An Approach of Supervised and Unsupervised Machine Learning Model for E-CRM Bank's Marketing

Abubakr Mustafa Eltahir[1], Professor. Tarig Mohamed Ahmed[2], Dr. Amgad Atta Abdelmageed Mohammed[3]*
Dr. Anwer Mustafa Mohamedsalih Hilal [4]
, Dr. Tarig Abdalkarim Abdalfadil[5],

1- Department of Information Technology, Faculty of Computer and Information Technology, University of Alnileen, Khartoum, Sudan. Numo Alaali for Training and Education, KSA.
2- Department of Information Technology, Faculty of Computing and Information Technology, King Abdul-Aziz University, Jeddah, Saudi Arabia.
3- Assistant Professor: Department of Computer and Self Development, Prince Sattam bin Abdulaziz University, Al-Kharj, Saudi Arabia *.
Assistant Professor: Ahfad University for Women School of Management Studies.
4- Faculty of Computer Science and Information Technology, Omdurman Islamic University, Omdurman, Sudan
5- Department of Computer Science, University of Alnileen, Khartoum, Sudan.

***Corresponding Author:**
Dr. Amgad Atta Abdelmageed Mohammed

Summary
This paper displays a combination data mining algorithm within a Bank's database system, which is started by providing a data mining application that characterizes a solution for one of the problems that exist in the bank's database systems, which is creating a new electronic means to increase the effectiveness of the mechanism to attract the community of depositors and their resources to banks. After that, it fits into the bank system to perform the solution inside the bank's database system. This method is then repeated every day at a specific time to improve and enhance the bank system by providing data mining methods that support different departments in the bank, particularly in the Marketing department and Decision-Makers employees, to execute a specific marketing decision. A hybrid model, of unsupervised machine learning, was designed, represented by the (K. means) algorithm and its data outputs were used in the supervising machine learning represented by the (multi-class decision jungle) algorithm, to determine which, cluster the customer belongs to and whether he will subscribe in a term deposit or not, and how much he will participate in deposit or not participate in a term deposit. which provides these applications with a conceptual and scientific approach to link the resulting application in the bank's database system, it is worth noting that this experience can be used in bank databases with data mining technologies. The researcher found that; data of data algorithms can be applied to the bank system as a smart banking system and to adapt data mining algorithms to any database system, regardless of the system's modern or old, in addition to that data mining algorithms contribute to producing important results that contribute effectively to make effective strategic decisions. In the conduct of the banking process.

Keywords:
Data Mining; Bank system; Database; combination; two-class boosted decision tree; Term deposit.

1. Introduction

In the business world, we found that banks play a vital role in the flow of cash from customers and there are challenges in banks to attract capital money, there are many ways and strategies that drive and motivate the customer to subscribe to a term deposit, and banks must work on the strategies that would seek to improve customer service and polarization. One of these strategies is the use of data mining techniques, which are now one of the best tools for data processing and results from extraction that turns into knowledge with the help of banking system experts and contribute to decision-making. We found that lately data mining is used in the field of business and banking to track customer behaviour and predict their potential operations of buying and selling stocks and subscribing to term deposits and other banking operations using certain questions such as what are the attractive facilities that make the customer participate in a term deposit, and what are the specifications the customer candidate to subscribe to a term deposit According to his education Level, Age, Job, Contact, Campaign, Duration, Education, Marital Status…etc.

We found that the use of data mining in the world of banks has found greater interest in light of the competition between banks which is seeking to attract capital and strive

Manuscript received April 5, 2022
Manuscript revised April 20, 2022
https://doi.org/10.22937/IJCSNS.2022.22.4.73

to serve the customers above what is expected by identifying the customer specifications and taking data (Attributes) from him/her that would help in the analysis and discover patterns that work to design offers that satisfy the customer and make him loyal to the bank.

This paper attempts to integrate data mining applications into the bank system, as it provides a thesis to integrate data mining algorithms in solving bank problems and then suggests a way to integrate the proposed solution with the existing bank system to attract customers and work to satisfy them and their loyalty.

We found that; there are many studies, projects and experiments that have been conducted in the field of analyzing the data of input, and renewed banks to investigate deeper into tracking and understanding the behaviour of the customer. To be served in a manner that satisfies the customer. Furthermore, to make him a loyal owner of the bank, and predicts the behaviour of his future attitude. For example: (B & Pooja, 2021), (Obiedat, 2020), (Krishna & Reddy, 2019b), (Krishna & Reddy, 2019a), (A.Elsalamony, 2014) argued that; predicting a customer subscribes to a term deposit or not, which is determined by (Yes / No) using data mining algorithms, that give distinguished accuracy in predicting customer's behaviour. However, (Raju et al., 2014), (Chitra & Subashini, 2013) authors of this studies provided clarification and enlightenment for data mining and its role in improving customer relationship management CRM, in the bank and making the customer loyal and retain during strong competition. Also, (Dinçer et al., 2019) there are studies conducted using the data mining approach on some Turkish banks to identify the level of customer satisfaction with the use of mobile banking applications and subscriptions in bank deposits.

In these studies, (Vrontis, 2022), (Hassani et al., 2018), (Abbas, 2015), (Baratzadeh et al., 2022) The authors wrote, about the importance of data mining in the era of big data, and its impact on the accuracy of the results. Discovering customer behaviour, and urging them to take advantage of banking services and support decision-makers, with useful information help them in their campaigns marketing and other sub-activities in banks.

Various banking applications are using data mining algorithms, that helped detect credit card fraud, and thus, the banks avoided losing millions of dollars, (Carneiro et al., 2017). We found that; data mining technology and applications have been used in banks, to solve the problem of loan defaulting remain, by predicting the future behaviour, of the customer (Nguyen, 2019). Moreover, we found that; the data mining algorithms were used to detect phishing criminals and unethical activities, targeting the banking system. For example, creating fabricated web pages exploited the increasing digital use by bank customers (Kanchana, M., Chavan, P., & Johari, 2020). This paper contributes to the field of E-CRM and its application in the banking system through the following methodology:

1- Create and adapt data mining algorithms with banking system database systems to serve the banking system and facilitate its electronic flow of operations.

2- Finding a framework approach that supports and develops the process of integrating data mining techniques with banking system databases.

3- Finding suitable solutions by analyzing databases of the banking system to predict if the customers subscribe to a term deposit or not.

4- Integrate solutions in the banking system.

2. Academic-State-of-the-Art:

Figure (1) is organized as a table to demonstrate the previous studies in seven-dimension, the Author, year of publication, the sample size, factors, algorithms used, Results and Recommendation.

Author	Year	Sample size	Target class	Factors	Algorithms used	Results and Recommendation
(Can et al., 2016)	DECEMBER 2016	41181 records	target Clients that would most likely make term deposit purchases	Twenty-one Factors	1\ Logistic Regression 2\ Binary Decision Tree	1\Running the Logistic Regression Model with all significant variables, we found that variables age, housing, campaign 2\ largest impact on odds ratios are euribor3m_cat and nr_employed_cat. **Recommendation:** It would be an interesting study to determine why. there is such a different impact in the levels of these social & economic variables.
(Abbas, 2015)	January 2015	4521 records	predict the deposit of the customer.	Seventeen Factors	1\ decision tree (DT). 2\ rough set theory(RST)	1\ The gain ratios showed that the "Duration "age "balance "features have the maximum gain ratio. 2\ decision tree also provides a huge number of decision and predictive rules. 3\ The gain ratios showed that the "The Duration "feature has the maximum gain ratio. 4\" age" feature was the 8th. 5\"balance" was the 10th.
(Krishna, 2019)	April 2019	1000 records	predict whether a client Deep subscribes fora term deposit.	Twenty-one Factors	1\ Decision Tree 2\ Naïve Bayes 3\ Support Vector Machines 4\ k-Nearest Neighbors 5\ Deep Neural Network Classifier	found that our proposed DNN classifier outperforms the other classifiers in terms of accuracy. **Recommendation:** improving the accuracy of the Deep Neural Network classifier

Figure (1) demonstrates the previous studies.

