International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

COPY RIGHT

ELSEVIER
SSRN

IJIEMR Transactions, online available on 31st Mar 2023. Link :http://www.ijiemr.org/downloads.php?vol=Volume-12&issue=Issue 03

10.48047/IJIEMR/V12/ISSUE 03/109

Title **IMPROVE PROFILING BANK CUSTOMER'S BEHAVIOUR USING MACHINE LEARNING**

Volume 12, ISSUE 03, Pages: 772-782

Paper Authors

B Srikanth, Anuhya Kolagani, Kiran Kumar Penugonda, Swapna Nissankarao

Gopi Musugu

USE THIS BARCODE TO ACCESS YOUR ONLINE PAPER

To Secure Your Paper As Per UGC Guidelines We Are Providing A Electronic Bar Code

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

IMPROVE PROFILING BANK CUSTOMER'S BEHAVIOUR USING MACHINE LEARNING

B Srikanth[1], Anuhya Kolagani[2], Kiran Kumar Penugonda[3], Swapna Nissankarao[4] and Gopi Musugu[5]

[1]Professor, Department of Computer Science Engineering, KHIT, Chowdavaram, Guntur.

[2]Student, Department of Computer Science Engineering, KHIT, Chowdavaram, Guntur.

[3]Student, Department of Computer Science Engineering, KHIT, Chowdavaram, Guntur.

[4]Student, Department of Computer Science Engineering, KHIT, Chowdavaram, Guntur.

[5]Student, Department of Computer Science Engineering, KHIT, Chowdavaram, Guntur.

ABSTRACT

The primary goal of this research is to highlight how credit card evolution is a notable development in the banking sector. Every financial system has a sizable dataset for credit card transactions by customers. As a result, banks would require customer profiling as bank customers are aware of the issuer's determinations regarding who should receive banking facilities and what credit limit should be offered. Additionally, it aids issuers in developing a deeper understanding of their existing and potential clients. In earlier studies, customer profiling mostly relied on transaction data or demographic data; however, in the current study, both data are combined to produce a more accurate result and reduce risk. Finding the most effective method improves accuracy and aids banks in growing. By focusing on the important client (businesses), which are regarded as the key engine in the bank's profitability, finding the optimum technique improves accuracy and aids banks in achieving higher profitability. This study attempts to use fuzzy c means, improved k means, neural networks, and the k mean. The primary goal of this study is to develop a new label as a target for neural network classification using the labelled dataset. This will speed up clustering execution and produce the highest accuracy results. The accuracy ratio comparison demonstrates that the neural network is the best clustering technique.

Introduction: In the current era of the banking industry, banks have vast datasets that contain information about their customers and the history of transactions they have made. In order to assess customer behavior and recommend the best course of action to get the most advantages and customer happiness to boost profitability, banks must break these enormous datasets into small clusters. Customer segmentation or customer profiling is utilized to accomplish this goal. Customer profiles created by profiling give banks a detailed account of their consumers based on a number of characteristics. Customer segmentation refers to identifying distinct groups of customers based on either their behavior or certain attributes (such as area, age, and income for demographic segmentation) (for behavioral segmentation) Yet, "profiling" and "client segmentation" are viewed as two sides of the same coin. In order to increase profitability and lower risk, banks must overcome numerous obstacles such default prediction, risk management, client retention, and consumer profiling. In order to overcome these obstacles, it is crucial to accurately identify customers.

The study of machine learning enables computers to act without being instructed. Today, machine learning is so commonplace that you probably utilize it numerous times every day without even realizing it. Machine learning teaches the computer how to learn, how to identify equations and functions that will not only satisfy its current set of examples but also satisfy future ones that it is not yet aware of. Machine learning plays a crucial part in anticipating consumer behavior based on a specific collection of occurrences or patterns that identify their future strategy, preparing to give tailored credit products to the customers. This helps businesses improve connection levels with present customers.

