A Machine Learning Approach for Predicting Bank Credit Worthiness

Regina Esi Turkson, *School of Computer Science and Engineering, University of Electronic Science and Technology of China, Email: regina_turkson@yahoo.com*
Edward Yeallakuor Baagyere, *School of Information and Software Engineering, University of Electronic Science and Technology of China, Email: ybaagyere@uds.edu.gh*
and Gideon Evans Wenya, *Institute of Fundamental and Frontier Sciences, University of Electronic Science and Technology of China, Email: gideonwenya@gmail.com*

***Abstract*—Machine learning is an emerging technique for building analytic models for machines to "learn" from data and be able to do predictive analysis. The ability of machines to "learn" and do predictive analysis is very important in this era of big data and it has a wide range of application areas. For instance, banks and financial institutions are sometimes faced with the challenge of what risk factors to consider when advancing credit/loans to customers. For several features/attributes of the customers are normally taken into consideration, but most of these features have little predictive effect on the credit worthiness or otherwise of the customer. Furthermore, a robust and effective automated bank credit risk score that can aid in the prediction of customer credit worthiness very accurately is still a major challenge facing many banks. In this paper, we examine a real bank credit data and conduct several machine learning algorithms on the data for comparative analysis and to choose which algorithms are the best fit for learning bank credit data. The algorithms gave over 80% accuracy in prediction. Furthermore, the most important features that determine whether a customer will default or otherwise in paying his/her credit the next month are extracted from a total of 23 features. We then applied these most important features on some selected machine learning algorithms and compare their predictive accuracy with the other algorithms that used all the 23 features. The results show no significant difference, signifying that these features can accurately determine the credit worthiness of the customers. Finally, we formulate a predictive model using the most important features to predict the credit worthiness of a given customer.**

***Index Terms*—machine learning, bank credit, classification, confusion matrix, predictive analysis.**

I. Introduction

THE growing volumes, varieties and velocity of data due to the emergence of the Internet in particular and the cheaper data sharing and storage facilities coupled with the cheaper but more powerful computational tools have opened a new frontier in the field of data science. And thus, there is currently an active ongoing research within the fields of data mining (discovering patterns in data) and machine learnings (building analytical models using algorithms for machine to "learn" from data), both aim at using algorithms and concepts to extract knowledge and pattern from data.

One of the major reasons for establishing banks is to advance loans to customers. But in order to stay in business, banks advance these loans to people who have the ability to pay back the money, thereby minimizing the risk of the non-payments of loans. However, risk management; knowing who is credit worthy is still an on-going challenge within the banking sector. The ability to identify a risk score of a customer base on some features such as occupation, age, marital status, salary range/amount of equity, credit history, etc. is an important step that banks go through before giving credit to customers. For the credit risk score helps the banks to decide on how much interest to charge on the loan, etc. However, these risk factors sometimes do not give an inform decision on the credit worthiness of customers. Moreover, many banks lack a central well integrated, automated finance and risk management system due to the inability to develop a robust and scalable risk management system to forecast risk score of customers.

Another nightmare faced by many banks these days is frauds. And the machine learning approach is seen and considered as the right tool that can be leveraged on in order to understand the banking transaction pattern of customers, by identifying pattern in customer data, so as to be able to distinguish between fraudulent activity from that of a normal one [1], [2], [3], [4]. Therefore, we leveraged it on the bank credit dataset in order to understand the key factors that influence the payment of bank loans. The dataset is obtained from the `UCL machine repository` [5]. We perform analysis and applied machine learning algorithms on the bank credit data, firstly to understand the nature of the data and the best algorithms suitable for learning bank credit data. Secondly, we determined among the 23 features of bank customers, which ones are the most important in determining the credit worthiness of a customer. Thirdly, we formulated a predictive model to determine the credit worthiness or otherwise of a given bank customer using a linear regression method.

The rest of the paper is organized into sections as follows: Section II gives the background information on the machine learning algorithms used in studying on the bank's credit dataset. In Section III, the related work is given. We outlined our methodology in Section IV. The experimental setup and model formulation is described in Section V. The paper is concluded in Section VI.

II. Background

The most important background information on machine learning algorithms and their theoretical formulation are outlined in this section. These algorithms are used in analyzing the bank credit data.

