



“DIFFERENT DATA MINING TECHNIQUES USED FOR DATA ANALYSIS”

¹ Mr. Dhanaji P. Bhanvase , ² Dr. Sushilkumar R. Kalmegh

¹ Research Scholar

² Associate Professor

¹ Department of Computer Science, ² PG Department of Computer Science & Engineering

¹ Vidnyan Mahavidyalaya Sangola, Sangola, India ² SGBAU, Amravati Amravati India

Abstract : Data mining techniques useful in research area i.e. researchers and individuals to extract valuable information from amount of data form different data set. The process of extracting information from huge sets of data to identify patterns, that data can useful in different business to take the decision. Data analysis is the process of simply converting the gathered data in to meaningful information. The huge amount of data available is in the Information Industry. This type of data is not useful until it is converted into useful form. These huge amounts of data are necessary to analyze and extract in to useful form. This paper work consists of analysis of data as well as different data mining techniques are summarized.

IndexTerms - Data mining, Data mining techniques, Analysis, KDD.

I. INTRODUCTION

Data Mining is the process of identifying hidden information in various areas for categorization into useful data, which are collected and assembled in particular areas such as data warehouses, data mining algorithm, analysis, decision making and other data requirement for cost-cutting and generating revenue. Data mining is the act of automatically searching for huge amount data and large storage of information to find patterns and trends that go beyond simple analysis procedures. [2]

1.1 KNOWLEDGE DISCOVERY IN DATABASE (KDD):

Data mining is also called Knowledge Discovery in Database (KDD). The knowledge discovery process includes Data mining, Data selection, Pattern evaluation, Data cleaning, Data integration, Data transformation, and Knowledge presentation. [1].

