_____

Forest Tree- An Efficient Proposal Approach for Data Mining

Anuradha Angurala
Computer Science &
Engineering Golden College
of Engineering and
Technology
Gurdaspur, India

Suraj Pal
Computer Science &
Engineering Golden College
of Engineering and
Technology
Gurdaspur, India

Mohit Angurala
Asst. Prof. GNDU
Regional Campus,
Gurdaspur,
Punjab, India

Kewal Krishan
Department of Computer
Science, R K Arya College,
Nawanshshr, Punjab
kkbanth11@gmail.com

Abstract— Data Mining (DM) is a way of looking on different models, summaries, & derived values from a given gathered data. DM itself work on the process of looking for analytical information in huge amount of available databases. An illustration of a predictive riddle is targeted marketing. There are many factors that influence the performance of mining on large data sets. In this paper we are going to use forest tree technique in order to improve the way of performance of how the data is to be fetched and when on implementation it will definitely overcome the performance of previous work which includes existing approach decision tree algorithm.

Keywords— DM, ITS, AI, DB.

_____*****_____

I. INTRODUCTION

Data mining derives its name from the similarities between searching for valuable business information in a large database For example, finding linked products in gigabytes of store scanner data and mining a mountain for a vein of valuable ore. Both processes require either sifting through an immense amount of material, or intelligently probing it to find exactly where the value resides. There are ample of key properties of Data Mining which is shown in below figure:
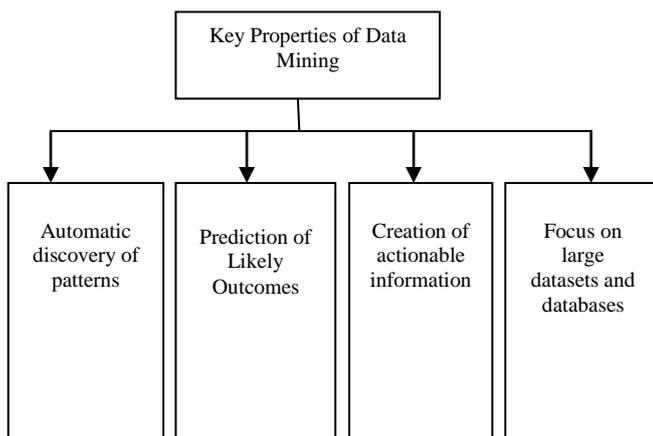

Figure 1: Key Features of Data Mining

II. LITERATURE REVIEW

Gyozo et al., (2007) Explained that the proficient administration of moving object databases has increased much enthusiasm in recent years because of the improvement of mobile communication and positioning advancements. A regular method of representing moving objects to utilize the trajectories. Much work has concentrated on the subjects of indexing, query processing and information mining of moving object directions, yet little consideration has been paid to the conservation of privacy in this setting. In numerous applications, for example intelligent transport systems (ITS) and fleet management, floating automobile data, i.e., tracked vehicle locations, are gathered, and utilized for mining movement patterns. Case in point, mining vehicle directions in urban transportation systems after some time can without much of a stretch distinguish thick zones and utilize this for predicting traffic blockage. By information mining the occasional movement patterns for individual drivers, customized, context-aware services can be conveyed. Be that as it may, uncovering area/direction information of moving objects to application servers can cause serious compromise to the location privacy of individual users. For instance, a service provider with access to direction information can concentrate on a client's personal habits to keep the client ID mystery, since basic areas, for example, the home and office address can be found by relating recorded directions, trailed by cross-referencing these areas with, e.g., Yellow Pages, to uncover user identity. Privacy–preserving information mining of moving object directions has not been tended to in the literature. The test of getting point by point, precise patterns from anonymized area and direction information is the inspiration for this paper.

