Application of Data Mining Techniques: Bank Marketing Dataset

Ganesh Kumar Chimakurthi
X16130961
MSc in Data Analytics
Team I

Sateesh Chandra Kammanethi
X16134800
MSc in Data Analytics
Team I

Vasu Mahendar Kumar Rajamandry
X16136195
MSc in Data Analytics
Team I

Abstract*— In recent times, data mining techniques have been emerged and developed rapidly. In this paper, we focused on many data mining techniques such as Association rule, Clustering and Classifiers techniques to explore bank marketing data for better purpose of marketing. Clustering techniques area also used to divide the data in to their group of segments. Association rules are applied to find the relation between the attributes in a data. Dataset has been analyzed by applying Association rules to find the patterns among the attributes in a dataset. Classification Algorithms such as Decision trees and Logistic algorithms were used to classify the data and to find the classification algorithm which is performing well on the dataset.

Keywords – Association rule, Logistic regression, Bank marketing, Clustering and Decision trees.

1. INTRODUCTION

Increase in rapid technology has given birth to vast amounts of data in various sectors. Data mining techniques in this area can help in decision making. The extraction of data for decision making by using data mining concepts is called Knowledge Discovery process (Xiao, 2012). Knowledge discovery is a logical process to find the patterns in data which is hidden. These patterns are helpful in developing the marketing and business.

There are three steps 1. Exploring 2. Pattern Identification and 3. Deployment. Exploring of data contains multiple steps involved in it such as finding out the data and transforming the data in to applicable method. Pattern identification is finding out the hidden patterns in attributes and its variable and finally deployment is implementing the patterns in appropriate methods (Ahmed and Hannan, 2012).

In this model, we propose four machine learning algorithms to find the pattern and performance of dataset using Association rules, Logistic regression, decision tree and clustering models. Association rules are efficient algorithms which finds the itemset that frequently appears among vast volumes of data. This type of finding hidden patterns helps in decision making among cross marketing and customer purchasing behavior. Association rules are generated based on support and confidence having less than 1. It also generates a measure called lift to measure the performance of rule in database. Association rules are often used to reduce transaction databases and to understanding the hidden concepts in a database. (Lee and Song, 2016).

Clustering is grouping of variables based on their patterns. A group in cluster contains similar variables. Using cluster methods, we can find the correlation and distribution pattern of attributes among data. We used cluster to group the variables in our dataset because classification techniques increase cost in grouping the variables (Bharathi and Ramageri, 2010).

Finally, we used Classification techniques to predict the performance of our model. These

classification techniques often employee's decision tree and neural networks for predicting the accuracy of model and performance. We used Decision tree model and Logistic regression in our model to classify the data and analyzing the data. The further implement is to predict the performance and best model in between them. In this paper, Section 2 describes the related models and section 3 states about the cleaning of dataset and section 4 focus on the Data mining techniques used for exploring the data.

I. Related Work

To classify and evaluate security investment funds we can apply Data mining technique Cluster Analysis. The author has taken 40 equity funds as sample data. They have used K- means cluster algorithm to process the data and the evaluated results from the cluster analysis is useful to guide and helpful in Decision making process where to invest (Zhang and Yang, 2016).

At present day by day traffic is increasing rapidly so there is chance of increasing the traffic accidents as well. So in order to decrease the traffic accidents we are using some decision tree rules. This author has concluded that to reduce the traffic accidents it is required to have a deep knowledge about the cause to avoid accidents. Decision tree applications with some CART algorithm rules we can try to avoid some accidents by following some rules. We can analyze some road safety point of view. By using this methodology data in some highways and in some countries we can keep the measures (da Cruz Figueira, Pitombo and Larocca, 2017).

In data mining we have much number of algorithms in that Association Rule algorithm is one of the powerful algorithm to find hidden relationships between the attributes in the large database and also it provides confidentiality of sensitive patterns. Privacy preserving is used to hide the association rules. We have many methods to implement here the author has used Heuristic approach in Data Distortion technique (Mistry, B.R. and Desai, 2015).

Logistic Regression is the one of the most popular method for clinical prediction model. There are mainly some problems in clinical model are Accuracy of medical problem, if we predict wrongly that may cause for any human severe effect, to predict correct medical prescription for the patient, prediction to avoid which medicine is not suits for particular persons, prediction model should allow doctors to take early decisions. The author has provided new method to give high accuracy of prediction (Taslimitehrani, V. and Dong, 2014).