Author	Year	Sample size	Target class	Factors	Algorithms used	Results and Recommendation
(Farooqi et al., n.d)	Jul 2019	41,188 records	study is on the way to prediction of term deposit subscription	20 input variables	1\ Decision Tree 2\Neural Network 3\Naive Bayes 4\ k-nearest Neighbor	1\ The overall better performance achieved by J48 decision tree which reported 91.2% 2\ data mining can be further used in to getting information form a vast set of data warehouse Recommendation: The experimental setup had be applied to other massive live banking data sets.
(Femina & Sudheep, 2015)	2015	45,211 records	To predict the behavior of customers to enhance the decision-making processes for retaining valued customers	16 input variables	1\. Naive Bayes 2\ Neural Networks	1\ MLPNN classifier model shows better accuracy (88.63%) among the two models experimented 2\ NB gives better values, ROC area (0.858). Recommendation: The experimental set up can be applying to other huge live banking datasets.
(Asare-frempong, 2017)	November 2017	45,147 records	1\ predict customer reaction to bank direct marketing. 2\ To determine the main factors of customers who have subscribed and are likely to later subscribe to term deposits	17 input variables	1\ Neural Network 2\ Decision Tree 3\ Logistic Regression 4\ Random Forest .	1\ Results from the study describe that Random Forest algorithm with an accuracy of 87% is the greatest classifier in terms of predictive capability. 2\ customers on whom more call durations who spent on present a higher probability of subscribing to the term deposit. 3\ customers with a minimum of secondary school education are good prospects to be targeted for the bank term deposit subscriptions.

Figures (2) demonstrates the previous studies

3. The Current Practice of the Bank's System:

In this study, we will conduct experiments, on the data of a Portuguese bank, in a marketing campaign, carried out by the bank. The bank's customers were contacted through e-mail, phone, and cell phone calls to urge them to subscribe to a term deposit, and with that, the study aims to predict whether customers will subscribe to a term deposit or not. Therefore, the total dataset used consisted of 41,181 records with 21 attributes. In general, 10 Attributes has selected for the study.
Table 1: Showed data that are extracted from a bank dataset (Can et al., 2016) after that we analyzed it and organized in the following table.

Table 1: **Bank marketing dataset**

A. No.	Attribute name	Attribute category	Description	Attribute Type
1	age	Client Data	Client's age at the time of the call	Continuous
2	job	Client Data	Client's type of job	Categorical
3	marital	Client Data	Client's Marital Status at time of the call	Categorical
4	education	Client Data	Client's educational background at the time of the call	Categorical
5	default	Client Data	Does the client have credit in default?	Categorical
6	housing	Client Data	Does the client have a house loan?	Categorical
7	loan	Client Data	Does the client have a personal loan?	Categorical
8	contact	Last Contact Info	Communication type with client	Categorical
9	month	Last Contact Info	Last contact month of the year with client	Categorical
10	day_of_week	Last Contact Info	Last contact day of the week with the client	Categorical
11	Duration	Last Contact Info	Last contact duration, in seconds to Client	Continuous
12	campaign	Other	The number of contacts made during this campaign for this client	Continuous
13	plays	Other	Number of days that passed by after the client was last contacted from a previous campaign	Continuous
14	previous	Other	Number of contacts performed before this campaign and for this client	Continuous
15	outcome	Other	Results of the previous marketing campaign	Categorical
16	emp_var_rate	Social & Economic	Employment Indicator of Variation Rate - Quarterly	Continuous

17	cons_price_idx	Social & Economic	Monthly Index of consumer prices	Continuous
18	cons_conf_idx	Social & Economic	Consumer Confidence Index – Monthly indicator	Continuous
19	nr_employed	Social & Economic	The three-month Euribor rate – Daily indicator; Euribor is short for Euro Interbank Offered Rat	Continuous
20	nr_employed	Social & Economic	Number of Employees – Quarterly indicator; Number of employed persons for a quarter.	Continuous
21	Y	Target/Response	Has the client subscribed a term deposit? - 'yes','no'	Categorical/ Binary

This paper took eight attributes to conduct in data mining Algorithms using Microsoft Machine Learning Studio (Azure Machine Learning designer) to analyze the dataset and data selected shown.

Table (2): Table (2) displays the Data that has been selected in our study.

A. No.	Attribute name	Attribute category	Description	Attribute Type
1	age	Client Data	Clients' age at the time of the call	Continuous
2	job	Client Data	Clients' type of job	Categorical
3	marital	Client Data	Clients' Marital Status at the time of the call	Categorical
4	education	Client Data	Clients' educational background at the time of the call	Categorical
5	default	Client Data	Does the client have credit in default?	Categorical

4. The Methodology:

In this study, we follow the experimental approach, in describing the construction of an automated Customer Relationship Management E-CRM model, which is integrated with the bank statute, that uses data mining algorithms, that select the influencing factors for the nomination of the customer, to subscribe to a term deposit. While employing the comparison approach, between the two systems as well, as taking advantage of the deductive approach, to build a model where it contributes to the development of banks, through their historical development.

In addition, to benefit from the research tools as required by the research stages of the inductive approach and the case study approach. To achieve the above objective, a dataset of a Portuguese bank was collected, which was collected during a marketing campaign, and then pre-processing data to get it free from errors, duplications and inconsistencies, and replace missing values to obtain clean data and thus the appropriate attributes. Then comes the process of entering the attributes selected from the dataset to enter the algorithms. At this stage, the experiments have been conducted. By using unsupervised machine learning k-means and data mining algorithms. Then the output of the unsupervised machine learning model was represented in the (k-means) algorithm the quality of each cluster was adjusted, and then a dataset was developed and used in the machine learning subject to supervision, and the model was trained on all algorithms with multi-class, taking the highest accuracy, three algorithms have achieved the highest level of accuracy, multi-class decision tree, multi-class decision forest, and multi-class decision Jungle, multi-class neural network to obtain a classification which the customer will subscribe a term deposit or not according to the Attributes (Factors) that were taken from the customer.

So the aim of the process we want to classify customers using a data mining model to predict and candidate the customer that more probable to subscribe to a term deposit, based on the customer's Factors. The figure below demonstrates the idea of the combination of Supervisor and unsupervised machine learning.

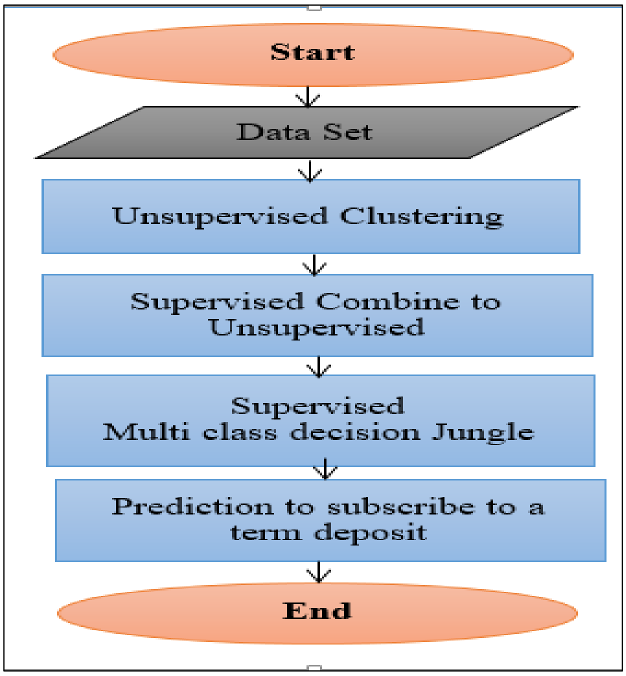


Figure (3) the data flow diagram of the idea of supervised combined with Unsupervised.

