

# Data Mining: Tools, Techniques and Application in the Context of Bangladesh

Mahfuza Mallika

DOI: <https://doi.org/10.51244/IJRSI.2024.11110064>

Received: 19 November 2024; Accepted: 30 November 2024; Published: 18 December 2024

## ABSTRACT

Data mining is a significant process for gaining better decision and predicts the result of datasets (future trends and behavior). This study focuses on variety of data mining tools and techniques used in data analysis. Its further analysis of data mining application in public and private sectors. This study also investigates the challenges around data mining application in the context of Bangladesh. The purposes of the study are to explore the issue of data analysis for better understanding of data mining process and application to the learners, researchers, teachers and policy makers. This paper further aims to identify possible and effective means of application of data mining in Bangladesh in order to make better policy on data mining in ICT sectors in the aspect of digitization of Bangladesh. Data mining will be helpful for different areas of our country if we know the proper use of it. I think this paper will be supportive for users who are interested to use data mining in their organizations. This is a doctrinal research follows with mix methodologies.

**Key Words:** Data mining Tools, Data mining Techniques, Data mining Application, Knowledge Discovery in Database (KDD), WEKA (Waikato Environment for knowledge Analysis), Rapid Miner, and SAS Enterprise Miner

## INTRODUCTION

Data are available in this entire world so it is impossible for human to process it and make sense it. The amount of raw data is stored in database is explosion. Now database is consisting of gigabyte or terabyte data (1 Terabytes = 1 Trillion bytes). Raw data by itself, but it does not make meaningful information. In today's competitive environment, companies or different organization need to rapidly change these terabytes of raw data into significant information into their customers for their various advantages. In this case, we can use a technology which gives us the proper meaningful database. Data mining is a new powerful technology which helps the companies or organizations focuses on the most important information in the raw data. Though data mining is still an not fully formed tools, companies are used this tools for various areas, like retail, finance, health care, manufacturing, transportation, communication; to get advantages of historical data (Alexander, 2023).

The main purpose of Data mining is to extract information from available data and discovers new information from enormous datasets. It analyzing data for recognizing the pattern of trends has a long history. Sometimes it referred to as "Knowledge Discovery in Database". Its foundation compromises three intertwined scientific disciplines: statistic (the numeric study of data relationship), artificial intelligence (human like intelligence display by software and/or machines) and machine learning (algorithm that can learn from data to make prediction) (Alexander, 2023). ICT is the backbone of digitization system and it covers the vast area of Information technology, Communication technology and Telecommunication technology. Data mining is helpful in ICT for extracting and discovering valuable and meaningful knowledge from huge amount of data (Ahmedet.al, 2013).

## Rationale of Study

This study simplifies Data Mining for readers, learners, researcher because most of the people of concern area of Bangladesh are not clear what is Data Mining and how can they use it. So, this paper explains Data Mining tools and techniques in a simple manner. This research helps those people in concern who wants to know about

Data Mining clearly. It also explains the application of Data Mining in different sectors so that different businesses or organizations of Bangladesh can use Data M in their proper analysis and accurate outcomes. It will be more effective for working with massive datasets and developing privacy.

### **Research Questions and Objectives:**

What is Data Mining? What are the tools of Data Mining? What are the techniques followed in Data Mining? What are the applications of Data Mining in various sectors of Bangladesh? What are the challenges of Data Mining in the context of Bangladesh? However, the research objectives are firstly, explicate the tools of Data Mining for analysis, secondly, explain the techniques of Data Mining for analysis, thirdly, examine an analysis the application in different sectors of Bangladesh and finally, clarify the challenges of data mining in the context of Bangladesh.

## **METHODOLOGY**

This is a doctrinal research follows with mix methodologies. This study consults primarily with the secondary sources of materials like published article as well as online resources. It further analysis the findings of the researches conducted in this area and improve the current study. This paper also consults with the data mining professionals in order to identify the difficulties facing by the users. The researcher undertakes one-to-one interview with data mining experts by unstructured questioners and analyze the opinion of experts in this study. Furthermore, this study conducts informal observations about difficulties in data mining by the graduate students at the university level and utilizes those observations accordingly. This study adopts theoretical framework and avoid complex statistical analysis to achieve the objectives of the study. Although there is no statistical analysis in this study, various diagrams and data tables are used to understand the use of data mining.