K-mean, enhanced k-mean, fuzzy c-mean, and artificial neural networks are the four machine learning approaches that are used in this study. Their applications are applied to a single real dataset from a Taiwanese bank, and the accuracy ratios of the techniques are compared. Utilized machine learning approaches focus on creating profiles of client behavior.

Profiling: Banks profile their customers by basing their services on factors like their transaction data or their demographic information. Profiling is the method used by the bank to give services to customers (Otherwise called as Segmentation).

Banking: Banking is a method of offering services to bank customers based on profile of the consumers' data.

Machine Learning: In terms of a class of tasks T and a performance measure P, a computer programme is said to learn from experience E if its performance at tasks in T, as measured by P, improves with experience E. Machine learning teaches the computer how to learn, how to create equations and functions that will not only work for the example it has, but will also work in the future for unknown ones.

2. Literature Survey

2.1 According to S. S.-Schwartz et al., who were cited in an article in "theory to algorithms" published by Cambridge, machine learning is one of the fields of computer science that is expanding the fastest and has the broadest range of applications. This textbook's objective is to provide a principled introduction to machine learning and the algorithmic paradigms it offers. The theoretical foundations of machine learning are described in the book, together with the mathematical derivations that turn these foundations into useful algorithms. Among them are a discussion of the convexity and stability concepts and the computational complexity of learning, as well as significant algorithmic paradigms like stochastic gradient descent, neural networks, and structured output learning, as well as recently developed theoretical ideas like the PAC-Bayes approach and compression-based bounds.

The text makes the fundamentals and techniques of machine learning understandable to students and non-expert readers in statistics, computer science, mathematics, and engineering. It is designed for advanced undergraduates or beginning graduates.

2.2 Using two steps and k-means clustering methods, Majid Sharahi et al. constructed a classification model for the dataset of Sepah

Bank Branches in Tehran in 2015. A type of demographic and behavioral segmentation was performed on 60 Sepah Bank clients in order to find the most devoted clients.

2.3 With 18 datasets from the UCI repository, Ali Arshad, S. Raiz, and L. Jiao et al. published a multi-class classification model in 2019. Clustering semi-supervised data was accomplished using semi-supervised deep fuzzy c-mean (DFCM-MC). By using fuzzy c-means, they added a new label to the unlabeled data. With the new label that retrieved the discriminatory information utilized for classification, they used the labelled data (supervised data) and the unlabeled data (unsupervised data). The DFCM-accuracy MC's rate was 80 82%, and its f-measure was 78 16%.

2.4 A Bansal, M. Sharma, and S. Goel presented a tweak to the k-means algorithm's clustering model. This adjustment is founded on normalization. The Cancer Dataset was used by the researcher to discover the findings. Although the original data were highly dimensional, only five qualities were ultimately taken into account due to limitations. This study demonstrated that the accuracy rate for the current algorithm was 57 14%, whereas the accuracy rate for the revised approach was 92 86%.

2.5 N. V. Dharwadkar and P. S. Patil: A prediction and classification model for two datasets of data from bank customers was built in 2017. In this model, they employed an Artificial Neural Network (ANN), after which they weighted the outcomes. By using the proposed model and the ANN algorithm, it is demonstrated that the ANN algorithm is effective for the two datasets. For dataset1 and dataset2, this algorithm's accuracy rate was 72% and 98%, respectively.

2.6 Using five clustering methods, Shenghui Yang and H Zhang produced a classification model for the credit card default data set at the Taiwanese bank. To determine the average area under the curve (AUC) and the proper rate of the model, 10-fold cross-validation was performed. The Microsoft Company's Lite GBM (high-performance Gradient Boosting framework) has the most accuracy. The accuracy ratio by F1-measure for the Lite GBM model was equal to 89 34%.