A. *Machine Learning Algorithms*

Machine learning techniques can be grouped broadly into two main categories. They include:

(i) **Supervised Learning:** The main feature of this algorithm consists of target or outcome variable (or dependent variable). The target variable is used to predict other features from a given set of predictors (independent variables). Furthermore, using the target variable, a function is generated that maps input to desired outputs. The training process then continues until the model achieves the desired level of accuracy on the training data. Supervised learning techniques are achieved using regression and classification algorithms or approaches that range from non-linear regression,generalized linear regression, discriminant analysis, Support Vector Machines (SVMs) to decision trees and ensemble methods.

(ii) **Unsupervised Learning:** In unsupervised learning, there is no target or outcome variable to predict or estimate. This algorithm is used mainly for segmenting or clustering entities in different groups for specific intervention. Examples of unsupervised learning algorithms include Apriori and K-means algorithms.

The various machine learning approaches and the algorithms that describe them are shown in Fig. 1

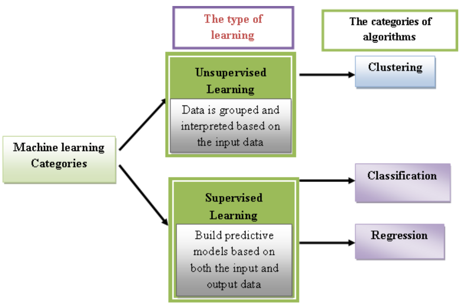

Fig. 1. Machine learning Tasks

Labeled data is known in the literature to be suitable for classification algorithms. The dataset used in this paper is a labeled data and is, therefore, suitable for doing classification analysis. And thus, we employed various classification algorithms described comprehensively in Section II-B. Some of the algorithms are implemented in MATLAB ® and some taken from the *Python scikit-learn package* [5] to predict the creditworthiness of bank customers with regards to their ability to pay their credit or otherwise within a given time frame.

B. *Classification Algorithms*

Classification algorithms work by predicting the best group to which a data point belongs to by "learning" from labeled observations. It uses a set of input features for the "learning" process. Classification algorithms are good for grouping data that are never seen before into their various groupings and are therefore extensively used in machine learning tasks. Some of the well-known classification algorithms used in this paper are briefly discussed below:

1) `Neural Networks:` The neural network supports both classification and regression algorithms and therefore, is very appropriate for studying the classification problem in this paper.
2) `Discriminant Analysis:` The discriminant analysis is based on the assumption that different classes of data are generated by using different Gaussian distributions. The main types of discriminant algorithms used for classification are the linear and the quadratic discriminant. We used the quadratic discriminant classifier in this paper.
3) `Naive Bayes:` This classification technique is based on Bayes' theorem that assumes independence between predictors, thus, the presence of a particular feature in a class is independence of another feature in a another class. Naive Bayes classification is therefore, based on estimating $P(X|Y)$, the probability or probability density of features X given class Y.
4) K`-Nearest Neighbor:` The KNN algorithm is used for both classification and regression problems. However, the KNN is more widely used in classification problems in the industry and thus will be used in doing classification and predictive analysis in this paper. The KNN is a simple algorithm that stores all available cases and classifies new cases by a majority vote of its k neighbors. The case being assigned to the class is most common amongst its K nearest neighbors measured by a distance function. The common distance functions used are the *Euclidean*, *Manhattan*, *Minkowski* and *Hamming distance*.
5) `Linear Regression:` It is used to estimate real values based on continuous variable(s). In linear regression, a relationship is established between independent and dependent variables by fitting the best line. This best fit line is known as regression line and represented by a linear equation $Y = a * X + b$, where Y is the dependent variable, a is the slope, X is the independent variable and b is the intercept. The coefficients a and b are derived based on minimizing the sum of squared difference of distance between data points and regression line.
6) `Ensemble Learning/method:` An example of ensemble learning method is the TreeBagger, where the bagging stands for bootstrap aggregation. Every tree in the ensemble is grown on an independently drawn sample of input data. To compute the prediction for the ensemble of trees, TreeBagger takes an average of predictions from individual trees (for regression) or takes votes from individual trees (for classification). Ensemble techniques such as bagging combine many weak learners to produce a strong learner.
7) `Decision Trees:` There are two kinds of decision trees; classification trees and regression trees. A decision tree can be described as a flow-chart like structure in which internal node represents test on an attribute, each branch represents outcome of the test and each leaf node represent decision taken after computing all attributes or a response after computing all given attributes.