### **Definition of Data mining**

We live in the world and it has many countries. Different countries have different dataset in various sectors for various purposes. Our birth place is Bangladesh. It has more data for its population and different areas; such as business, society, science, engineering, medicine, health care, insurance, bank etc. Now those data grow day by day and it makes a vast amount of datasets. Analyze is important to discover the knowledge from datasets. Data mining is the process of analyzing a large volume of data, taking out useful intelligence to help organizations or companies solve problems, recognize patterns of trends, alleviate risk and find new opportunities. Simply data mining is a system that involves searching, collecting, filtering and analyzing data (Simplilearn, 2023). Governments, private companies, large organization and all businesses are collected a large volume of data for the purposes of research and business development. The data/information is stored for future use. But it will take time for finding and searching information from website, databases and other internet sources. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam email filtering or even to detect the opinion of users.

### **Importance of Data Mining:**

Data mining helps companies to gather reliable information. Data mining is the procedure of capturing large set of data to identify the insight and vision of data. With this technique, we analyze the data and then convert it into meaningful information. This helps the business to take accurate and better decision. t helps to develop smart market decision, run accurate campaigns, and make predictions. With the help of Data mining, we can analyze customer behaviors and their approaching. This help to get vast success and data-driven business. It's efficient to find the cost-effective solution compared to other data application. Data scientists can use the information to delete fraud, build risk models and improve product safety (Huria, 2021)

### **Processes of Data mining:**

CRISP-DM is a reliable data mining process model consisting of six phases. It is a cyclical process that provides a structural approach to the data mining process. The six phases can be implemented in any order but sometimes it needs the previous steps and repetition of actions.

### Step 1: Business Understanding

It is important to understand the entity and the project of business. At first, think about the current situation of business and the goal of that business. The goal of the company is trying to achieve by data mining. Data mining process try to understand the business and then start the process after looking any data to achieve the goal (Alexandra, 2023).

### Step 2: Data Understanding

In this step, think about data. Data mining expert gather information, evaluate the outcome of data and explore the requirements or project. They try to understand the problems, challenges and convert them to metadata. Data mining statistics are used to identify and convert the data patterns (Alexandra, 2023).

### Step 3: Data Preparation

In this step, data preparation covers all activities to construct the final dataset from the initial raw data. It involves selecting the appropriate data, cleaning, constructing attributes from data, integrating data from multiple databases. Data preparation tasks are performed in multiple times. During this stage of data mining, the data are checked for size as an overbearing collection of information because of unnecessary slow computations and analysis (Jackson, 2002).

### Step 4: Build the Model

In this step, various modeling techniques are selected to use and calibrated to optimize value data. Several techniques are applied to a specific form of data to get predictions. Expert builds the model from datasets and assessing the build model to get the future outcome. This modeling step will be back to the previous step (data preparation) is often need. The purpose of building model is to use predictions to make more informed business decisions (Huria, 2021).

### Step 5: Evaluation

In this stage, the project model builds with high quality from data analysis perspective. It is more important to evaluate the model thoroughly and review the steps to construct the model, to achieve the business objectives. A key objective is if there is some important issue of business that will not be considerable (Huria, 2021).

### Step 6: Deployment and Change Implement

In this step, a deployment plan is made that is strategy to monitor and maintaining the data mining model results to check its usefulness, final reports are made. Then review the whole process and find any mistake. If there is any mistake then any step will be repeated (Jackson, 2002).