II. Dataset & data cleaning

Bank marketing dataset has been taken from UCI Machine Learning repository which contains multiple attributes such as Age, gender, address, loan amount, housing, education and marital status of person. It contains 45211 instances and 17 attributes in it.

Data cleaning and Data preprocessing task is important and crucial step before implementation of algorithms. It had to be make sure that dataset is clean and should not contain any missing values or duplicate values in it. We used Rapid miner for better understanding the data and RStudio for cleaning the data.

age	job	marital	education	default	balance	housing	loan	contact	day	month	duration	campaign	pdays	previous	poutcome	y
30	unemploy	married	primary	no	1787	no	no	cellular	19	oct	79	1	-1	0	unknown	0
33	services	married	secondary	no	4789	yes	yes	cellular	11	may	220	1	339	4	failure	0
35	managem	single	tertiary	no	1350	yes	no	cellular	16	apr	185	1	330	1	failure	0
30	managem	married	tertiary	no	1476	yes	yes	unknown	3	jun	199	4	-1	0	unknown	0
59	blue-colla	married	secondary	no	0	yes	no	unknown	5	may	226	1	-1	0	unknown	0
35	managem	single	tertiary	no	747	no	no	cellular	23	feb	141	2	176	3	failure	0
36	self-empl	married	tertiary	no	307	yes	no	cellular	14	may	341	1	330	2	other	0
39	technician	married	secondary	no	147	yes	no	cellular	6	may	151	2	-1	0	unknown	0
41	entrepren	married	tertiary	no	221	yes	no	unknown	14	may	57	2	-1	0	unknown	0
43	services	married	primary	no	-88	yes	yes	cellular	17	apr	313	1	147	2	failure	0
39	services	married	secondary	no	9374	yes	no	unknown	20	may	273	1	-1	0	unknown	0
43	admin.	married	secondary	no	264	yes	no	cellular	17	apr	113	2	-1	0	unknown	0
36	technician	married	tertiary	no	1109	no	no	cellular	13	aug	328	2	-1	0	unknown	0
20	student	single	secondary	no	502	no	no	cellular	30	apr	261	1	-1	0	unknown	1
31	blue-colla	married	secondary	no	360	yes	yes	cellular	29	jan	89	1	241	1	failure	0
40	managem	married	tertiary	no	194	no	yes	cellular	29	aug	189	2	-1	0	unknown	0

Fig no1 Dataset and Attributes

Environmental setup:

Rapid miner and RStudio are two platforms used in our Implementation. Rapid miner is a Data analytics software which consists Machine Learning Algorithms in it. It is easy to use and prominent

results will be displayed. Clustering and classification algorithms are used in Rapid miner in our implementation.

RStudio is Integrated Development software which runs on programming. It expects user to provide code for respective algorithm. It includes multiple packages built in it which helps in providing the appropriate results. Apriori algorithm in Association rules is implemented in this algorithm.

III. DATA MINING TECHNIQUES USED IN THIS PROJECT

*A. **Association Rule***

Association rule mining is one of the important techniques in data mining. It finds the relation pattern between the attributes in a dataset. Association rules reduce the size of database by compressing the data. Compressing of data is done by finding relations in between. Market Basket Analysis is one of the most efficient analysis in reducing the size. Generally, Association rules gives a pattern that if A and B events occurs, then there is a likely of event C to be occur.

For example, if a customer purchases bread and jam, then there is a chance of customer purchasing butter too.

{bread, jam} -> {butter}

Implementation of Apriori Algorithm:

Apriori algorithm helps in reducing the transactional databases. It starts finding subsets from itemset by bottom up approach. It also follows Breadth first search following the iterative search of layer by layer search. (Kaung, 2010). Apriori rules are implemented in RStudio by using package called “library (arules)”. Apriori algorithm is inbuilt with most of the relation finding patterns, so that we can simply use Apriori(dataset) to execute the rules. Apriori algorithm is applied on Loan, Job, Marital and Education attributes in our dataset.

Support and confidence are passed as arguments to determine the efficiency of rule.

Support: Support can be calculated by fractioning the transaction that contain item X with total number of transactions.