ISBN: 978-1-4673-9187-0 ©2016 IEEE

III. Related Work

The related work on the application of machine learning and data approaches to study financial data are comprehensively described below. Li *et al.* [6] conducted research on using attributes of customers to assess credit risk by using a weighted-selected attribute bagging method. They benchmarked their result experimentally by using two credit databases and reported outstanding performance both in term of prediction accuracy and stability as compared with another state of the art methods. A data mining approach is also proposed by Moro *et al.* [7] to predict the success or otherwise of a Portuguese retail bank in telemarketing. They applied various data mining models on the bank telemarketing data and reported that the neural network data mining method was the best for analyzing the data. The role of machine learning techniques in business data mining is outlined by [8]. Their work described the strengths and weaknesses of various machine learning techniques within the context of business data mining approach. Their analysis revealed that Rule Induction Technique was the best approach in mining business data, followed by that of the neural network approach. C. Tsai and M. Chen [9] used a hybrid machine learning approach to study credit rating by comparing four different types of hybrid machine learning techniques. They showed experimentally that 'classification + classification' hybrid model based on a combination of logistic regression and neural networks provides the highest prediction accuracy and also maximize the profit. Bank default data was used by [10] to model bank failure predictions using neural network approach. They compared their result with other machine learning approaches and concluded that the neural network approach is a promising method in terms of predictive accuracy, adaptability, and robustness. [11] proposed a generalize switching hybrid recommendation algorithm by combining machine learning classifiers and collaborative filtering recommender systems. They experiment with the hybrid recommendation algorithms on two sets of data and reported high scalability and better performance in terms of accuracy and coverage. A hybrid online sequential extreme learning machine with the simplified hidden layer is proposed by [12]. The algorithm is a combination of the Online Sequential Extreme Learning Machine and the Minimal Resource Allocation Network. Their experimental results showed that the algorithm has a comparable performance as that of the original online sequential extreme learning machine but with a reduced number of hidden layers.

Our approach used in this paper is complimentary but different in many ways. We employed diverse machine learning approaches to predict the creditworthiness of a bank credit data. Comparative analysis is carried out to understand the best fit algorithms for predicting the creditworthiness of a bank credit data. Secondly, we extracted the 5 most important features that determine the credit worthiness of the data. Thirdly, we developed a predictive model using ordinary linear regression approach. These approaches offer a better perspective on doing holistic analysis on bank credit data.

IV. Methodology

In this section, we employed various techniques on the dataset. These techniques are discussed below.

A. Description of the dataset

The dataset used in this paper is taken from the UCI `machine learning data repository` [13] submitted by I-Cheng Yeh. The attributes of the dataset are described in the repository.

The "default payment next month" coded as 'no' or 'yes' in the dataset is treated as the response variable represented as `y` in this paper.

We conducted data exploratory techniques on the dataset in order to understand the nature of the dataset. The exploratory analysis revealed that there seems to be some relationship between Age of the customers, their bank balance and their ability to pay their credit in the following month. Banks customer between the ages of 20 and 60 years with small limited bank balance are seen to be the highest defaulters in paying their bank loans. This is shown in Fig. 2. Further exploratory analysis revealed that the dataset is unbalanced in term of the response variable, for about 6636 (22.12%) records are related with 'yes' (client will default in paying credit in the following month) and 23364 (77.88%) records are related with 'no' (client will not default in paying credit in the following month). The response variable composition is shown Fig. 4. Also, there seems to be no relationship between the 23 feature variables as shown by the correlation plot shown in Fig. 3. Though there is seemingly no pattern or relationship in this dataset, the machine learning approach employed reveals several interesting and important patterns within the data. Thus, machine learning approach/analytics is proven to be an effective tool in extracting hidden pattern and relationships within data.

