IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9665

# DATA MINING: HISTORY APPLICATIONS AND CHALLENGES

**MADHUSHREE. D. S[1], Vidya. S[2]**

[1] Student, Department of MCA, Bangalore Institute of Technology

[2]Asst. Professor, Department of MCA, Bangalore Institute of technology.

**Abstract:** The Universe is made up of different kinds of data. The different types of data include Bigdata, Structured data, Unstructured data, Timestamp data Machine data, Open data, Dark data, Real time data and other forms of data. Analysis, Classifying and Summarizing all these kinds of data by humans is not possible due to the large increase of data in the present stage.This research paper provides the information regarding analysis of big data, Data Mining and Data Mugging, Challenges involved in Data Mining and Research issues regarding the Data Mining

.

**keywords** : Data, Data Mining, Machine Learning, Artificial Intelligence.

## I. INTRODUCTION

### WHAT IS DATA AND DATA MINING?

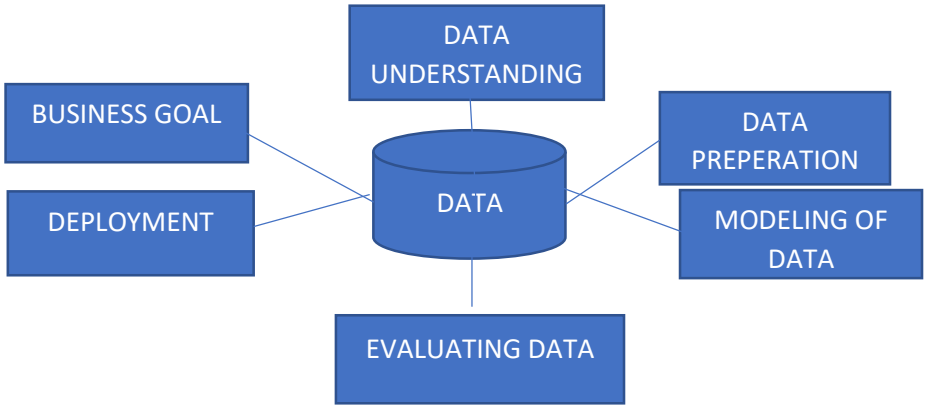
### A. Data

Data is nothing but the facts, statistics or information. In today's computing world data is some kind of information that can be converted into binary digital form. The large volume of data is known as Bigdata and is measured in terms of terabytes, yottabytes, petabytes, and zettabytes.

### B. DATA MINING

Data Science is an emerging field in recent years that has grabbed more attention due to the large increase of data in recent years[4].

In general Data Mining is a process used to extract the data and discover the patterns in big or large volume of data. The Data Mining involves different methods used for intersecting Statistics, Machine Learning and various Database System.[3] It is also used to find anomalies, co-relations and to predict outcomes or results.




## III. HISTORY OF DATA MINING

The term "Data Mining" was introduced in the 1990's. It has now become a great evolution sector in the history of Data Science.

In 1770's they made use of Bay's theorem for Data Mining as one of the techniques. In 1880's they made use of Regression Theorem for Data Mining as another technique. The rapid growth of data in terms of size and its complexity

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9665

level in computer science field has led to the discovery of many techniques such as neural networks, clustering, genetic algorithms(1950's), decision tree (1960), vector machines(1990's).

The Data Mining origin can be defined in three families which include [1] Classical Statistics [2] Artificial Intelligence [3] Machine Learning.
Classical Statistics are the basic technology for mining the data. The different Statistical techniques used for Data Analytics and Data Connection are Standard Deviation, Variance, Distribution, Intervals.
Artificial Intelligence is completely opposite to the Statistics. In Artificial Intelligence Human thoughts are used for processing problems on the statistics.
Relational Database Management System is the special concept used for solving the queries.
Machine Learning is the concept which includes both statistics as well as the Artificial Intelligence. It is an evolution of the Artificial Intelligence. It gives knowledge about the data that the computer programs will make use. By examining the characters of data, the decision is made by the computer programs.