Support = Number of transactions that contains item / number of transactions

Confidence: Confidence is calculated by fractioning the transaction that contains item X in Y with total number of transactions that contains X and Y.

Confidence = Number of transactions that contain x in y/ total number of transactions that contains both X and Y

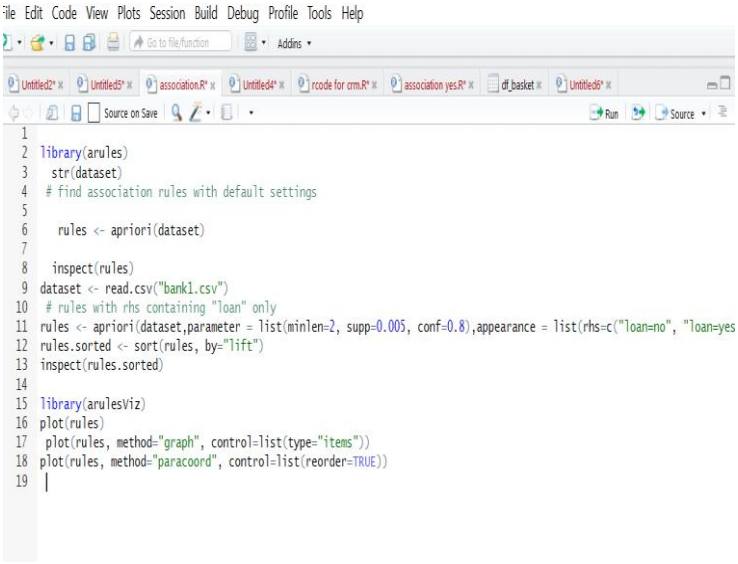

Fig 2 Apriori Algorithm

From the above figure, we can see that code for generating rules where dataset is read first by using “read.csv”. Rules are created by using Apriori (rules) and visualization of rules are done with package arulesviz.

Association rule Interpretation:

The following figures shows the rules formed by Association rules. Here, we got two outputs

- Consisting rules when loan is not taken from bank.
- Consisting rules when loan is taken from bank.

```
Console ~/
[67] {marital=single,education=secondary}          => {loan=no} 0.11455525 0.8489527  1.0020952
[68] {job=blue-collar,marital=divorced}            => {loan=no} 0.01481973 0.8481013  1.0011138
[69] {job=management,education=primary}            => {loan=no} 0.00729927 0.8461538  0.9988150
[70] {job=technician}                              => {loan=no} 0.14355231 0.8450521  0.9975145
[71] {job=admin.,marital=single}                   => {loan=no} 0.02654280 0.8391608  0.9905604
[72] {job=retired,education=tertiary}              => {loan=no} 0.00575094 0.8387097  0.9900278
[73] {job=blue-collar,marital=married}             => {loan=no} 0.12851139 0.8383838  0.9896432
[74] {marital=married}                             => {loan=no} 0.51846937 0.8380408  0.9892382
[75] {job=admin.,marital=single,education=secondary} => {loan=no} 0.02167662 0.8376068  0.9887260
[76] {job=self-employed}                           => {loan=no} 0.03384207 0.8360656  0.9869066
[77] {job=blue-collar}                             => {loan=no} 0.17474010 0.8350951  0.9857611
[78] {job=technician,marital=married}              => {loan=no} 0.07586817 0.8345499  0.9851175
[79] {marital=divorced}                            => {loan=no} 0.09688122 0.8295455  0.9792102
[80] {job=self-employed,marital=married}           => {loan=no} 0.02322495 0.8267717  0.9759359
[81] {job=technician,education=secondary}          => {loan=no} 0.09466932 0.8230769  0.9715746
[82] {job=services}                                => {loan=no} 0.07586817 0.8225420  0.9709431
[83] {education=secondary}                         => {loan=no} 0.41804910 0.8196010  0.9674716
[84] {job=retired,marital=married,education=secondary} => {loan=no} 0.01504092 0.8192771  0.9670892
[85] {job=management,marital=married,education=primary} => {loan=no} 0.00597213 0.8181818  0.9657963
[86] {job=blue-collar,marital=single}              => {loan=no} 0.03140898 0.8160920  0.9633294
[87] {job=technician,marital=married,education=secondary} => {loan=no} 0.05374917 0.8127090  0.9593362
[88] {job=services,education=secondary}            => {loan=no} 0.06525105 0.8126722  0.9592927
```

