THINK INDIA JOURNAL

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019

Data Mining with Big data

Mrs Harmanpreet kaur

Research Scholar Rayat Bahra University

Abstract In these days with the advancement emerging technologies to enter IT mainstream are big data with data mining. These two technologies are coming together to provide powerful results and making intelligent decisions in businesses. Big data is a blessing for an organisation that wants to store huge amount of structured ,unstructured and semi structured data. It has ability to store and process different types of data at a high speed which is stored at different locations. In this paper the data mining techniques are which help in organizations growth by finding different patterns to decide future trends has been discussed. Although there are couple of challenges and issues that come cross the path of integration between data mining and big data but we are able to cope up with these problems. This paper presents an overview of both the technologies and also introduces the data mining techniques, applications, type of big data, issues and challenges of data mining with big data.

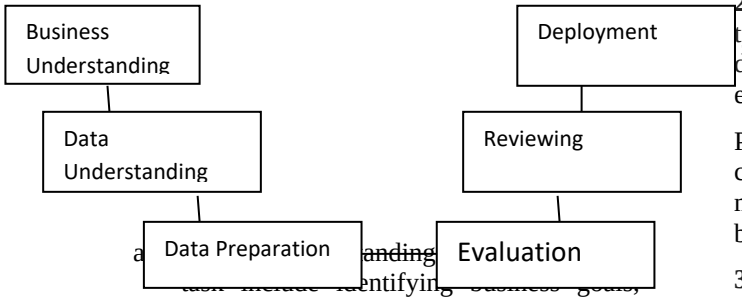
***Keywords*:** Big data, Data Mining, types of data, Techniques, Applications, and Challenges.

1 .Introduction

With the development and advancement of Information technology which leads to generation of huge amount of data and databases. The research in this field has given rise to manipulate and store important data for decision making. Data mining is basically a process of extracting important data from data warehouse. The result of extraction is the knowledge that we gain but not the data. Big Data is a term used to describe a collection of structured, unstructured and semi-structured data that is huge in size and growing exponentially with time. Big data is defined using five data characteristics: volume, velocity, variety, veracity and value.

Data mining is also termed as knowledge discovery process, analysis of pattern and knowledge extraction. It helps in predicting the hidden patterns, behaviours and helps business organizations in making critical decisions. It can be applied to various types of data such as database of time-series, multimedia data, data warehouse, World Wide Web etc.

The phases of data mining involves:



a) Business understanding: this phase task include identifying business goals, defining data mining goals etc.
b) Data understanding: in this phase includes the task such as describing the data, gathering, exploring and verifying the quality of data.
c) Data preparation includes the task such as selecting, clearing, constructing and integrating data.
d) Evaluation in this phase task includes evaluation of the result,
e) Reviewing the whole process and determining the next step.
f) Deployment it is the last phase which includes reviewing and reporting the final results.

This paper is organized as follows Section 2 provides data mining techniques; Section 3 discusses its applications; Section 4 presents types of big data; Section 5 presents issues and challenges; Section 6 presents the conclusion.

2. Data Mining Techniques

1. Classification: It is mostly applied technique which is used to retrieve important information about metadata. This method is used to classify data in different classes. Types of classification model Neural network, Support vector Machine, Decision tree, Bayesian classification.

2. Clustering: Clustering means to identify similar type of classes of objects. It helps to understand the differences and similarities between the data. For example, based on

Purchasing patterns we can make group of customers. Its methods includes Dense based method, Grid based, Partitioning Methods, Modal based etc.

3. Regression: This method of data mining is used to identify the relationship between variables and to identify the likelihood of a specific variable, given the presence of other variables. Methods include Linear regression, Non Linear regression, Multivariate Linear regression etc.

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019

4. Association Rules: Association and correlation is used to find frequent item set from large data sets. It helps businesses organization to make important decisions such as cross marketing and analysis the behaviour of customer shopping.

5. Prediction: It is the most important data mining technique. It is used to predict type of data that will be seen in future. Prediction has used a combination of the other data mining techniques like trends, sequential patterns, clustering, classification, etc. It uses regression technique which is used to build relationship between independent variables and dependent variables. Here, the independent variables are the attributes that we know and what we want to predict is the response variables.

Data Mining Techniques	
Classification	Neural network, Support vector Machine, Decision tree, Bayesian classification, Genetic Algorithm.
Clustering	Dense based method, Grid based, Partitioning Methods, Modal based.
Regression and Prediction	Support vector machine, Decision Tree , Rule Induction
Association rules	Association Rule Mining

3. Data Mining Applications

Data mining is used by many companies on regular basis with strong customer focus. It is a new technology that is not full matured. Despite this many organizations are using it that includes retail stores, hospitals, banks, telecommunication industries and insurance companies.

