Engineering and Technology Journal
Journal homepage: engtechjournal.org

Determination Efficient Classification Algorithm for Credit Card Owners: Comparative Study

Raghad A. Azeez [a] *

[a] University of Baghdad, Collage of Education for Human Science-ibn rushed Baghdad, Iraq
raghad.azeez@ircoedu.uobaghdad.edu.iq

*Corresponding author.

Submitted: 21/01/2020 Accepted: 07/02/2020 Published: 25/03/2021

K E Y W O R D S

Artificial Neural Network, Bayesian, C5.0 algorithm, Classification, Credit Card, Data Mining, Decision Tree, WEKA.

A B S T R A C T

Today in the business world, significant loss can happen when the borrowers ignore paying their loans. Convenient credit-risk management represents a necessity for lending institutions. In most times, some persons prefer to late their monthly payments, otherwise, they may face difficulties in the loan payment process to the financial institution. Mainly, most fiscal organizations are considered managed and refined client classification systems, scanning a valid client from invalid ones. This paper produces the data mining idea, specifically the classification technique of data mining and builds a system of data mining process structure. The credit scoring problem will be applied using the Taiwan bank dataset. Besides that, three classification methods are adopted, Naïve Bayesian, Decision Tree (C5.0), and Artificial Neural Network. These classifiers are implemented in the WEKA machine learning application. The results show that the C5.0 algorithm is the best among them, it achieves 0.93 accuracy rates, 0.94 detection rates, 0.96 precision rates, and 0.95 F-Measure which is higher than Naïve Bayesian and Artificial Neural Network; also, the False Positive Rate in C5.0 algorithm achieves 0.1 which is less than Artificial Neural Network and Naïve Bayesian.

How to cite this article: R. A. Azeez; "Determination Efficient Classification Algorithm for Credit Card Owners: Comparative Study," Engineering and Technology Journal, Vol. 39, Part B, No. 01, pp. 21-29-10, 2021.
DOI: https://doi.org/10.30684/etj.v39i1B.1577

1. Introduction

Credit card fraud is a growing problem that affects cardholders around the world. The credit models include the methods that are called today the techniques of data mining. Classifications are one of the most common goals in the online-based transactional activities that used in data mining techniques and it applied in the domain of credit models to predict the default probabilities of credit holders. Many techniques, such as the nearest neighbor method, decision trees, neural networks, and others have been used for growing credit scoring models. In general, credit card fraud detection has been known as the process of identifying whether transactions are genuine or fraudulent [1].

Disclosing the hidden information in big data via Data Mining techniques has become a new scope and final target for a wide range of future studies. The subject of huge data, banking has been a popular implementation field for researchers with data mining skills over the past decades of the information science revolution. Banks have confirmed that data mining instead of financial resources is the new biggest asset [2]. Data mining is nominated enforcement of particular algorithms that extracting features from data as the following processes:

1) Extract features from the data,
2) Prepare and Preprocess the data,
3) Select the data,
4) Clean the data,
5) And the association of suitable preceding knowledge [3].

An effective role in the data mining having by extract important information used in decision making of a decision support system has been the attractive scope of research in the last two-three decades. Integration of data mining and decision support systems can yield to develop the performance and give new types of solutions for the problems [4]. Artificial neural network techniques are expanding the dimensions of decision support; it furnishes various machine learning techniques to supply data mining.

One of the main methods for building a learning model from data mining is the classification technique. It can propose different types of classification. Algorithms have been determined the predictive based on invisible data. There is no single algorithm found to be best overall others for all data sets. Different types of performance criteria are recommended such as training time, error rate, and predictive accuracy to build the model of data mining, this model must have ability to robustness and scalability are essential [5]. Appropriate interpretation of performance of mining emphasizes that beneficial knowledge originates from the data. Data mining algorithms consist of three components which are [6]:

- Pattern. There are two important relevant points: The mathematical form of the pattern (e.g., Gaussian probability and a linear function of multiple variables) and the technicalities of the patterns (e.g., characterization, and clustering, classification). A pattern contains various factors that are to be confirmed from the data.
- Vantage Criterion: the given data have a great effect on the vantage of one pattern or set of variable factors over another. The vantage is often some form of adaptive force function of the pattern to the data, a smoothing term may be used to avoid overfitting, or creating a pattern with too many degrees of a facility to be compelled by the given data.
- Search Algorithm: The characteristics of an algorithm for finding a specific pattern and variable factor, a suitable evaluation, pattern (or family of patterns), and a given data.