Fig no 3 Association rules for loan = No

From the above figure, we can see the list of cases where there is chance of taking no loan. By observing rule 71, we can predict that case that when job is admin and marital is single, then there is chance of taking no loan with support of 0.02 and confidence of 0.8. Similarly, other rules are stating the cases loan = no.

```
[74] {job=housemaid,marital=married,education=primary}       => {loan=yes} 0.008683068 1        1
[75] {job=unemployed,marital=married,education=primary}      => {loan=yes} 0.007235890 1        1
[76] {job=self-employed,marital=married,education=tertiary}  => {loan=yes} 0.007235890 1        1
[77] {job=self-employed,marital=married,education=secondary} => {loan=yes} 0.021707670 1        1
[78] {job=retired,marital=divorced,education=secondary}      => {loan=yes} 0.005788712 1        1
[79] {job=retired,marital=married,education=primary}         => {loan=yes} 0.007235890 1        1
[80] {job=retired,marital=married,education=secondary}       => {loan=yes} 0.021707670 1        1
[81] {job=entrepreneur,marital=married,education=primary}    => {loan=yes} 0.005788712 1        1
[82] {job=entrepreneur,marital=single,education=tertiary}    => {loan=yes} 0.005788712 1        1
[83] {job=entrepreneur,marital=married,education=tertiary}   => {loan=yes} 0.017366136 1        1
[84] {job=entrepreneur,marital=married,education=secondary}  => {loan=yes} 0.018813314 1        1
[85] {job=services,marital=divorced,education=secondary}     => {loan=yes} 0.010130246 1        1
[86] {job=services,marital=single,education=secondary}       => {loan=yes} 0.021707670 1        1
[87] {job=services,marital=married,education=secondary}      => {loan=yes} 0.066570188 1        1
[88] {job=admin.,marital=divorced,education=secondary}       => {loan=yes} 0.018813314 1        1
[89] {job=blue-collar,marital=divorced,education=primary}    => {loan=yes} 0.010130246 1        1
[90] {job=technician,marital=divorced,education=secondary}   => {loan=yes} 0.020260492 1        1
[91] {job=management,marital=divorced,education=tertiary}    => {loan=yes} 0.031837916 1        1
[92] {job=blue-collar,marital=divorced,education=secondary}  => {loan=yes} 0.007235890 1        1
[93] {job=admin.,marital=single,education=secondary}         => {loan=yes} 0.027496382 1        1
[94] {job=admin.,marital=married,education=tertiary}         => {loan=yes} 0.005788712 1        1
```

Fig no 4 Association rules for loan =yes

Fig no 4 states rules which are chances of taking loan. Rule 84 states when job is entrepreneur then there is chance of taking loan with support 0.01 and confidence of 1. Similarly, as above, all rules are directing when loan = yes.

Visualization of rules:

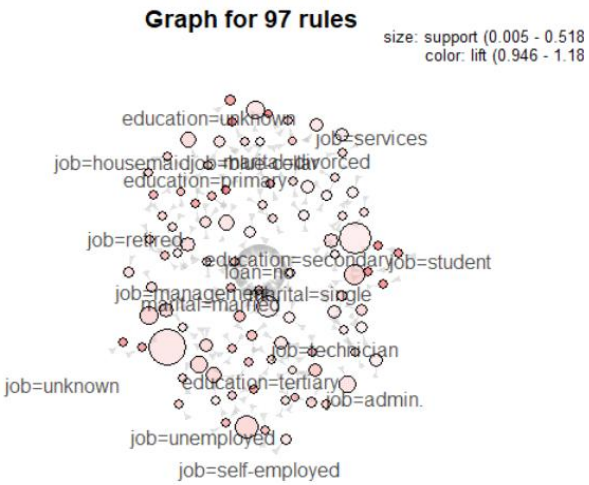

Fig no 5 Graph for 97 Rules

Above figures shows the graph for the rules created by association rules. It shows different kind of jobs that persons related to. Size of the cirlce indicates support of rule which indicates the efficiency of rule.Color of cirlce indiactes the lift of rule which is measure of rule. If the lift is high, then the occurrence of rule is high.