Isaac Cano et al., (2009) clarified that the computerized age has empowered global access to gathering of information. While there are a few pros to ubiquitous access to information, there is additionally the potential for breaching the privacy of people. Information perturbation is a traditional way for taking care of the issue of empowering access to information and saving their privacy at the same time. As information datasets contains sensitive data about the respondents, their transfer to

_____

_____

outsiders requires the utilization of systems to guarantee information privacy. Data loss measures assess in what degree the secured information is still substantial for analysis, and disclosure risk measures assess in what degree information fulfill the privacy limitations.

Isaac Cano et al., (2010) Explained that the data age has empowered numerous associations to collect huge volumes of information. But the handiness of this information is near to zero if relevant information or knowledge can't be separated from it. Data mining, also called learning discovery, endeavors to answer this need. Rather than standard measurable strategies, data mining systems look for intriguing information without demanding earlier theories.

A key issue that emerges in any collection of data is that of privacy. The need for confidentiality is now and then because of law or can be inspired by business interests. Notwithstanding, there are circumstances where the sharing of information can prompt mutual gain. All things considered, for this situation, privacy issues ought to be checked and information mining algorithms ought to be revaluated starting here of perspective. That is, privacy should be safeguarded.

Jingjing Qiang et al., (2011) Explained that data mining is a very important and widely used technology. Classification is an important information mining issue applicable in numerous different areas. Support vector machine is one of the most actively developed methods for classification and has proven to produce high performance in many real-world applications. To get a SVM classifier, a basic assumption is that we should know all data.

III. PROBLEM FORMULATION

Data Mining is the process of discovering interesting Patterns, pattern evaluation and knowledge presentation that allow the user to analyst data from different dimension, categorize it and summarize the relationship which identified during data mining. Quality of Education is still very challenging problem. Data mining Techniques can be used in educational field to enhance our understanding of studying process to emphasize on identification, extracting & evaluation variable related to the study Process of students.

IV. PROPOSED WORK

Data mining is the process of discovering interesting patterns, pattern evaluation and knowledge presentation that allow the users to analyze data from the different dimension, categorize it and summarize the relationships which identified during the data mining process. We will carry out research in following steps:

➢ Use Python Script Tool
➢ Take Student Dataset
➢ Select the processor from Data Set
➢ Implement Random Forest & Tree Algorithm on dataset

➢ Display the predictions from input Data Set.
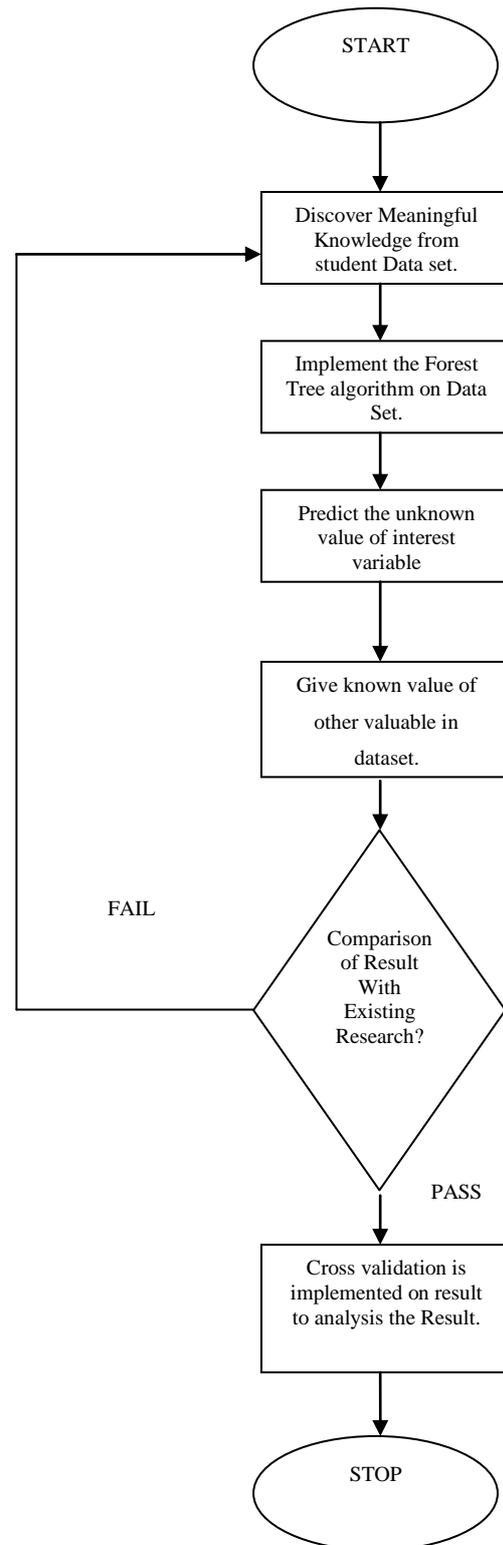➢ Cross validation on Implemented result

