

Data Mining in Different Fields: A Study

Golam Kaderye¹; Md. Ahsan Arif²

Department of Computer Science & Engineering
University of Scholars, Dhaka, Bangladesh

Ronjon Kundu³

The University of Western Australia
Australia

Abstract:- A revolutionary technique for managing massive amounts of data from databases is data mining. These days, information is more commonplace in all types of businesses and industries. For a range of business objectives, fields choose data mining approaches, and data mining is what makes it possible for fields to establish the ideal customer baseline and cultivate long-lasting relationships. Statistical analysis, machine learning, predictive modelling, and database approaches are all combined in knowledge discovery. This review research focuses on and quickly discusses the various domains in which data mining is applied. Here, we discuss several sectors where data mining is relevant in order to comprehend its significance. We also discuss the numerous technologies used in related fields, which depend on user desire. To learn about it, many algorithms are described.

Keywords:- Data, Information, Knowledge, Data Mining, Domains, Techniques, Tools.

I. INTRODUCTION

In all disciplines, data is being gathered and accumulated quickly. A new generation of computational theories and tools are urgently required, according to a report, to assist humans in extracting knowledge from the rapidly growing volumes of digital information [1]. The process of removing useful information and patterns from massive volumes of data is known as data mining. This technique is referred to by the name's knowledge discovery, data mining, knowledge extraction, and data analysis [2]. Massive volumes of data can be analysed and explored using a technique called data mining to uncover information that is helpful. Similar to how "knowledge mining from data" should have been used to describe data mining, it was sad that it wasn't [3]. Most academics agree that Fayyad introduced data mining for the first time at the Knowledge Discovery Conference in 1995. He viewed data mining as a challenging process that automatically or semi-automatically identifies useful, pertinent, potentially beneficial, and easily comprehensible data models from a set of data [4].

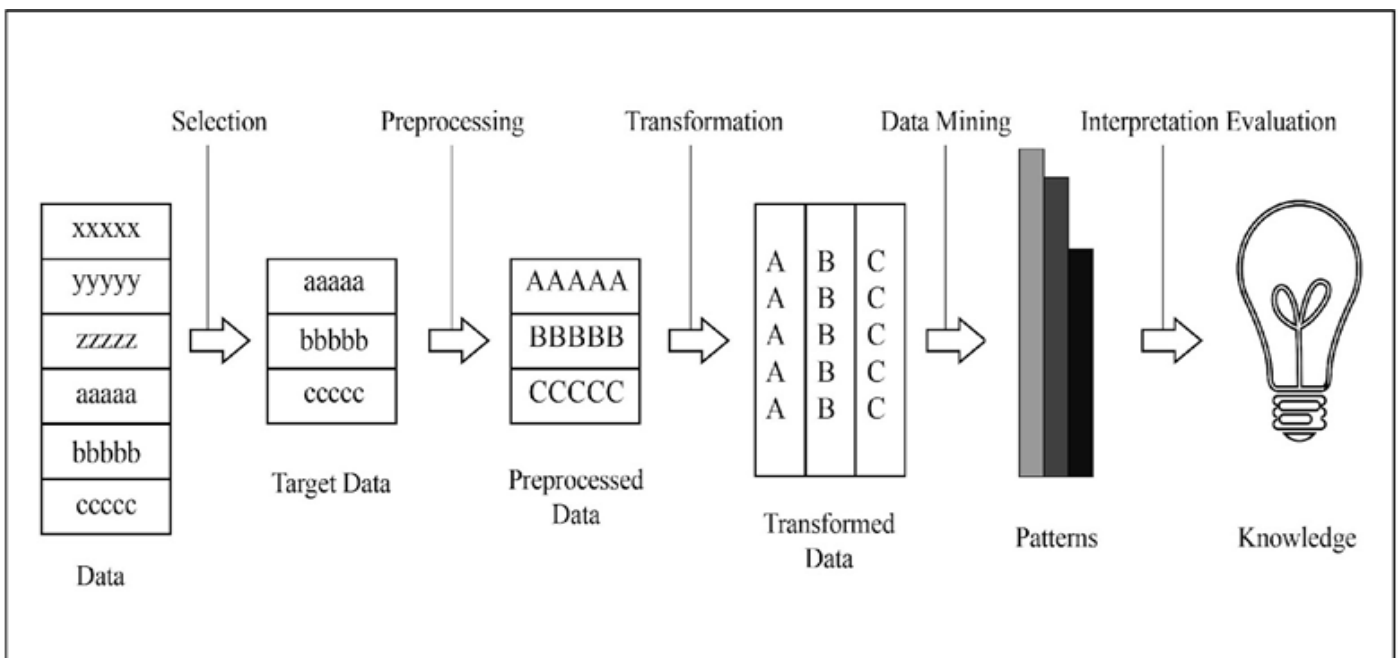


Fig 1 Data Mining Process

Figure 1 illustrates how data is processed using data mining techniques. A collection of datasets can yield a variety of information that can be used for a number of different things. An exponential growth in data production and storage is observed. We'll examine how data mining, a current topic, might be applied in several fields to glean

useful information from the data that surrounds us. Those who master this technology and its methods will benefit greatly and acquire an advantage over competitors. This cutting-edge data mining technique is presently used in a variety of areas, including education, healthcare, social media, retail sales, and crime investigation.

II. PURPOSE

To turn unstructured data into useful information, businesses use data mining. Employing algorithms to look for trends in enormous amounts of data might help businesses understand their customers better. These data patterns help in predicting business or information trends and in figuring out what to do in response among the several fields are:

➤ Education

Using educational data mining, it is possible to categorize and forecast student performance, dropout rates, and instructor performance. It can help students with course selection and educational administration, as well as teachers in tracking academic success to improve the teaching process.

➤ Crime Analysis

Data mining gives us a number of useful and practical tools to evaluate substantial and varied sets of information. It aids in revealing concealed information from the sizable database of criminal records in order to support organizations' and users' efforts to control, investigate, and prevent crime.

➤ Social Media

Data mining is used to uncover hidden patterns and trends on social media platforms like Twitter, LinkedIn, Facebook, and others. As a result, numerous papers have been written that investigate and evaluate various fields on a case-by-case basis. How and why data mining is applied in diverse sectors are included in our review article. This review paper's goal is briefly explained.