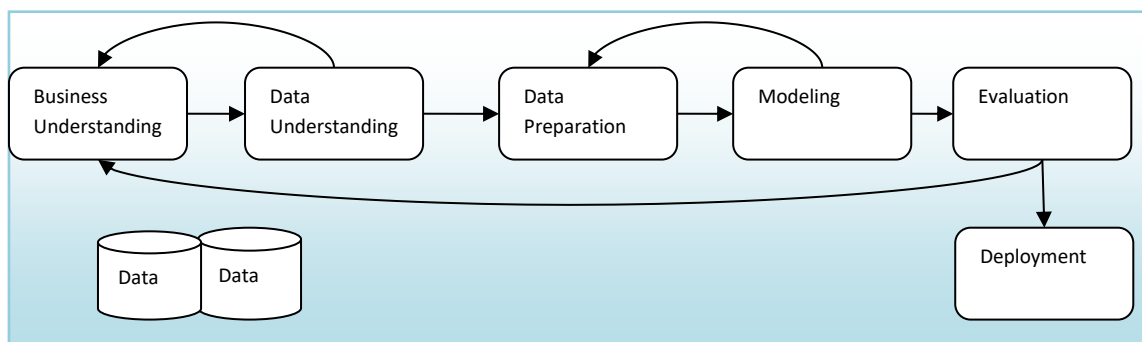


Figure 1: CRIPS-DM Model (Huria, 2021)

### Tools of Data mining

The top most five data mining tools are defining there for learning how data mining software does and more importantly what can do for business. Data analysts do their research and determine that these data mining tools currently on the market.

---

## Rapid Miner

Rapid Miner is a free-open-source data science platform that features hundreds of algorithms for data preparation, machine learning, deep learning, text mining and predictive analysis. It is drag-and-drop interface and pre-built models that allow non-programmers to intuitively create predictive workflows for specific uses of cases, like fraud detection and customer chum. Meanwhile, programmers can take advantages of Rapid Miner's R and Python extensions. Rapid Miner helps user to analyze data and visualize the result. It also finds out the pattern's outliers and trends in data. The benefits are free version, process optimizing, and database processing, interactive data preparation, drag-and-drop Interface (Sharma et.al., 2023).

## Oracle Data Mining:

Oracle Data Mining is a component of oracle Advanced Analytic that enables data analysts to build up and implement predictive models. It contains several data mining algorithms for task like classification, regression, anomaly detection, prediction and more. It can build models that help to predict customer behavior, segment customer profiles, detect fraud and identify the best proposal to target. Developers can use a Java API to integrate these models into business intelligence applications to help them discover new trends and patterns. The benefits are reducing operation cost, execute quick backup and recovery, and offer multiple database, better identity management and user control, autonomous data warehouse (Mikut et.al, 2011).

## IBM SPSS Modeler:

IBM SPSS Modeler is a data mining solution which allows data scientists to speed up and visualize the data mining process. Users can use advanced algorithms to build predictive models in a drag-and-drop interface without programming experience. With IBM SPSS Modeler, data scientist can import vast amount of data from multiple sources and rearrange it to uncover trends and patterns. The standard version of this tool works with numeric data from spreadsheets and relational database. The benefits are ease to used but need knowledge of statistics, many data sources, automatic and preparation, powerful graphics engine, visual analysis streams, automated modeling, range of algorithm, Text analysis and so on (Mikut et.al, 2011).

## WEKA

WEKA (Waikato Environment for knowledge Analysis) is an open source machine learning software with a vast collection of algorithms for data mining. It was developed by the University of Waikato in New Zealand and it's written in JavaScript. It supports different data mining task, like pre-processing, classification, regression, clustering and visualization in a graphical user interface that makes it easy to use. For each of these tasks, WEKA provides built-in machine learning algorithms which allow users to quickly test your ideas and deploy models without writing any code. Users need to take knowledge of different algorithms so that they can choose the right one for their task. The benefits are free availability under the GNU (General Public License), a comprehensive collection of data, Java Programming Language, modern computing platform, comparing different approaches and so on.

## SAS Enterprise Miner

SAS Enterprise Miner is an analytic and data management platform. Its goal is to simplify the DM Process to analytic professional's turn's large volumes of data into insights. It is a GUI based tool. Users can generate data mining models fast and solves critical business problems by SAS Enterprise Miner. SAS provides a rich set of algorithms for preparing and exploring data. It also used for building advanced predictive and descriptive models. Companies can use SAS Enterprise mining for fraud detection, resource planning and increase response rates on marketing campaigns among other applications. The benefits are distributed in-memory processing, build more models faster with an easy to use for GUI, enhance accuracy of predictions, easily share results through the unique model repository, Cooperation and so on (Mikut et.al, 2011).