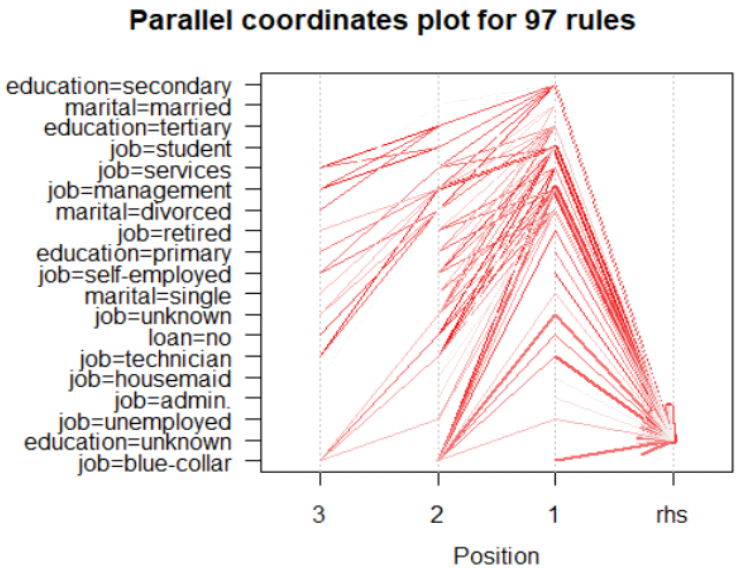

Fig no 6 Parallel coordinates for rules

The rules sorted by lift in shown in above figure. The rules which lies in plot 3 has high lift than rest of rules.

B. ***Decision Tree Algorithm:***

Decision tree has the flow chart or tree like structure. Decision tree is one of the most popular supervised learning technique. By using decision tree we can handle both numerical and categorical variables effectively. Decision tree structure contains root node and leaf nodes. Root node represents test of the attribute and leaf node (child node) represents class labels. The paths from the root node to child node or leaf nodes are the rules. Attribute value pairs are used to represent the Decision tree. One of the main reason most people prefer is it can perform efficiently for both numerical and categorical variables.

The decision tree algorithm has the capability of taking required attribute from the dataset accordingly into smaller groups. The main idea of decision tree is to convert the data into tree and decision rules for any business purpose or future use (Thariqa, Sitanggang and Syaufina, 2016).

Decision tree is easy to implement and identify the key characteristics of any attributes. Here a bank dataset has been taken from the UCI Repository which consists of Age, Marital status, Housing, Loan, Duration and whether the client deposit with a term deposit or not etc.

- Platform used – RapidMiner Software

Implementation and interpretation:

First need to load the data into RapidMiner and the data need to be drag into operation process and we need to give set role there we should mention attribute for which we are going to analyze. After set role need to give split validation next decision tree is applied to it and we should give performance to measure the accuracy.

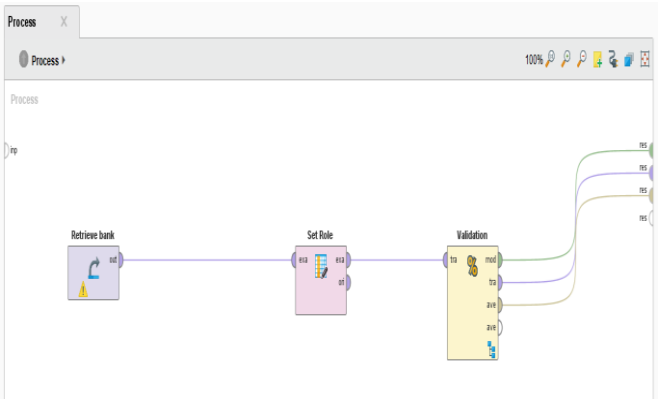

Fig (1): Validation for the data

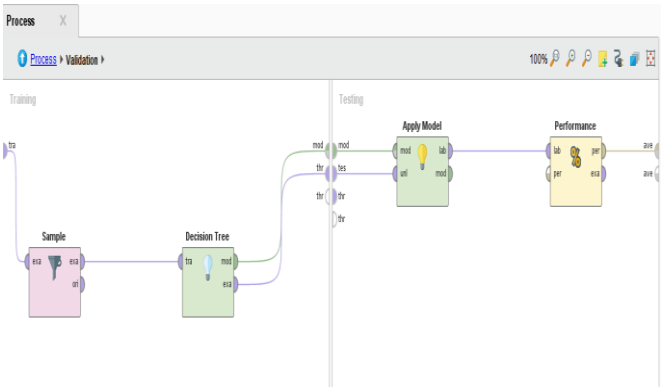

Fig (2): Decision tree and performance is applied

Table View | Plot View

accuracy: 89.16%

	true no	true yes	class precision
pred. no	1186	118	90.95%
pred. yes	29	23	44.23%
class recall	97.61%	16.31%	

