



DATAMINING AND ITS APPLICATIONS

¹Athira Raj

Assistant Professor,

¹Department of Computer Science
JPM Arts and Science College, Labakkada, India

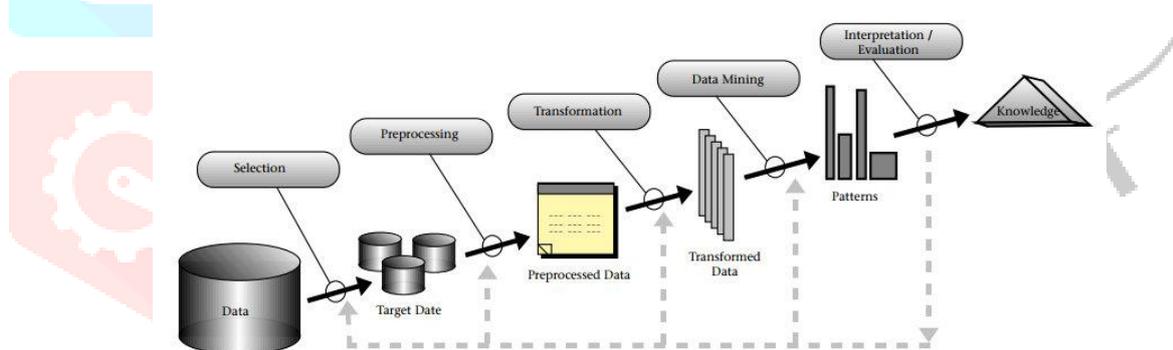
Abstract: - Data mining is a process that finds useful patterns from a large amount of data. The paper discusses a few of the data mining techniques and algorithms

Index Terms - Data mining Techniques, Datamining algorithms, Data mining applications

I. INTRODUCTION

The development of Information Technology has generated a large number of databases and huge data in various areas. The research in databases and information technology has given rise to an approach to store and manipulate this precious data for further decision making. Data mining is the process to extract useful knowledge from a large set of data. Data mining is also known as Knowledge Discovery in Database (KDD). It has many applications in the field of science, mathematics, marketing, etc.

Figure 1. Knowledge discovery Process



Datamining aims to find previously unknown patterns. Once these patterns are found they can further be used to make certain decisions for the development of their businesses.

There are mainly three steps involved in this process

- Exploration
- Pattern identification
- Deployment

1.1 Exploration:

This is the initial step of data mining. It analyses missing data, outlier treatment, etc

1.2 Pattern Identification:

After the exploration of data, the second step is to form pattern identification. Identify and choose the patterns which make the best prediction.

1.3 Deployment:

Patterns are deployed for desired outcome.

II.MAJOR DATAMING ALGORITHMS AND TECHNIQUES

Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithms are major techniques used for knowledge discovery from databases.

2.1. Classification

Classification is the most commonly used data mining technique. It classifies data based on the training set and values in a classifying attribute and uses it for classifying new data.

For example, the classification model can predict the credit risk of loan applications as low, medium, or high.

Types of classification models:

- Classification by decision tree induction
- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Associations

2.2. Clustering

Clustering is the process of grouping similar objects together. The main advantage of clustering over-classification is that it is adaptable to changes. clustering is used in many applications such as market research, data analysis, image processing, etc.

Types of clustering methods

- Partitioning Methods
- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based methods

2.3. Predication

Prediction is the process of recognizing the missing or not available numerical data for a process. in prediction the sample of data can be called a predictor and the authenticity depends on how the predictor. The major method of prediction is regression. Types of regression methods are,

- Linear Regression
- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

2.4. Association rule

Association rule mining finds interesting associations and relationships among large sets of data items. This rule shows how frequently an itemset occurs in a transaction. A typical example is a market-based analysis

Types of association rule

- Multilevel association rule
- Multidimensional association rule
- Quantitative association rule

2.5. Neural networks

A neural network is a set of connected input/output units and each connection has a weight present with it. Neural networks have the ability to derive meaning from complicated data and can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. These are well suited for continuous-valued inputs and outputs. For example, handwritten character reorganization and many real-world business problems Neural networks are best at identifying patterns or trends in data and well suited for prediction or forecasting needs.

Types of neural networks

- Back Propagation

III.APPLICATIONS OF DATA MINING

Data mining offers many applications in various fields. Such as retail stores, hospitals, banks, and insurance companies, the telecommunication industry, biological data analysis, other scientific applications, etc. Data mining can be used to find patterns and connections that would otherwise be difficult to find. This technology is popular with many businesses because it allows them to learn more about their customers and make good marketing decisions

IV.CONCLUSION

Data mining is the process used to extract unknown knowledge from a larger set of data. Data mining involves effective data collection and warehousing. The major steps involved in a data mining process are extracted, transform and load data into a data warehouse, store and manage data in a multidimensional database, provide data access to business analysts using application software and present analyzed data in easily understandable forms.

ACKNOWLEDGEMENT

Gratitude is a feeling which is more than words. It gives me immense of pleasure and satisfaction in submitting this paper **DATAMINING AND ITS APPLICATION** . In under taking this paper publication I need the direction, Assistance , cooperation of various Individuals. I would like to express my sincere gratitude to my college management , parents and my friends for their support.

REFERENCES

- [1]. The graph taken from KDD process, <https://www.google.com/data-science-process-rediscovered.html>
- [2]. Points based on the observation from www.researchgate.net
- [3]. Association rules taken from <http://www.geeksforgeek.com>
- [4]. Neural networks taken from www.getsmarter.com