## Techniques of Data mining

Data mining techniques or algorithms are used for generating enormous amount of data to get new information and discover knowledge from large database. There are many data mining techniques or algorithms have been

developing and using in data mining projects. Each technique has its rules and methods which identify the problem of project and work out it properly (Osman, 2019).

Some of them are described in below:

### **Association**

Association is one of the most uses data mining techniques than others. In this technique, a transaction and the relationship between variables are used to identify a pattern. It also refers a relation technique. This technique helps to identify some hidden patterns within the data and concurrence the different variables that are appeared very frequently in the datasets. Association rules are useful for examining and forecasting customer behavior. This technique is very helpful for retailers who can study the buying habits of customers. Retailer can study past sales data and lookout the products that customers buy together regularly. Then retailers put those products in their shop for customer's advantages so that they save their time and it increases their sales. This technique is used to determine shopping basket data analysis, product clustering, catalog design and store layout. Data mining technique adopt two step processes- a) Find out all the repeatedly occurring datasets and b) develop strong association rules from the recurrent datasets (Sharma, 2022).

### **Classification Analysis**

Classification analysis is used to retrieve important and appropriate information about data and metadata. It is used to classify different data into different classes. Recording data into different segments is called classes. Classification is used to develop software and it became capable of classifying items in a datasets into different classes. It uses linear programming, statistics, decision trees and artificial neural network in data mining among other techniques. For instance, we can use it to classify all the candidates who attended an interview into two groups. The first group is the list of those who are selected candidates for the job and the second group is the list of those candidates who are rejected (Osman, 2019, Sharma, 2022).

### **Clustering**

Another data mining methodology is clustering. The cluster is a collection of data object; these objects have same characteristics. The similar objects are in the same group or cluster and another similar object are in the other group or cluster. That means each group contains same characteristics data. Cluster analysis is the process to clarify similarities and differences between the data. It identifies group or cluster between data that belong to the same group. For example, a bookshelf is full of books on different topics. Now challenges are to organize those books for the readers so that they don't have any problem to find out a particular book. We can keep similar books in one shelf with a meaningful name using clustering process. When readers searching their particular books, they will go to the bookshelf and find out their particular topic's book from the shelf easily (Sharma, 2022).

### **Regression Analysis**

Regression analysis is the statistical term which is used to identify and analyze the nature of relationship among variables in a datasets. The characteristic of dependent variable will be change; if the characteristic of independent variable is varied. That means one variable is dependent on another variable. Regression analysis helps to provide the exact relationship between two or more variables in a datasets. It's generally used for prediction, forecasting and data modeling (Sharma, 2022).

### **Sequential Patterns**

Sequential pattern is one of the data mining technique that pursue to determine or recognize associated pattern, fixed events and trends in transaction data over a period of time. In historical transaction data, businesses can recognize a set of data items that customers bought together different things in a year. Then companies deal with those customers who are regular to buy their product based on their historical sale data transaction (Sharma, 2022).



## Prediction

Prediction is a combination of data mining technique which is used for complex data that determines the association between independent variables and correlation between dependent and independent variables. It analyzes past events or actions to predict an event in the right sequence. For example, the prediction analysis technique can be used in the sale to calculate the future income. Consider if the sale is an independent variable, the income would be dependent variable. Then this technique draws a fixed regression curve for predict information based on historical sales and earnings dataset.

### Induction Decision Tree Analysis:

Induction decision tree refers to a tree and predictive data mining models. It helps users to understand that how the input data affect the outputs. In this technique, every branch of tree carries data and these data are observed by classification queries. The tree's leaves are partitioned by classification rule. The different leaves carry different variables in each side. Every data is falling under a segment and contain similarities attribute with the information which is already predicted. Decision tree is an understandable technique and provide result very clearly. This is used for data pre-processing, exploration analysis and prediction analysis. So, it is a versatile data mining method. For example, use the following decision tree to determine whether a person has eligible for vote or not for club of university (MPHIL, D. R. M., 2017).