2.7 According to S Hyang and H Zhang et al, artificial neural networks continuously construct meaningful correlations between input and output using nonlinear mathematical equations. output variables obtained through the learning process Back propagation neural networks employ feed forward topology and supervised learning to classify data. A back propagation network's structure typically comprises of an input layer, one or more hidden layers, and an output layer, with each layer made up of multiple neurons. The non-linearities and interactions of the explanatory factors can be handled by artificial neural networks with ease. Artificial neural networks' primary drawback is that they cannot produce the results of a straightforward classification probability calculation. SVM is a statistical learning theory-based pattern recognition technique that uses the kernel function to map the input space's data X into a high-dimensional feature space, and then at high-level After obtaining the generalized optimal classification surface in the dimensions space, it is possible to linearly classify the data that was linearly indivisible in the original space in the high dimensional space.

2.8 According to J. Wang and Su Xialong et al. Data clustering is the process of organizing the data into groups based on their intrinsic characteristics; each group's elements should have a similar nature. The disparity between clusters ought to be as great as possible. Although numerous data cluster algorithms, like the K-means method, the Db scan algorithm, the Wave-cluster algorithm, etc., have been presented for a while, however, all of these algorithms are inefficient when collecting data on a group of people, which severely restricts their use in related domains.

K-mean clustering algorithm is one of many clustering algorithms that can be used, including task decomposition of heterogeneous neural network structure, pre-processing of system modelling with radial basis function networks, and image and audio data compressing.

2.9 B K Singh, K Verma, and other Feature extraction and feature normalization is a crucial pre-processing approach that is typically used before classification, according to A S Thoke et al. The process of feature normalization is helpful for limiting the values of all features to specific ranges. Since applying normalization to the input could modify the structure of data and affect the results of multivariate analysis and calibration used in data mining and pattern recognition challenges, choosing the right normalization approach and normalization range is a crucial issue. Here, some common feature normalization strategies are examined and evaluated, and their effects on classifier performance are studied with reference to the classification of breast tumors using ultrasound images. Back-propagation artificial neural network [BPANN] and support vector machine are used to evaluate the feature normalization strategies. Models for [SVM] classifiers are employed. The classification accuracy is demonstrated by the results to be significantly impacted by the normalization of features. According to statistics, breast cancer is the second most common cancer among women. As its frequency has increased over the past few years, it has emerged as one of the most pressing health concerns. Breast cancer's early identification is crucial for lowering death rates because the disease's causes are yet unknown. In addition to early detection, a precise and trustworthy diagnosis that can differentiate between benign and malignant tumors is crucial for preserving human life. Computer assisted diagnosis (CAD) has developed as a technique that has been clinically demonstrated to help doctors discover and diagnose breast cancer. Important steps in CAD systems include feature extraction, feature normalization, and feature classification. Both positive and negative effects of normalization on classifier performance may exist.

3. Existing System

- Auto Regression:

A time-varying process is represented by an autoregressive (AR) model, which is used to describe some time-varying processes in nature, economics, etc. The stochastic difference equation is the result of the autoregressive model's specification that the output variable relies linearly on its own prior values and on a stochastic term (an imperfectly predictable term).

- Vector Auto Regression:

A VAR model only considers previous values when describing the evolution of a set of k variables over the same sample period (t = 1, T). The variables are gathered in the vector ak * 1 vector yt, whose i th member is the value of the GDP at time t. The formula for a pth order VAR, abbreviated as VAR(p), is

yt = A3 yt3 + A2 et2 + Aet1 + et

where the l-periods back observation yt1 is referred to as the lth lag of y, c is a k*1 vector of constants (intercepts), Ai is a time-invariant.

4.Proposed System

Our suggested model's major goal is to use various machine learning approaches to better profile bank customers' behavior. The data set used to create this model was collected from the UCI machine learning repository. Following that, data goes via the data pre-processing stage. After that, the consumer profile is built using machine learning algorithms.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

The profiling stage of machine learning identifies the elements in a collection and assigns them to target categories. This examines the accuracy rate of techniques using the Gini coefficient for the unsupervised techniques, uses the results as input for the supervised technique (Artificial Neural Network), and compares the results to determine which technique is most accurate.