2. Related Work

Cheng L.H., et al. in [7] present two credit datasets in the UCI database and select as the experimental data to demonstrate the accuracy of the SVM classifier, a comparison is done with neural networks, genetic programming, and decision tree classifiers, experimental results show that the accuracy of C4.5 is 87.06%. Cheng Y., et al. in [8] have described payment data from Taiwan bank and the targets were credit cardholders of the bank, he explained the differences between using the evaluation metrics (area ratio and error rate), area ratio can give better results for comparing the accomplishment of different patterns than the error rate, the studies show that there is little variation in error rates among the data mining techniques whose used, while, there are relatively big variations in area ratio among the data mining techniques whose used. It is clear that area ratio is more susceptible and is suitable to measure the classification accuracy of patterns, the accuracy rate of area ratio 55% in training rate and 54% in validation rate in ANN.

R. El-Bialy, et al. in [9] have examined the results from the StatLog project on classification algorithms. This project contains inclusive differences between seventeen algorithms from statistics, neural networks, and symbolic learning on twelve classification tasks. They have accomplished that there is no single optimal algorithm. The optimal algorithm for a specialized dataset depends crucially on attributes of that dataset, such attributes are:

1) Symbolic algorithms are the best option for raising accuracy.

2) Nearest neighbor algorithms are the best option for accuracy and cost.
3)Back-propagation may be chosen as a situation needing enormous machine resources.
4) Bayes algorithms should not usually be handled for classification only if the dataset has a very degraded correlation (near zero) and few other complexities.

Shigeyuki H., et al., in [10] have tested default loan from a database in Taiwan and study the classification ability of three learning models (boosting, bagging, and random forest) with different activation functions, and compare them with eight neural- networks techniques and compute the prediction accuracy of each one, the results show that the classification power of boosting is optimal among the other two learning models, and the performance of machine learning neural-networks depends on the number of middle layers and the choice of the activation function. The maximum accuracy ratio of the original data for the training and testing set is 71.01% and 69.59% respectively. Whereas, the maximum accuracy ratio of the normalized data for the training and testing set is 71.14% and 68.75% respectively.

3. Data Set

The objective of this paper is the credit card owner of Taiwan bank dataset was having 30000 observations [11]. In this research, a binary variable that describes the state of good payment is used to indicate the response state 1 for good payment (Normal) and 0 for delay payment (Abnormal). in addition to the financial information, the dataset has also containing personal information about past bank customers as shown in Table 1. The attribute label of this dataset contains 23364 for 0 (Abnormal), which represents 77.88%, and 6636 for 1(Normal), which represents 22.12%.

F6–F11: The status monthly of payback for each month is filled as follows:

-1: payback on time.
1: late payback for one month.
2: late payback for two months.
3: late payback for three months.
4: late payback for four months.
5: late payback for five months.
6: late payback for six months.
7: late payback for seven months.
8: late payback for eight months.
9: late payback for nine months and above.

4. Classification

We purposed to split the dataset into two parts randomly, one part for training and the other for testing the pattern. Most times, the error rate is used to measure the accuracy of the classification of patterns. The records of the dataset of Taiwan credit card holders are not critical (87.88%), then the error rate is not dangerous to the classification of patterns [8].

Credit register can seem like the type of a classification technique of data mining. Meanwhile, its practical implementations related to many techniques relevant to the credit industry. Due to the hard decision process credit register has always been based on a realistic approach: A solution cannot be the best one for everywhere, only for specific states. The operation of the credit register is not standardized. A real problem with the nonstandard pattern constructing process is purposeless, repeated, and expensive data analysis operations that cannot yet ensure the best model solution [1].

