



INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X

Impact factor: 4.295

(Volume 5, Issue 5)

Available online at: www.ijariit.com

Hybrid model for customer behavioural mining

Aabid Ud Din Wani

waniaabid369@gmail.com

CT Institute of Engineering Management and
Technology, Lambri, Punjab

Prince Verma

princeverma.cse@gmail.com

CT Institute of Engineering Management and
Technology, Lambri, Punjab

ABSTRACT

Prediction is a crucial facet within the medical domain. This paper encourages the utilization of Artificial Intelligence for prediction in medical life as technological support. There's a powerful impact upon all activities because of health, and human consultants should have the power to choose the adequate treatment which is able to be the evolution of the patient throughout the treatment. Since the mechanisms of AI have many advantages that are appropriate for this, it assists human intelligence in decision-making/prediction. The preciseness of mathematics and also the power of current technologies are the two robust qualities used in the system when they are combined to produce positive options as a result.

Keywords— Data mining, Clustering, K-mean, Customer churning

1. INTRODUCTION

According to William J Frawley, and Christopher J Matheus "The non-trivial removal of inferred, formerly unknown, and potentially practical facts from data". Data mining finds precious information concealed in huge volumes of data.

Data mining is the study of data and the use of software techniques for finding customs and regularities in sets of data. The computer in control of finding the custom by identifying the underlying rules and quality in the data. The establishment of data mining and its growth has become an evolution of the intelligent network management system an unavoidable trend. It can be used for marketing purpose for collecting information and data.

1.1 Role of Data Mining

- In the web-based business, the data mining is useful of the discovery to exchange improvement propensity, of the right decision by making the undertaking.
- Data Mining also helps regarding applications of Marketing Strategy formulation.
- The use of data mining in healthcare, is in interest of data mining techniques over traditional technique, special attribute of health data, and regular health condition mysteries have made data mining very crucial for health data analysis.it can play role in health care also.

1.2 Customer Churning

Customer Churning business Study and Customer Relationship Management (CRM) analysers requires to know the causes for stir customers, as well as, way of behaving patterns from the existing stir customers' data. Here, churning refers to shifting of customer from one product to another for seek of satisfaction.

In India, numerous media transmission clients moved to JIO telecom for better and modest administrations. Since clients have numerous choices as better and more affordable administrations. A definitive objective of telecom organizations is to expand their benefit and remain alive in an aggressive commercial centre. A customer churning or agitate happens when a huge level of customers is not happy with the administrations of any telecom organization. It brings about administration movement of clients who start changing to other specialist organizations.

To deal with churning, customer churning can be limited by examining the previous history of the potential clients of an organization methodically. Also, chiefs are constantly looked at loose activity the board issue. Henceforth the requirement for expectation component for stir the executives and persistent updating of fitting techniques have turned out to be increasingly significant the present aggressive world.

Section 1 includes introduction regarding data mining and KDD process in detail.

Section 2 includes reviews by the authors who have already researched on this technique.it is section of related work.

Section 3 consist of proposed work including proposed methodology and techniques.

Section 4 includes results observed after implementing proposed work.

Section 5 includes conclusion showing accuracy in existing and proposed work.

2. RELATED WORK

Ibrahim et al [1] used the classification methods including, Decision Tree, Logistic Relapse, SVM, KNN, Random Forest, and Naive Bayes procedure on the huge data. At that point, we break down the outcomes by ascertaining the exactness,

accuracy, review, and F-measure estimations of the outcomes. In addition, we execute a component choice technique on the information and investigate the outcomes with past ones. The outcomes will lead organizations to foresee their workers' beat status and thusly help them to decrease their human asset costs.

Leidvs et al [2] portrays the consequences of utilizing data mining procedures to investigate the conduct of clients of a fashion firm in Instagram social network. The purpose was CRISP-DM through which the illustrative models utilizing the procedures of clustering and affiliation srules were assessed. The outcomes demonstrate that the proposed models can give valuable data to structuring advertising methodologies suitable as indicated by client inclinations.