Fig (3): Accuracy measure

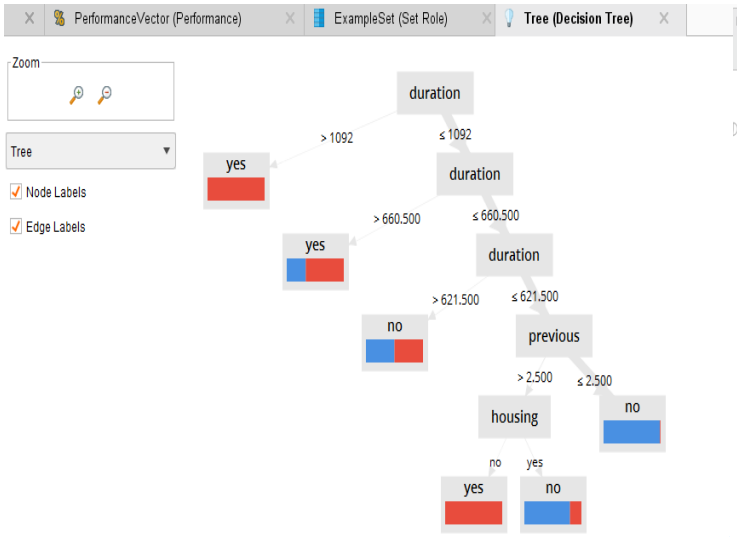

Fig (4): Decision Tree

The accuracy in this model is 89.16%. Whether the client has subscribed to term deposit or not will depends on duration and housing loan also

*C. **Logistic Regression:***

By using Logistic Regression we will get efficient results but to apply Logistic Regression we need categorical dependent variable. i.e either 0 or 1. Logistic Regression is mostly using in many Machine Learning techniques, medical field and social sciences. Logistic Regression is used to predict the future results efficiently.

Wireless heart beat monitors are increasing present day, It requires battery to work. In such kind of monitors RFID machines are using and noting the patient's heart beat and pulse rate etc. here by using Logistic Regression we can predict efficiently at what conditions the patient might get heart strokes (Vora and Kurzweg, 2016)

Implementation and Interpretation:

The loaded data need to drag into process and connection needs to give for set role after set role split operator we need to give then we need to apply Logistic Regression and Performance. Performance is used to detect the accuracy of the model. Here by using Logistic Regression we will get 88.05% accuracy.

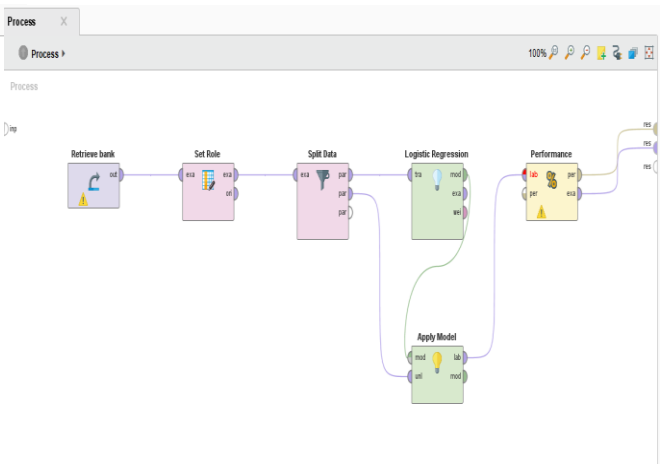

Fig (5): Applying Logistic Regression

xampleSet (1356 examples, 4 special attributes, 16 regular attributes)