In order to anticipate future events, supervised machine learning algorithms can use labelled examples to apply what they have learned in the past to fresh data. The learning algorithm creates an inferred function from the study of a known training data set in order to forecast the output values. The learning process can also identify flaws in the output and compare it to the intended, correct output to adjust the model as necessary.

Contrarily, unsupervised machine learning methods are utilized when the training data is neither categorical nor labelled. Unsupervised learning investigates how systems might extrapolate a function from unlabeled data to describe a hidden structure. Although the system is unable to determine the proper output, it explores the data and can use data sets to infer hidden structures from unlabeled data. With this design, large amounts of data may be analyzed. While it typically produces quicker, more accurate findings to spot lucrative opportunities or risky situations, it may also need more time and resources to be properly trained. Processing massive amounts of data can be improved by combining machine learning with neural networks and cognitive technologies.

This is incredibly helpful since it enables programmers to design software with a certain goal and purpose in mind instead of having to pay attention to every step of how it does that. Machine learning enables software to be goal-oriented and also enables machine developed processes that programmers may not have even considered. Machines have a much wider scope of data processing capability than humans do, and can organize and scan data for important information much more quickly than any person could. It produces more effective software in addition to more helpful software.

The proposed system's workflow is displayed in order to better profile bank customers' behavior utilizing various machine learning approaches. The data set used to create this model was collected from the UCI machine learning repository. The process framework focuses on bank customer data and later applies pre-processing techniques to the data. Processing of the data is done with the help of demographic data, which includes some variables like education, limit balance, marital status, and behavioral data, which includes customer transactions.

The following unsupervised methods are employed:

1) K-Means algorithm
2) A K-Means algorithm improvement
3) Fuzzy C-Means Algorithm

Artificial neural networks are employed as a supervised learning technique.

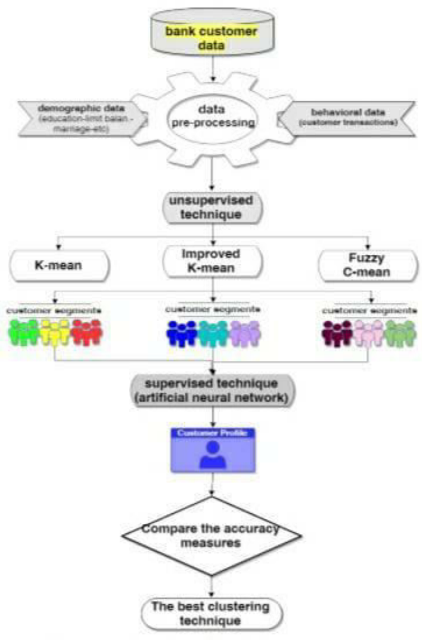


Figure1: Proposed frame work for profiling of the data

4.1 Methods used for profiling

a) Un supervised learning techniques:

K-Means algorithm: Because of its stability and Mac Queen's proposed simplicity, the K-mean clustering technique has been one of the most often utilised procedures for years. A partition-based cluster analysis technique was developed in 1967 using the K-Means clustering algorithm. K-means conduct partition of items into clusters that are “similar” amongst them and “dissimilar” to the objects belongs to another cluster. The K-means technique offers superior efficiency and scalability and converges quickly when working with huge data sets, which is why it is extensively employed in cluster analysis. When you have unlabelled data, you can utilise a type of unsupervised learning called K-means clustering (i.e., data without defined categories or groups).

Improved K-Means Algorithm: The k-means clustering algorithm was improved since it can automatically determine the number of clusters and assign the necessary cluster to unclustered points. The suggested modification enables excellent accuracy and shortens the clustering time for the cluster member. a more effective dissimilarity-based k-means clustering technique.