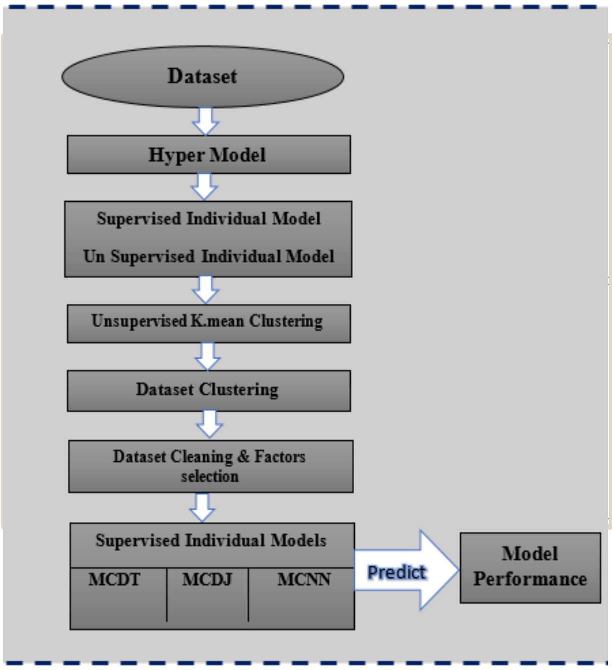


Figure (4): demonstrating modelling system of integration of unsupervised and supervised machine.

Figure (4) determining the experimental process of combining the unsupervised machine learning & supervised machine and the model performance.

5. A Novel Framework for E-CRM:

Transforming the model of the traditional CRM system, into an automated system integrated with the bank statute E-CRM, by contributing to reducing cost, increasing efficiency, reducing time-consuming, simplifying procedures, and selecting the required technology tools for work. This section describes the process of integrating data mining algorithms with the banking databases system in two ways.

1- The existing Databases in the bank system.

2- The process of integrating data mining algorithms into the banking system. The main banking system data flow diagram quoted from a previous study that is used in the current banking system, as shown in Figure 4 showed the proposed process that the research wants to classify customers using a data mining model to predict and candidate the customer that more probable to subscribe in a term deposit. The proposed model contains ECRM to classify customers and predict that customers will subscribe to a term deposit according to their factors.

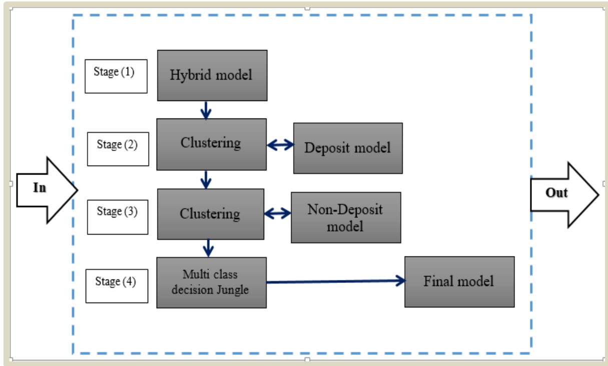


Figure (5) determining the experimental process of combining the unsupervised machine learning & supervised machine and the model performance.

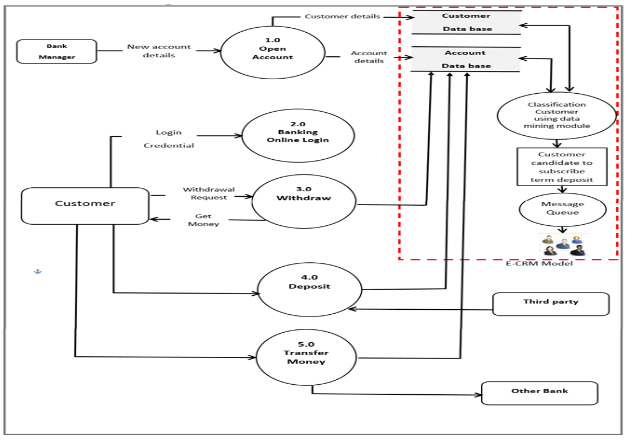


Figure (6) shows the study Framework: An Approach of Supervised and Unsupervised Machine Learning Model for E-CRM Bank system.

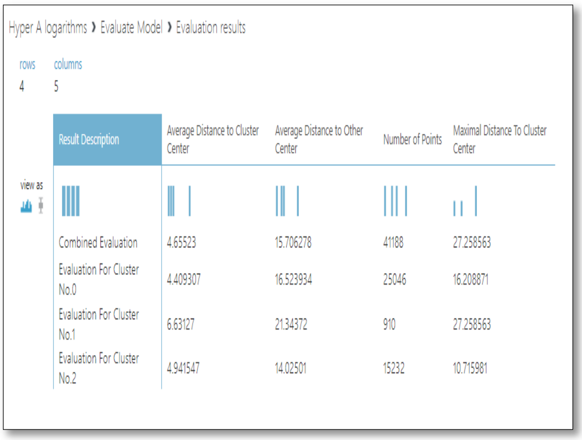
The proposed main banking system and E-CRM model. Shows the process

we want to classify the customers using the data-mining model to predict customers, which are potential to be subscribed in a term deposit.

6. The Experiment:

This section presents the experiment of the suggested solution, the experiment will be conducted in a marketing campaign, the dataset has been taken from a Portuguese bank, the experiment will analyze, examine, and predict the customer who are potential to subscribe to a term deposit and who is not. Applying unsupervised machine learning k- means clustering and the output of the dataset from clustering. Using to predict supervised machine learning, Multi-class Decision Forest, Multi-class Decision Jungle, Multi-class logistic regression, Multi-class neural network, and this experiment can integrate one algorithm into one application with a user interface and advice decision-makers that the customer will subscribe to a deposit who are not. In addition, it sends messages via automated media of communication to attract customers to subscribe to a term deposit. Furthermore, this test will be conducted inside the bank's system. Then maybe apply to all banking systems, the development of the last application follows the algorithm that is shown in figure (6) above. In addition, the researcher used an algorithm K-means to distinguish customers who are similar in a certain characteristic and put them in clustered groups. So, that it is easier to deal with them, according to the marketing strategies, followed until it becomes clear, to decision-makers and their vision is clear to them, regarding the customers most expected to subscribe in a term deposit. Then, focus on them, in the bank's upcoming marketing campaigns. After that, clustering algorithms look for an increase in the similarity and decrease in the multi-dimensional distance, among dataset locations, in the same cluster. Firstly, the researchers clean the dataset, so that it is in an appropriate structure for clustering. Secondly, we divide the dataset into three different clusters, using K-Means Clustering. Thirdly, in this step, we examine our model results when we right-click the Train clustering model. We can visualize the assigned data to a certain cluster either cluster (0), cluster (1), or cluster (2). As shown in the cluster diagrams below. In the following example, you will see three clusters in figure (7).

Clustering is the greatest famous algorithm form of unsupervised machine learning, we do not have any labels in clustering, only a set of factors for reflection and your goal is to create clusters that have similar interpretations clubbed together and dissimilar observations were as far as possible. Assessing the performance of a clustering algorithm is not as trivial as calculating the number of mistakes or the precision and recall as in the briefcase of supervised machine learning algorithms.


Hyper A logarithms › Evaluate Model › Evaluation results

rows 4 columns 5

Result Description	Average Distance to Cluster Center	Average Distance to Other Center	Number of Points	Maximal Distance To Cluster Center
Combined Evaluation	4.65523	15.706278	41188	27.258563
Evaluation For Cluster No.0	4.409307	16.523934	25046	16.208871
Evaluation For Cluster No.1	6.63127	21.34372	910	27.258563
Evaluation For Cluster No.2	4.941547	14.02501	15232	10.715981

Figure (7) displays the assigned dataset to three clusters Based on the similarity and dissimilarity

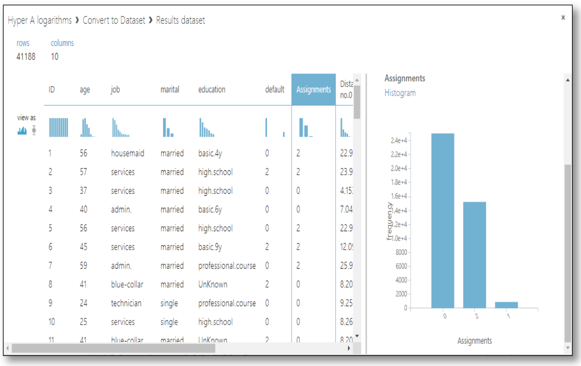
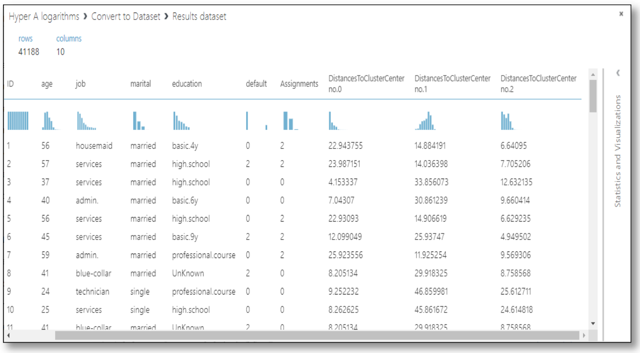


Figure (8) exhibit evaluate the model and evaluate the result and average distance to the cluster centre

The above figure presents the results of the three Clusters and shows that (Cluster 0) with an average distance to the cluster centre (4.4) is the closest to the lowest average, and it contains the largest number of customers wishing to subscribe to a term deposit and those who do not want it, except for both. Next is (Cluster 2) with an average distance to the cluster centre (4.9) and then the last (Cluster 1) with an average distance to the cluster centre (6.6).