Row No.	y	prediction(y)	confidence(...	confidence(...	age	job	marital	education
1	no	no	0.880	0.120	35	management	single	tertiary
2	no	no	0.990	0.010	30	management	married	tertiary
3	no	no	0.937	0.063	43	services	married	primary
4	no	no	0.949	0.051	43	admin.	married	secondary
5	no	no	0.838	0.162	36	technician	married	tertiary
6	no	no	0.976	0.024	44	entrepreneur	married	secondary
7	no	no	0.962	0.038	26	housemaid	married	tertiary
8	no	no	0.957	0.043	41	management	married	tertiary
9	no	no	0.983	0.017	56	self-employed	married	secondary
10	yes	no	0.986	0.014	32	blue-collar	married	secondary
11	yes	yes	0.552	0.448	33	management	married	secondary
12	no	no	0.998	0.002	23	services	single	tertiary
13	no	no	0.935	0.065	36	management	single	tertiary
14	no	no	0.933	0.067	32	technician	married	tertiary
15	no	no	0.983	0.017	32	technician	single	tertiary
16	no	no	0.920	0.080	38	technician	single	secondary
17	yes	yes	0.526	0.474	34	technician	married	tertiary
18	no	no	0.971	0.029	53	blue-collar	married	secondary
19	no	no	0.946	0.054	41	blue-collar	married	primary
20	no	yes	0.568	0.432	63	retired	married	secondary
21	no	no	0.902	0.098	48	management	married	tertiary

Fig (6): prediction values

accuracy: 88.05%

	true no	true yes	class precision
pred. no	1105	52	95.51%
pred. yes	110	89	44.72%
class recall	90.95%	63.12%	

Fig (7): Performance Accuracy

.

D. Clustering:

Clustering is grouping the set of attributes, the attributes belongs to the same cluster are more similar compare to the attributes belongs to other

clusters. In clustering we have different types are there like Hierarchal clustering, Centroid- based clustering (K- means clustering) and Density- based clustering etc.

Kisore and Koteswaraiah, (2017) did their research on improving density of ATM's using cluster analysis. They have divided the clusters based on density of the people in areas. Here they have taken density of the people how they are using the ATM services. They concluded that ATM placement has great impact in banking business. In this they have solved identifying the places to keep ATM's successfully.

Implementation and Interpretation:

First we need to take read excel and export the data into read excel in RapidMiner. While uploading the data the type of the variable should be real and attribute. Next we need to take the K-means clustering algorithm and performance operator is also included to get accuracy percent. Here in K- means cluster we can select how many clusters we want.

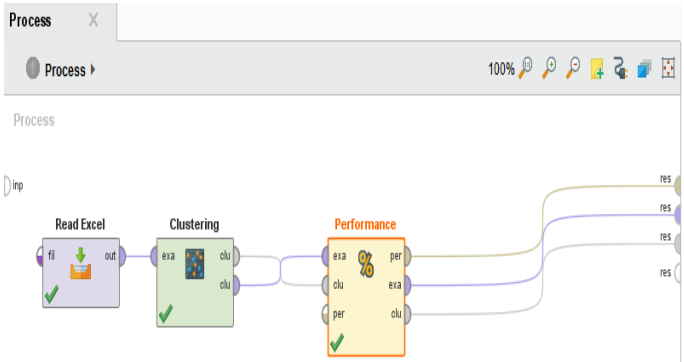

Fig (8): clustering implementation

ExampleSet (4521 examples, 2 special attributes, 7 regular attributes) Filter (4,521 / 4,521 examples): all

Row No.	job	cluster	age	balance	duration	loan	campaign	pdays	prev
1	unemployed	cluster_3	30	1787	79	0	1	-1	0
2	services	cluster_0	33	4789	220	1	1	339	4
3	management	cluster_2	35	1350	185	0	1	330	1
4	management	cluster_2	30	1476	199	1	4	-1	0
5	blue-collar	cluster_2	59	0	226	0	1	-1	0
6	management	cluster_2	35	747	141	0	2	176	3
7	self-employed	cluster_2	36	307	341	0	1	330	2
8	technician	cluster_2	39	147	151	0	2	-1	0
9	entrepreneur	cluster_2	41	221	57	0	2	-1	0
10	services	cluster_2	43	-88	313	1	1	147	2
11	services	cluster_0	39	9374	273	0	1	-1	0

Fig (9): Clusters

Cluster Model

```
Cluster 0: 266 items
Cluster 1: 73 items
Cluster 2: 3402 items
Cluster 3: 761 items
Cluster 4: 19 items
Total number of items: 4521
```