TABLE I
The composition of the dataset

Bank Credit Data		
Value	**Count**	**Percentage**
No	23364	77.88%
Yes	6636	22.12%
Training Dataset		
No	14092	78.29%
Yes	3908	21.71%
Test dataset		
No	9272	77.27%
Yes	2728	22.73%

TABLE II
Layout of Confusion Matrix

		Predicted class	
		no	*yes*
Actual class	*no*	**TP**	**FP**
	yes	**FN**	**TN**

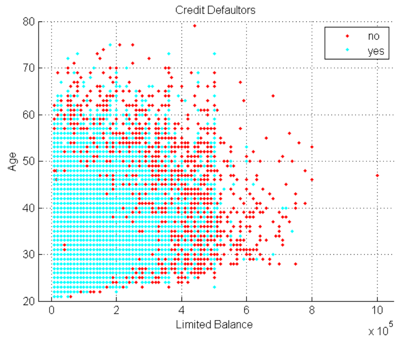

Fig. 2. A scatter plot of Limited Balance, Age and the response variable

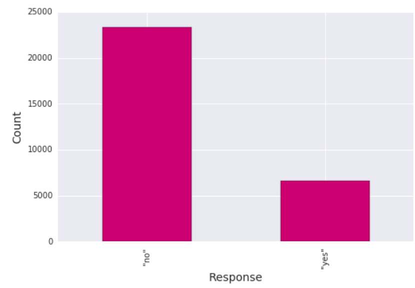

Fig. 4. The responses

B. *The Mathematical Framework*

The training dataset denoted as ω can be described as a matrix with $|\omega|$ instances and each instance has a numeric/categorical feature vector, $\vec{v}_i$., where $i \in \{1, 2, 3, \ldots, |\omega|\}$. The instance vector $\vec{v}$ has a corresponding feature value β_i that is coded as *yes* and *no* to denote credit defaulters and non-defaulters respectively. If T and N denotes the number of clients that will not default the credit payment in the following month and clients that will default in the payment of their credit in the following month respectively, then the total number of the dataset is expressed as $T + N$. Furthermore, TP (True Positive) and TN (True Negative) represent the total positive and negative cases/instances that are rightly classified, respectively. The FP and FN also denote the number of predicted/classified instances that are incorrectly predicted *yes* when it is actually *no* and the number of instances that are predicted *no* when it actually *yes*, respectively. These constitute the entries to the `confusion matrix` shown in Table II and these entries are used to define and evaluate the performance of the machine learning algorithms discussed in this paper. These metrics and their evaluations are shown in Table IV and V.

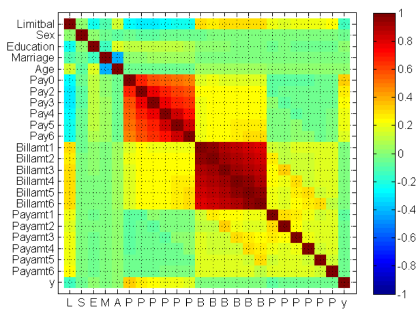

Fig. 3. A correlation plot of the dataset

V. Experiment setup

The major steps we employed in developing the machine learning tasks/algorithms are outlined in Fig. 5. These steps are further discussed below:

- Step 1: *Collect the data*: The dataset used in this paper is from the University of California Irvine (UCI) `machine learning resiportory` [13].
- Step 2: *Prepare the input data*: This step was done by the original owners of the dataset. And the composition of the dataset is shown in Table I.
- Step 3: *Analyze the input data*: The dataset was explored in MATLAB® and Python programming language to understand the relationship (if any) among the 23 different features. A plot of the core features and the entire dataset are shown in Figs. 2, 3 and 4. The dataset is further split into 2/3 for training and 1/3 for testing the algorithms. Furthermore, in order to obtain a representative sample, each class in the full dataset is represented in about the right proportion in both the training and testing datasets. The various proportions of the training and testing datasets used in the paper are shown in Table I.
- Step 4: *Train the algorithm*: The various classification algorithms are trained using a different set of data. The training dataset is shown in Table I
- Step 5: *Test the algorithm*: The various algorithms are used to predict the effectiveness of the algorithm on the test dataset. A 10-fold cross-validation was done on the algorithms to estimate how accurately these algorithms will perform in practice on a different set of data and their corresponding confusion matrices and classification accuracies measured. In evaluating the performance of the classification algorithms, we adopt the commonly used metrics in the literature [3]. They include accuracy, precision, recall, specificity and F-measure (F1-measure). These values are calculated using the Python scikit-learn tool [5] and MATLAB® with input values as the entities of the confusion matrix shown in Fig. II. The formula for the various evaluating metrics is shown in III, with their definitions. In this paper, a 'positive' instance refers to *no*(signifying there will not be a default in the payment of the loan in the next month) whereas the 'negative' instance refers to *yes* (signifying there will be a default in the payment of the loan in the next month). The experimental procedure is depicted in Fig. 5.