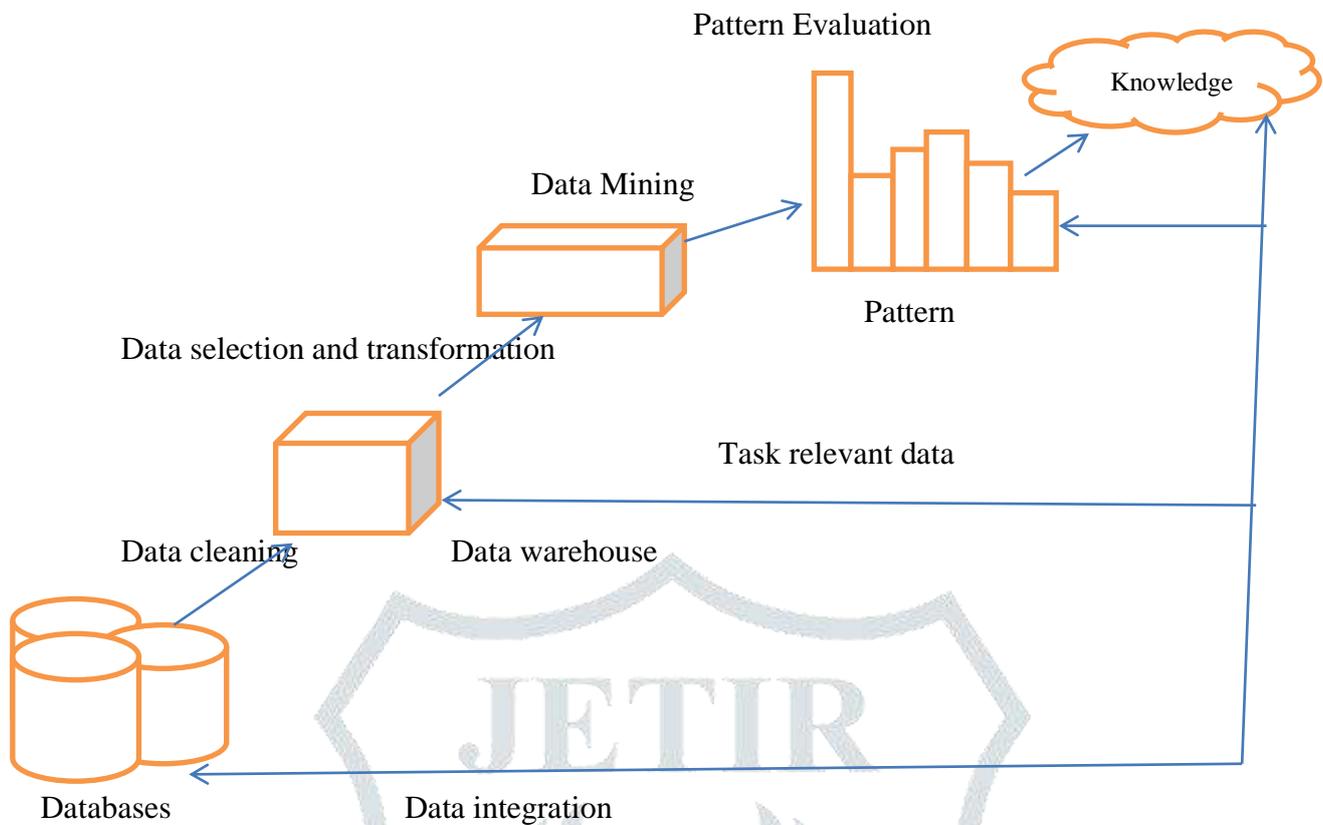


Diagram: Knowledge Discovery in Database (KDD) process

- I. **Data Cleaning:** Data cleaning is defined as removal of irrelevant data from collection.
- II. **Data Integration:** Data integration is defined as heterogeneous data from multiple sources combined in a common source.
- III. **Data Selection:** Data selection is defined as the process where data relevant to the analysis is decided and retrieved from the data collection.
- IV. **Data Transformation:** It is defined as the process of transforming data into its appropriate form required by mining procedure.
- V. **Data Mining:** data mining is the techniques that are applied to extract data patterns useful form.
- VI. **Pattern Evaluation:** Pattern Evaluation is representing knowledge based on given measures defined as identifying strictly increasing patterns.
- VII. **Knowledge representation:** Knowledge representation is the technique which Utilizes visualization tools to represent data mining results

II. Data Analysis

Data analysis is a method that can be used to investigate, analyze, and demonstrate data to find useful information. The data analysis output is a verified by using different hypothesis based on the data. There are different types of data analysis such as - text analytics, predictive analysis, data mining, etc. analytical methods are responsible for developing models and testing. [7]

2.1 Methods of Data Analysis

There are two methods of data analysis: qualitative and quantitative.

Qualitative research:

Qualitative research describes the product characteristics. It does not utilize any number. It emphasizes the quality of the product.

Quantitative research:

Quantitative research is the inverse of qualitative research because its primary focus is numbers. Quantitative research is all about quantity.

III. Data Mining Techniques

Different data mining techniques includes the utilization of purified data analysis tools to find previously unknown, relationships and valid patterns in huge data sets. These tools can include statistical models, machine learning techniques, and mathematical algorithms, such as neural networks or decision trees. Thus, data mining includes analysis and prediction. Depending on different methods and technologies related to machine learning, database management, and statistics, professionals in data mining have devoted their careers to better understanding how to process and make conclusions from the huge amount of data. In recent data mining projects, different data mining techniques has been developed and used, such as association, classification, clustering, prediction, sequential patterns, and regression. [7]

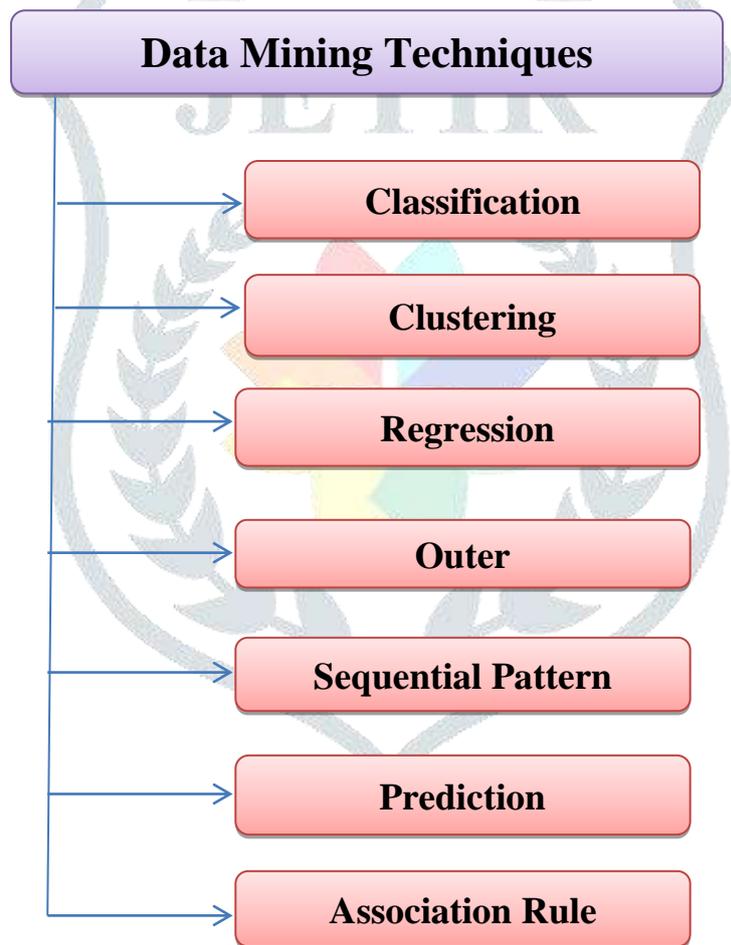


Diagram: - Data Mining Techniques

3.1 Classification:

It is used to obtain important and meaningful information about data and metadata. To classify data in different classes data mining technique are used. Classification techniques are used to classify different data records into one set of predefined classes. The different types of classification models can be used for Classification as like decision tree induction, Bayesian Classification, Neural Networks, Support Vector Machines (SVM), and Classification Based on Associations etc. Classification techniques are used in different applications such as detection, credit card fraud detection, and speech recognition and computer vision etc.