1. **Health care:** In health care sector data mining uses analytics and data to identify the best practise in order to reduce cost and improve the service. It can be used to predict the volume of patients in different categories and to discover the relationship between diseases and treatments to ensure patients receive timely and proper care.
2. **Education:** Education data mining is a new emerging field. The main goal of EDM is to predict future learning behaviour of students', effect of educational support and increasing the knowledge. Institutions use it in order to take correct decisions and to predict result of students. By this institution can focus on how and what to teach.
3. **Telecommunication:** Due to the development of new communication technologies, the telecommunication industry is expanding day by day. This industry provides various services such as internet, fax, cellular, email etc. Data mining helps in identifying the telecommunication patterns, detect the fraudulent activities, make use of the resources in better way, and improve service quality.
4. **Retail industry:** Data Mining has important role to play in Retail Industry as it collects huge amount of data from sales, purchasing history of customer, transportation of goods and services. It can help in identifying buying patterns of customer and trends which can lead to improved quality of services and customer satisfaction.
5. **Financial Data Analysis:** With computerised banking huge amount of data is generated with every new transaction. Data mining can help in solving problems in banking and finance by finding patterns, and correlations in business information and market prices that are not immediately apparent to managers because the volume data is too large. The managers use this information for targeting, retaining, acquiring and maintaining a profitable customer.

4. Types of Big data

1. **Structured-** It can be stored, accessed and processed in the fixed format. It refers to highly organised data that can be stored and accessed from a database by simple search engine algorithms.
   Example the employee table contains employees details such as salary, designation, Emp Id , personal details etc. in a company db will structured
2. **Unstructured**- Data with unknown form is classified as unstructured data. It very difficult and time consuming to process and analyze unstructured data. A typical example of unstructured data is a heterogeneous data source containing a combination of simple text files, images, videos etc.
   Semi-Structured- Semi-structured data contain both the forms of data. We can see semi-structured data as a structured in form but it is actually not defined with e.g. a table definition in relational DBMS. Example of semi-structured data is a data represented in an XML file.

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019

5. **Issues and Challenge of data mining with big data**

The main issues and challenges of data mining in big data are:

a) The size of data is inadequate and the data is noisy.
b) The algorithms that are used in data mining are not so effective.
c) Processing of unstructured data into structured data is very difficult task.
d) The security violation includes modifying information by unauthorized user, release of information through unauthorized user and last one is the denial of the resources.
e) Another major challenge is the increased communication cost as compared to the data processing cost.
f) It is becoming very difficult task to find user-friendly visualizations.

6. Conclusion

Data is going to continue growing every year and will be becoming more large and complex. So, different techniques are proposed by researchers' in order to solve big data challenges. We need high speed computing paradigm for data mining to solve the problem of big data.

Data mining techniques such as classification, clustering, prediction, association rules and regression are discussed in this paper which will help in business growth by finding different patterns to decide future trends. So, the big data mining will help us to discover knowledge that no one has discovered.

g) Unauthorized release of information, unauthorized modification of information and denial of
h) resources are the three categories of security violation
i) Unauthorized release of information, unauthorized modification of information and denial of
j) resources are the three categories of security violation

References

[1] Madden Sam • Massachusetts Institute of Technology , "*Database to Big Data*",IEEE,2012.

[2] *https://www.allbusiness.com/Technology /computer-software-data-management/*, last retrieved on 15th Aug 2010.

[3] Michael Katina, Miller Keith W., " *Big Data: New Opportunities and New Challenges*", IEEE Computer Society 0018-9162/13/$31.00 © 2013 IEEE.

[4] Shan Suthaharan, "*Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning"*.

[5] Xingquan Zhu, Ian Davidson, "*Knowledge Discovery and Data Mining: Challenges and Realities*", ISBN 978-1-59904-252, Hershey, New York, 2007.

[6] Berry, M., and Linoff, G. S. (2004). *Data Mining Techniques*

for Marketing, Sales, and Customer Relationship Management. Wiley.

Author's Profile

Mrs Harmanpreet kaur received the M.C.A degree in 2013 from Panjab University, Chandigarh, India and cleared UGC-NET exam in January 2018. She works as a Resource Person in Post Graduate Government College for Girls Sector 11, Chandigarh, India in the Department of BSc Computer Science. Presently, I am pursing Phd degree in Computer application.

ISSN:0971-1260

Vol-22-Special-Issue-27-December-2019