➤ Accurate Information

Large datasets are created and kept in this digital age. These datasets are typically unstructured, have a lot of noise, and contain information that isn't relevant. In order to locate the precise information, they need and can employ to satisfy their objectives, enterprises use data mining.

➤ Identifying Correlation

After gathering the necessary data, the very next step for industries is to examine the patterns. Numerous methods, including classification, clustering, and ANN algorithms, can be used to carry out these processes. After this review paper is finished, it will be evident which tools and procedures are most commonly employed by industries. Figure 2 demonstrates the correlation of related data that has been created. (a) Initial stage of data, (b) Pre-processing of data and (c) Similar group of data.

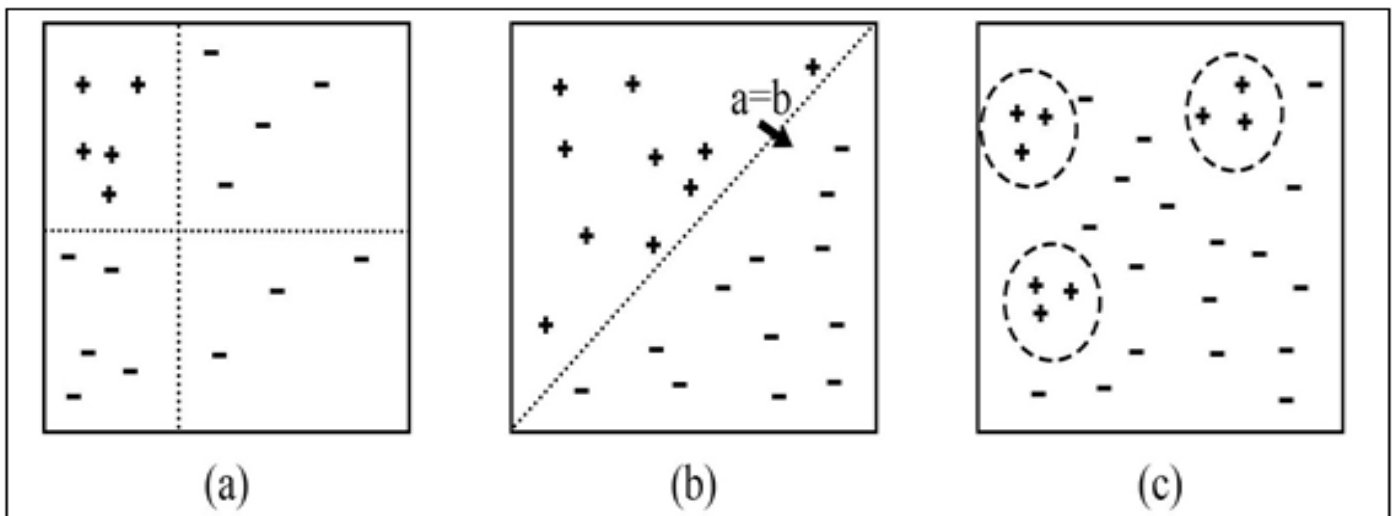


Fig 2 Correlation Among Data

➤ Predictive Analysis

Thirdly, it will be simpler to forecast and move on to the next phase once the correlation has been discovered. Industries create predictions based on the information they

have, and for a better outcome, this prediction process needed that certain criterion be followed. Table 1 in [5] Various algorithms with the capacity for prediction have been mentioned.

Table 1 Function for Algorithm Prediction

| Learning Method | Tasks Handled | Expressive Power | Training Time | Testing Time | Model Comprehensibility |
|------------------|------------------------------|------------------|---------------|--------------|---|
| Decision | Classification | Average | Fast | Fast | Good |
| Rule-Based | Classification | Average | Fast | Fast | Good |
| ANN | Classification Regression | Good | Slow | Fast | Average |
| Nearest-Neighbor | Classification Regression | Good | No Time | Slow | Predictions are explicable despite the lack of a model. |
| Naïve Bayesian | Classification | Good | Fast | Fast | Average |

➤ *Take Sensible Action*

In decision support systems, data mining techniques are used. Management needs knowledge to make decisions. Data analysis can lead to issues when working with large amounts of data, which is where knowledge extraction comes in helpful. Data is evaluated automatically in order to make wise decisions since a data mining technique is necessary.

➤ *Effective Techniques*

Industries involved in data mining must adhere to certain strategies, and in order to have these strategies, data mining requires the usage of data mining tools for analysis. Finally, industries will be able to select from a variety of strategies.

III. OBJECTIVES

Data mining has created a world of new opportunities for businesses. By comparing millions of isolated pieces of data, businesses may learn about and predict consumer behavior using this area of computational statistics. It aims to create new commercial opportunities. Data is transformed into knowledge through data mining.

➤ *Relations with Customers*

Data mining techniques are the most popular and effective approach for customer behaviour analysis.

Determine how well data mining tools complement the conventional approach by first identifying a consumer's behaviour, psychological state, and decision-making during a transaction. For use in purchased data in a retailer's store or organization, necessary consumer rules are extracted using various association criteria. Afterward, the necessary rule can be used in other industries, such as online commerce, medical, and so forth [7]. in order to develop a relationship with the client.

➤ *Revenue Growth*

Once the first two goals are verified, every organization can instantly boost their revenue by modifying these data mining approaches.

IV. ASSESSMENT

Our attention was on various articles that took into account the industries of Education, Healthcare, Finance, social media, Retail Marketing, Telecommunication, and Fraud Detection. We noticed that different data mining approaches, including clustering, decision trees, k-means, etc., are applied in different industries to examine data that has been extracted from datasets in accordance with their needs. In the article, the author describes Educational Data Mining (EDM) as a study area aimed at examining the specific forms of data that arise in educational contexts in order to address educational research issues [6].

Data mining in healthcare is mostly used to foresee various illnesses and support physicians' decision-making. Data mining has several applications in the healthcare

sector, including medical research, drugs, medical gear, hospital management, health insurance, and fraud detection and prevention. Benefits of data mining, which are significant for patients, healthcare organizations, researchers, and insurers, are growing in the healthcare sector. As a second opinion for a doctor's treatment choice, predictive models based on data mining can be used. It could help people with similar illnesses or health issues be grouped together [7].

