Research Review
The Refereed & Peer Review International Journal
www.researchreviewonline.com
ISSN: 2321- 4708
Nov. 2021, Year - 8 (103)
Paper ID: RRJ737147

DATA MINING

Concepts and Techniques

By
Mr. Junaid Hussain Wani
Contractual Lecturer
Govt. Degree College D. H. Pora
Anantnag, Jammu & Kashmir

Abstract

In the digital era we are overwhelmed with data. The growth of data in the world seems ever-increasing and there's no end in sight. Therefore, Data mining is a method of extracting information through from big data to produce an information pattern which can be used for further analysis to help in decision making.

Keywords: data mining; data mining techniques; Challenges, issues.

I. Overview of Data Mining

The development of Information Technology has generated large amount of databases and huge data in various areas. The technology has given rise to an approach to store and manipulate this precious data for further decision making. This Precious data can have extracted from large amount of database with the help of data mining technology to turn raw data into useful information. Data mining is used to analyze and explore large amounts of data to find a valuable result from the extraction.

There are a lot of algorithms and techniques which can be used here. The goal of this technique is to find patterns that were previously unknown. These patterns can further be used to make certain decisions for development of their businesses.

Three steps involved are

- Exploration
- Pattern identification
- Deployment

Exploration: In the first step of data exploration data is cleaned and transformed into another form, and important variables and then nature of data based on the problem are determined.

Pattern Identification: Once data is explored, refined and defined for the specific variables the second step is to form pattern identification. Identify and choose the patterns which make the best prediction.

Deployment: Patterns are deployed for desired outcome.

Publishing URL: : http://www.researchreviewonline.com/issues/volume-8-issue-103-noember-2021/RRJ737147

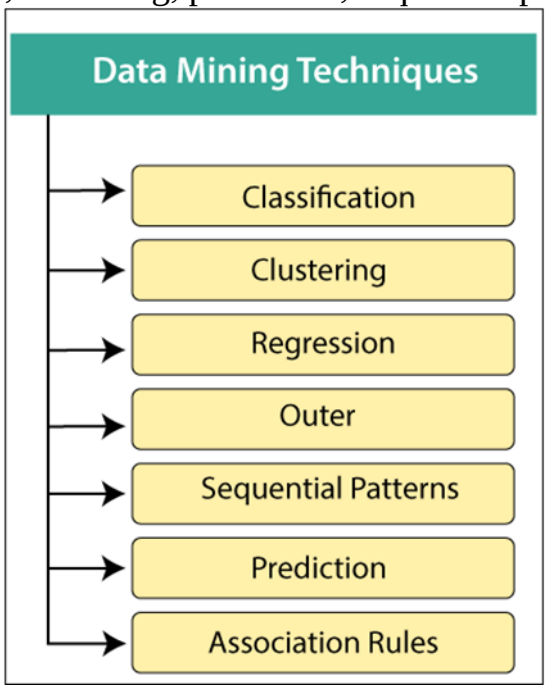
ISSN: 2321- 4708
Nov. 2021, Year - 8 (103)
Paper ID: RRJ737147

The high level primary goals of data mining are as follows.

1. The descriptive function deals with the general properties of data in the database such as Class Description, Frequent Patterns, Associations, Correlations and Clusters as well.
2. Classification is that process for finding a model that describes the classes and concepts of data. This model used to predict the class of objects whose label is unknown.
3. Prediction is used to predict missing and unavailable numerical data values rather than class labels during data mining process

Data Mining Techniques:

In recent data mining following techniques have been developed and used, including association, classification, clustering, prediction, sequential patterns, and regression.



Classification analysis:

Classification is the most commonly applied data mining technique, which employs a set of pre-classified to develop a model that can classify the population of records at large. This analysis is used to retrieve important and relevant information about data, and metadata. It is used to classify different data in different classes.

Data Cleaning:

Data cleaning and preparation is a vital part of the data mining process. Raw data must be cleansed and formatted to be useful in different analytic methods.

Publishing URL: : http://www.researchreviewonline.com/issues/volume-8-issue-103-noember-2021/RRJ737147

ISSN: 2321- 4708
Nov. 2021, Year - 8 (103)
Paper ID: RRJ737147

Clustering:

Clustering involves grouping chunks of data together based on their similarities. This technique helps to recognize the differences and similarities between the data. Clustering mechanisms use graphics to show where the distribution of data is in relation to different types of metrics. By using clustering techniques we can further identify dense and sparse regions in object space and can discover overall distribution pattern and correlations among data attributes.

Regression:

Regression techniques are useful for identifying the nature of the relationship between variables in a dataset. Regression technique can be adapted for predication. Regression analysis can be used to model the relationship between one or more independent variables and dependent variables.

Outlier detection

Outlier detection determines any anomalies in datasets. Once organizations find aberrations in their data, it becomes easier to understand why these anomalies happen and prepare for any future occurrences to best achieve business objectives.

Outlier detection is valuable in numerous fields like network interruption identification, credit or debit card fraud detection, detecting outlying in wireless sensor network data, etc.

Sequential Patterns:

this technique of data mining helps to discover or recognize similar patterns in transaction data over some time. Understanding sequential patterns can help organizations recommend additional items to customers to spur sales.

Prediction:

Data mining that is done for the purpose of using business intelligence or other data to forecast or predict trends. Prediction is a very powerful aspect of data mining that represents one of four branches of analytics. Predictive analytics use patterns found in current or historical data to extend them into the future. Thus, it gives organizations insight into what trends will happen next in their data.