In the following figure, the basic data is displayed with its assignment in the particular cluster based on the proximity of the data to the centre of the group, as the closest distance from the centre of the cluster is included in it, In the sense of the customer whose characteristics are similar to the cluster, he is closer to it in terms of distance, and therefore is belong to the cluster.


Hyper A logarithms › Convert to Dataset › Results dataset

rows 41188 columns 10

ID	age	job	marital	education	default	Assignments	DistancesToClusterCenter no.0	DistancesToClusterCenter no.1	DistancesToClusterCenter no.2
1	56	housemaid	married	basic.4y	0	2	22.943755	14.884191	6.64095
2	57	services	married	high.school	2	2	23.987151	14.036398	7.705206
3	37	services	married	high.school	0	0	4.153337	33.856073	12.632135
4	40	admin.	married	basic.6y	0	0	7.04307	30.861239	9.660414
5	56	services	married	high.school	0	2	22.93093	14.906619	6.629235
6	45	services	married	basic.9y	2	2	12.099049	25.93747	4.949502
7	59	admin.	married	professional.course	0	2	25.923556	11.925254	9.569306
8	41	blue-collar	married	UnKnown	2	0	8.205134	29.918325	8.758568
9	24	technician	single	professional.course	0	0	9.252232	46.859981	25.612711
10	25	services	single	high.school	0	0	8.262625	45.861672	24.614818
11	41	blue-collar	married	UnKnown	2	0	8.205134	29.918325	8.758568

Statistics and Visualizations

Figure (9) present assign customer in a certain cluster & distances to the cluster

- The above screenshot displays the customer dataset as follows (age =56),(Job = housemaid),(marital status = married),(education = basic.4y),(default = 0) and this customer has been assigned to the cluster (2) and it has been assigned according to the nearest distance from the centre of the cluster(2), where cluster (2) was the distance to cluster centre was (6.6), While we find that the distance of the customer dataset from distance to cluster centre (0) was (22.9), As for the customer dataset from distance to cluster centre (1) was (14.8).

- By looking at the distance of the customer's dataset to the center of the three clusters, we find that it is closer to cluster (2), concerning its nearness to cluster (2), so the customer was assigned to this cluster (2).

- Figure (9) shows the clusters chart and shows that the cluster (0) was the most fortunate in assigning customer data with similar characteristics.

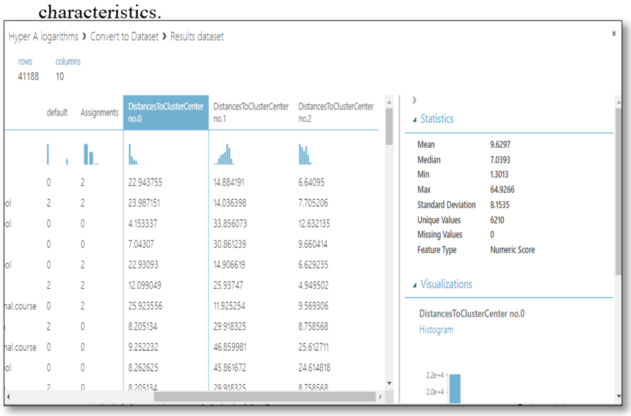


Figure (10) presented assignments diagram.

By entering the details of the dimensions of the distances between the dataset points and then from the cluster centre, we found that the total mean of the distance to cluster centre no (0) was (9.6297) shown in figure 11. furthermore, the total mean of the distance to cluster centre no (1) was (31.005) as shown in figure 13, additionally the total mean of the distance to cluster centre no (2) was (12.3471) as shown in figure 15

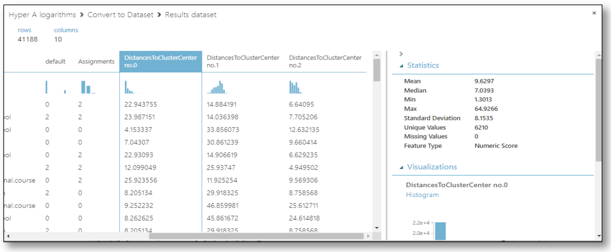


Figure (11) demonstrates the mean of the distances to cluster centre (0)

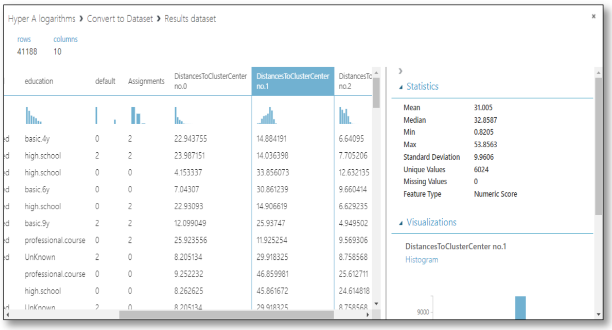
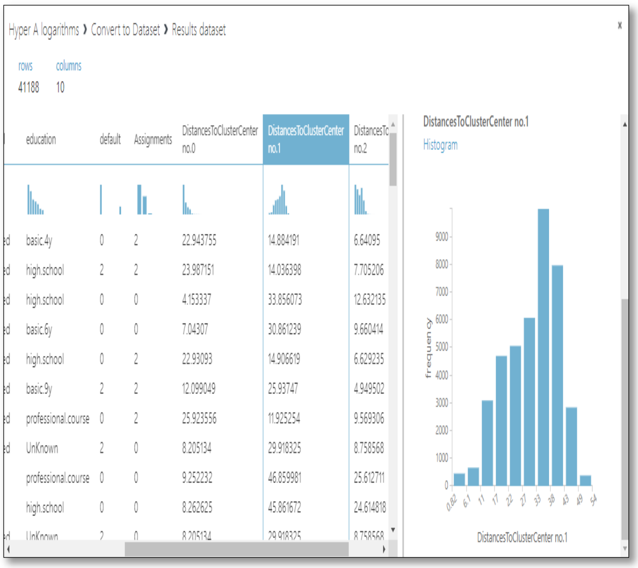


Figure (13) explains the mean of the distances to cluster centre (1)



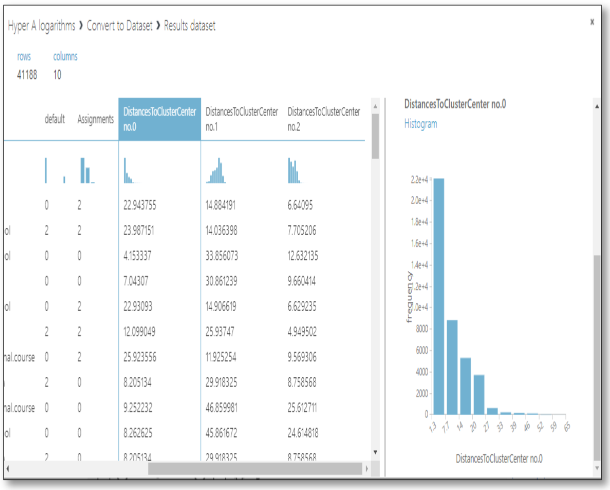


Figure (12) demonstrates the mean of the distances to cluster centre (0) diagram. **Figure (14)** reveals the mean of the distances to the cluster centre (1) diagram.