One of the most significant financial operations is the forecasting of the stock market, currency exchange rates, bankruptcies, trade futures, financial risk management, credit ratings, loan management, bank client profiling, and money laundering studies. According to the author, data mining for consumer profiling in other industries and some of these jobs, such bank customer profiling, have a lot in common. There has been a lot of investigation into the long-term advantages of overcoming these challenges [8].

Sentiment analysis models have received a lot of attention from researchers who are interested in using social network data to draw statistical conclusions. Bo Pang and Lilliam Lee presented a thorough talk on sentiment analysis. They examined the proportion of motivational affirmations to total words to gauge their point of view [9].

Social media platforms like Facebook, Twitter, and Instagram are examples of places where data mining is used to analyze user activity and locate target customers using techniques like clustering, k-mean analysis, and the priority algorithm. For every firm, it aids in the development of web marketing.

The use of data mining tools has made it possible for retailers to operate in a dynamic and competitive market. As globalization and competitiveness increase, businesses are seeking more advertising and promotion. A bank or other financial institution can utilize data mining to examine loan payments, customer credit policies, target marketing, and look into money laundering and other financial crimes [10].

We found that there is very little study on the use of data mining technology in the retail industry to target customers for marketing efforts. Most shops struggle to execute successful campaigns and acquire acceptable customers.

Crime analysis is the process of systematically identifying and assessing patterns and trends in criminal activity. In the areas of crime data mining and criminal investigation data mining, numerous researchers have conducted studies. To lessen, prevent, and enhance the quality of corruption procedures, researchers are continually working. The authors used a data mining approach, a study of crime frequency reports, and evidence analysis to explain the causes of crime and criminal activities [11].

In a number of situations and ways, the telecoms industry benefited from data mining tools. Given the extensive customer data that telecom companies maintain

and the intense competition in the market, there is a strong motivation to investigate consumer data for multiple objectives [12].

Fraud detection techniques are always being improved to prevent fraudsters from changing their tactics. The creation of current fraud detection techniques has gotten harder because of the considerable restrictions on the sharing of ideas in that field [13].

Overall, it is clear that data mining has a significant impact on practically every well-known industry. Industries employ data mining techniques including anomaly detection, clustering, classification, discriminant analysis, decision trees, etc. to evaluate data for forecasting, forming appropriate strategies, and maximizing commercial profit.

V. DATA MINING - FIELDS

We focused on eight industries where data mining is used, and we looked at how these industries are able to grow both their business and activities thanks to data mining. According to their needs, the eight industries of education, healthcare, finance, social media, retail marketing, crime analysis, telemarketing, and fraud detection use data mining tools and methodologies. The following is a succinct description.

➤ *Education*

The term "educational data mining" (EDM) refers to the use of data mining techniques to educational data, such as student information, academic records, exam results, academic achievement predictions, classroom engagement, and the frequency of student questioning. In order to help with the outcome, educational data mining uses a number of optimization techniques to increase educational attainment. EDM is a process that transforms unprocessed educational system data into crucial information that can be used to make statistical judgments.

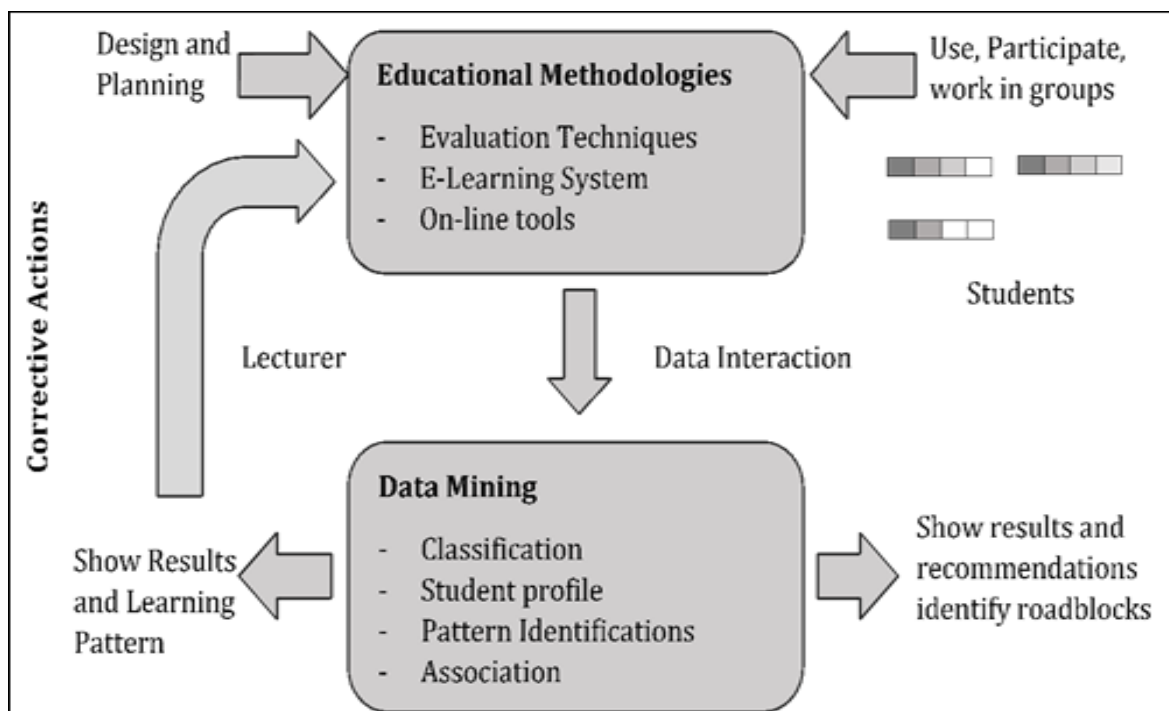


Fig 3 Illustrates how Students' Learning can be Determined using Data Mining [14]

➤ *Healthcare*

Healthcare generates enormous volumes of data, which are constantly being added by administrative reports, electronic medical records, benchmark findings, etc. By using data mining in the healthcare industry, we can largely forecast many ailments, including helping doctors make clinical diagnoses.

enhance healthcare organization, and quickly identify patients who require specialized intensive care and critical sickness conditions. Hierarchical clustering, partitioned clustering, most algorithms utilized in the healthcare field include density-based clustering, Bayesian classifiers, support vector machines (SVM), K-NN, and decision trees [14].

