A Review on Data mining techniques

Twinkle Chawla

School of Computer Science & Engineering,
Lovely Professional University, Punjab, India.

Abstract- In this paper, the idea of data mining was sum up and its importance towards its methodologies was described along with their algorithms. In this paper the techniques and the applications of data mining was surveyed in detail. In this paper we also survey that how you can extract large data sets with the help of data mining. It has been seen that methods of data mining vary according to their data sets. In this paper our main focus on techniques that how you can apply different techniques on huge data sets and extract the fruitful information from that data.

1. Introduction

Data mining is searching for unseen data, proper and possibly patterns in immense informational indexes. Information Mining is tied in with finding unspecified connections among the information. It infers investigating information design or pattern in enormous group of information utilizing at least one of the programming software like R tool. It has different applications in dissimilar fields, similar to science and research. Mining of data includes powerful information collection and warehousing just as computer processing. The additional name of Data mining is Knowledge Discovery in Data (KDD).

Some Essential features of Data mining are

- Automatic pattern prediction are build on trend and their behaviour.
- Prediction dependent on their likely results.
- The creation of information or data is decision oriented and that is very useful to all.
- The data mining concentrate on their huge data set and their databases for investigating so that meaningful information can be extracted.
- Clustering dependent on finding and visually archived gathering of facts not recently known.

Process of data mining: The entire procedure of information mining can’t be finished in a single step. As such, you can’t get the necessary data from the huge volumes of information. It is a very perplexing procedure. The procedure of this include various things like cleaning of information, integration of data, the choice of information, how you can transform the data and finally the evaluation of different pattern that comes from huge data sets and these steps are to be completed in the same order so as to get the best results.

Types of data mining process

Various process of information mining can be categorized into two types: the first one is preparation of data and the second one is pre-processing of huge data sets. In fact, the cleaning of data, integration of data, the choice of information and the transformation of data are considered as the preparation of data. And then the mining of data and the evaluation of different patterns are integrated into one process that is called as data mining.
a) **Data Cleaning**: Information cleaning is where the information gets cleaned. Information in reality is regularly insufficient, boisterous and conflicting of data. The accessible of information from various different sources may be inadequate along with their properties. So main our goal is the cleaning of data to get the effective results. If the data is not clean then the results would neither be valuable nor accurate.

b) **Data Integration**: The integration of data is where information from various sources are incorporated into one. Different information lies in various arrangements in various areas. That gathered information can be put in various databases, spreadsheets, records etc. It is very complicated process because we are collecting the data from various sources.

c) **Data Selection**: This process requires huge volumes of verifiable information for examination. From the accessible information, our interests of information should be chosen and stored. After that investigation of that data is done.

d) **Data transformation**: The transformation of data is the way to change the information into various structures that are appropriate for mining.

e) **Data mining**: Information mining is the centre procedure where various techniques are applied to extricate designs from information. Different techniques like classification, clustering, association, neural networks and many more techniques.

2. **Literature Survey**

<table>
<thead>
<tr>
<th>Sr.no</th>
<th>Year</th>
<th>Authors</th>
<th>Topic</th>
<th>Methods adopted</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>2015</td>
<td>Bo, Guo, Rui Zhang, Guang Xu, Chuangming Shi and Li Yang</td>
<td>Predicting Student performance in data mining</td>
<td>A study of prediction system is used by deep learning approach</td>
</tr>
<tr>
<td>2.</td>
<td>2013</td>
<td>Nikita Jain, Vishal Srivastava</td>
<td>Data Mining Techniques</td>
<td>A study of different techniques like neural networks, decision trees, Genetic Algorithm and Rule extraction.</td>
</tr>
</tbody>
</table>
3. Methodologies of Data Mining

**Classification**: This analysis is utilized to retrieve significant and the relevant data about information and metadata. This information mining strategy assists with ordering information in various classes. The objective of classification is to precisely predict the objective class for each case in the information.

**Clustering**: This technique is used to put the information components into their related gatherings. Clustering is the way towards dividing the information or items into similar class and the information in one class is more like each other than to those in other group and the way of dividing the information objects into their subclass is called as group.

**Regression**: This method of mining is used to find a range of numeric qualities and they are additionally called as continuous values, given a specific data set. Regression is the data digging method for recognizing and dismembering the association between factors. It is utilized to recognize the likelihood of a specific variable.

**Outer**: Outer detection is a sort of information mining framework that insinuates impression of valuable things in a huge data sets which don’t coordinate the expected pattern of different set of data. This method can be used in a number of ways like intrusion, detection or fault detection, etc.

**Sequential patterns**: The function of sequential pattern mining is a information mining task for scrutinize continuous data to locate useful pattern of gathered information or huge data sets. The main goal of this process is finding those words which are used again and again and then they are put in one group so as to get the better results.
**Prediction:** Prediction is not anything but it is the process of discovery or some of the repeatable pattern from the large volumes of information. This process is basically the aggregation of many techniques and then they figure out the past events and used them with the help of different algorithms to predict the future event.

**Association Rules:** This technique is basically trying to figure out the correlation within the items and it mainly discovers the undisclosed pattern of the huge set of information. This type of process is used in our daily life so as to get more benefit. Real life examples like market-based analysis and many more.

4. **Conclusion**

Data mining is an extremely wide area that include various types of techniques and algorithms for finding the fruitful information from the large volumes of data. In this paper our main focus on the understanding the techniques of information mining so that they can be put in a different volume of data. The techniques can be applied in market-based analysis, finding the student performance, used in healthcare system, web based applications and many more.

**References**