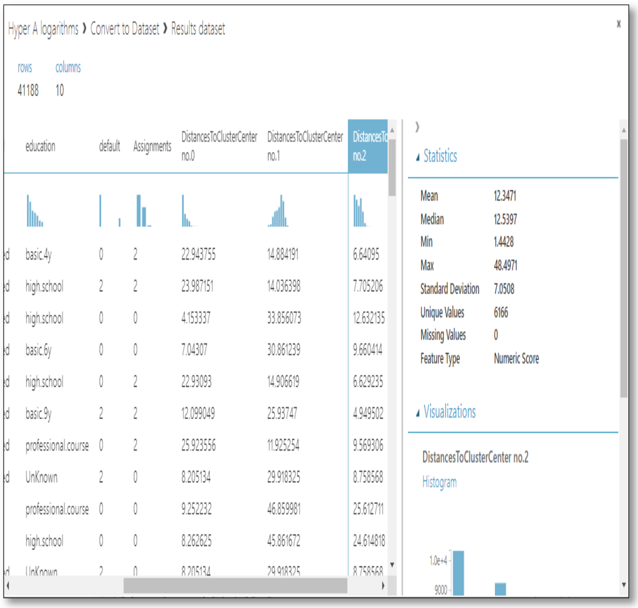


Figure (12) demonstrates the mean of the distances to cluster centre (0) diagram.

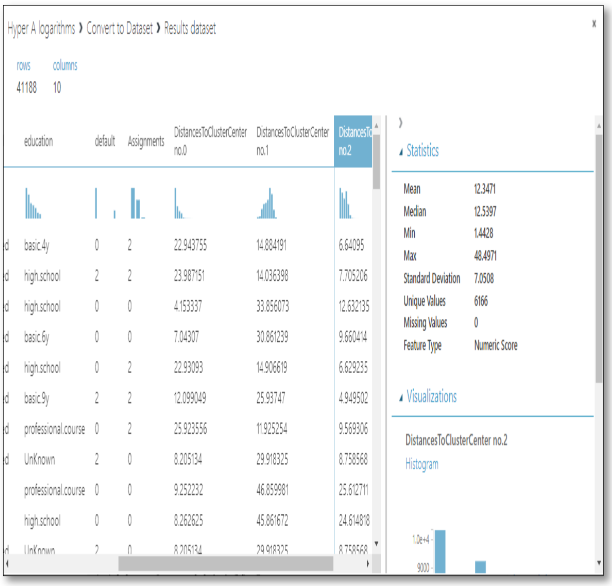


Figure (15) illuminates the mean of the distances to the cluster centre (2).

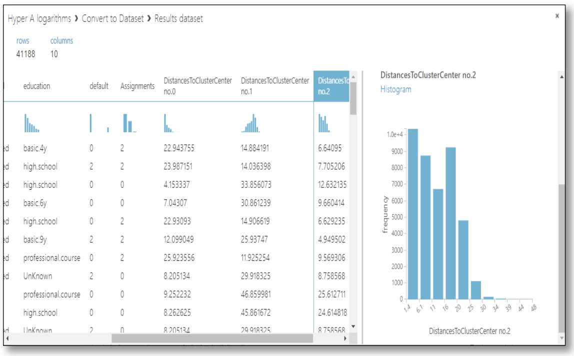


Figure (16) expresses the mean of the distances to the cluster centre (2) diagram.

Customers were divided into three clusters according to their similar characteristics clusters (0,1,2), and then clusters were checked in terms of quality and the researcher found that (cluster 0) is the best in terms of whether the customer deposited or not, then (cluster 2) and (cluster 1). The customers were classified into clusters and it was determined whether the customer deposited a deposit or not made a deposit based on the original data, and from here we have improved data by recognizing the distance from the centroid of the cluster, from the unsupervised machine learning data output (k-means) the average accuracy of the calculation was calculated Whether the customer deposited a deposit at a certain percentage, or did he not deposit a deposit with a specific percentage. This is the new column in the improved dataset that was used in the prediction process in the new model of supervised machine learning using multi-class algorithms

The following figures show the steps for implementing the model design, the graph, and the outputs of the model results.

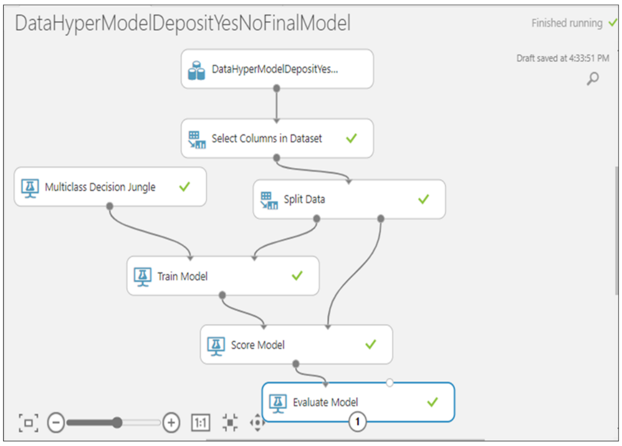
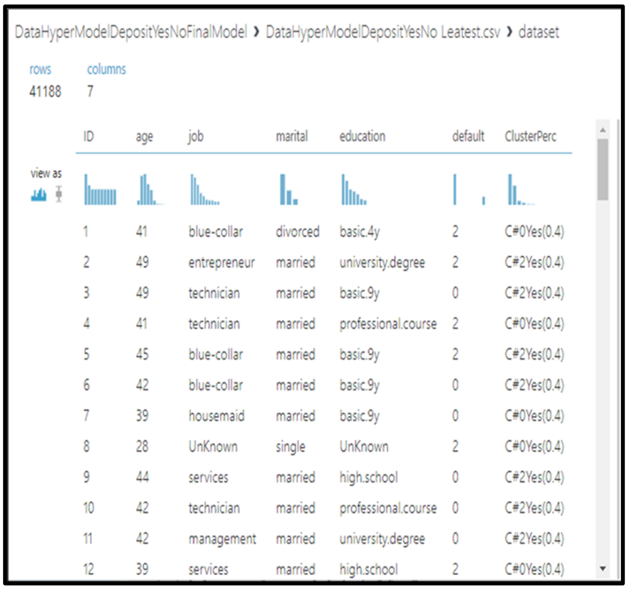


Figure (17) demonstrates the diagram of dataset in the model of Multi class algorithms.

The model operates by supervised machine learning, and multi-class algorithms and the model has been trained on the (multi-class decision jungle) algorithm.

The following figure displays the data used in the model, which was the real dataset, plus the output dataset of running an algorithm model (K. means) in the last column, which contains a determination that the customer belongs to any cluster, and that the customer will subscribe in a term deposit or not and the average percentage of his subscription in the deposit or the average percentage Not subscribing in the deposit



ID	age	job	marital	education	default	ClusterPerc
1	41	blue-collar	divorced	basic.4y	2	C#0Yes(0.4)
2	49	entrepreneur	married	university.degree	2	C#2Yes(0.4)
3	49	technician	married	basic.9y	0	C#2Yes(0.4)
4	41	technician	married	professional.course	2	C#0Yes(0.4)
5	45	blue-collar	married	basic.9y	2	C#2Yes(0.4)
6	42	blue-collar	married	basic.9y	0	C#2Yes(0.4)
7	39	housemaid	married	basic.9y	0	C#0Yes(0.4)
8	28	UnKnown	single	UnKnown	2	C#0Yes(0.4)
9	44	services	married	high.school	0	C#2Yes(0.4)
10	42	technician	married	professional.course	0	C#2Yes(0.4)
11	42	management	married	university.degree	0	C#2Yes(0.4)
12	39	services	married	high.school	2	C#0Yes(0.4)

Figure (18) determine the real dataset including the column of cluster assign and average percentage

