we submit a case for the Cameroonian banking sector 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 Cameroonian banking sector to provide credit not only at lower cost to themselves but also more expeditiously and to a much larger population.

TABLE (4) HERE

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 over the traditional, judgemental approaches to credit forecasting. We explore the case for the more sophisticated scoring techniques through two stages. At the investigative stage, we find that traditional, 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 CART is the best scoring model based on the overall sample achieving a 96.66% ACC rate. Furthermore, in terms of predictive accuracy, CCNN is superior to LR and CART models as a classifier. Our results suggest that the default rate from 15.69% under the current approach would drop to 5.01% (100% - 94.99%) under CCNN (see Table 3). In addition ROC curves and GINI coefficients show that CCNN is more powerfully predictive than the other scoring models applied in this paper. From our sensitivity analysis, we find that the five key variables, based upon the three modelling techniques are BAF, POC, GRT, CON and LPE. Of these, Previous Occupation, Guarantees and Account Functioning Borrower in particular are highlighted for their importance in the cultural and economic environment of Cameroonian banking. We consider this to be of critical interest to bankers.

Future research could be conducted again on a larger sample. 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 and to the other members of BEAC. Further research could investigate the socio-economic benefits of shifting the risk from the current Tontine system to formal banking.

References

Abdou, H. (2009). Genetic programming for credit scoring: The case of the Egyptian public sector banks. *Expert systems with applications*, 36 (9), 11402-11417.

Abdou, H. & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: a review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18 (2-3), 59-88.

Baesens, B., Gestel, T. V., Viaene, S., Stepanova, M., Suykens, J., & Vanthienen, J. (2003). Benchmarking State-of-the-Art Classification Algorithms for Credit Scoring. *Journal of the Operational Research Society*, 54 (6), 627-635.

BEAC, Banque des Etats de l'Afrique Centrale (2010). l'institut d'emission de l'afrique centrale a travers le xxe siecle. Available at: http://www.beac.int/histbeac.htm (Accessed January, 2010).

Bellotti, T. & Crook, J. (2012). Loss given default models incorporating macroeconomic variables for credit cards. *International Journal of Forecasting*. 28 (1), 171-182.

Central Intelligence Agency (CIA) (2010). The world FACTBOOK, Cameroon (hitting 'WORLD FACTBOOK', 'Cameroon'. Available at: https://www.cia.gov/library/publications/the-world-factbook/geos/cm.html (Accessed February, 2010).

Chandra, F. & Varghese, P. (2009). Fuzzifying Gini Index based decision trees. *Expert Systems with Applications*, 36 (4), 8549-8559.

Chen, M., & Huang, S. (2003). Credit scoring and rejected instances reassigning through evolutionary computation techniques. *Expert Systems with Applications*, 24(4), 433-441.

Chuang, C-L. & Lin, R-H. (2009). Constructing a reassigning credit scoring model. *Expert Systems with Applications*, 36 (2, 1), 1685-1694.

COBAC (2010). La Commission Bancaire de l'Afrique Centrale (COBAC). Aailable at: http://www.beac.int/cobac/cbcobac.html (Accessed January, 2010).

COBAC (2008). Annual Report. Available at: http://www.beac.int/cobac/Publications/rapcobac2008.pdf (Accessed March, 2010).

Crone. S. & Finlay, S. (2012). Instance sampling in credit scoring: An empirical study of sample size and balancing. *International Journal of Forecasting*. 28 (1), 224-238.

Crook, J. & Banasik, J. (2012). Forecasting and explaining aggregate consumer credit delinquency behaviour. *International Journal of Forecasting*. 28 (1), 145-160.

Crook, J., Edelman D. & Thomas, L. (2007). Recent developments in consumer credit risk assessment. *European Journal of Operational Research*, 183 (3), 1447-1465.

Da Silva, J. D. S. (no date). 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, 2010).

Damgaard, C. & Weiner, J. (2000). Describing inequality in plant size or fecundity. *Ecology*, 81 (4), 1139-1142.