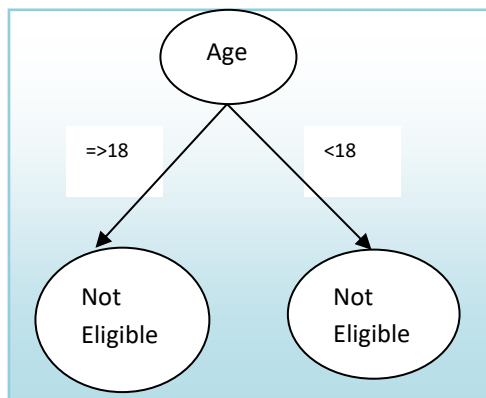


Figure 2: Decision Tree (MPHIL, D. R. M., 2017)<sup>9</sup>

### Application of Data mining in Bangladesh:

Data mining is the computational process by which different perspective data analyzing and transfer it into meaningful information. Data mining can be applied any type of data, such as Data Warehouse, Transaction Databases, Relational Databases, Multimedia Database, Spatial Databases, Time-Series Database, World Wide Web data. Data mining provides competitive advantages in the knowledge economy. It gives measurable benefits in various application areas.

### Scientific Analysis

Scientific simulation is generating a vast amount of data day by day. Data are collected from nuclear laboratories, human psychology etc. These data are analyzing by data mining techniques. Now it is possible to keep and capture more new data faster than the old data accumulated. It is used for analysis in Bio-informatics, Classification of astronomical objects, Medical decision Support etc. (Ahmed et.al. 2013). Bangladesh Institute of Nuclear Agriculture can use this data mining tools and techniques for analyzing the large amount of data properly and get more accurate result. BINA can also classify undiscovered data using data mining techniques; such as classification analysis, and can organizes the datasets separately which will be helpful to understand for all of BINA's members.

### Business Transaction

Business Transaction is a financial transaction system between two or more parties that involves the exchange of goods, money or services. It can occur between two parties, such as a business entity and a customer for

their benefits. The transaction of business is usually time-related internal business and external business operation. The transaction data are most important to make decision and solve problem for the business in this competitive world. Data mining helps to analyze business transaction, solving problem, decision making and identify marketing approach. For example, borrowing money from a bank, purchasing goods from a vendor, paying rent and other utilities, sale of goods, and paying interest etc. (Ahmed et.al., 2013).

**Market Basket Analysis**

Market basket analysis is a data mining technique used by retailers to increase sale by better understanding customer purchasing patterns. Using this data mining technique, retailer can identify those customers who are regular to purchase product and make a predictive pattern. Retailer can also optimize product information, special offer deals and create new product bundles to promote sale. Data mining concepts for sales and marketing to provide better customer services, cross-selling opportunities, to increase messenger response rapidly. Data mining find out the companies which are regular to customer response. In Bangladesh, some companies use this market basket analysis technique for better services from different super shop; understand customer purchasing model and sale increases. For example, Amazon’s website is a well-known example of market basket analysis. In Bangladesh, Aarong, Agora, Le Re Ve, Mina Bazar etc. are online shop. These are also example of market basket analysis (Ahmed et.al. 2013).

**Education**

Data mining in education is known as Educational Data Mining (EDM). It is concerned with developing methods for exploring the unique types of data that come from different educational environments (Romero, 2013). Educational data mining method can be used to classify and predict the performance of students go away as well as teacher’s presentation. It can help learners and educators to follow academic progress to improve the teaching process and it can also help students for selecting the courses. Educational data mining can solve more educational task, such as predicting student’s admission in higher education, predicting students profiling, predicting student’s performance, teachers teaching, curriculum development, predicting student placement opportunities etc. In Bangladesh, the education systems develop as e-learning or web-based education in which large amount of data about teaching-learning interaction are endlessly generated from COVID 19 (Romero et. al. 2010). Now most of the school, college and universities are uses Educational data mining (EDM) model to develop their educational management more efficient and effective in our country (Ahmed et.al. 2013).