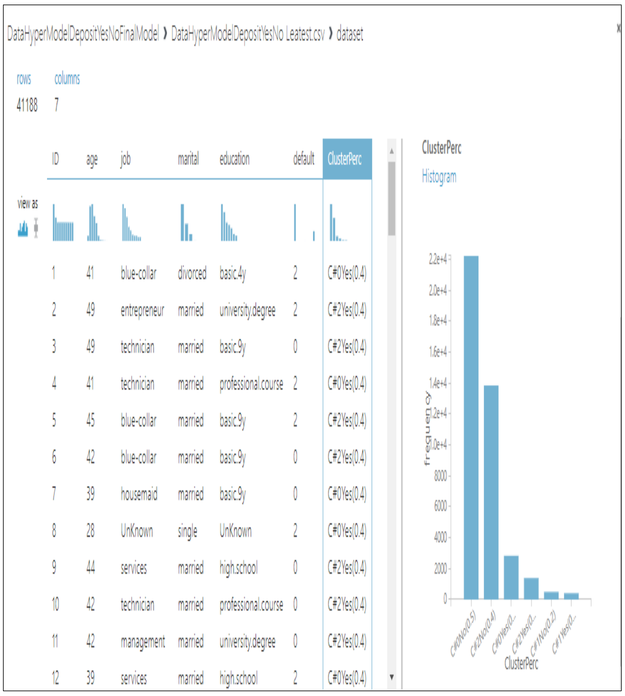


Figure (19) define the real dataset including the column of cluster assign and average percentage diagram.

7. Result and Discussions:

The results of the model demonstrated great percentages; The highest accuracy was 92%, which is considered a high percentage because we were able to determine what the customer's cluster is and whether he will subscribe or not, and what is the average percentage of his subscription or the average percentage of non-subscription to focus on effective customers and target them by the bank's marketing decision makers.

Table demonstrates the accuracy results of the different algorithms used in this model.

Table (5) explain the data mining algorithms

Techniques	Overall Accuracy	Micro averaged Precision	Micro averaged Recall
Multi-class Decision Forest	0.92	0.92	0.92
Multi-class Decision Jungle	0.92	0.92	0.92
Multi-class logistic regression	0.90	0.90	0.90
Multi-class neural network	0.92	0.92	0.92

In the flowing figures, we demonstrate the confusion metrics and classification reposts of all the algorithms that were conducted in the experiment in the model.

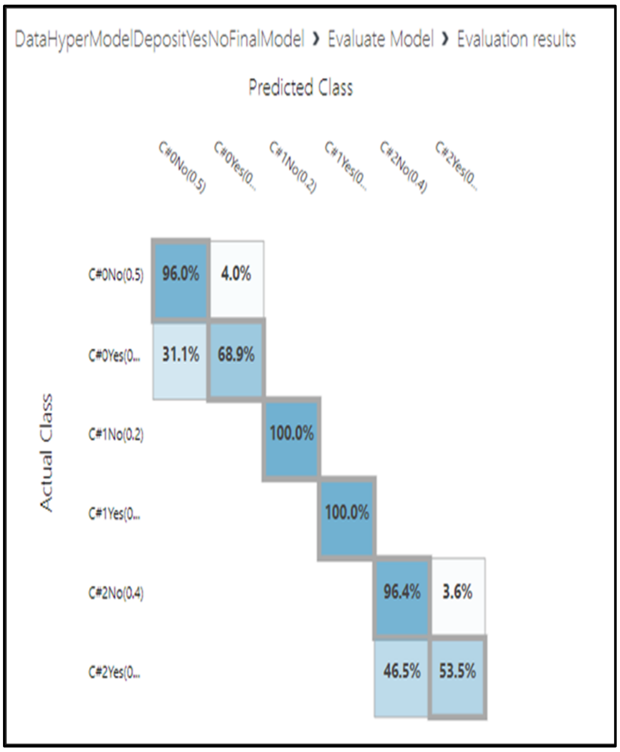


Figure 20) explain the classification report of the multi-class Decision jungle

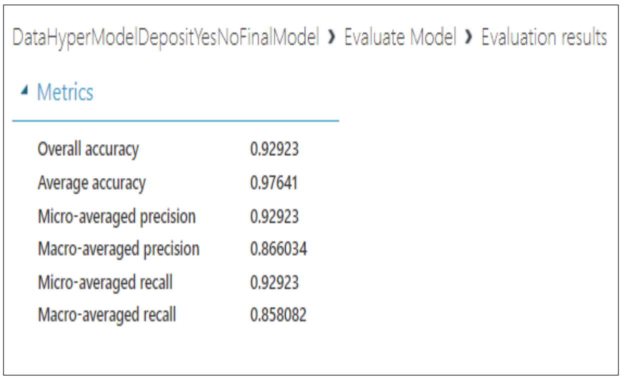


Figure (21) explain the confusion metrics of multi-class Decision jungle

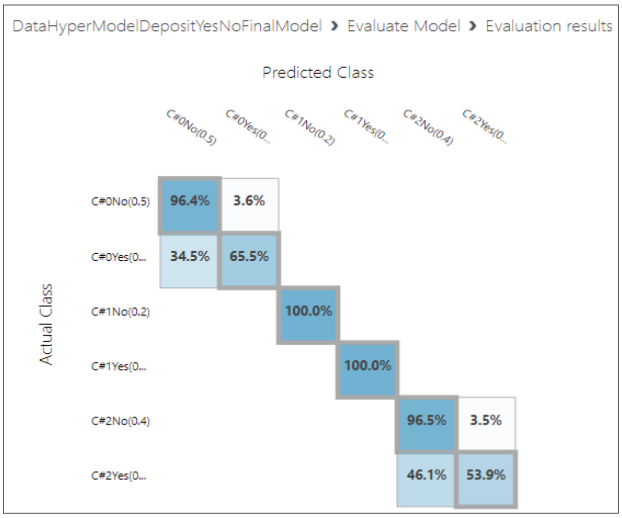


Figure (22) explain the classification report of multi-class Decision Forest

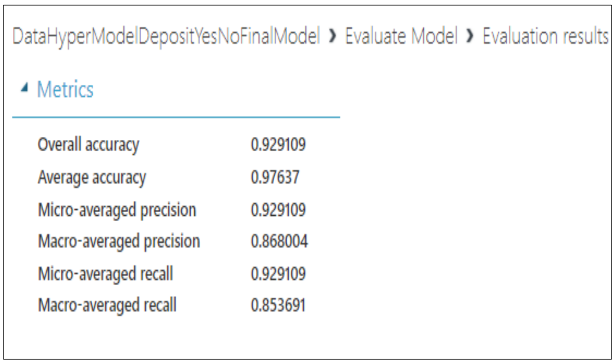


Figure (23) describe the confusion metrics of multi-class Decision Forest

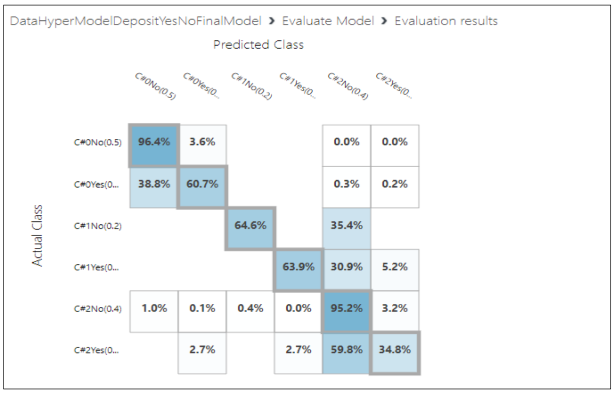
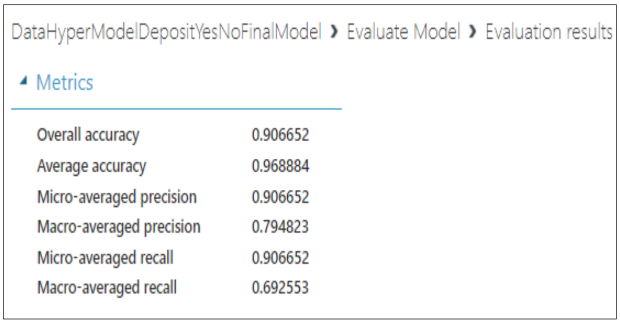