A. *Extracting the Importance Features for Predicting Credit Defaulters*

The total number of features within the bank credit defaulters dataset are 23. However, not all the 23 features have significant influence in determining the ability of a given customer in paying his/her loan or not. We, therefore, seek to find out the increasing order of importance

TABLE III
EVALUATION METRICS AND THEIR DEFINITIONS

Metric	Governed Equation	Definition
Accuracy	$\frac{TP+TN}{P+N}$	Proportion of the total number of predictions that are correct
Precision	$\frac{TP}{TP+FP}$	Proportion of the predicted postive cases that are correct
Recall/Sensivity	$\frac{TP}{TP+FN}$	Proportion of positive cases that are correctly identified
Specifity	$\frac{TN}{FP+TN}$	Proportion of negative cases that are correctly identified
F -measure (F1)	$\frac{2\times\text{precision}\times recall}{\text{precision + recall}}$	Weighted harmonic mean of precision and recall

TABLE IV
PERFORMANCE EVALUATION OF THE VARIOUS ALGORITHMS

Ensemble Algorithms					
Algorithm	Accuracy	Precision	Recall	Specificity	F-Measure(F1)
ExtraTree	0.80(+/- 0.02)	0.83	0.7	0.47	0.75
RandomForest	0.81(+/- 0.02)	0.84	0.94	0.37	0.88
RandomForest(with only 5 features)	0.81(+/- 0.02)	0.84	0.94	0.33	0.89
AdaBoost	0.81(+/- 0.02)	0.84	0.96	0.33	0.89
GradientBoosting	0.81(+/- 0.02)	0.84	0.92	0.37	0.88
RandomTreesEmbedding	0.80(+/- 0.02)	0.84	0.92	0.37	0.88
Voting	0.81 (+/-0.02)	0.84	0.94	0.37	0.88
Other Algorithms					
LogisticRegression	0.78(+/- 0.00)	0.78	1	0.0004	0.88
SVM	0.78(+/- 0.00)	0.79	1	0.01	0.88
CART	0.73(+/- 0.02)	0.83	0.81	0.37	0.82
NearestCentroid	0.54(+/- 0.05)	0.85	0.5	0.68	0.63
KNeighbors	0.76(+/- 0.01)	0.8	0.91	0.19	0.85
GaussianNB	0.38(+/- 0.02)	0.84	0.94	0.33	0.88
Neural networks	0.80(+/- 0.01)	0.84	0.92	0.37	0.88
LinearDiscriminantAnalysis	0.81(+/- 0.01)	0.82	0.97	0.24	0.89
Bagging	0.98(+/- 0.02)	0.84	0.94	0.34	0.89

TABLE V
PERFORMANCE EVALUATION OF 5 ALGORITHMS USING ONLY 5 OUT OF THE 23 FEATURES

Using the 5 most important features for Predicting					
Algorithm	Accuracy	Precision	Recall	Specificity	F-Measure(F1)
ExtraTree	0.80(+/- 0.02)	0.83	0.94	0.3	0.88
CART	0.72(+/- 0.02)	0.83	0.83	0.37	0.83
AdaBoost	0.82(+/- 0.02)	0.84	0.96	0.32	0.9
RandomForest	0.80(+/- 0.02)	0.83	0.95	0.3	0.88
LogisticRegression	0.81(+/- 0.02)	0.82	0.98	0.22	0.89