Davis, R. H., Delman, D. B. & Gammerman, A. J. (1992). Machine learning algorithms for credit-card applications. *IMA Journal of Mathematics Applied in Business and Industry*, 4 (4), 43-51.

Desai, V. S., Crook, J. N. and Overstreet, G. A. (1996). A Comparison of Neural Networks and Linear Scoring Models in the Credit Union Environment. *European Journal of Operational Research*, 95 (1), 24-37.

Dinh, T. H. T. & Kleimeier, S. (2007). A credit scoring model for Vietnam's retail banking market. *International Review of Financial Analysis*, 16 (5), 471–495.

Durand, D. (1941). *Risk Elements in Consumer Instalment Financing, Studies in Consumer Instalment Financing*. New York: National Bureau of Economic Research.

Fahlman, S. E. (1988) "Faster-Learning Variations on Back-Propagation: An Empirical Study" in *Proceedings of the 1988 Connectionist Models Summer School*, Morgan Kaufmann.

Fahlman, S. (1991). The Recurrent Cascade-Correlation Architecture. Available at: http://pi.314159.ru/fahlman1.pdf (Accessed April, 2010).

Fahlman, S. & Lebiere, C. (1991). The Cascade-Correlation Learning Architecture. Available at: http://www.cs.iastate.edu/~honavar/fahlman.pdf (Accessed April, 2010).

Fawcett, T. (2005). An introduction to ROC analysis. *Pattern Recognition Letters*, 27, 861-874.

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, 7 (2), 179-188.

Glorfeld, L. W. & Hardgrave, B. C. (1996). An improved method for developing neural networks: The case of evaluating commercial loan creditworthiness. *Computers & Operations Research*, 23 (10), 933-944.

Hand, D. J. & Jacka, S. D. (1998). *Statistics in Finance*, Arnold Applications of Statistics: London.

Henry, A. (2003). Using Tontines to run the economy. Available at: http://ecole.org/seminaires/FS3/SEM105/VC190603-ENG.pdf/view (Accessed March, 2010).

Hsieh, N-C. & Hung, L-P. (2010). A data driven ensemble classifier for credit scoring analysis. *Expert Systems with Applications*, 37(1), 534-545.

Huang, J., Tzeng, G. & Ong, C. (2006). Two-stage genetic programming (2SGP) for the credit scoring model. *Applied Mathematics and Computation*. 174 (2), 1039-1053.

Kouassi, A., Akpapuna, J. & Soededje, H. (no date). Cameroon. Available at: http://fic.wharton.upenn.edu/fic/africa/Cameroon%20Final.pdf (Accessed March, 2010).

Lachenbruch, P. A. & Goldstein, M. (1979). Discriminant Analysis. *Biometrics*, 35 (1), 69-85.

Larivière, B. & Poel, V-D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, 29 (2), 472-484.

Lee, T., Chiu, C. Lu, C. & Chen, I. (2002). Credit Scoring Using the Hybrid Neural Discriminant Technique. *Expert Systems with Applications*, 23 (3), 245-254.

Lee, T. & Chen I. (2005). A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression spines. *Expert Systems with Applications*, 28 (4), 743-752.

Lee, T., Chiu, C., Chou, Y., & Lu, C. (2006). Mining the customer credit using classification and regression tree and multivariate adaptive regression spines. *Computational Statistics & Data Analysis*, 50 (4), 1113-1130.

Lin, S. L. (2009). A new two-stage hybrid approach of credit risk in banking industry. *Expert Systems with Applications*, 36 (4), 8333-8341.

Malhotra, R, & Malhotra, D. K. (2003). Evaluating consumer loans using Neural Networks. *Omega the International Journal of Management Science*, 31 (2), 83-96.

Ong, C., Huang, J. & Tzeng, G. (2005). Building Credit Scoring Models Using Genetic Programming. *Expert Systems with Applications*, 29 (1), 41-47.

Sarlija, N., Bensic, M. & Zekic-Susac, M. (2009). Comparison procedure of predicting the time to default in behavioural scoring. *Expert Systems with Applications*, 36 (5), 8778-8788.