**Healthcare and Insurance**

Data mining techniques are applied to undiscovered medical datasets for finding accurate results and outlines (Sohail, et.al. 2018). Several data mining tools and techniques have been applied to set of selected diseases to find the accuracy level in dissimilar healthcare issues (Mia,M. R. et.al.,2018). There are many healthcare centers in Bangladesh and vast amount of data of various patient are stored in each hospital. Data mining helps doctors to create more accurate diagnosis by collecting every patient’s medical history, physical examination results, medication and treatment patterns. It identifies and stores claim of medical procedure. It also discovers effective medical therapies for diverse illnesses. On the other hand, data mining helps Insurance sector to find out the customers who will buy new policies, identify behavior pattern of risky customers and fraudulent behavior of customers. (Ahmedet.al, 2013) One example is given below to understand that how data mining tools and techniques are used to determine the accuracy level in dissimilar healthcare issues (MPHIL, D. R. M., 2017).

Table 1: Tools and Techniques in Healthcare (MPHIL, D. R. M., 2017)

SL. No.	Types of Diseases	DM Tools	DM Techniques	Algorithms	Accuracy in Level in %
	Heart Diseases	ODND NCC2	Classification	Naïve	60
	Cancer	WEKA	Classification	Rule Decision Table	97.77

	HIV/AIDS	WEKA 3.5	Classification and Association Rule mining	J48	81.88
	Blood Bank Sector	WEKA	Classification	J48	89.9
	Brain Cancer	K-means Clustering	Clustering	MAFIA	85
	Tuberculosis	WEKA	Naïve Bayes Classifier	KNN	78
	Diabetes Mellitus	ANN	Classification	C4.5	82.6
	Kidney Dialysis	RST	Classification	Decision Making	75.97
	Dengue	SPSS Modeler		C5.0	80
	In vitro Fertilization (IVF)	ANN, RST	Classification		91
	Hepatitis C	SND	Information gain	Decision Rule	73.20

The above table 1 shows the various tools and techniques are used to find the accuracy level of various diseases. (MPHIL, D. R. M., 2017)

### Financial and Banking Sectors

Data mining tools and techniques is primal for the banking sector which aims to discover valuable information from the volume of data and reach better strategic management and customer satisfaction (Hassani,et.al., 2018). Banking sectors have overwhelming amount of data of customer are stored in different sides, such as credit card. Identify those customers who are most interested for a new credit card. This vast amount of transaction data is analyzed by data mining technique and find out the pattern of these data. Also, it can find out the fraudulent customers and doing our transaction properly. In Bangladesh, most of our banks are online bank. Data mining will be convenient for this running period and future; such as data mining techniques help bank in purchasing transaction and card transactions, analyzing customer finance data. It also helps the bank to understand the online customer’s behavior preferences which will be helpful to design a new marketing campaign. Financial sector is not secure system in Bangladesh. These are risky companies but there is large amount of data of customers. Data mining techniques will be helpful to financial institutions for loan information and credit reporting task. So, financial sectors can also use data mining techniques to get pattern and find fraudulent on their behavior (Ahmed,et.al, 2013)(Islam, M. R., 2015)

### Customer Service

Customer satisfaction may be caused for a variety of reasons. For example, a company wants to ship their goods or product to customers. But customers may become unhappy with ship time, shipping quality or communication on shipment expectations. So, customer may become frustrated with long time telephone wait time or slow e-mail responses. Data mining can gather operational information about customer interactions and summarizes findings to determine weak points and company’s work (Twin, 2023). Now days, companies are doing well with customer interaction in Bangladesh. In future, they have to know about data mining process to apply this technology more in the online company because of large amount of data.

### Challenges of Data mining in Bangladesh:

Now days, data mining is a fundamental technology for business, bio-informatics and researchers in various domains. We should establish it in our country (Bangladesh) to develop our task in different areas to get perfect solution and error detection which help to make our decision properly for organizations. But there are some pending challenges have to be solved for establish the data mining process in our organizations.



Some of these challenges are described in below with solution:

### **Data capture**

In Bangladesh, data are increased day by day. For this reason, datasets are becoming larger and more difficult to maintain it using traditional database tools. That's why, it is more difficult to capture data, store it, manage and analysis data properly in time. In this situation, large amount of data needs new infrastructure and new economic system gets proper function. Data mining ensures an organization that the collecting of data is reliable. It also captures the location of data.