Data mining is a framework and a collection of tactics for converting raw data into information that may be used to make statistical choices about things like managing healthcare, managing doctor-patient relationships, and detecting fraud and abuse. Reduce hospital admissions,

➤ *Finance*

Business operations have greatly improved thanks to data mining services for the banking sector. Target marketing, stock market forecasting, customer credit policies, money laundering detection, loan repayment,

transaction analysis, bankruptcy forecasting, and other financial services heavily rely on data mining. The three financial statements that financial analysts look at are the cash balance, the statement of cash flow, and retained earnings.

Data mining has a significant positive impact on the financial sector since it allows large shareholders to increase their profits by strategically exchanging assets at the right periods. The finance sector uses classification, clustering, k-means, and HC algorithms the most [15].

➤ *Social Media*

Social media mining entails examining and gathering unprocessed data from these channels. Data mining gathers or extracts user behavior, feelings, wishes, mindset, personality, needs, likes-dislikes, etc. patterns, or correlations, from social media platforms like Facebook, Instagram, Twitter, TikTok, LinkedIn, YouTube, etc. By examining user age, gender, job title, geography, frequently visited websites, content sharing, search history, comments, likes, and clicks from social media networks, raw social media data can offer applicable services or recommendations. Trend analysis, social spam detection, e-commerce, digital media, research, and other areas all employ social media mining.

Most commonly employed in social media mining are decision trees, naive bayes classifiers, K-nearest neighbors, and clustering [16]. AdaBoost Artificial Neural Network (ANN), Apriori, Bayesian Networks (BN), Decision Trees (DT), Density Based Algorithm (DBA), Fuzzy, Genetic Algorithm (GA), Hierarchical Clustering (HC), Linear Discriminant Analysis (LDA), Linear-Regression (Lin-R), Logistic Regression (LR), Markov, Maximum Entropy (ME), Novel, Support Vector Machine (SVM), and Wrapper [17].

➤ *Retail Marketing*

Worldwide, a lot of businesses try to remake and rebuild themselves. They also benefit greatly from data mining. Retail data mining in the retail sector identifies customer behavior, shopping patterns and trends, customer acquisition and retention, customer buying preferences, delivery performance, reduces business costs, conducts shopping cart analysis, risk management, fraud detection, and improves customer service for retailers' prediction and proactive decision-making from a large amount of sales record and shopping history.

➤ *Crime Analysis*

Modern methods and high-tech procedures provide criminals an advantage, and the crime rate is rising daily as a result. In order to gather accurate information, we can gather crime statistics from various news sources, blogs, and websites and enter them into a crime database.

Crime analysis and prediction take a methodical approach to locating, examining patterns and trends. The

police agency and victims are the main beneficiaries in this field [11].

To extract from a big amount of data the information the investigator or police need is the core aim of crime analysis. Criminal analysis is a challenging endeavor since a variety of factors, including emotional instability, fury, envy, retaliation, identity crisis, ego, gain, and harm to the reputation or property of others, may have contributed to the commission of any crime, shortcuts to financial success, etc. So, in addition to reports, knowledge of outside circumstances is necessary to understand a crime. The best algorithms must be taken into account when a computer is required to carry out this kind of intelligent task. Crime analysis is difficult in general because it necessitates collecting and reporting vast amounts of data [18].

➤ *Telecommunication*

Data produced and stored by the telecoms sector is enormous. These data comprise customer data, which represents telecommunication customers, network data, which describes the condition of the hardware and software components in the network, and call detail data, which details the calls that travel over telecommunication networks [12].

The telecommunications industry has quickly expanded from providing unique and long-distance phone services to supplying a wide range of other data services. Many different types of communication and computation, including telecommunications, computer networks, the Internet, and others, are being merged. A call detail record is created [15] to keep the call information.

Telecommunications data is fundamentally multidimensional and includes elements like calling time, frequency, caller location, and call mode. A mixture of information from such data can be used to identify and compare data traffic, system load, resource management, consumer group action, and earnings.

➤ *Fraud Detection*

The goal of fraud detection is to stop people from obtaining money or other items under false pretences. High-level skill is required for fraud detection. The most popular AI methods for fraud detection use data mining to categorize, cluster, and segment data and automatically find linkages and rules in the data that may imply interesting patterns, including those associated to fraud.

Financial fraud is a problem that is particularly prevalent in sectors that deal with massive amounts of data, such banking, telecommunications, insurance, and credit cards. Data mining can be used to identify fraud involving credit cards, consumers, health care, charities, natural disasters, business and investments, e-commerce, banking, etc.