Data mining techniques can be classified by different ways, as follows:

3.1.1 Classification of Data mining frameworks as per the type of data sources mined:

It can be classified as per the type of data handled. For example, World Wide Web, text data, time-series data, multimedia, spatial data and so on.

3.1.2 Classification of data mining frameworks as per the database involved:

This classification depends upon different data model involved. For example. Relational Database, Object-oriented database, transactional database and so on.

3.1.3 Classification of data mining frameworks as per the kind of knowledge discovered:

This classification based on the types of knowledge discovered or data mining functionalities. For example, clustering, discrimination, classification, characterization, etc. some frameworks offering the extensive frameworks of few functionalities of data mining.

3.1.4 Classification of data mining frameworks according to data mining techniques used:

The data analysis approach utilized by the classification as per, such as machine learning, genetic algorithms, neural networks, data warehouse-oriented or database-oriented visualization, statistics, etc.

3.2. Clustering:

This data mining technique used for to identify the similar data. This technique used for to distinguish the differences and similarities between the data. This technique is very similar to the classification technique, but it can be involved piece of information together based on their similarities. Clustering can be defined as the task of organizing data into groups known as clusters such that the data objects that are similar to each other are put in the same cluster. There is no one correct basis of clustering; there could be many different ways to categorize data objects. Clustering is a form of unsupervised learning in which no class labels are provided. [3]

3.3 Regression:

Regression analysis is the process of data mining which is used to identification and analysis of the relationship between variables because it might be presence of the other factor. This analysis used to define the probability of the specific variable. Primarily, regression can use for planning and modeling. For example, it uses the project certain costs, also depending upon the other factors such as consumer demand, availability, and competition. Primarily it contains relationship between two or more variables in the given data set.

3.4 Association Rules:

The Association Rules technique used to identify the link between two or more items. This technique used for finds out the hidden pattern in the data set. These technique are uses if-then statements, that statement support to show the probability between data items within large data sets in different types of databases.

3.5 Outer detection:

These types of techniques are used for the different observation of data items in the data set; it cannot be match an expected pattern or expected behavior. This type of technique may be used in different domains as like fraud detection, detection, intrusion, etc. this technique is also known as Outlier Analysis or Outlier mining. The outlier is a data point that separate the too much from the rest of the dataset. The huge amounts of the real-world datasets have an outlier. it plays an important role in the data mining field. Outlier detection is valuable in number of fields such as detecting outlying in wireless sensor network data, network interruption identification, credit or debit card fraud detection, etc.

3.6 Sequential Patterns:

To discover sequential patterns for evaluating sequential data uses different data mining techniques. It compares to find out the interesting subsequences in a set of sequences, where the sequence can be measured in terms of different criteria such as occurrence frequency and length. In other words, this data mining technique helps to find or realize similar patterns in transaction data over some time.

3.7 Prediction:

In data mining Prediction techniques are used for combination of trends, clustering, classification, etc. It can use for analysis of past events or instances in the right sequence to predict a future event. This technique show how certain attributes within the data will behave in future. For example, depends on the analysis of buying transactions by customers. Regression is used to map a data item to a real valued prediction variable [5] Regression analysis can be used to model the relationship between one or more independent variables and dependent variables. Basically prediction models are continuous valued functions that are used to predict missing or unavailable numerical data values rather than class labels. Prediction also encloses the identification of distribution trends based on the available data. Regression analysis is a statistical methodology that is most often used for numeric prediction. Various types of regression methods are used like Linear Regression, Multivariate Linear Regression, Nonlinear Regression, and Multivariate Nonlinear Regression [3]

IV. Conclusion

- I. Data mining definition: The term data mining is all about explaining the past data and predicting the future data via Data analysis.
- II. Different data mining techniques helps to extract information from large sets of data. It is the procedure of mining knowledge from data.
- III. Different important Data mining techniques are Clustering, Association rules, Classification, Sequential Patterns, Regression, Outer detection, and prediction.
- IV. Data mining technique helps to different origination to get knowledge-based information.
- V. Data mining is used in different industries such as Communications, Insurance, Education, Manufacturing, Banking, Retail, Service providers, e-Commerce, Supermarkets Bioinformatics, Agriculture, weather forecasting.

References

1. Jiawei Han and Micheline Kamber, Data Mining Concepts and Techniques, published by Morgan Kaufman, 2nd e, (2006).
2. Han. J, Kamber. M, Pei. J, “ Data Mining Concepts and Techniques”, Third edition The Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers, July 2011
3. Anshu , International Journal of Innovative Research in Computer Science & Technology (IJIRCST)ISSN:2347-5552,Volume-7,Issue-2,March2019DOI: 0.21276/ijircst.2019.7.2.4
4. Dr. M.H.Dunham, “Data Mining, Introductory and Advanced Topics”, Prentice Hall, 2002.
5. Fathimath Zuha Maksood, Geetha Achuthan, International Journal of Computer Applications (0975 – 8887)Volume 140 – No.3, April 2016
6. Bharati M. Ramageri, Indian Journal of Computer Science and Engineering Vol. 1 No. 4 301-305
7. <https://www.javatpoint.com/data-analysis>