Imran et al [3] present a prediction churn model is assessed utilizing measurements, for example, exactness, accuracy, review, f-measure, and accepting working qualities (ROC) area. The outcomes demonstrate that our proposed agitate expectation model delivered better beat characterization utilizing the RF calculation and client profiling utilizing k-means clustering. Moreover, it likewise gives factors behind the churning of churn clients through the standards created by utilizing the characteristic chose classifier calculation.

Semrl et al [4] present a series of trial that aim to forecast customer way of action, in order to get larger gym inflection, and customer retention. They use two off-the-shelf machine learning stage, so that we can assess whether these stages, used by non ML specialist can help companies improve their services.

Liu et al [5] mainly considers a lining up model with stranding and recurrent utilization conduct for call centres. As a general rule, customers regulate their way of conduct mainly based on their recognition of waiting time. We certain to take into account customer contentment with waiting participation in model that queuing system with endogenic customer arriving processes.

Lei Zuo et al [6] study customer categorization issue based on customer benefit outlier data, a customer categorization model based on outlier data examination concerning customer advantage is constructed successfully.

Rohit et al [7] focuses on the ways of different data mining techniques used for mined out for the raw data. However, these techniques are cumbersome and can be optimised using Linked Data. Hence, we discuss the data mining skilfulness with associated data that may play a crucial role in future in bring out meaningful information from unstructured or semi-structured data.

3. PROPOSED METHODOLOGY

The proposed technique comprises of classification with clustering model. The following steps are taken for defining the methodology of proposed technique.

Choose a value of k for the data.
The arithmetic mean of each part is taken as the centroid point.

$$\text{Mean} = \{ (\sum d_i) / n \}$$

Where $\sum d_i$ = sum of all data instances or records and n = total number of data instances in each partition of data set.

Compute the Euclidean distance of each data point d_i to all the centroids weighted distance of data record 1 to data record 2 is carried out utilizing the following formula:

$$d = \text{sqrt}(\text{stddev} \sum_{i=1}^n (x_i - y_i)^2);$$

Where,

$$\text{stddev} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad \bar{x} \text{ is the mean of attributes}$$

For each d_i , examine the closest centroid and assign d_i to that centroid.

- Set the $\text{Near_edist}[i] = \text{edist}(d_i, C_j) // C_j$: to nearest centroid.
- For each of the cluster j, recompute the centroids.
- Repeat
- For each of the data-point d_i
- Compute its distance with new centroid of the present closest cluster.
- If this distance is equal to or less than the previous distance, the data-point will stay in the cluster,
- Else
- Compute $\text{edist}(d_i, C_j)$ from all of the cluster centroids; Ended for.
- Assign the data-point d_i to the cluster with the closest Centroid.
- Set $\text{Near_edist}[i] = \text{edist}(d_i, C_j)$; End for loop.
- Take best average to sum of all Euclidean distances and get the final output.
- Train the classifier with artificial neural network.