TABLE I: Dataset Description

Feature	Feature Name	Feature Description
F1	LIMIT_BAL	Amount of money in the credit card
F2	Gender	male =1; female = 2
F3	Education	Graduate school = 1; university = 2; high school = 3; others=4.
F4	Married	Marital status (married = 1; single = 2; others= 3)
F5	Age	Age (in year)
F6	PAY_1	Payback status from borrower in September

F7	PAY_2	payback status from borrower in August
F8	PAY_3	payback status from borrower in July
F9	PAY_4	Payback status from borrower in June
F10	PAY_5	Payback status from borrower in May
F11	PAY_6	Payback status from borrower in April
F12	BILL_AMT1	the invoice amount in September
F13	BILL_AMT2	the invoice amount in August
F14	BILL_AMT3	the invoice amount in July
F15	BILL_AMT4	the invoice amount in June
F16	BILL_AMT5	the invoice amount in May
F17	BILL_AMT6	the invoice amount in April
F18	PAY_AMT1	Payback from borrower in September
F19	PAY_AMT2	Payback from borrower in August
F20	PAY_AMT3	Payback from borrower in July
F21	PAY_AMT4	Payback from borrower in June
F22	PAY_AMT5	Payback from borrower in May
F23	PAY_AMT6	Payback from borrower in April
F24	Label	1= Normal, 0 = Abnormal

5. Comparative Model

In this paper, the Taiwan credit card dataset is used for training and testing the patterns to determine the best classification algorithm. A number of operations will be conducted on this data as has been illustrated in Figure 1. Firstly, the data will pre-process. Secondly, select the pre-processed data by competent methods. Finally, classify the selected data, via using the methods of classification field.

I. Preprocess the Data

The gained data from the credit dataset are normalized, which is an important role in preprocessing to improve the performance to obtain the best result. Normalization related to statistical basics; transformed the data to another new scale of specific range between (0 and 1). (Ming L.K.) has pointed out that, some of these normalization methods are the Decimal scaling method, Min-Max method, and standard deviation. In the decimal scaling method, there is some problem when the range is in narrow subinterval; this problem is solved when using Min-Max normalization or Standard Deviation normalization [12]. In this paper, Standard Deviation is applied to describe data and obtained statistical results, see Eq. (1), where x1, x2, x3, …, x_n are the data samples, n=sample size [13].

$$\text{SD}=\sqrt{\frac{\sum_{i=1}^{n}(\bar{x}-x_i)^2}{n-1}} \quad (1)$$

Algorithm (1): Proposed System.
Input: Credit Card Dataset.
Output: An Efficient Classifier.

Begin
Step1: Load the Dataset
Step2: Preprocess the data by (normalized the data via applying the standard deviation) using equation (1).
Step3: Select the features using the random forest method.
Step4: Classify the data which are selected in step (3) by using the following classifiers:
Bayesian classifier via equation (2).
Decision Tree (C5.0).
Artificial Neural Networks ANN
Step5: Extract the confusion matrix for each classifier in step (4).
Step6: Evaluate the following criteria on each classifier
Accuracy rate via equation (3).
Detection rate via equation (4).
Fale Positive Rate via equation (5).
Precision rate via equation (6).
Step8: Check which classifier has the highest rate of accuracy, detection false positive and precision and which classifier have the closest F-Measure rate to1. This is the best classifier.
Step9: End

Figure 1: General Algorithm

II. Feature Selection

To obtain efficient, accurate data reduction, a feature selection method must be used. The selected features have the capability to retain the original meaning of a huge of original data. The random forest has obtained a subset of the dataset which random samplings of variables to create a set of decision trees, the node of the tree is growing according to a limited set of randomly chosen features. When classifying a model, each tree produces its report as a proposal and an overall report is determined by assembling proposals. The concept of using the random forest for feature quality assessment is based on the variation between classifier activity on the original data set and the activity on the changed data set in which the algorithm randomly rearranges values of the observed feature between examples. When measuring the activity before and after the changes of data for each tree in the forest, the algorithm merges these variations into an important assessment [14]. The advantage of using random forest selection gives more stable results when there are more features than examples are obtained from a subset of the dataset [15]. Table 2 obtains the best 10 feature selections using the random forest importance algorithm.