Fuzzy C-Means Algorithm: Each data point can be a member of multiple clusters in fuzzy clustering, which is also known as soft clustering. Data points may potentially belong to more than one cluster in fuzzy clustering. The Fuzzy C means Clustering (FCM) Algorithm is one of the most popular fuzzy clustering algorithms. J.C. Dunn created (FCM) clustering in 1973, and J.C. Bezdek enhanced it in 1981. The algorithm ignores noise and outliers in favour of optimising clustering or centroid computation.

b) Supervised learning technique:

Artificial Neural Networks: A neural network is occasionally a streamlined model of how the human brain processes information. The neural network functions by resembling the internal connections that exist between neurons. Threshold logic, developed by Warren McCulloch and Walter Pitts in 1943, is a computer neural network model based on mathematics and algorithms. This paradigm set the path for the division of neural network research into two schools of thought. The focus of one strategy was on the biological functions of the brain, whereas the focus of the other was on the use of neural networks in artificial intelligence. The term "ANN model" is frequently used to define a class of these functions (where members of the class are obtained by varying parameters, connection weights, or specifics of the architecture such as the number of neurons or their connectivity). With this technology, it is possible to build a variety of various structures based on

- Number of layers
- Selection of activation function
- The number of perceptron’s
- Normalization layers
- Dropout adjustments

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

5. Modules

Broadly speaking, this study proposes 2 modules. Every module is implemented in order to identify spam or ham emails.
As follows:
5.1 Pre-processing of Data
5.2 Classification using algorithms for machine learning
5.1 Data Pre-processing
The first crucial step in the data mining process is data pre-treatment. When there is a lot of noise and unreliable data, or if there is a lot of irrelevant and extraneous information present, the analysis of the data may not yield accurate results if it has not been thoroughly examined for these issues. Consequently, before doing the analysis, the data's quality and representation are crucial. Data preprocessing has frequently been the most crucial stage of our machine learning research. First, the database confirms the normalization procedure. To conveniently compare data from various sources, most problems require first eliminating the units of measurement for the data before normalizing it. One of the most used techniques for normalizing data is:

- Re-scaling: - Re-scaling data to have values between 0 and 1. This is usually called feature scaling One possible formula to achieve this is

$$Xnew = \frac{\text{X-Xmin}}{\text{Xmax-Xmin}}$$

5.2 Data Classification using machine learning algorithms
The final training set is produced as a result of data preprocessing. then using the final training set to apply the four machine learning approaches. The K-means algorithm was used as the first technique. On the basis of the researcher's prior information, the number of clusters is established. So, in this work, the researcher identified three groups. The second classifier uses an enhanced version of k-mean to establish that there are five clusters.

1.use the intra-cluster distance measure, defined as the average of all distances between a point and its cluster center

$$\text{intra}=\frac{1}{N}\sum_{k}^{i=1}\sum_{C_i}||x-Z_i||^2$$

Where
N is the total number of pixels in the image,
K is the total number of clusters, and
Zi is the center of cluster Ci.
We certainly wish to reduce this action.
2.The following step is to scale back this measure. measuring the maximum feasible inter-cluster distance, or the distance between clusters Then, using the formulas
Use the smaller of these two values, to determine the distance between cluster centers.

$$inter = min(\|zi \text{ -}zj\|^2), i = 1,2,\ldots,$$
$$k-1 \text{ and } j = i+1,..,k$$

Where,
cluster centers are zi' and zj
K is the number of clusters
3. Only taking the minimum of this value, the smallest of this distance to be maximized, and the other larger values will automatically be bigger than this value
4. Finally, calculate the ratio of inter and intra which defined as validity:

$$\text{Validty}=\frac{Intra}{Inter}$$

5. Thus, clustering, which provides a minimum value for the validity measure, informs us of the optimal value of K. (number of clusters). A fuzzy c-mean classifier, applied to the data set using five clusters, is the third classifier.
The following step is to determine which of the three unsupervised methods has the highest Gini coefficient and the best accuracy for profiling the dataset.
The outputs of the unsupervised techniques are then applied to a neural network as a target in order to determine its accuracy. We add a new label to the dataset by using our K-means, improved

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

K-means, and fuzzy C-means results as targets. After that, test them out and determine their correctness by comparing seven accuracy measures. The
highest accuracy classifiers are the best for enhancing bank clients' profiling.