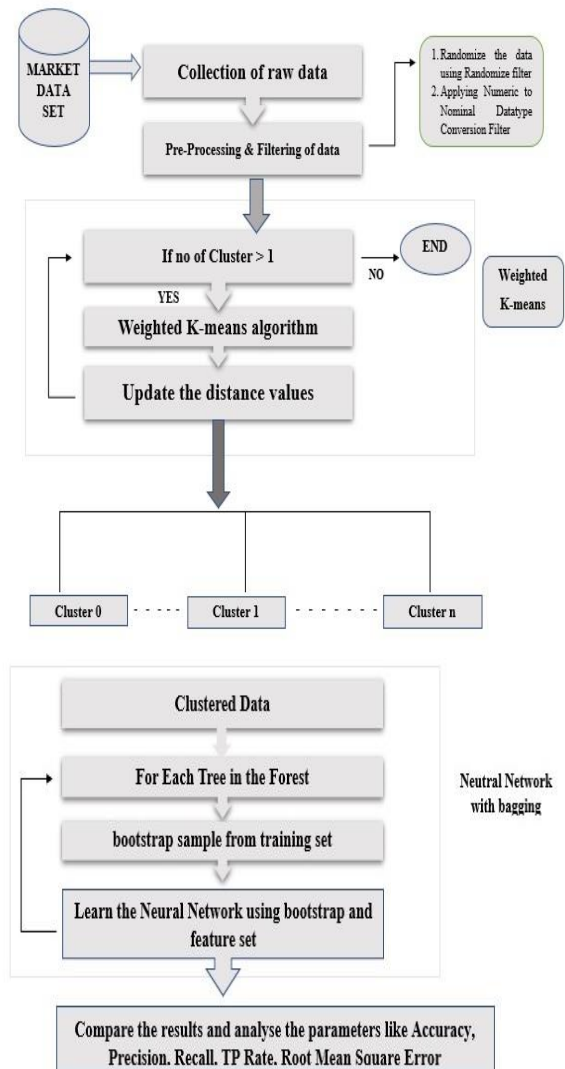


Fig.1: Flow diagram

Evaluating the performance of the proposed method with the basis of accuracy, precision, recall, Root mean squared error.