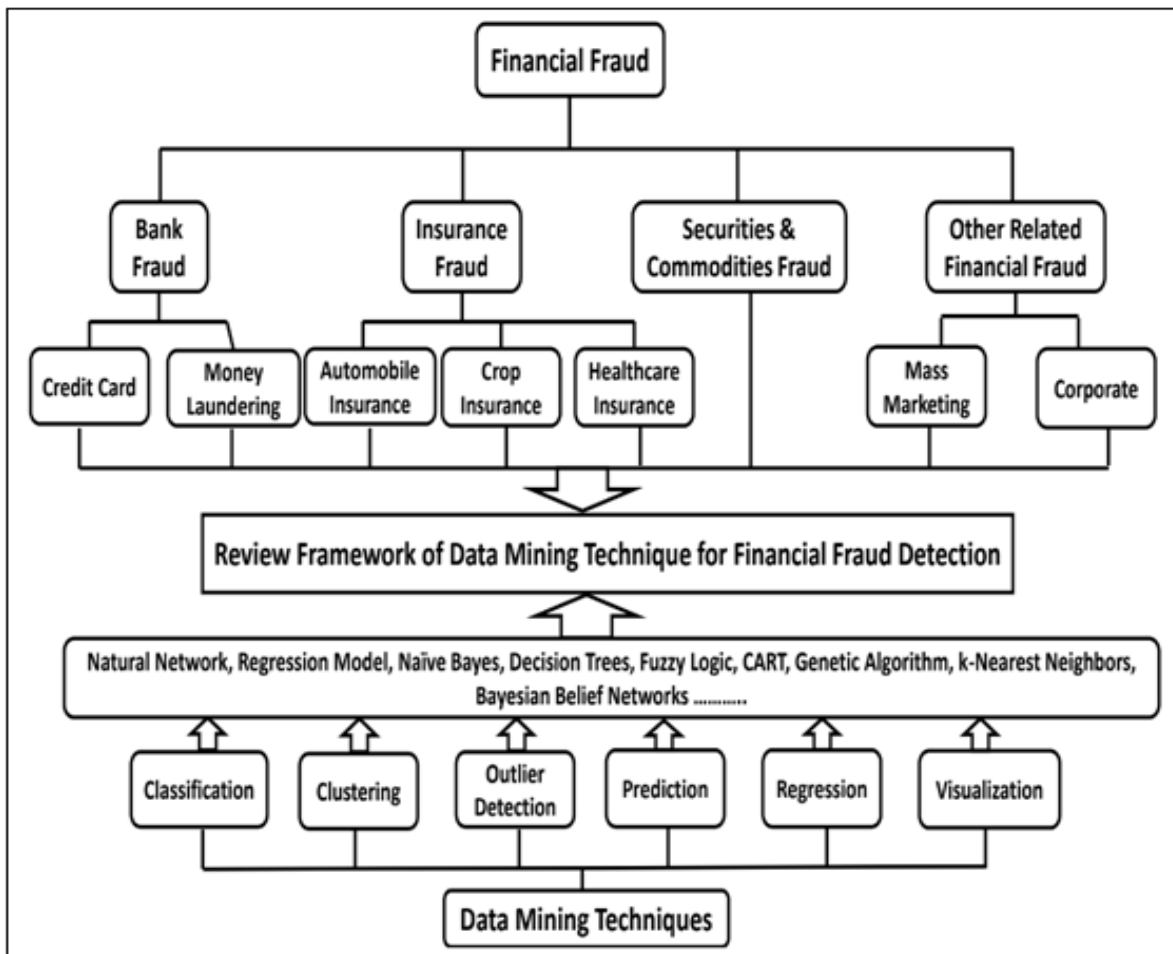


Fig 4 We can use Data Mining to Lower this Rate because some Popular Types of Fraud Increased Across Industries Globally from 2020 to 2021 [13]

VI. TOOLS

The purpose of data mining techniques is to turn vast data sets into more specialized information by identifying patterns, trends, and groups within them. In order to help firms, make better business decisions, data mining technologies forecast upcoming trends and behaviors.

Data mining can be used to identify sales trends, create more effective marketing initiatives, and precisely anticipate

consumer loyalty [19]. It might be difficult to extract relevant and useful information from repositories. This issue is addressed by the computer science subfield known as data mining. And a lot of software is employed in mining. Data mining tools include KNIME, ORANGE, IBM SPSS, WEKA, and others [20].

We combine the with an accurate comparison in the table below. Below is a list of several well-known open-source data mining tools.

Table 2 Uses of Tools

| Tools | RapidMiner | KEEL | Solver | SPSS | Alteryx | Orange | Board | KNIME | Spark | K-core | Datawatch | Weka |
|-------------------|------------|------|--------|------|---------|--------|-------|-------|-------|--------|-----------|------|
| Fields | | | | | | | | | | | | |
| Education | ✓ | ✓ | | ✓ | | ✓ | | ✓ | | | | ✓ |
| Healthcare | ✓ | | | | | | | ✓ | ✓ | | | |
| Finance | | | | | ✓ | ✓ | | | | | ✓ | |
| Social media | ✓ | | | | | ✓ | | | | ✓ | | |
| Retail Marketing | | | | ✓ | ✓ | ✓ | | | | | ✓ | |
| Crime Analysis | | | ✓ | | | | | | | | ✓ | |
| Telecommunication | | | | | | ✓ | | | | | | ✓ |
| Fraud Detection | | | | ✓ | | | ✓ | | ✓ | | | |

➤ *Data Mining Tools are Described*

- *R*
Programming tool R is available for free. Data cleansing, analysis, and graphing are all possible with R's statistical computing and graphical capabilities. Data processing and storage are made easy with the R tool [25]. The most potent and widely used programming language for data research, visualization, and computational statistics is R.
- *Rapidminer*
The data mining software RapidMiner is available without cost. With it, you can prepare data, learn algorithms, and deploy models. Numerous options for creating fresh data mining activities and evaluating potential setups are included in this free data mining program. Its drag-and-drop interface and pre-built models enable non-programmers to create predictive processes for specific use cases like fraud detection and customer attrition. Rapid Miner's R and Python extensions allow programmers to customize their data mining in the interim. Powerful data mining software called RapidMiner Studio supports everything from model deployment to model operations.
- *Weka*
Weka is open-source mining program built on the Java platform. It was created at the Waikato University in New Zealand. It is a situation where the necessary steps in data mining can be completed. Weka is a collection of machine learning algorithms for data mining jobs. It includes tools for data preparation, categorization, regression, clustering, association rule mining, and visualization [21].
- *Orange*
Orange is a strong platform for data analysis and visualization, allowing users to see data flow and increase productivity. With a superior debugger that improves code debugging, Orange is the tool that is the simplest to learn. Python can be used to mine data by developers, on the other hand. This platform includes tools for data analytics, text mining, bioinformatics, and machine learning.
- *Knime*
Data scientists can build their own independent apps and services using the robust drag-and-drop drag-and-drop data mining tool KNIME [25]. Data science is produced using the free source KNIME Analytics Platform. By being straightforward, open, and constantly incorporating new technologies, KNIME enables understanding data and creating data science workflows and reusable components accessible to everyone.
- *Keel*
KEEL is a Java open-source software tool that enables users to assess the performance of evolutionary learning and Soft Computing based algorithms for a range of data mining problems, such as regression, classification, clustering, pattern mining, etc. For constructing experiments with diverse datasets and soft computing techniques to assess the behavior of the algorithms, KEEL offers a straightforward data stream GUI.
- *Solver*
Solver data mining is a group of tools that displays precise data in a number of ways. It provides methods for classification, prediction, affinity analysis, data exploration, and reduction in context using statistical and machine-learning techniques. Excel data mining, forecasting, and visualization are all made simple using Solver's XLminer, a professional data mining program. It includes a wide variety of data preparation tools for gathering and cleansing user data.
- *Spark*
Spark tools, which are utilized for effective and reliable data processing in big data analytics, are the framework's primary software components. Open-sourcing of the Spark framework is done under the Apache license. For interactive searches, machine learning, and real-time workloads, use the free and open-source Spark platform.
- *Datawatch*
An industrial intelligence and data mining system is Datawatch Desktop. It enables the user to concentrate on real-time data visualization. Without writing a single line of code, it offers customers capabilities to construct and maintain monitoring and analytical systems. Renters, building owners, and property managers can all rely on Datawatch Systems for cutting-edge technology solutions that provide complete asset security.
- *K-Core*
Network analysis, spam detection, and biology are three fields where data mining algorithms regularly use the k-core feature. Compared to other density-based measures like the densest sub graph, it has the advantage of giving a score to every node in the network. The highest number of items that are individually related to at least k other entities in the collection are gathered by a k-core tool. On a network, small, interconnected core sites are referred to as K-Core.