Association rule

Association and correlation is usually to find frequent item set findings among large data sets. This type of finding helps businesses to make certain decisions, such as catalogue design, cross marketing and customer shopping behavior analysis. Association Rule algorithms need to be able to generate rules with confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high proportion of the rules are usually of little (if any) value.

Types of association rule

· Multilevel association rule

Publishing URL: : http://www.researchreviewonline.com/issues/volume-8-issue-103-noember-2021/RRJ737147

ISSN: 2321- 4708
Nov. 2021, Year - 8 (103)
Paper ID: RRJ737147

· Multidimensional association rule
· Quantitative association rule

DATA MINING APPLICATIONS

The predictive capacity of data mining has changed the design of business strategies. This technology is popular with many businesses because it allows them to learn more about their customers and make smart marketing decisions. Here is some examples of data mining applications.

1. BANKS:

Banks use data mining to better understand market risks. It is commonly applied to credit ratings and to intelligent anti-fraud systems to analysis transactions, card transactions, purchasing patterns and customer financial data.

Data mining also helps banks better understand their customers' online habits and preferences, which helps when designing a new marketing campaign.

2. Marketing:

Data mining is used to explore increasingly large databases and to improve market segmentation. By analyzing the relationships between parameters such as customer age, gender, tastes, etc., it is possible to guess their behavior in order to direct personalized loyalty campaigns. Data mining in marketing **also predicts which users are likely to unsubscribe from a service,** what interests them based on their searches, or what a mailing list should include to achieve a higher response rate.

3. Healthcare:

Having all of the patient's information, such as medical records, physical examinations, and treatment patterns, allows more effective treatments to be prescribed. Data mining helps doctors create more accurate diagnoses by bringing together every patient's medical history, physical examination results, medications, and treatment patterns.

4. E-commerce

Many E-commerce companies use Data Mining and Business Intelligence to offer cross-sells and up-sells through their websites. One of the most famous of these is, of course, Amazon, who use sophisticated mining techniques to drive their, 'People who viewed that product, also liked this' functionality.

Concept of Big Data

Big Data is the source, variety, volume of the data and how to store and process this amount of data. Big Data analytics and data mining are not the same. Both of them involves the use of large data sets, handling the collection of the data or reporting of the data which is mostly used by business. When data is quite large preferably in the scale of petabyte then four important aspects are associated with it get predominantly high.

Volume

Quantity of data generated. With big data, you'll have to process high volumes

Publishing URL: : http://www.researchreviewonline.com/issues/volume-8-issue-103-noember-2021/RRJ737147

ISSN: 2321- 4708
Nov. 2021, Year - 8 (103)
Paper ID: RRJ737147

of unstructured data. This can be data of unknown value, such as Twitter data feeds, clickstreams on a webpage or a mobile app, or sensor-enabled equipment.

• Velocity

Velocity is the rate at which data is generated like Social media posts generated every sec.

• Variety

Variety refers to data generated in different formats. Traditional data types were structured and fit neatly in a relational database. With the rise of big data, data comes in new unstructured data types. Unstructured and semi structured data types, such as text, audio, and video require additional preprocessing to derive meaning and support metadata.

CHALLENGES AND ISSUES

A. Challenges

Efficient and effective data mining in large data bases poses numerous requirements and great challenges to researchers and developers.

The issues involved include data mining methodology, user interaction, performance and scalability, and the processing of a large variety of data types. Other issues include the exploration of data mining applications and their social impacts.

*B. **Issues***

- ***Mining Methodology***
  - ✓ Mining different kind of knowledge from diverse data type. Because different users may be interested in different kind of knowledge.
  - ✓ Visualization and representation of data should be easily understandable.
  - ✓ Handling noise and incomplete data.

• Information poorness

▪The abundance of data, coupled with the need for powerful data analysis tools, has been described as data rich but information poor situation

▪Data collected in large data repositories become “data tombs”

• Decision Making

• The decision maker does not have the tools to extract the valuable knowledge embedded in the vast amounts of data.

CONCLUTION

The main goal and the contribution is to demonstration and the implementation of the data mining approach in the field of business. The data mining approach enables to make strategic decisions and data mining technology play a significant role to improve business performance. Data mining has wide application domain almost in every industry where the data is generated that’s why data mining is considered one of the most important frontiers in database and information systems and one of the most

Publishing URL: : http://www.researchreviewonline.com/issues/volume-8-issue-103-noember-2021/RRJ737147

ISSN: 2321- 4708
Nov. 2021, Year - 8 (103)
Paper ID: RRJ737147

promising interdisciplinary developments in Information Technology

REFERENCES

[1] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining: Concepts and Techniques", published by Morgan Kauffman, Wyman Street, Waltham,
[2] Ian H. Witten, Eibe Frank, "Data Mining Practical Machine Learning Tools and Techniques", 500 Sansome Street, Suite 400, San Francisco,
[3] Dr. Gary Parker, vol 7, 2004, Data Mining: Modules in emerging fields.
[5] Nirmal Kaur, Gurpinder Singh," A Review Paper On Data Mining And Big Data", ISSN No. 0976-5697, Jalandhar, Punjab, India, 2017
[4] Hossin, M. and Sulaiman, M. N (2015) International journal of data mining and knowledge management process.