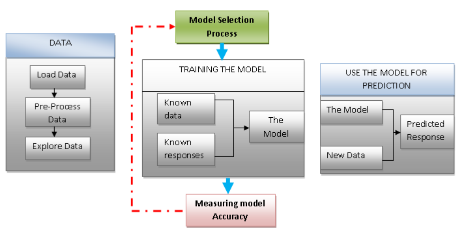

Fig. 5. Work Flow in Machine learning

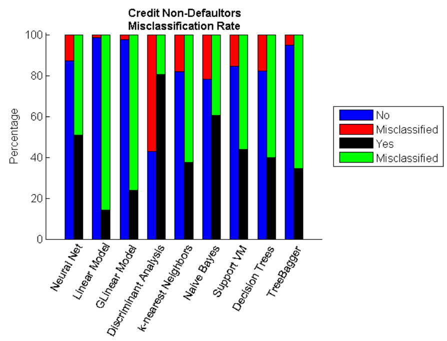

Fig. 6. Classifications Accuracy

of these 23 features. The order of importance of the 23 features is shown in Fig. 7. It is observed that only about 5 features are enough to determine the creditworthiness of a bank customer. To ascertain this, we extracted the 5 most important features from the list and used them on selected machine learning algorithms. The results show no

ISBN: 978-1-4673-9187-0 ©2016 IEEE

significance difference between using all the 23 features and using only the 5 most important features. This can be verified in Tables IV and V. That is, the 5 most important features that can accurately predict whether a customer will default or otherwise in paying his/her loan are the customer/client credit payment history, age, bank limited balance, and history of Bill payment amount.

B. The Predictive Model

We leveraged on the 3 most important features (since these features are the outstanding ones as shown in Fig. 7) to develop a predictive model using ordinary linear regression model. This model is shown in Eqn. 1, where y is the response variable coded as *no* and *yes* in this dataset and been transformed into 1/0 respectively. This can serve as a tool in determining the credit worthiness of bank clients because these are among the main features taken into consideration by most banks in advancing loans to customers.

$$y = 0.4504 + 0.0987 PAY\_0 + 0.0018 AGE - 0.0227(\log(LIMIT\_BAL)) \quad (1)$$

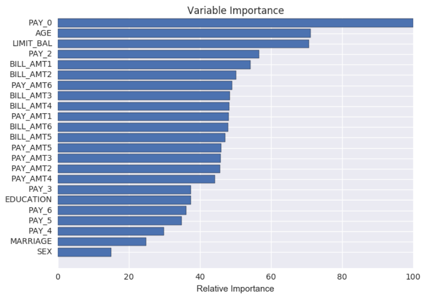

Fig. 7. Relative Importance of each input Attribute/Features

C. Discussion of Results

In this paper, we employed 15 machine learning approaches to study on the bank credit dataset. The algorithms performed relatively well. Apart from the Nearest Centroid and Gaussian Naive Bayes, the rest performed credibly well on the dataset with a prediction accuracy ranging between 76% and 98%. A graphical representation of the prediction accuracy of non-defaulters is shown in Fig. 6. The precision, recall, specificity and the F-measure values are extremely good for all the algorithms. These are shown in Tables IV and V. Thus, these algorithms are very suitable for bank data analytics and prediction of credit non-defaulters in particular. Furthermore, only about 5 features are enough to predict the creditworthiness or otherwise of bank customers. These include the first previous payment status(PAY0), the age of the client(AGE), the amount of money in the person's account at the time of taking the credit(LIMIT_BAL), the second previous payment status(PAY2), and the first amount of bill of statement (BILLAMT1) in the first

VI. Conclusion

In this paper, we applied machine learning approach to study bank credit dataset in order to predict customers' credit worthiness (their ability to pay their loan in the next month). We employed 15 different machine learning algorithms on the dataset in order to determine which algorithms are the best fit for studying bank credit dataset. The experiment revealed that, apart from the Nearest Centroid and Gaussian Naive Bayes, the rest of the algorithms perform credibly well in term of their accuracy and other performance evaluation metrics. Each of these algorithms achieved an accuracy rate between 76% to over 80%. We also determined the most important features that influence the credit worthiness of customers. These most important features are then used on some selected algorithms and their performance accuracy compared with the instance of using all the 23 features. The experimental results showed no significance difference in their predictive accuracy and other metrics. We further formulated a predictive model using linear regression, that composed of the 3 most important features, for predicting customers credit worthiness. These findings have a lot of implications. The model can be used as a tool to advise banks as which factors are important in determining the credit worthiness of customers. Furthermore, the result showed which machine learning algorithms are not suitable for studying bank credit dataset. We intend to develop a hybrid machine learning system that will incorporate the most important features that determine credit worthiness of customers in order to formulate banks' risk automated system.