Scorto (2007). Scorto Credit Decision – User Manual. Scorto™ Cooperation.

Steenackers, A., & Goovaerts, M. J. (1989). A Credit Scoring Model for Personal Loans. *Insurance: Mathematics and Economics*, 8 (8), 31-34.

Šušteršic, M., Mramor, D. & Zupan, J. (2009). Consumer credit scoring models with limited data. *Expert Systems with Applications*, 36 (3), 4736–4744.

Tape, T. G. (2010). Interpreting Diagnostic tests. Available at: http://gim.unmc.edu/dxtests/roc3.htm (Accessed April, 2010).

Thomas, L. C. (2000). A survey of credit and behavioural scoring: forecasting financial risk of lending to consumers. *International Journal of Forecasting*, 16 (2), 149-172.

Thomas, L. C. (2009). Modelling the Credit Risk for Portfolios of Consumer Loans: Analogies with corporate loan models. *Mathematics and Computers in Simulation*, 79 (8), 2525-2534.

Thomas, L. C., Edelman, D. B. & Crook, L. N. (2002). *Credit Scoring and Its Applications*. Philadelphia: Society for Industrial and Applied Mathematics.

West, D. (2000). Neural Network Credit Scoring Models. *Computers & Operations Research*, 27 (11-12), 1131-1152.

Zekic-Susac, M., Sarlija, N., & Bensic, M. (2004). *Small Business Credit Scoring: A Comparison of Logistic Regression, Neural Networks, and Decision Tree Models.* 26th International Conference on Information Technology Interfaces. Croatia.

Zhang, J. & Thomas, L. (2012). Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD. *International Journal of Forecasting*. 28 (1), 204-215.

Zhang, D., Zhou, X., Leung, S.C.H. & Zheng, J. (2010). Vertical bagging decision trees model for credit scoring. *Expert Systems with Applications*, 37 (12), 7838-7843.

Appendix

List of Bank in Cameroon as per COBAC annual report 2008

Bank name	Short name	Capital (million CFA)	Capital distribution (%)		Number of branches
Afriland First Bank	First Bank	9 000	Foreign Private	56.45 43.55	14
Amity Bank Cameroon PLC	Amity	7 400	Foreign Private	6.75 93.25	9
Banque Internationale du Cameroun pour l'Epargne et le Crédit	BICEC	6 000	Foreign Public	82.5 17.5	27
Commercial Bank of Cameroon	CBC Bank	7 000	Foreign Private	33.66 66.44	9
Citibank N.A. Cameroon	Citibank	5 684	Foreign	100	2
Ecobank Cameroun	Ecobank	5 000	Foreign Private	86.05 13.95	15
CA SCB Cameroun	CLC	6000	Foreign Public	65.00 35.00	15
Société Générale de Banques au Cameroun	SGBC	6 250	Foreign Public	74.40 25.60	21
Standard Chartered Bank Cameroon	SCBC	7 000	Foreign Private	99.99 00.01	2
Union Bank of Cameroon PLC	UBC Plc	20 000	Foreign Private Public	54.00 11.45 34.55	5
National Financial Credit Bank	NFC Bank	3 317	Private	100	8
Union Bank of Africa	UBA	5000	Foreign Private	99.99 00.01	2
TOTAL = 12 Banks		87651			128 branches

TABLES

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 people, relying on the borrower for financial support
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; Unknown = 2	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	N/A	-	Not required by the bank

enquiries			
Loan status*	LST	Bad = 0; Good = 1	Quality of the loan

*Variables are finally selected in building the scoring models

Table 2: Classification results for the scoring models, namely, LR, CART and CCNN

Model		Training set				Testing set				Overall set			
		G	B	T	%	G	B	T	%	G	B	T	%
LR													
	G	403	4	407	99.02	80	18	98	81.63	483	22	505	95.64
	B	26	47	73	64.38	9	12	21	57.14	35	59	94	62.77
	T			480	93.75			119	77.31			599	90.48
CART													
	G	407	0	407	100	98	0	98	100	505	0	505	100
	B	1	71	73	97.26	19	2	21	9.52	20	74	94	78.72
	T			480	99.58			119	84.03			599	96.66
CCNN													
	G	397	10	407	97.54	88	10	98	89.80	485	20	505	96.04
	B	4	69	73	94.52	6	15	21	71.43	10	84	94	89.36
	T			480	97.08			119	86.56			599	94.99

Note: G is good; B is bad and T is total.