VII. APPLICATION

➤ *Each Field can Benefit Fundamentally from Data Mining. Several of them are Listed below.*

- *Education*
Educational data mining is a method for turning unstructured data from enormous educational databases into useful and insightful information (EDM). The advantages of data mining for the education sector are: Identifying student performance and making performance predictions. Predicting dropouts' academic achievement. To monitor the academic progress of the kids and the effectiveness of their teachers. To improve both the teaching process and teacher performance.

- *Health Care*

Data mining methodologies and technologies in healthcare can be useful for decision-making in disease diagnosis and prediction, predictive medicine, physician's best practices, healthcare administration, and gauging the efficacy of particular treatments [22]. Health systems can employ data collection to create comprehensive patient profiles, customize therapies, advance treatment techniques, improve patient-doctor communication, and boost health outcomes.

- *Finance*

Financial institutions that provide services to their clients or members collect financial data. Finance is one of the most fascinating data mining application fields in these new technologies. For managing massive amounts of data, organizational effectiveness, and business intelligence, data mining and machine learning are essential. Data mining for the financial sector has the following advantages: It helps with threat and fraud detection.

- *Social Media*

To uncover hidden trends and patterns, data mining is employed across social media platforms including Twitter, LinkedIn, Facebook, and others. Social media data mining aims to gather useful information from users, spot patterns and trends, and derive business conclusions. The advantages of social media data mining include: To identify the most pertinent facts, gather, arrange, and process a lot of data. Utilizing automated data analysis allows for time, resource, and labour savings.

- *Retail Marketing*

Ever-growing datasets are being searched through with data mining to better categorize the retail market. By analyzing the relationships between elements like consumer age, gender, and preferences, for instance. Retail data mining can help in identifying user buying patterns, identifying user shopping themes and patterns, improving user service, increasing user happiness and retention rates, raising ratios of items consumed, and creating more effective logistics and transportation plans for goods. Market basket analysis is a data mining approach used by merchants to boost sales by tracking and examining client buying trends or habits from big databases of past purchases.

- *Crime Analysis*

Data mining applications let businesses and users find hidden information in a vast database of criminal records in order to monitor, track, and prevent crime. Data mining for the criminal justice sector has the following advantages: Used to represent problems in criminal detection. Find hidden relationships in data by using artificial intelligence tools. Bolster criminal analysis, contribute to crime prevention, and reduce crime Aid in the analysis of crime data by the police.

- *Telecommunication*

The communications industry was a pioneer in data mining, deploying a number of big data applications [22]. There are many data mining applications because the telecommunications industry was one of the first to adopt data mining technology. Data mining can be used to find network problems, improve corporate effectiveness, and uncover telecom fraud. In the telecoms industry, telecommunication organizations maintain the utmost confidentiality regarding information like who each customer calls. Privacy worries have so far been limited because the majority of telecommunications companies utilize this information responsibly. The telecoms industry is more worried about specific regulatory restrictions on the usage of data [23].

- *Fraud Detection*

Fraud detection is a process that involves a number of procedures and analysis to assist businesses find and stop fraudulent financial activities. This can include, among other things, insurance fraud, identity theft, credit card fraud, and cybercrime. Websites, policies, employee training programs, and other security measures can all be connected with fraud detection. Data mining for the Fraud Detection business has the following advantages: By classifying and segmenting data groups that can accommodate millions of transactions, one can look for trends and fraud by doing so. Using neural networks, suspicious patterns are learned and applied to find additional repeats.

VIII. ALGORITHMS

An algorithm is a set of presumptions and calculations used in data mining to create a model from data. The industries' most practical algorithms are briefly described.

- *Decision Tree*

Decision trees use a variety of ways to determine whether to split a node into two or more sub-nodes. With the development of sub-nodes, the consistency of the generated sub-nodes improves. Building Classification Models makes use of a particular data mining technique called Decision Tree Mining. The intended outcome is predetermined in supervised learning. Both qualitative and numerical data can be used with decision trees. Decision trees employ inductive inference procedures to estimate a goal function that will yield discrete values. It is widely used, resistant to noisy data, and regarded as an effective technique for learning disjunctive phrases [19].

- *Clustering*

A number of abstract things are grouped together into classes of related objects through the process of clustering. Statistical data processing techniques include cluster analysis. Grouping a set of items so that they are more similar to one another than to those in other groups is the work of cluster analysis, often known as clustering. The clustering data mining approach is employed to find data that is similar. The procedure helps to comprehend data's similarities and differences. This method is somewhat similar to classification; however, the former requires

grouping together sets of data based on similarity. Clustering is the process of grouping the population or data points so that they are more similar to one another within their own group than they are to those in other groupings [12].