5.1.1 Dataset to be used

In this paper we are going to use "default of credit card clients" which is obtained from the archive of the UCI (University of California, Irvine) Machine Learning Repository

The data set consists of:

30000 observations and 23 variables and there are no missing values in it All explanatory variables were normalized.

Attribute no.	Attribute name	Description
X1	Limit_ BAL	Amount of the given credit (NT dollar)
X2	Sex	Gender (1 = male; 2 = female).
X3	Education	Education (1 = graduate school; 2 = university; 3 = high school; 4 = others).
X4	Marital status	Marital status (1 = married; 2 = single; 3 = others).
X5	Age	Age (year).
X6-X11	Pay_0 to Pay_6	April to September
X12-X17	Bill_AMT1 to BILL_AMT6	Amount of bill statement (NT dollar)
X18-X23	Pay_AMT1 to PAY_AMT6	Amount of previous payment (NT dollar)
X24	Y	Default payment (Yes = 1, No = 0)

Table1: represents the default of credit card clients data attributes and their description

5.1.2 Analysis and comparison

Performance of classification utilizing various unsupervised machine learning classifier counts. Gini coefficient was then used to calculate accuracy. Using this new label as a target for the artificial neural network method and using these results as a new label for the dataset as opposed to the previous label to carry out the following phase of our experiment.

Machine learning technique	Best Gini Obtained	Rank
Unsupervised (K-Means)	26.37%	3
Unsupervised (Improved K-Means)	37.61%	1
Fuzzy C-Means	29.04%	2

Table 2 describes the results of applying the unsupervised three techniques on the dataset after evaluating the performance with Gini co-efficient. It shows that improved k-means are the best accuracy technique equal to 37.61%.

5.1.3 Evaluation Metrics

Accuracy =(TP+TN)/(TP+FP+TN+FN)

Sensitivity=TP/(TP+FN)

Specificity =TN/(TN+FP)

Precision =TP/(TP+FP)

Recall =TP/(TP+FN)

F-measure F= 2*P*R/(P+R)

where,

• True Positive (TP): Observation is positive, and it is projected to be positive.
False Negative (FN): Positive observation with an expected negative result.
• True Negative (TN): The observation and prediction are both negative.
False Positive (FP): When an observation is negative but a positive result is predicted.

International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

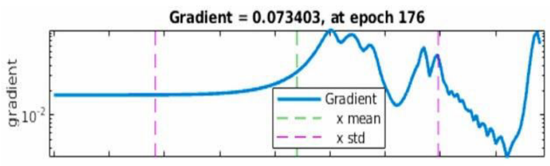
TABLE 3. The evaluation of the proposed neural network model.

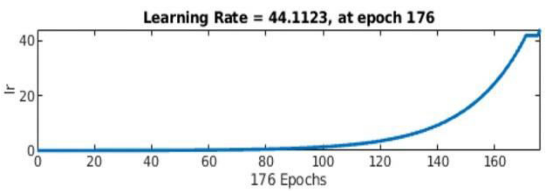
Measure	Value
Accuracy rate	0.9808
Sensitivity	0.9777
Specificity	0.9816
Precision	0.9275
Recall	0.9777
F-Measure	0.9519
G-mean	0.9796

6. Neural network evaluation

During this stage, the neural network algorithm will use the enhanced k-mean clustering method's highly accurate output as its aim. According to table 3's results, the neural network had the highest accuracy rate when it came to categorizing the dataset. Hence, by using unsupervised machine learning techniques to generate a new label, we were able to accomplish the goal of this experiment, which was to improve the profile of bank customer’s behavior.