4. RESULTS

4.1 Choosing the dataset

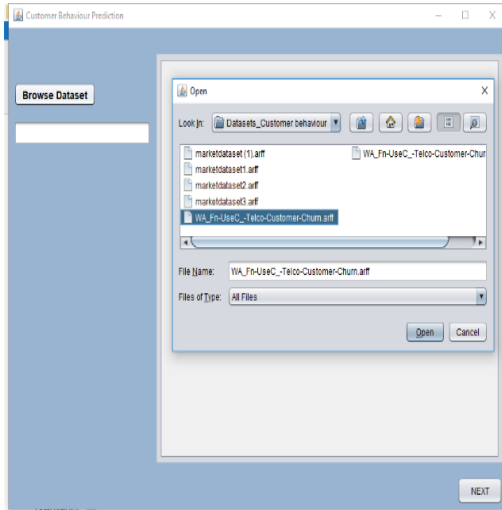


Fig. 2: Choosing the dataset

4.2 Showing the dataset content

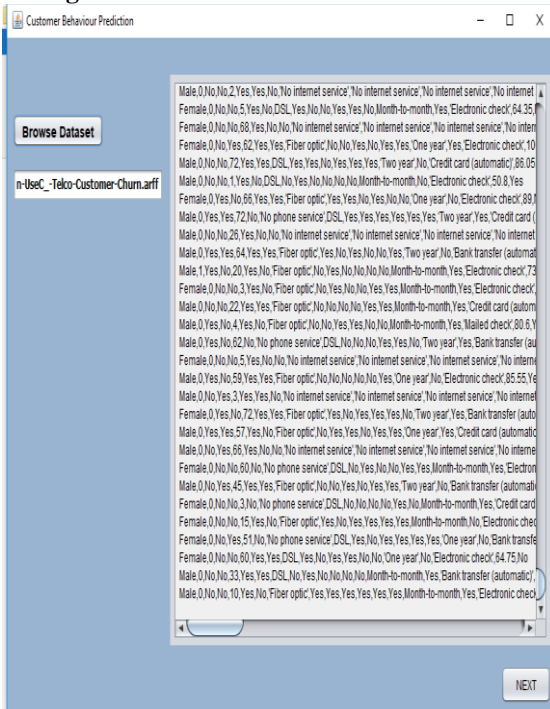


Fig. 3: Showing the dataset content

4.3 Replacing the missing values from data

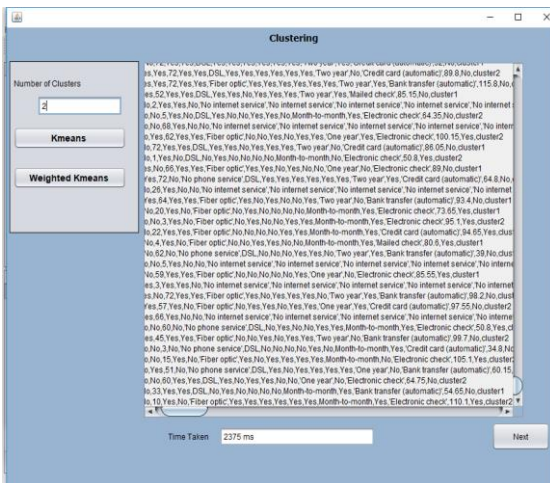


Fig. 4: Replacing the missing values

4.4 Applying K-means clustering

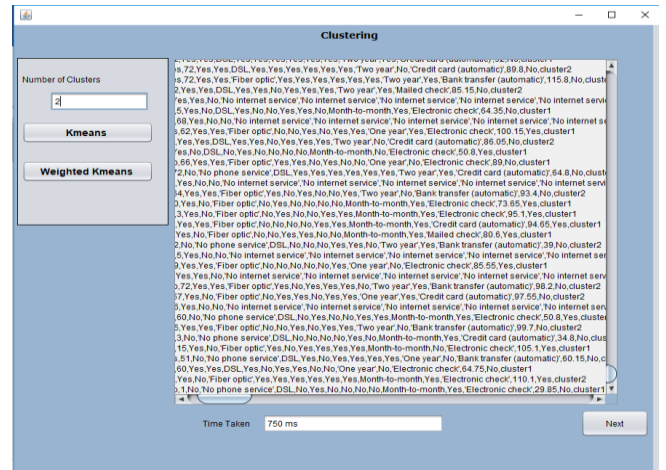


Fig. 5: Applying K-means clustering

4.5 Applying the proposed weighted enhanced K-means.

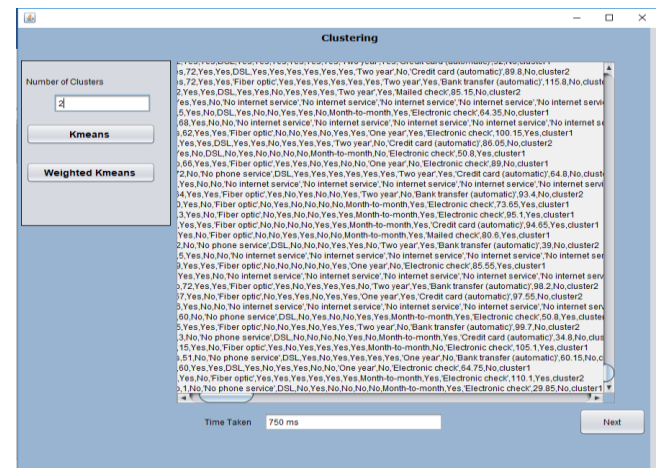


Fig. 6: Applying the proposed weighted enhanced K-means

4.6 Decision tree with K-means Clustered data base paper work.

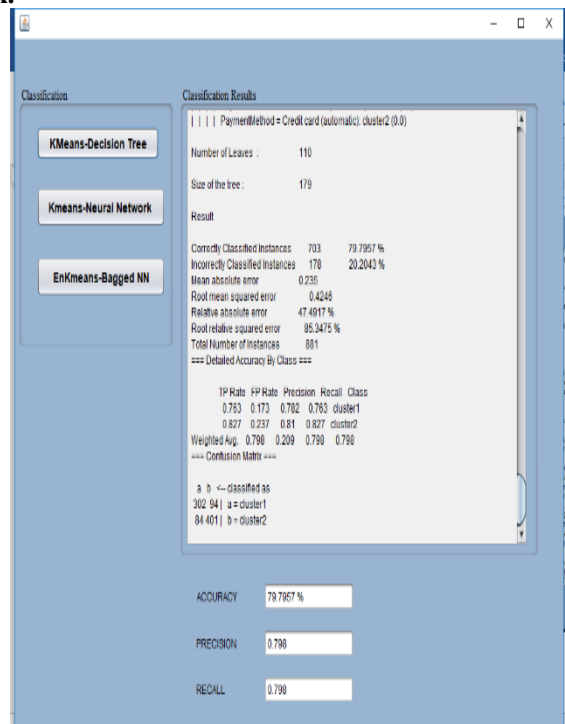


Fig. 7: Decision tree with K-means Clustered data base Paper work

4.7 Neural network with K-means clustering data

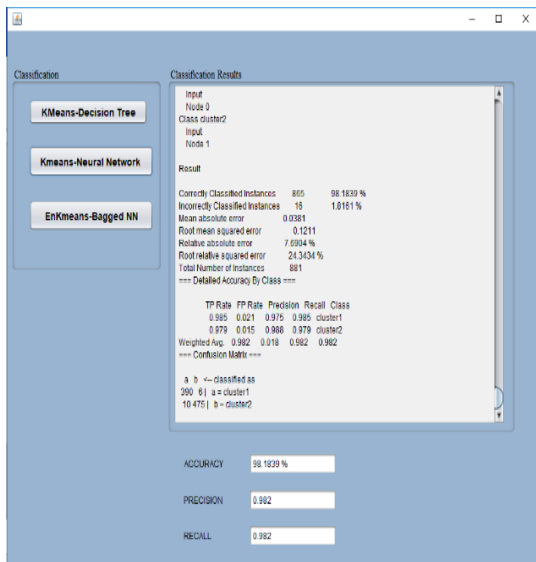


Fig. 8: Neural network with K-means clustering data

4.8 Bagged neural network with weighted Enhanced K-means Clustering data (Proposed Work)

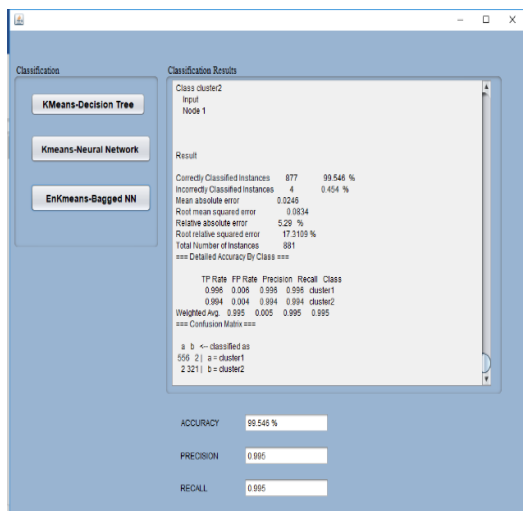


Fig. 9: Bagged neural network with weighted Enhanced K-means Clustering data (Proposed Work)

5. CONCLUSION

The motive of this paper is to clarify techniques which will be used to identify churn customers and effective marketing strategies could be planned for this group of customers. It will describe bagged weighted k mean clustering technique with neural network. Data mining proposed technique to be used to extract useful information regarding customer. In this research we observed that proposed technique has more accuracy than existing technique.

6. REFERENCES