### **Noisy and Incomplete Data**

Data mining is the way to acquire the information from the large volume of datasets. It represents the reality of information such as noisy, incomplete and heterogeneous. Huge amount of data are unreliable or inaccurate and these issues are done by human's mistake, blunders or error in the instruments. Avoid noisy and incomplete data for analysis and do the work sincerely. But there are more noisy data in our country and it will take time to minimize those large data properly.

### **Distributed Database**

Data are stored in different locations using distributing process. It may be on website, individual system or database of organization. But it is very difficult to carry all the data from various sources because of technical and organizational problem. Tools and algorithm need to be developed for mining distributing data (Twin,2023).<sup>13</sup> It's already established in different sectors of Bangladesh (such as Bank, Government offices, Educational organization, Hospitals etc) but we have no more knowledge about distributed database.

### **Complex data**

Most of the data are heterogeneous such as natural language text, time series, spatial data, audio and video data, image etc. It is very hard to maintain and identify these data. New system and instruments would be needed to manage and separate the data for the organization (Simplilearn, 2023).

### **Data Infrastructure**

Our data are not well-organized in different location database. There need well-structured data for operating those data to get the services and facilities for an economic function. For example, our online services companies need well-structured data so that they can use data mining tools for their development. Otherwise it will be a great problem to use the data mining tools. Most of the organization, data are not properly organized in our country to use data mining process.

### **Organizational Structure change**

Traditional organizational systems are not more comfortable to access their data properly. Data mining is structured process which finds a problem, gathers all problems of data set and tries to formulate a solution. Data mining tools are well organized to do the work properly. So, it needs well organized organization to understand any task of organization so that data mining can help an organization to become more profitable, efficient, or operationally stronger (Lakshmi et.al, 2011). IT infra-structured organization can use this data mining tools and technique but others companies can't use this technique in our country.

### **User Interface**

If data mining tools are interesting and understandable for users, then it will be helpful to discover knowledge. Otherwise it will be great problem to understand the large data. Data mining tools helps to better understand and interpret the data easily. Many research carrying out big datasets that display and manipulate mined knowledge (Simplilearn, 2023).

## **Various types of Knowledge in database**

Different users need different knowledge of data mining tools to operate it. Data mining can cover a wide range of knowledge discovery task. Sometimes different knowledge may irritate the interest of users. Need to know the knowledge of database for all data managers.

## **Scalability and Efficiency of data mining algorithm**

Data mining algorithm may also be most effective when using huge data sets. However, these data sets must be stored and require heavy computational power to analyze. Data mining algorithm should be scalable and efficient to recover the information from vast amount of data into a datasets.

## **Background Knowledge and Interaction multiple levels of abstraction**

Previous information can be used to express discovered pattern not only short term but also at multiple levels of abstraction. The data mining process should be interactive with multiple levels of abstraction to focus on searching for patterns, providing results and refining data.

## **Data mining tools and improve algorithm**

Data mining tools are complex and challenges to use for users. So, data analysis needs training and knowledge to use the tools effectively. Also, data mining tools should be easy and understandable to access for end users so that the data mining tools will be acceptable for all end users. Improve the algorithm of data mining to work properly. Otherwise it will be difficult to work also.

## **Communication Challenges**

The communication cost is too much than the processing cost of the data. Here, the challenge of communication cost is minimizing while the data storage and requirement is fulfilled for data processing. The bandwidth and latency are the two features of network that will affect the communication between the clients and the cloud server. We will find the solution for data processing at Distributed File System (DFS) and User Interaction and Learning System (UILS) (Jadhav, 2013). Communication is a great problem for our country because of its complexity, cost and electricity.

## **Privacy and Security**

Data mining needs more privacy and security for analyzing data properly, though IT infrastructure may be costly as well. Otherwise, it will be risky for the information of organization. For example, it is seen that an online retailer gives their product information on the page to purchase product. But customers see the uncovered information and access that information without permission of retailer. It destroys the privacy of information of organization. Need strong security and privacy so that nobody can access the data set. Privacy and security is as well immense issue for our country to save information. So, we have to develop our organizational security and privacy.