➤ Classification

The data mining function of classification divides up the objects in a collection into specific groups or classifications. To group data into a class or category, classification techniques are utilized. Both structured and unstructured data may be used in its execution. Generally speaking, a classification algorithm is a function that divides a class into positive and negative values based on the input features [24].

➤ Association Rule

Data mining function called association determines the likelihood of elements in a collection occurring together. For finding intriguing relationships between variables in sizable databases, association rule learning is a rule-based machine learning technique [12]. Large volumes of data are analyzed using association rule mining to uncover intriguing linkages and relationships. It searches for the rules governing how or why many things are frequently purchased together in a same transaction [24].

➤ Artificial Neural Network

A neural network is a programming paradigm that mimics the function of brain nerve cells. A computational model called an artificial neural network (ANN) is made up of numerous processing elements that take inputs and outputs based on a specified perceptron [17].

➤ Naïve Bayes

The Naive Bayes algorithm, a supervised learning method, is based on the Bayes theorem. A probabilistic computational model that can be used for a variety of classification applications is the Naive Bayes method. Useful applications include text classification, spam filtering, and sentiment analysis.

➤ Support Vector

Both classification and regression problems can be resolved using the sophisticated Supervised Learning method known as the Support Vector Machine. A particular kind of statistical model is a technology known as a support vector machine. It applies the kernel trick, a technique for transforming user data, to get the best margin between potential values. One of the best and most effective classification algorithms is support vector machine. For both linear and nonlinear data, it is a new classification technique.

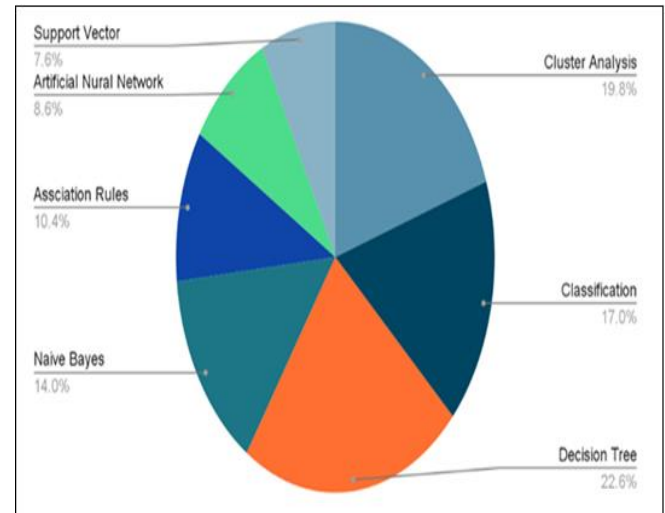


Fig 5 Specific Popular Algorithms

(<https://link.springer.com/article/10.1007/s10916-018-1018-2>)

IX. DISCUSSION

Large unsupervised data sets can be mined for valuable information using data mining techniques. It is a technology that uncovers patterns and correlations in a sizable dataset and then analyses those patterns to generate helpful data for additional study. Nearly all industries, including those in education, healthcare, banking, social media, retail, criminal analytics, telecommunication, fraud detection, and others, have greatly benefited from its application. It has sped up the creation of management and analytical tasks and cut down on the amount of time and human effort needed to organize massive data sets.

Data mining offers many advantages, but it also has some disadvantages, much like every instrument. Dependence on this method can occasionally neglect the value of human intervention because it is a strong strategy that relies heavily on computers. Therefore, if someone is familiar with the method and the variables used in data mining, he or she can input false data in a way that the method won't be able to tell apart from the true data, leading to the production of misleading information. It's important to acknowledge some of these disadvantages.

➤ Education

One such technology that deals with educational data for educational institutions or the entire educational system is called educational data mining (EDM). Any student's personal information, academic history, test scores, attendance in class, and other details are used. Utilizing the book information also contributes to enhancing the library system. As a result, there are certain issues in this situation. Any official with access to the database system can change a student's data, including their name, contact information, grades, and attendance.

➤ *Health Care*

Data mining is also essential in the healthcare industry. A health care system has a wealth of information regarding individuals, illnesses, medications, and other research projects. These data sets require careful analysis since they are so complex. A significant impact on public health could result from tampering with this data due to ignorance or negligence. There may be many different drawbacks associated with it.

➤ *Finance*

One industry that deals with a lot of data is the financing sector. Without a data mining tool, it is hard to keep track of the information on any financial institution or entire system because the volume of data is so large. But like many industries, data mining in this one has some significant disadvantages. Since financial data is primarily concerned with money, this data can paint a broad picture of economic activity. Therefore, a mistake in this data could depict the local or national economy incorrectly. The majority of transactions are now made online. These transactional data are conveniently stored in databases. These statistics can be used to calculate cash flow, which industries account for the majority of the economy, and how supply and demand for money are changing.

➤ *Social Media*

Social networking is become a hub for each person's personal information. Social media institutions now have easy access to data about individual behaviour, lifestyle, mental health, and other inner issues that are typically very difficult to extract. AI's use of social media for data mining advances the process. From the massive volume of unsupervised data, this AI method extracts relevant information. Even so, some of its disadvantages are more serious than others. Social media uses user data to better accurately forecast each user's behavior patterns and can occasionally have an unintended consequence. This, it would appear, converts a social media user to a product. Some social networking sites sell users' personal data to third parties, which in some situations could put users at risk.

➤ *Retail Marketing*

In retail marketing, data mining is also quite important. Data in the retail sector primarily consists of sales and product information. Therefore, this information is utilized to determine whether a specific product is sold in a specific location. This facilitates decisions on the distribution of goods in various regions. Along with retail businesses, traders also examine their profits for several product categories. Therefore, the issues with these data are: Lack of experience is the biggest issue with data mining in the retail industry. The user of the retail market data mining technique does not fully understand the process he is operating in.

➤ *Crime Analysis and Fraud Detection*

Security-related topics like crime analysis and pattern detection employ data mining. Any person's specific actions are followed on social media and in real life. Through several effective data mining techniques, the information

gathered through this procedure was eventually organized and classified. This output format aids in identifying area and individual criminal patterns. Sometimes crime analysis is followed by an automated process with no human involvement. As a result, we only have partial information about this crime's commission and its perpetrator.