Figure (24) describe the classification report of multi-class logistic regression



Overall accuracy	0.906652
Average accuracy	0.968884
Micro-averaged precision	0.906652
Macro-averaged precision	0.794823
Micro-averaged recall	0.906652
Macro-averaged recall	0.692553

Figure (25) illustrate the confusion metrics of multi-class logistic regression

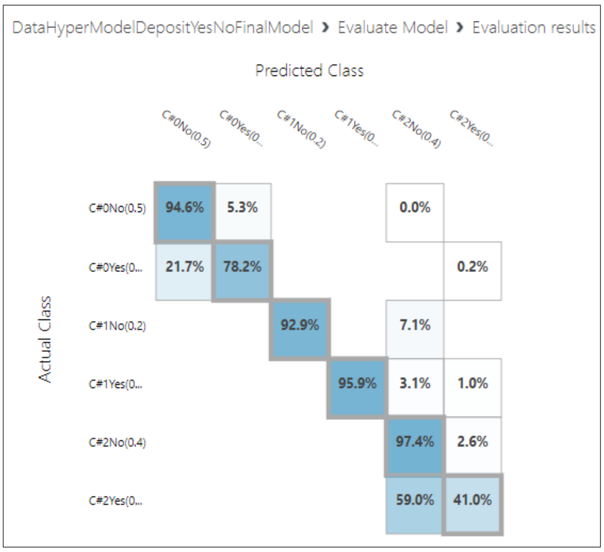
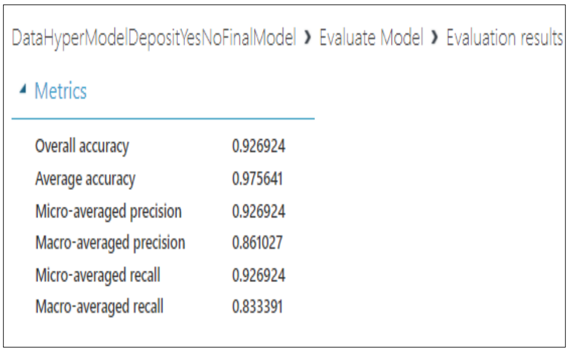


Figure (26) define the classification report of multi-class neural network



Overall accuracy	0.926924
Average accuracy	0.975641
Micro-averaged precision	0.926924
Macro-averaged precision	0.861027
Micro-averaged recall	0.926924
Macro-averaged recall	0.833391

Figure (27) presented the confusion metrics of multi-class neural network

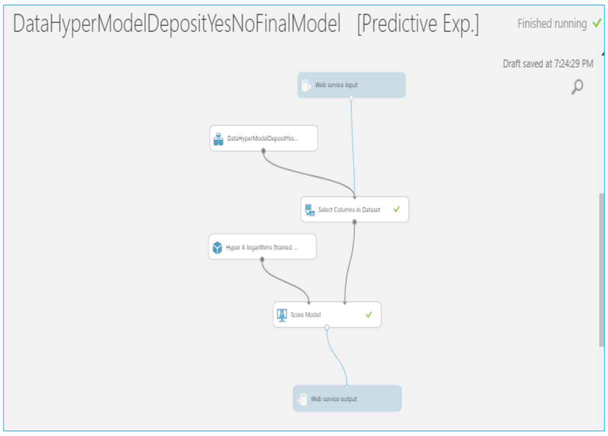
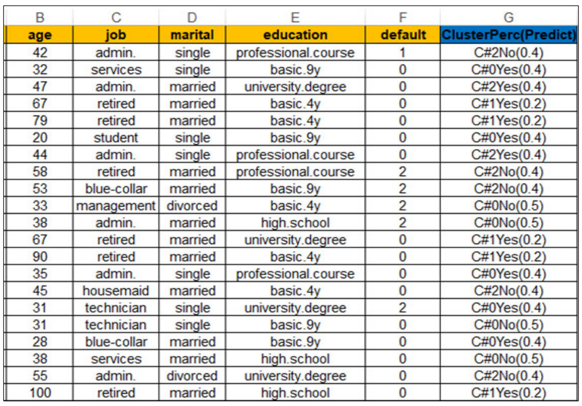


Figure (28) presented the predictive experiment diagram

After designing the model, the researcher tested the model with real data from the original data, then the test was done on similar data as new customers to predict the customer's classification in which cluster and probability of the customer will subscribe in a deposit for a term or not and the average percentage of the customer's subscription in a deposit, or the average customer's non-subscription in a deposit.



B	C	D	E	F	G
age	job	marital	education	default	ClusterPerc(Predict)
42	admin.	single	professional.course	1	C#2No(0.4)
32	services	single	basic.9y	0	C#0Yes(0.4)
47	admin.	married	university.degree	0	C#2Yes(0.4)
67	retired	married	basic.4y	0	C#1Yes(0.2)
79	retired	married	basic.4y	0	C#1Yes(0.2)
20	student	single	basic.9y	0	C#0Yes(0.4)
44	admin.	single	professional.course	0	C#2Yes(0.4)
58	retired	married	professional.course	2	C#2No(0.4)
53	blue-collar	married	basic.9y	2	C#2No(0.4)
33	management	divorced	basic.4y	2	C#0No(0.5)
38	admin.	married	high.school	2	C#0No(0.5)
67	retired	married	university.degree	0	C#1Yes(0.2)
90	retired	married	basic.4y	0	C#1Yes(0.2)
35	admin.	single	professional.course	0	C#0Yes(0.4)
45	housemaid	married	basic.4y	0	C#2No(0.4)
31	technician	single	university.degree	2	C#0Yes(0.4)
31	technician	single	basic.9y	0	C#0No(0.5)
28	blue-collar	married	basic.9y	0	C#0Yes(0.4)
38	services	married	high.school	0	C#0No(0.5)
55	admin.	divorced	university.degree	0	C#2No(0.4)
100	retired	married	high.school	0	C#1Yes(0.2)

Figure (29) described the sample of original real dataset that we want to use to test the model

B	C	D	E	F	G
age	job	marital	education	default	ClusterPerc
41	blue-collar	divorced	basic.4y	2	C#0Yes(0.4)
49	entrepreneur	married	university.degree	2	C#2Yes(0.4)
49	technician	married	basic.9y	0	C#2Yes(0.4)
39	housemaid	married	basic.9y	0	C#0Yes(0.4)
28	UnKnown	single	UnKnown	2	C#0Yes(0.4)
44	services	married	high.school	0	C#2Yes(0.4)
66	retired	married	basic.4y	0	C#1Yes(0.2)
67	retired	married	professional.course	0	C#1Yes(0.2)
37	admin.	married	university.degree	0	C#0Yes(0.4)
32	admin.	married	university.degree	0	C#0Yes(0.4)
33	student	married	professional.course	0	C#0Yes(0.4)
31	admin.	single	university.degree	0	C#0Yes(0.4)
32	admin.	married	university.degree	0	C#0No(0.5)
38	entrepreneur	married	university.degree	0	C#0No(0.5)
40	management	divorced	university.degree	0	C#0No(0.5)
60	blue-collar	married	basic.4y	0	C#2No(0.4)
57	retired	married	professional.course	0	C#2No(0.4)
46	blue-collar	married	professional.course	0	C#2No(0.4)
56	retired	married	university.degree	0	C#2No(0.4)

Figure (30) illustrated the sample of the new external data for the prediction in the column in blue

In our results, the study showed that it is possible to predict that the customer will be classified in any cluster, and not only that, but the model determines whether he subscribed to a deposit for a term or not, and not that, but the model also determines the average percentage of the customer in subscribing to a deposit or the average percentage of the customer not subscribing in a deposit Yes, and this is in contrast to previous studies that focused on accuracy and how to raise it.

9. Conclusion

Generally, it was nice to apply both unsupervised and supervised machine learning techniques to research the problem that gets up in the banking sector in how to attract premium customers to sign up for deposits. After much effort, the researcher reached to cleaning the data, which would lead to the treatment and classification of the customer according to his characteristics in the particular cluster, and then determining whether he will subscribe to a deposit for a term or not, and then determining the average percentage of the customer's subscription in a term deposit or the average percentage of non-deposit The customer did not subscribe in a deposit. So, the model can be integrated into the bank system (ECRM) in an integrated way to distinguish customers to participate in a deposit and to predict the average subscription rate, or the average non-subscription rate in a deposit.