TABLE II: Results of Feature Selection Attributes

Feature name	Importance Attribute
PAY_1	269.039648
BILL_AMT4	82.319835
BILL_AMT5	73.341931
BILL_AMT3	73.043003
PAY_2	72.979171
PAY_6	68.750318
BILL_AMT6	67.537093
BILL_AMT2	65.893180
PAY_4	64.534368
PAY_AMT6	63.567784

III. Data Mining Classification

The aim of classification is to classify a model to a class based on the value of several attributes. Many ways to classification try to explicitly build a function from the common set of values of the attributes to class specifics. An example of such classifiers includes decision trees (C5.0), Bayesian and ANN Artificial Neural Networks. In this study, three algorithms are used, Bayesian, decision tree (C5.0) and ANN Artificial Neural Networks**.**

A. Bayesian Classification

Bayesian classification is used to treat the classification problem by teaching the distribution of models given different class values by approximating the common probability distribution of the class and attributes. After building such an estimator, new models are classified by testing conditional probability given the specific feature values and get back the class which is most probable. The Bayesian formula is in Eq. (2) [16]:

Let $X_1, X_2, ..., X_k$ be events that partition the sample space Ψ,(i.e. $\Psi = X_1 \cup X_2 \cup ... \cup X_k$ and $X_i \cap X_j = \emptyset$ when $i \neq j$) and let Y an event on that space for which Pr(Y) > 0. Then Bayes' theorem is:

$$P(X_j|Y) = \frac{Pr(X_j)Pr(Y|X_j)}{\sum_{j=1}^{K} Pr(X_j)Pr(Y|X_j)} \quad (2)$$

This formula can be used to reverse conditional probabilities. If one knows the probabilities of the events X_j and the conditional probabilities $Pr\ (Y|X_j),\ j = 1, ..., k$, the formula can be used to compute the conditional probabilities $Pr\ (X_j|Y)$.

B. Decision Tree

One of the classification techniques is the decision tree, which is the best choice in most researchers' opinion when the traditional methods are failed in taken a decision. If the trees remain growing without limit, then they take a long time for built and became unintelligible. It is possible to control the size of the tree through:

1) Determine the maximum depth at which the tree can grow.
2) Build a limited number of restrictions in the node without making any external branches.
3) The programmer can interpose to prune the tree by cutting off inconsequential nodes; CART can do this through Cross-checking to see if precision improvements can balance increasingly important nodes.

The decision tree can manipulate non-numeric data in a good manner, this facility accepts critical data, reduce the amount of transient data and explosion predictive natural variables in neural networks.

C. C5.0 Algorithm

This algorithm used the fragmentation concept, according to the common features among samples, and then separate fragments that have maximum information gain. By repeating the fragmented process, it will obtain Subsamples and this process is repeated until the subsamples cannot be fragmented further, finally, if the lowest level of subsample is not divisible, then re-tested these subsamples and, if it is not attractive, then prune out it.

D. Artificial Neural Network (ANN)

When discussing data mining algorithms, at first, it comes to mind Decision Tree or ANN's. Neural networks were the focus of attention during the formation of data mining technology, there is some drawback in using ANN's, like easy to use and spread, but Some of the most important benefits at the top of the list of benefits are models in high-precision prediction, which can be applied to a large number of different types of problems.

In order to retain the genius of biological neural systems, artificial neuron is defined as follows:

1)It receives a number of entries (either from the original data or from other neurotransmitters in the neural network). Each entry comes through a connection that has weight. These weights correspond to the efficiency of the biological neuron and each nerve cell also has a single permitting value (threshold) that constitutes the total input weight. Permitting is allowed, to form the activation of the neuron.

2)Pass the activation signal during the activation function to produce the neuron output. In a similar way to the biological neuron system, when the activation function is used, small changes to the input value sometimes induce significant changes in output, and sometimes significant changes in the input value have an insignificant effect on the output.