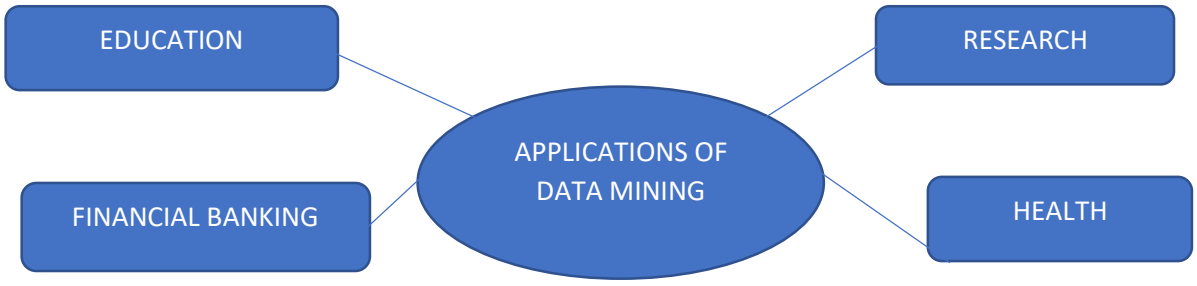
## IV. TASKS INVOLVED IN DATA MINING

There are two types of Data Mining tasks namely1. Predictive and2. Descriptive. Predictive task involves Classification, Prediction and Time Series Analysis [1]. Descriptive task includes Association, Clustering and Summarization.
Predictive task helps to predict the future values of another dataset. Descriptive task helps to describe the new patterns with the available dataset.
Classification is used to identify class of an object based on its attributes.[3] This is commonly used in Direct Marketing.
Prediction is a process where the future data is predicted by using the missing values. The development model can be done by using the available data or the existing data. For example, Fraud Detection.
Time Series Analysis is a series of events where one event is determined by the next event. The different methods are used to extract trends, patterns, rules and statistics.
Association is used to identify the relationship between any objects. Commonly used for designing the catalogues, advertisements, managing commodities. Clustering is a process where the data objects which are similar to each other are identified by using clustering.[2] It uses different factor for identifying the similar data. Summarization is generalization of the data. The results obtained here is in the smaller set that provides the aggregated data.

## V. CHALLENGES WORKING WITH THE DATA

**1. HETEROGENEOUS DATA (DIFFERENT FORMS OF DATA)**
Data collected can be of different forms. It can be low quality data or high-quality data.[2] The data can be adulterated also. This problem is due to collection or accumulation of big data from different sources. For example, Survey made on common people such that people may incorrectly submit the information such as age, date of birth or e-mail.
**2.DATA SECURITY**
It is one of the biggest challenges in Data Mining because the data is from either the ethical source or if it is protected also the data is mainly used for data mining and data mugging. The data theft can happen through weak encryption, weak encryption, invisible data and data password leaks.
**3.DOMAIN KNOWLEDGE**
This is one of the major challenge id Data Mining because without any background knowledge it is hard to dig the data.
**4.RESEARCH ISSUES IN DATA MINING**
The main important thing to focus here is observation, focus on different methods, algorithms and techniques by which the problems can be solved. Background knowledge is also necessary for solving different research problems.

## APPLICATIONS OF DATA MINING



IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 7.105 ⁜ Vol. 9, Issue 6, June 2022

DOI: 10.17148/IARJSET.2022.9665

**1.RESEARCH**

In Data Mining the research is done by using train test model. It is used for grouping, clustering and predicting the data.

**2. HEALTH CARE AND INSURANCE**

Pharmacy sector uses its new ideas and results. In Insurance sector data mining is used to predict the customer behaviour.

**3. TRANSPORTATION**

There will be diversity in transportation. Data Mining will analyse the loading patterns in Transportation.

4. FINANCIAL/BANKING

Data Mining helps to provide information regarding customers their loyalty and their credit card spending.

## RESULTS

In this paper we have made a survey on data and data mining. The techniques involved in the data mining, steps for performing data mining. It also gives information regarding the Research issues involved in data mining. The research issues include collecting the proper data, having proper domain knowledge. Data Security, Data Privacy has been a major challenge in the Data Mining. The survey is also made on different applications of Data Mining.

## CONCLUSION

This paper provides general information about the data and data mining. It also gives information regarding the history of data mining, the old techniques used. The paper also speaks about challenges involved in data mining. Future research will include the development of new techniques to overcome all these challenges.

## REFERENCES

1. Our Wolfson, A. Prasad Sista, Sam Chamberlain, and Yelena Yeshe, Updating and Querying Databases that Track Mobile Units, MA: Kluwer Academic Publishers, 1999.
2. Reynold Cheng, Dmitri Kalashnikov, Sunil Prabhakar, Evaluating Probabilistic Queries over Imprecise Data, UK: Elsevier Science Ltd, 2007.
3. Russell, S; Norvig, P. Artificial intelligence. A modern approach. 3rd ed. Upper Saddle River, NJ: Prentice-Hall; 2010.
4. Bibel, W.; Ertel, W.; Kruse, R.: Grandeurs Unsliced Intelligent. Eine praxisorientierte Einfühlung. 3rd ed.Wiesbaden: Springer; 2013.
5. Hennig, C., Meila, M., Murtagh, F., Rocci, R.: Handbook of Cluster Analysis. Chapman & Hall, London (2015)
6. Klein, H.U., Schäfer, M., Porse, B.T., Hasemann, M.S., Ickstadt, K., Dugas, M.: Integrative analysis of histone chip-seq and transcription data using Bayesian mixture models. Bioinformatics 30(8), 1154–1162 (2014)