The gradient coefficient varies with relation to the number of epochs, as seen in the figures below. The test is terminated at epoch number 176 after the mistakes have occurred six times after epoch number 170, as indicated in the Figure. The gradient coefficient's ultimate value at epoch 176 is 0.073403, which is roughly close to zero. Network training and testing will benefit from the gradient coefficient having a minimal value.





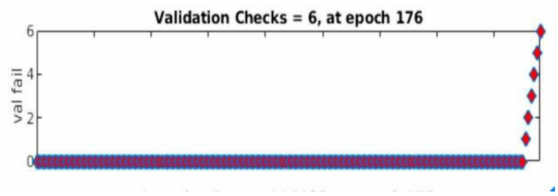
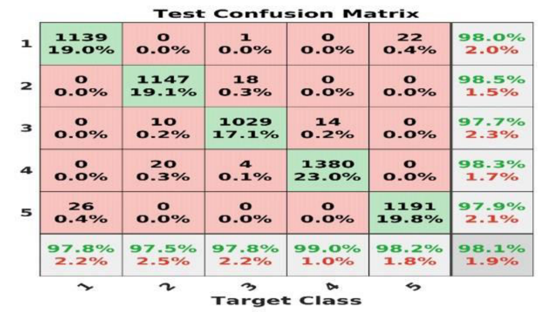


Fig6: (1)(2)(3) Training state of the proposed model

The confusion matrix, which is depicted in Figure 4, is a table that illustrates how well our classification model, also known as the "classifier ANN," performed on a set of test data in order to display the true values. This matrix allowed for the visualisation of the algorithmic performance. It results in an effortless determination of class confusion. This confusion matrix is used to determine the performance metrics.



7. Results

TABLE 4. The clusters result from the proposed neural network model.

Cluster	Cluster name	N. customer	Details
1	Platinum	5765	Top class
2	Golden	5580	2nd rank
3	Bronze	5171	3rd rank
4	Silver	6832	4th rank
5	Classic	5858	5th rank

7.1 Results of earlier founding researches

The accuracy rate for the neural network in Matlab is 98.08% after scanning the confusion matrix of the network. There are five clusters, each with a distinct amount of clients, according to this confusion matrix. We may categorise them as

TABLE 5. the results of the earlier founding researches:

Name of researcher	Date of publishing	Used technique	Accuracy rate
Shenghui Yang[22]	Sept. 2018	neural network	88.83
Sharjeel Imtiaz[23]	2017	neural network	90.99
Furrakh Shahzad[24]	2017	Neural network (MLP)	81.7
Vladislav Pyzhov[25]	May 2017	neural network	81.1

8. Conclusion

Banks have been able to develop interactive relationships based on humanistic experience and trust thanks to profiling. Using clustering algorithms, massive data sets are divided into clusters. Proposed. The accuracy level and computation time required to cluster the data set, which are the two main shortcomings of K-Means clustering, have been eliminated by modification. To ensure effective and efficient segmentation of the bank's customer base and to help build its service and product offerings to achieve customer loyalty and satisfaction, detailed analysis of the profiling environment should be made. The supervised machine learning showed high accurate results of profiling than the unsupervised technique by creating a new label target for the data set. The artificial neural network showed the highest accuracy by seven different measures. So that any bank in the future can use this model and technique to improve profiling of its customer, get high profitability, and reduce the risk.

References

[1] S. Ben-David and S. S.-Schwartz. From theory to algorithms, machine learning is understood. 2014; Cambridge University Press.

[2] M. Aligholi and M. Sharahi. Classify the Data of Bank Clients Using Data Mining and Clustering Approaches. February 11, 2015, Journal of Applied Environmental and Biological Sciences.

[3] "Customer segmentation based on CLV model and neural network," by M. Ayoubi. IJCSI 13.2 (2016): 31. International Journal of Computer Science Issues.