Finally, at the time of our analysis, we understand that only a select few industries are currently using data mining. The most common industries to use data mining techniques are education and healthcare. However, crime analysis might be a fantastic area for data mining study. In terms of timing, locations, conduct, etc., it can assist in identifying crime patterns. It will improve government system efficiency and increase the security of social life.

X. CONCLUSION

Data mining is an emerging technology. Simple quantitative data and category data are the current emphasis of data mining. Future developments in data mining will include increasingly intricate data types. In addition, by examining additional variables and their relationships, any model that has already been developed can be improved even further. Data mining research will lead to the creation of fresh methods for identifying the most fascinating aspects of data. Models can be used as an enrolment management tool as they are developed and implemented.

As mentioned above, data mining helps industries increase revenue in a variety of ways. Data mining should be used in other fields, nevertheless, in order to handle huge datasets. We aim to determine how other firms might benefit from data mining and increase their volume of business. Crime analysis, fraud detection, and other disciplines have already started using the data mining method, where there is still much opportunity for development.

ACKNOWLEDGEMENTS

- *The Authors would Like to Thank all of My Students.*

REFERENCES

- [1]. N. Jothi, N. A. Rashid and W. Husain, "Data Mining in Healthcare – A Review," *Procedia Computer Science*, vol. 72, pp. 306-313, 2015.
- [2]. S.-H. Liao, P.-H. Chu and P.-Y. Hsiao, "Data mining techniques and applications – A decade review from 2000 to 2011," *Expert Systems with Applications*, vol. 39, no. 12, pp. 11303-11311, 2012.
- [3]. P. F. B. S. and B. Sugiantoro, "A review paper on big data and dataMining Concept and Techniques," *International Journal on Informatics for Development*, vol. 7, no. 1, pp. 36-38, 2018.
- [4]. X. Teng and Y. Gong, "Research on Application of Machine Learning in Data Mining," *IOP Conference Series: Materials Science and Engineering*, vol. 392, no. 6, 2018.

- [5]. G. M. Weiss and B. D. Davison, "Data Mining," To appear in the Handbook of Technology Management, 2010.
- [6]. C. Romero and S. Ventura, "Data mining in education," WIREs Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12-27, 2012.
- [7]. M. N. O. Sadlku, K. G. Eze and S. M. Musa, "Data Mining in Healthcare," International Journal of Advances in Scientific Research and Engineering, vol. 4, no. 9, pp. 90-92, 2018.
- [8]. E. Vityaev and B. Kovalerchuk, "Data mining in finance: From extremes to realism," Journal of Financial Transformation, vol. 11, pp. 81-89, 2004.
- [9]. K. C. Khatib, T. D. Kamble, R. R. Chendake and G. N. Sonavane, "Social Media Data Mining For Sentiment Analysis," International Research Journal of Engineering and Technology (IRJET), vol. 3, no. 4, 2016.
- [10]. B. M. Ramageri and D. B. L. Desai, "ROLE OF DATA MINING IN RETAIL," International Journal on Computer Science and Engineering, vol. 1, p. 5, 2013.
- [11]. N. Jabeen and P. Agarwal, "Data Mining in Crime Analysis," in Proceedings of Second International Conference on Smart Energy and Communication, 2021.
- [12]. N. Petrovic, "Adopting Data Mining Techniques in Telecommunications Industry: Call Center Case Study," in IEEEESTEC - 11th Student Projects Conference, Serbia, 2018.
- [13]. R. G. Bhati, "A Review on Present Technologies for Fraud," International Journal of Applied Engineering Research, vol. 14, no. 7, 2019.
- [14]. R. Liorente and M. Morant, "Data mining in higher education," New fundamental technologies in data mining, pp. 201-220, 2011.
- [15]. I. Ijarcet, "A Review on Data Mining in Healthcare," International Journal of Advanced Research in Computer Engineering & Technology, 2018.
- [16]. H. A. A. Hafez, "Mining Big Data in Telecommunications Industry: Challenges, Techniques, and Revenue Opportunity," International Journal of Computer and Information Engineering, vol. 10, no. 1, 2016.
- [17]. P. A. Joshi and S. N. Bhamare, "A Study of Social Media Data and Data Mining Techniques," International Journal of Engineering Research and Application, vol. 7, no. 4, pp. 46-52, 2017.
- [18]. M. Injadat, F. Salo and A. B. Nassif, "Data mining techniques in social media: A survey," Neurocomputing, vol. 214, pp. 654-670, 2016.
- [19]. S. Sathyadevan, M. S. Devan and S. S. Gangadharan, "Crime analysis and prediction using data mining," in First international conference on networks & soft computing (ICNSC2014), 2014.
- [20]. B. Khalid and N. Abdelwahab, "A Comparative Study of Various Data Mining Techniques: Statistics, Decision Trees and Neural Networks," International Journal of Computer Applications Technology and Research, vol. 5, no. 3, pp. 172-175, 2016.
- [21]. R. Ratra, P. Gulia and N. S. Gill, "Performance Analysis of Classification Techniques in Data Mining using WEKA," in Proceedings of the International Conference on Innovative Computing & Communication (ICICC), 2021.
- [22]. R. Ranjan, S. Agarwal and D. S. Venkatesan, "Detailed Analysis of Data Mining Tools," International Journal of Engineering Research & Technology, vol. 6, no. 5, 2017.
- [23]. S. Mirza, D. S. Mittal and D. M. Zaman, "A Review of Data Mining Literature," International Journal of Computer Science and Information Security (IJCSIS), vol. 14, no. 11, 2016.
- [24]. G. Weiss, "Data Mining in the Telecommunications Industry," Networking and Telecommunications: Concepts, Methodologies, Tools, and Applications, p. 8, 2010.
- [25]. G. M. Weiss, "Data Mining in Telecommunications," Data Mining and Knowledge Discovery Handbook, pp. 1189-1201, 2005.
- [26]. S. Jones and O. K. Gupta, "Web Data Mining: A Case Study," Communications of the IIMA, vol. 6, no. 4, 2006.
- [27]. M. Bansal and M. Kaur, "Analysis and Comparison of Data Mining Tools Using Case Study of Library Management System," International Journal of Information and Electronics Engineering, vol. 3(5), pp. 466-469, 2013.