So, the model can be integrated into the bank system (ECRM) in an integrated way to distinguish customers to participate in a deposit and to predict the average subscription rate, or the average non-subscription rate in a deposit.

- According to previous analysis, the most responsive customers enjoy these features:

- Cluster (0) contained the highest percentage in terms of the number of customers who subscribed to a term deposits and those who did not subscribe to a term deposit.

- Cluster (2) ranked second to terms of both those who subscribed in a term deposit and those who did not subscribe to a term deposit alike.

- Cluster (1) ranked third in terms of both those who subscribed in a term deposit and those who did not subscribe in a term deposit alike.

- By conducting the multi-class decision jungle algorithm, the classification and estimation model was successfully built. with this model, the bank will be able to predict the customer's response to the telemarketing campaign before contacting that customer. In this way, the bank can dedicate more marketing efforts to customers classified as most likely to accept term deposits and reduce contact with those who are not likely to make term deposits.

- The study proved that the elderly is more responsive to subscribing to a term deposit, perhaps because they have become retired and less mobile and do not have the strength and youthful determination to build their own business, and therefore subscribing to a deposit is the best way to secure their needs.

- The study proved that the elderly is more responsive to subscribing to a term deposit, perhaps because they have become retired and less mobile and do not have the strength and youthful determination to build their own business, and therefore subscribing to a deposit is the best way to secure their needs.

- The study proves that customers who have obtained higher education or professional certificates can operate their own business, and prefer not to subscribe to a term deposit, perhaps because they have the mental and physical ability to operate their own business

10. Recommendation:

- By targeting the best customers, the bank will have more and more positive responses, and the rating algorithm (multi-class decision jungle) will eventually eliminate the imbalance in the original data set and more accurate information will be provided to the bank to improve subscriptions.

- The bank can better meet the demand of its customers by providing banking services to the right customer at the right time. By tracking the social and age status and behavior of customers, paying attention to the factors that the customer needs, and targeting the most expected to subscribe in a term deposit.

- Banks should focus on customers who have similar characteristics to the customers predicted by the model to participate in deposits.

Acknowledgment

This publication was supported by the Deanship of Scientific Research at Prince Sattam bin Abdulaziz University, Alkharj, Saudi Arabia.

11. References: -

[1] A.Elsalamony, H. (2014). Bank Direct Marketing Analysis of Data Mining Techniques. International Journal of Computer Applications, 85(7), 12–22. https://doi.org/10.5120/14852-3218

[2] Abbas, S. (2015). Deposit subscribe Prediction using Data Mining Techniques based Real Marketing Dataset. International Journal of Computer Applications, 110(3), 1–7. https://doi.org/10.5120/19293-0725

[3] Asare-frempong, J. (2017). Predicting Customer Response to Bank Direct Telemarketing Campaign. September, 10–14. https://doi.org/10.1109/ICE2T.2017.8215961

[4] B, S. S., & Pooja, A. (2021). Bank Marketing Data Classification Using Machine Learning. 8(2), 182–185. https://doi.org/10.17148/IARJSET.2021.8226

[5] Baratzadeh, F., Mohammad, S., & Hasheminejad, H. (2022). Customer Behavior Analysis to Improve Detection of Fraudulent

Transactions using Deep Learning. 10(1), 87–101. https://doi.org/10.22044/JADM.2022.10124.2151

[6] Can, B., Other, G., & December, M. (2016). FACTORS DETERMINING TERM How a Bank Can Get Other People ’ s Money.

[7] Carneiro, N., Figueira, G., & Costa, M. (2017). A data mining based system for credit-card fraud detection in e-tail. Decision Support Systems, 95(January), 91–101. https://doi.org/10.1016/j.dss.2017.01.002

[8] Chitra, K., & Subashini, B. (2013). Data Mining Techniques and its Applications in Banking Sector. International Journal of Emerging Technology and …, 3(8), 219–226. http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Data+Mining+Techniques+and+its+Applications+in+Banking+Sector#0

[9] Dinçer, H., Yüksel, S., Canbolat, Z. N., & Pinarbaşi, F. (2019). Data mining-based evaluating the customer satisfaction for the mobile applications: An analysis on turkish banking sector by using IT2 fuzzy dematel. Tools and Techniques for Implementing International E-Trading Tactics for Competitive Advantage, 320–339. https://doi.org/10.4018/978-1-7998-0035-4.ch015

[10] Farooqi, R., Iqbal, N., Farooqi, R., & Iqbal, N. (n.d.). Performance Evaluation for Competency of Bank Telemarketing Prediction using Data Mining Techniques Performance Evaluation for Competency of Bank Telemarketing Prediction using Data Mining Techniques. https://doi.org/10.35940/ijrte.A1269.078219

[11] Femina, B. T., & Sudheep, E. M. (2015). An efficient CRM-data mining framework for the prediction of customer behaviour. Procedia Computer Science, 46(Icict 2014), 725–731. https://doi.org/10.1016/j.procs.2015.02.136

[12] Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. Big Data and Cognitive Computing, 2(3), 1–13. https://doi.org/10.3390/bdcc2030018

[13] Improving KNN Model for Direct Marketing Prediction in Smart Cities Improving KNN Model for Direct Marketing Prediction in Smart Cities. (2021). June. https://doi.org/10.1007/978-3-030-72065-0

[14] Kanchana, M., Chavan, P., & Johari, A. (2020). Detecting Banking Phishing Websites Using Data Mining Classifiers. EasyChair.

[15] Krishna, C. L. (2019). An Efficient Deep Neural Network Classifier in Banking System. 6(2), 511–517.

[16] Krishna, C. L., & Reddy, P. V. S. (2019a). An Efficient Deep Neural Network Multilayer Perceptron Based Classifier in Healthcare System. 2019 Proceedings of the 3rd International Conference on Computing and Communications Technologies, ICCCT 2019, April, 1–6. https://doi.org/10.1109/ICCCT2.2019.8824913

[17] Krishna, C. L., & Reddy, P. V. S. (2019b). Deep neural networks for the classification of bank marketing data using data reduction techniques. International Journal of Recent Technology and Engineering, 8(3), 4373–4378. https://doi.org/10.35940/ijrte.C5522.098319

[18] Nguyen, C. (2019). The credit risk evaluation models: an application of data mining techniques. SAIS 2019 Proceedings, 1–7. https://aisel.aisnet.org/sais2019/36/

[19] Obiedat, R. (2020). Developing Decision Tree based Models in Combination with Filter Feature Selection Methods for Direct Marketing. 11(1).

[20] Patwary, M. J. A., Akter, S., Bin Alam, M. S., & Rezaul Karim, A. N. M. (2021). Bank Deposit Prediction Using Ensemble Learning. Artificial Intelligence Evolution, August, 42–51. https://doi.org/10.37256/aie.222021880

[21] Raju, P. S., Bai, V. R., & Chaitanya, G. K. (2014). Data mining: Techniques for Enhancing Customer Relationship Management in Banking and Retail Industries. International Journal of Innovative Research in Computer and Communication Engineering, 2014, 2650–2657.

[22] Vrontis, D. (2022). Big data analytics in strategic sales performance : mediating role of CRM capability and moderating role of leadership support. February. https://doi.org/10.1108/EMJB-07-2021-0105

Abubakr Mustafa Eltahir received the B.E. and M.E. degrees, from Alnileen Univ. in 2003, 2009 respectively. Ongoing Ph.D degree from Alnileen Univ. in 2015. After working as, a research assistant (from Dept. of Computer and Information Technology.

AMGAD ATTA BDELMAGEED Mohammed received BSc. Computer Science, University of Science and Technology 2007.
Master of Science in Information Systems University of Science and Technology 2013.
Respectively. He received the Ph.D degree from Sudan University of Science &Technology 2019, Philosophy of Computer Science and Information Technology, Sudan University of Science & Technology.