[4] "Customer Profiling Using Classifier Method for Bank Telemarketing," S. Palaniappan, A. Mustapha, et al. International Journal on Informatics Visualization, Volume 14, No 2, 2017, Pages 214–217.

[5] M. Sharma, S. Goel, and A. Bansal. K-means clustering method improved for data mining classification approach prediction analysis. 157.6 (2017): 0975-8887 International Journal of Computer Applications.

[6] P. S. Patil and N. V. Dharwadkar, "Analysis of banking data using machine learning," 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics, and Cloud) (I-SMAC), Palladam, pp. 876-881.

[7] "Comparison of Various Data Mining Techniques in Credit Card Default Prediction," S. Yang and H. Zhang. 10.05 (2018): 115 for Intelligent Information Management.

[8] M. A. I. Navid and N. H. Niloy. Naive Bayesian Classifier with Classification Trees for the Accurate Prediction of the Chance of Default Credit Card Customers. 3.1 (2018) of the American Journal of Data Mining and Knowledge Discovery: 1.

[9] S. Riaz, L. Jiao, and A. Arshad. "Deep Fuzzy CMean Clustering under Semi-Supervision for Imbalanced Multi-class Classification." Access IEEE (2019).

[10] Ahram Online, "Egypt's Bank Misr conditionally suspends card usage abroad amid currency crisis," -Economy-Business–Ahramonline. [Online].Easily accessible at:

English.Ahram.org.eg/News/246079.aspx. [Accessed:10-Apr2019]

[11] N. M. El Agroudy, F. A. Shafiq, and S. Mokhtar. "The effect of the rise in the dollar rate on the Egyptian economy." 509–514 in Science 5.02 (2015).

[12] H. Hassan and A. Jreisat, "Is bank productivity important? Egypt as an example 6.2 (2016): 473-478 International Journal of Economics and Financial Problems.

[13] T. Hafez, ''IN DEPTH-The ups and downs of the Egyptian pound,'' AmCham. [Retrieved: 09-April-2019]

[14] "Artificial intelligence and decision support systems," T. Perraju 2.4 (2013): 17-26 – International Journal of Advanced Research in IT and Engineering.

[15] International Journal of Innovative Research in Science, Engineering, and Technology article by M. and N. Kaur titled "Adaptive K-Means Clustering Approaches For Data Clustering" (2014).

[16] J.Wang and Su. Xiaolong "An improved K-Means clustering algorithm." Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on. IEEE, 2011.

[17] F. BASER, S. GOKTEN, and P. O. GOKTEN. "Using fuzzy c means clustering algorithm in financial health scoring." Audit Financier 15.147 (2017): 385-394.

[18] S. Deb, "Application of Artificial Neural Networks (ANN)-In Developing SODEPUS (Study of Dynamic Earth Processes using Software),"

[19] Credit card clients' default data set, UCI machine learning repository.

[20] B. K. Singh, K. Verma, and A. S. Thompson. investigations of the effectiveness of feature normalization methods in classifying breast tumors. the 116.19 issue of International Journal of Computer Applications (2015).

[21] Rose H. Turi and S. Ray. "Determining the number of clusters in the K-means clustering and using it in color image segmentation." The 4th international conference on advancements in pattern recognition and digital approaches proceedings were published in 1999.

[22] S. Yang as well as H. Zhang. The study is titled "Comparison of Several Data Mining Techniques in Credit Card Default Prediction." 10.05 (2018z): 115 for Intelligent Information Management.

[23] A Better Comparative Overview of Credit Scoring Classification, S. Imtiaz and A. J. Brimicombe, 23. 1-4 in International Journal of Advanced Computer Science and Applications 8.7 (2017).

[24] Performance comparison of data mining techniques for the predicted accuracy of credit card defaulters, M. Pasha et al., Int. J. Comp. Sci. Netw. Secure 17.3 (2017): 178–183.

[25] "Comparison of methods of data mining approaches for the predicted accuracy," V. and S. Pyzhov. (2017)