- [1] Ibrahim Onuralp Yiğit ; Hamed Shourabizadeh, “ An approach for predicting employee churn by using data mining”, 2017 International Artificial Intelligence and Data Processing Symposium (IDAP), November 2017
- [2] Leidys del Carmen Contreras Chinchil; Kevin Andrey Rosales Ferreira, “Analysis of the behavior of customers in the social networks using data mining techniques” :2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM) November 2016
- [3] Irfan Ullah ; Basit Raza ; Ahmad Kamran Malik ; Muhammad Imran ; Saif Ul Islam ; Sung Won kim; “ A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector”,
- [4] Jas Semrl ; “ Alexandru Matei, Churn prediction model for effective gym customer retention”, 2017 International Conference on Behavioral, Economic,” Socio-cultural Computing (BESC), January 2018
- [5] Shu-an Liu ; Jun Gong ; Luping Ma ; Miao Yu, “ Influence of waiting times on customer loyalty and queueing behavior in call centers”, 2017 29th Chinese Control And Decision Conference (CCDC), July 2017
- [6] Lei Zuo ; Junfeng Guo,” Customer Classification of Discrete Data Concerning Customer Assets Based on Data Mining”, 2019 International Conference on Intelligent Transportation, Big Data & Smart City (ICITBS)”, Jan. 2019
- [7] Rohit Beniwal; Vikas Gupta; Manish Rawat; “Rishabh Aggarwal, Data Mining with Linked Data: Past, Present, and Future”, 2018 Second International Conference on Computing Methodologies and Communication (ICCMC), October 2018.

BIOGRAPHY



Aabid Ud Din Wani

CT Institute of Engineering Management and Technology, Lambri, Punjab, India



Prince Verma

CT Institute of Engineering Management and Technology, Lambri, Punjab, India