## **CONCLUSION**

Bangladesh is an over populated country. So many data are stored in different organizations; such as Govt. office, Bank, Medical Science, Research centre, NGO, Agriculture, Telecom industries, Market analysis and many others areas. Data mining will be helpful to analyze data properly in these sectors. Data mining is a new opportunity which is used to supports fraud detection, risk management, cyber security planning and many other serious commerce tasks. We can use it to increase information, reduce cost, improve customer relationship management, and reduce economic trends and others.

Now we have learned the proper knowledge of data mining. All tools and techniques have limitation. In the same way data mining tools have some limitation which will be challenges for us to use these tools. Data mining doesn't always assurance proper outcome for its limitations. A company may perform statistical

analysis, make conclusion based on strong data, implement changes, and not gather any benefits. But through accurate findings and appropriate data populations, data mining can guide better decisions and ensure outcomes. So, we have to use it to promote results for our organization and we have to rise above challenges of data mining though it will take time to establish in Bangladesh.

## REFERENCES

1. Alexander, Doug, Data Mining, 2023, <https://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/> (Accessed March 15, 2023).
2. Ahmed, K., Habib, M. A., Jesmin, T., Rahman, M. Z., & Miah, M. B. A. (2013). Prediction of breast cancer risk level with risk factors in perspective to bangladeshi women using data Mining. *International Journal of Computer Applications*, 82(4).
3. Huria, Rahul, Data Mining and Its Importance, 2021, <https://www.loginworks.com/blogs/tag/data-mining-importance/>(Accessed March 15, 2023).
4. Jackson, J. (2002). Data mining; a conceptual overview. *Communications of the Association for Information Systems*, 8(1), 19.
5. Mikut, R., & Reischl, M. (2011). Data mining tools. *Wiley interdisciplinary reviews: data mining and knowledge discovery*, 1(5), 431-443.
6. MPHIL, D. R. M. (2017). A Survey on Data Mining Tools and Techniques in Medical Field. *International Journal of Advanced Networking and Applications*, 8(5), 51-54.
7. Osman, A. S. (2019). Data mining techniques, *IJDSR*, Voume2, Issue1 June2019, Al-Madinah International University, Malaysia
8. Simplilearn, what is Data Mining: Definition, Benefits, Applications, Top Techniques and More, <https://www.simplilearn.com/what-is-data-mining-article> (Accessed March 15, 2023).
9. Sharma, Rohit, Data Mining Techniques: Types of Data, Methods, Applications, <https://www.upgrad.com/blog/data-mining-techniques/> (Accessed 15 March, 2023).
10. Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data mining and knowledge discovery*, 3(1), 12-27.
11. Lakshmi, B. N., & Raghunandhan, G. H. (2011, February). A conceptual overview of data mining. In *2011 National Conference on Innovations in Emerging Technology* (pp. 27-32). IEEE.
12. Twin, Alexandra, What Is Data Mining? How It Works, Benefits, Techniques, and Examples, 2023, <https://www.investopedia.com/terms/d/datamining.asp>, (Accessed March 15, 2023)
13. Sohail, M. N., Jiadong, R., Irshad, M., Uba, M. M., & Abir, S. I. (2018). Data mining techniques for Medical Growth: A Contribution of Researcher reviews. *Int. J. Comput. Sci. Netw. Secur*, 18, 5-10.
14. Romero, C., & Ventura, S. (2010). Educational data mining: a review of the state of the art. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (applications and reviews)*, 40(6), 601-618.
15. Mia, M. R., Hossain, S. A., Chhoton, A. C., & Chakraborty, N. R. (2018, February). A comprehensive study of data mining techniques in health-care, medical, and bioinformatics. In *2018 International Conference on Computer, Communication, Chemical, Material and Electronic Engineering (IC4ME2)* (pp. 1-4). IEEE.
16. Hassani, H., Huang, X., & Silva, E. (2018). Digitalisation and big data mining in banking. *Big Data and Cognitive Computing*, 2(3), 18.
17. Islam, M. R., & Habib, M. A. (2015). A data mining approach to predict prospective business sectors for lending in retail banking using decision tree. *arXiv preprint arXiv:1504.02018*.