6. PERFORMANCE EVALUATION

There are different criterions of performance, which defined as the expression of the confusion matrix variables. These criterions yield some numeric values that are simply distinguishable, and are briefly explained in subsequent paragraphs [17]:

1) Accuracy rate: It is defined as the rate of correctly classified cases and the total number of cases.

$$\text{Accuracy} = (\text{TP}+\text{TN}) \,/\, (\text{TP}+\text{FP}+\text{FN}+\text{TN}) \tag{3}$$

2) Detection Rate (DR) or Recall: It is calculated as the rate between the number of correctly revealed abnormal and the total number of abnormalities.

$$\text{DR}=\text{TP} \,/\, (\text{TP}+\text{FN}) \tag{4}$$

3) False Positive Rate (FPR): It is defined as the rate between the numbers of normal cases revealed as abnormal and the total number of normal cases.

$$\text{FPR} = \text{FP} \,/\, (\text{FP}+\text{TN}) \tag{5}$$

4) Precision (PR): It is the fraction of data cases predicted as positive which are in fact, it is positive.

$$\text{PR} = \text{TP} \,/\, (\text{TP}+\text{FP}) \tag{6}$$

5) F-Measure (FM) or F-score: It is the harmonic average of the precision and recall, where an F1 score reaches its best value at 1 (perfect precision and recall) and worst at 0.

$$\text{FM} = 2 * (\text{Recall} * \text{PR}) \,/\, (\text{Recall}+ \text{PR}) \tag{7}$$

In Eq. (3-7): TP is the number of correctly detected, TN is the number of correctly identified, FP is the number of wrongly identified, FN is the number of wrongly detected.

7. RESULTS

The experiments are implemented using knowledge analysis WEKA data mining tool; it is an open gate Java-based program containing a different set of machine learning algorithms for data mining functions. The algorithms can either be applied directly to a dataset or invited from a Java program. WEKA includes gadgets for data preparation, clustering, regression, classification, association rules, and visualization. It can be used to reveal the cover from paradigms in the dataset and find the most determining factors out of many. WEKA only processes dataset in Attribute-Relation File Format (ARFF) format. Therefore, once the data preparation being done, we convert the dataset into an ARFF file with an extension of ARFF.

Table 3 shows the confusion matrix when we perform the naïve Bayesian algorithm with an 80% training dataset and 20% testing dataset.

TABLE III: Confusion Matrix of Naïve Bayesian Algorithm

Naïve Bayesian		Predicted Class	
		Abnormal	***Normal***
Actual Class	***Abnormal***	3919	344
	Normal	527	1210

Where normal and abnormal represent the good and laggard payments respectively in the label attribute of the dataset. Table 4 shows the confusion matrix when we perform the C5.0 decision tree algorithm with an 80% training dataset and 20% testing dataset.

TABLE IV:Confusion Matrix of C5.0 Algorithm

C5.0		Predicted Class	
		Abnormal	***Normal***
Actual Class	Abnormal	4428	143
	Normal	258	1171

Table 5 shows the confusion matrix when we perform the Artificial Neural Network algorithm with an 80% training dataset and 20% testing dataset.

TABLE V: Confusion Matrix of ANN Algorithm

ANN		Predicted Class	
		Abnormal	***Normal***
Actual Class	Abnormal	4161	273
	Normal	325	1241

According to confusion matrices of Tables 3-5 with performance metric Eq. (3-7), the results are abstracted in Table 6.

TABLE VI:Final Results

Classifier	Accuracy	DR	FPR	PR	FM
Naïve Bayesian	0.8548	0.8814	0.2213	0.9193	0.8999
C5.0	0.9331	0.9449	0.1088	0.9687	0.9505
ANN	0.9033	0.9275	0.1803	0.9384	0.9205

Usually, the FPR metric is better if it does not exceed 10%, as shown in Table 6, the C5.0 algorithm achieved the lowest FPR rate among the other classifiers. Commonly F-Measure metric is better if it's closer to 1, Table 6 indicates that C5.0 achieves a higher value in F-Measure among the other classifiers. Table 7 shows the comparison of accuracy measurement between the proposed work and the related works which are listed in this paper.