Table 3: Comparing classification results, error rates, AUC values and GINI coefficients

	Classifications results			**Error results**		**Evaluation Criteria**	
CSMs	GG	BB	ACC rate	Type I	Type II	AUC	GINI
LR	95.64%	62.76%	90.48%	4.36%	37.24%	0.8940	0.788
CART	100%	78.72%	96.66%	0.00%	21.28%	0.9210	0.842
CCNN	96.03%	89.36%	94.99%	3.97%	10.64%	0.9475	0.895

Note: GG is % good correctly classified as good; BB is % bad correctly classified as bad; Type I is % good misclassified as bad; Type II is % bad misclassified as good.

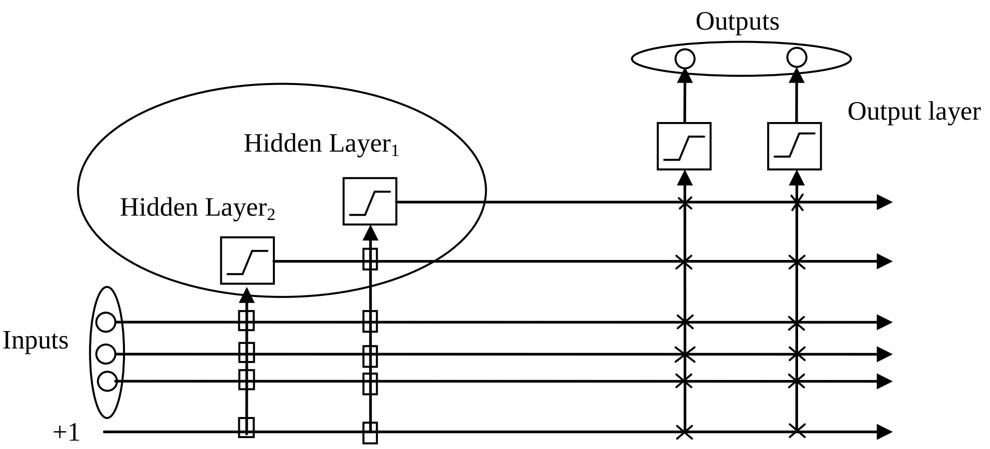
Table 4: Importance of the variables under each model

LR		**CART**		**CCNN**	
Variable	Contribution weight	Variable	Contribution weight	Variable	Contribution weight
POC	0.289	BAF	0.087	BAF	0.109
GRT	0.181	P OC	0.086	LOB	0.109
BAF	0.119	CON	0.066	POC	0.108
LOB	0.115	GRT	0.063	GRT	0.093
LPE	0.073	LPE	0.063	MCR	0.093
TPN	0.049	LAT	0.062	CON	0.085
MNC	0.048	MST	0.061	MNC	0.069
MST	0.046	EDN	0.054	TPN	0.069
MCR	0.037	GNR	0.054	HST	0.069
JOB	0.021	MCR	0.053	EDN	0.043
CON	0.012	JOB	0.051	LAT	0.030

GNR	0.010	AGE	0.049	NDP	0.029
HST	0.000	MNC	0.048	LPE	0.028
EDN	0.000	TPN	0.043	JOB	0.023
NDP	0.000	HST	0.043	GNR	0.018
AGE	0.000	LDN	0.043	AGE	0.018
LDN	0.000	NDP	0.038	LDN	0.004
LAT	0.000	LOB	0.036	MST	0.003
∑	1.000	∑	1.000	∑	1.000

FIGURES

Figure 1: CCNN structure



Source: Fahlman & Lebiere (1991, p. 4) & Fahlman (1991, p. 2), modified.

Figure 2: ROC curves and GINI coefficients for different scoring models

LR **CART** **CCNN**