Fig (10): Cluster Model

Attribute	cluster_0	cluster_1	cluster_2	cluster_3	cluster_4
age	43.147	43.123	40.628	42.583	46.526
balance	6501.620	13674.192	335.776	2673.219	27766.316
duration	255.831	214.630	263.129	277.661	167.684
loan	0.094	0.055	0.168	0.118	0.105
campaign	2.887	2.808	2.823	2.633	2.526
pdays	45.515	32.890	39.095	40.803	64.526
previous	0.767	0.589	0.509	0.607	0.684

Fig (11): Centroid Table

Here for our bank dataset if we apply K-means algorithm for five clusters we will get the above Fig (10) result. i.e we have total 4521 items it will be separated into five groups, cluster 0, cluster 1, cluster 2, cluster 3 and cluster 4 with 266, 73, 3402, 761 and 91 respectively. By seeing the results we can clearly observe that cluster 2 has comparatively more number of items than the remaining clusters. From the Fig (11) Centroid table what are the attributes are falling under which cluster.

Conclusion:

Here we have used different data mining techniques like Association Rules, Decision Tree algorithm, Logistic Regression algorithm and Cluster Analysis. Association rule is used for to get the hidden relationship between the attributes and Association Rules are created by using Apriori algorithm in Rstudio. By using Decision Tree algorithm we can get the accuracy of 89.16% and in Logistic Regression we will get the accuracy of 88.05% here in Decision Tree algorithm we will get more accuracy compare to Logistic Regression. The entire data is separated into small parts based on

similar data called clusters and processed to get efficient results and we can measure the performance using Avg. within centroid distance.

REFERENCE:

[1] Thariqa, P., Sitanggang, I.S. and Syaufina, L., 2016. Comparative Analysis of Spatial Decision Tree Algorithms for Burned Area of Peatland in Rokan Hilir Riau. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *14*(2), pp.684-691.

[2] Vora, S. and Kurzweg, T., 2016, February. Modified logistic regression algorithm for accurate determination of heart beats from noisy passive RFID tag data. In *Biomedical and Health Informatics (BHI), 2016 IEEE-EMBS International Conference on* (pp. 29-32). IEEE.

[3] Kisore, N.R. and Koteswaraiah, C.B., 2017. Improving ATM coverage area using density based clustering algorithm and voronoi diagrams. *Information Sciences*, *376*, pp.1-20.

[4] Zhang, J. and Yang, K., 2016. Clustering analysis in the evaluation of securities investment funds. *Journal of Intelligent & Fuzzy Systems*, *31*(2), pp.949-956.

[5] da Cruz Figueira, A., Pitombo, C.S. and Larocca, A.P.C., 2017. Identification of rules induced through decision tree algorithm for detection of traffic accidents with victims: A study case from Brazil. *Case Studies on Transport Policy*.

[6] Mistry, B.R. and Desai, A., 2015, March. Privacy preserving heuristic approach for association rule mining in distributed database. In *Innovations in Information, Embedded and Communication Systems (ICIIECS), 2015 International Conference on* (pp. 1-7). IEEE.

[7] Taslimitehrani, V. and Dong, G., 2014, November. A new CPXR based logistic regression method and clinical prognostic modeling results using the method on traumatic brain injury. In *Bioinformatics and Bioengineering (BIBE), 2014 IEEE International Conference on* (pp. 283-290). IEEE.

[8] Gu, X.F., Hou, X.J., Ma, C.X., Wang, A.G., Zhang, H.B., Wu, X.H. and Wang, X.M., 2015, December. Comparison and improvement of association rule mining algorithm. In *Wavelet Active Media Technology and Information Processing (ICCWAMTIP), 2015 12th International Computer Conference on*(pp. 383-386). IEEE.

[9] Bharati, M. and Ramageri, M., 2010. Data mining techniques and applications.

[10] Song, K. and Lee, K., 2017. Predictability-based collective class association rule mining. *Expert Systems with Applications*, *79*, pp.1-7.

[11]Ahmed, A. and Hannan, S.A., 2012. Data Mining Techniques to Find Out Heart Diseases: An Overview. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, *1*(4), pp.18-23.

[12] Lili KUANG. Discussion on Apriori algorithm and FP-tree algorithm. Huaibei Coal Industry Teachers College (NATURAL SClENCE EDITION), 2010, 31 (2).