TABLE VII: Compression between the proposed system and the related work

	Proposed work	**Cheng L.H[7]**	**Cheng Y[8]**	**Shigeyuki H [10]**
Dataset	***Taiwan***	***Australian***	***Taiwan***	***Taiwan***
Classifier	C5.0	C4.5	ANN	Boosting
Accuracy	0.9331	87.06	0.89	71.01

8. CONCLUSIONS

In this paper, three classifiers are used in data mining and compare the performance of them, there are little differences in all metrics which has been applied, the results show that the C5.0 algorithm has the lowest error rate and high rate of predictive accuracy, detection, precision, and F-Measure among the two other methods. C5.0 algorithm does not exceed 10% in False Positive Rate and it's the lowest among the other classifiers. Also, the C5.0 algorithm is close to 1 in F-Measure metrics. Therefore, the C5.0 algorithm should be employed to score clients instead of other data mining classifiers, such as an Artificial Neural Network, Naïve Bayesian classifiers.

REFERENCES

[1] O. S. Yee, S. Sagadevan, N. H. A. H. Malim, Credit card fraud detection using machine learning as data mining technique, J. Telecommun. Electron. Comput. Eng., 10 (2018) 23–27.

[2] H. Hassani, X. Huang, and E. Silva, Digitalisation and big data mining in banking, Big Data Cogn. Comput., 2 (2018) 18. **https://doi.org/10.3390/bdcc2030018**

[3] K. Walunj Swapnil, H. Yadav Anil, S. Gupta, Big Data: Characteristics, Challenges and Data Mining, Proceedings on International Conference on Advances in Information Technology and Management, 2016.

[4] R. Sahara, Y. Jumaryadi, A. Kartika, "Decision Support System For The Best Teacher Election With Simple Additive Weighting Method Based On Web (Case Study On Al-Ijtihat Vocational School)," development, vol. 5, no. 03, 2018. https://doi.org/10.26562/IRJCS.2018.MRCS10082

[5] P. Kumar, V. K. Sehgal, D. S. Chauhan, A benchmark to select data mining based classification algorithms for business intelligence and decision support systems, Int. j. data min. knowl. manag. process., 2 (2012). https://doi.org/10.5121/ijdkp.2012.2503

[6] P. Fournier-Viger, J. C.W. Lin, R. U. Kiran, Y. S. Koh, R. Thomas, A survey of sequential pattern mining, Data Sci. Pattern Recognit., 1(2017) 54–77.

[7] C.L. Huang, M.C. Chen, C.J. Wang, Credit scoring with a data mining approach based on support vector machines, Expert Syst. Appl., 33 (2007) 847–856. https://doi.org/10.1016/j.eswa.2006.07.007

[8] I.C. Yeh, C. Lien, The comparisons of data mining techniques for the predictive accuracy of probability of default of credit card clients, Expert Syst. Appl., 36 (2009) 2473–2480. https://doi.org/10.1016/j.eswa.2007.12.020

[9] R. El-Bialy, M. A. Salamay, O. H. Karam, M. E. Khalifa, Feature analysis of coronary artery heart disease data sets, Procedia Comput. Sci., 65 (2015) 459–468. https://doi.org/10.1016/j.procs.2015.09.132

[10] S. Hamori, M. Kawai, T. Kume, Y. Murakami, C. Watanabe, Ensemble learning or deep learning? Application to default risk analysis, J. Risk Financ. Manag., 11 (2018) 12. https://doi.org/10.3390/jrfm11010012

[11] Y. S. Lee, S. J. Yen, C. H. Lin, Y. N. Tseng, L. Y. Ma, A data mining approach to constructing probability of default scoring model, in Proceedings of 10th conference on information management and implementation, 2004, pp. 1799–1813.

[12] K. M. Leung, "Preparing the Data." Polytechnic University. Department of Computer Science, Finance and Risk, 2007.

[13] H. Jasiulewicz W. Kordecki, Additive versus multiplicative parameters-applications in economics and finance, arXiv Prepr. arXiv1306.4994,1 (2013).

[14] L. Čehovin, Z. Bosnić, Empirical evaluation of feature selection methods in classification, Intell. data Anal., 14 (2010) 265–281. 10.3233/IDA-2010-0421

[15] Z. Reitermanov´, Information theory methods for Feature selection, Univ. Charles, Prague, (2010).

[16] A. Luoma, "Introduction to Bayesian analysis," Univ. Tampere, Finl., 2014.

[17] A. K. Santra, C. J. Christy, Genetic algorithm and confusion matrix for document clustering, Int. J. Comput. Sci., 9 (2012) 322 - 328.