References

[1] G. McLachlan, K.-A. Do, and C. Ambroise, *Analyzing microarray gene expression data*, vol. 422. John Wiley & Sons, 2005.

[2] E. Ngai, Y. Hu, Y. Wong, Y. Chen, and X. Sun, "The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature," *Decision Support Systems*, vol. 50, no. 3, pp. 559–569, 2011.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. John Wiley & Sons, 2012.

[4] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2005.

[5] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[6] J. Li, H. Wei, and W. Hao, "Weight-selected attribute bagging for credit scoring," *Mathematical Problems in Engineering*, vol. 2013, 2013.

[7] S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems*, vol. 62, pp. 22–31, 2014.

[8] I. Bose and R. K. Mahapatra, "Business data mininga machine learning perspective," *Information & management*, vol. 39, no. 3, pp. 211–225, 2001.

[9] C.-F. Tsai and M.-L. Chen, "Credit rating by hybrid machine learning techniques," *Applied soft computing*, vol. 10, no. 2, pp. 374–380, 2010.

[10] K. Y. Tam and M. Y. Kiang, "Managerial applications of neural networks: the case of bank failure predictions," *Management science*, vol. 38, no. 7, pp. 926–947, 1992.

[11] M. Ghazanfar and A. Prugel-Bennett, "Building switching hybrid recommender system using machine learning classifiers and collaborative filtering," *IAENG International Journal of Computer Science*, vol. 37, no. 3, 2010.

[12] M. Er, L. Zhai, X. Li, and L. San, "A hybrid online sequential extreme learning machine with simplified hidden network," *IAENG International Journal of Computer Science*, vol. 39, no. 1, pp. 1–9, 2012.

[13] M. Lichman, "UCI machine learning repository," 2013.

Acknowledgements

This work is supported in part by the National Science Foundation of China (Grant No. 61133016, 61300191 and 61370026), the National High Technology Joint Research Program of China (Grant No. 2011AA010706), the Ministry of Education - China Mobile Research Foundation (Grant No. MCM20121041), and the Fundamental Research Funds for the Central Universities (Grant No.ZYGX2012J067).

TURKSON Regina Esi had her BSc. degree in Information Technology at the University of Cape Coast in 2011. She is currently pursuing her MSc. degree in Computer Science and Technology in the School of Computer Science and Engineering at the University of Electronic Science and Technology of China (UESTC). Miss Turkson's research interest areas are Machine learning, Artificial Intelligence and Information Security.

BAAGYERE Edward Yellakuor received his BSc. degree (Hons) in Computer Science from the University for Development Studies (UDS), Tamale, Ghana in 2006, and MPhil. degree in Computer Engineering from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana in 2011. Dr. Baagyere received his D.Eng degree in computer science and Technology from the University of Electronic Science and Technology of China (UESTC), in June 2016. Currently, Dr. Baagyere is a post doctoral fellow in the school of Information and Software Engineering at UESTC. His research interests include Social Networks, Computer Arithmetic, Cryptography, Residue Number System and its Applications.

WENYA Gideon Evans was born in Bawku, Ghana, in 1984. He received his B. E. degree in Mathematics Education from the University of Education, Winneba, Ghana, in 2010 , and the M.Eng degree in Electronic Science and Technology in the University of Electronic Science and Technology of China (UESTC) Chengdu, Institute of Fundamental Frontier Sciences(IFFS), China, in 2015. He is currently pursuing Doctoral degree in Electronic Science and Technology in the same university (UESTC, IFFS). His current research interests include calculations of growth and nanofabrication of metal nanoparticles in thin films , Nanomaterials for Energy, which look into Processing, Properties, Structure, Performance thin film materials materials (dielectrics and composites)