Figure 2: Flowchart for proposed methodology

V. CONCLUSION

Data mining means obtaining or mining knowledge from huge level of data. The term is actually a misnomer. Therefore, data mining is more correctly also known as Knowledge mining

_____

_____

which focuses on obtaining from huge level of data. It is the computational process of discovering patterns in large data sets involving methods at the intersection of artificial intelligence, machine learning, statistics, and database systems. In this paper we have taken a scenario or proposal on how we are going to work on student data set to predict the unknown values from interest value by using random forest tree algorithm. Therefore overall in this paper we have given flowchart of our work that if implemented tree algorithm to predict the unknown values from student data set.

REFERENCES

[1]. Gyozo Gidofalvi (2007). Privacy-Preserving Data Mining on Moving Object Trajectories in International Conference on Mobile Data Management. 60-68.

[2]. Isaac Cano (2009). Generation of Synthetic Data by means of fuzzy c-Regression in FUZZ-IEEE. 20-24.

[3]. Isaac Canoand Susana Ladra (2010). Evaluation of Information Loss for Privacy Preserving Data Mining through comparison of Fuzzy Partitions inIEEE International Conference on Fuzzy SystemS. 1-10.

[4]. Weijia Yang (2010). A novel anonymization algorithm: Privacy protection and knowledge preservation in Expert Systems with Applications 37 (1). 756–766.

[5]. Jingjing Qiang (2011). Privacy-Preserving SVM of Horizontally Partitioned Data for Linear Classification in International Congress on Image and Signal Processing. 23-27.

[6]. Khaled Alotaibi (2012). Non-linear Dimensionality Reduction for Privacy-Preserving Data Classification in ASE/IEEE International Conference on Social Computing and 2012 ASE/IEEE International Conference on Privacy, Security, Risk and Trust. 334-337.

[7]. Kshitij Pathak (2012). Privacy Preserving Association Rule Mining by Introducing Concept of Impact Factor in IEEE Conference on Industrial Electronics and Applications. 1458-1461.

[8]. Majid Bashir Malik (2012). Privacy Preserving Data Mining Techniques: Current Scenario and Future Prospects in Third International Conference on Computer and Communication Technology. 26-32.

[9]. Lei Xu (2014). Information Security in Big Data: Privacy and Data Mining in IEEE Publications 2 (1). 1149-1176.

[10]. P.Usha (2014). Sensitive Attribute based Non-Homogeneous Anonymization for Privacy Preserving Data Mining in International Conference on Information Communication and Embedded Systems. 1-5.

[11]. N P Nethravathi (2015). CBTS: Correlation Based Transformation Strategy for Privacy Preserving Data Mining in IEEE International WIE Conference on Electrical and Computer Engineering. 190-194.

[12]. Qi Jia (2016). Privacy-preserving Data Classification and Similarity Evaluation for Distributed Systems in IEEE 36th International Conference on Distributed Computing Systems. 690-699.

[13]. Sunil Dutt Jha. 5 data mining techniques for optimal results. http://www.computerweekly.com/tip/5-data-mining-techniques-for-optimal-results. Accessed on July, 2016.

_____