

Exploring Data Mining Applications and Techniques: A Comprehensive Research Survey

M. Vasumathy ¹, Dr. Ujjwal Agarwal ², G. Divyamrutha ³, Prof (Dr.) Harikumar Pallathadka ⁴, Dr. Dolpriya Devi Manoharmayum ⁵

¹ Assistant Professor, The American College, Madurai, Tamil Nadu, India.
vasumathy@americancollege.edu.in

² Lecture (Information Technology), College of Computing & Information Sciences, University of Technology and Applied Sciences, Salalah, Oman. drujjwalagarwal@gmail.com

³ Assistant Professor, Electronics and Communications Engineering Department, Mallareddy Engineering College, India. divya.gnits@gmail.com

⁴ Professor and Director, Manipur International University, Imphal, Manipur, India,
harikumar@miu.edu.in, <https://orcid.org/0000-0002-0705-9035>

⁵ Assistant Professor, KL College of Agriculture, Koneru Lakshmaiah Education Foundation - KLEF (Deemed To Be University), Guntur, Andhra Pradesh, India. dolpriya.ag@gmail.com,
<https://orcid.org/my-orcid?orcid=0000-0003-1188-9450>

Abstract: Every second, huge amount of data is generated and accumulated. This data could possibly be used in forecasting the future. Data mining uses this data and generates valuable information which can be transformed into relevant knowledge. Data mining is a technique of identifying outliers, behaviours, trends of patterns and relationship among huge datasets. It is hugely associated with the skill of decision making. The knowledge on a relevant subject will help in understanding future trends. This survey paper supplies the overview of data mining, the processes involved, the scope it can offer, its different techniques and multiple applications. Data mining is a great model of using data efficiently.

Keywords: Data Mining, KDD, Applications, Techniques, Machine Learning, Artificial Intelligence.

1. INTRODUCTION

The present world of rapid technological globalization contains a massive amount of data. This data would end up in waste if not converted into useful information. The research done in the field of information technology arrived at a solution to manipulate the precious data. Data mining is a method where the ocean of unfiltered data is used to obtain required knowledge and spot the pattern in order to interpret the future trends. The increasingly accumulating database is the fodder for creating new technologies based on conclusions arrived after data mining. Since, the datamining is accurately in relation with converting the information to knowledge, it is also called as KDD (knowledge discovery in database)

2. LITERATURE SURVEY

K. Vanishree et.al elaborates various applications involved in data Mining [1]. S. S. Al-Rifai et.al list out various components exists in data mining applications [7]. E. S. Priya et.al done a survey about a basic descriptive overview about data mining. The author also elaborated about the principal processes of data mining starting from data preparation to knowledge extraction The ideas of the components of data mining are drawn out precisely [8]. N. Padhy

et.al describes how the Data Mining Applications and Feature Scope Keywords Data mining task, Data mining life cycle, Visualization of the data mining model, Data mining Methods, Data mining applications have been executed in these modern environment [2]. The Author has given a sketch about the methods under data mining and attempts for visualizing as a proposed model. S. Mukherjee et.al discussed the various applications and techniques has been implemented in data mining [24]. A. H. Mohamed et.al list out the various data mining techniques approved and executed successfully [20]. S. D. GHEWARE et.al introduced various Task Tools Techniques and Applications in data mining [14]. H. A. Madni et.al introduces new techniques and applications which suits for data processing and storage. The author also briefs about the techniques and its classifications [13]. The article evaluates the versatility in data mining by exploring each technique and its utilisations and also an overview about the applications of data mining has been given. Data mining is used in various fields for important purposes. These applications in certain fields are remarkable. This survey also mentioned papers elaborate the applications of data mining in the booming educational field which portray the advantages and the problems data mining has solved in educational sector. T. Devasia et.al surveyed Prediction of student's performance using Educational Data Mining [6]. S. Lakshmi Prabha and A. R. M. Shanavas list out the various data mining applications related with education [9]. O. A. Santana has performed the research about Teacher in Brazil for Social Mobility and Elements for Federalism Strategies [25]. A. Iyer et.al proposed Data mining techniques for Diagnosis of Diabetes Using Classification Mining [10]. S. Anwar Lashari et.al introduced applications of Data Mining Techniques for Medical Data Classifications [23]. P. G. Subbarao et.al performed a Case Study on Data Mining Application in Health Care Monitoring Systems, which summarises the considerable amount of contribution done by data mining in health sector [18]. The above articles emphasize new inventions, solutions, evolutions and ideas brought by data mining applications. H. Hassani et.al made a review of data mining applications in crime [6]. Q. Chang and J. Hu proposed research about the applications of the Data Mining Technology in Economic Intelligence System [5]. L. Ren and Y. Chen, has performed the research on the application of Data Mining Technology in Military Audit which discusses about the economic, military, criminal and investigational applications of data mining [4]. Thus, our study demonstrate about how powerfully and brightly data mining could be used to solve modern issues.

3. DATA MINING PROCESSES

As to obtain the goal of acquiring knowledge from data mining, few steps have to be followed. Knowledge extraction is done in seven stages:

DATA CLEANING: It is the process where we clean out the irrelevant data from the raw data

DATA INTEGRATION: In this step, data is gathered from many resources and combined into one single form of data

DATA SELECTION: We select the data most related with our purpose in which we will do our work in the further steps

DATA TRANSFORMATION: Data transformation is the step where we convert or transform our data into the standard forms for mining.

DATA MINING: In this important step, knowledge is extracted from data by following some techniques.

PATTERN EVALUATION: It is the step where we identify the pattern out of data mining

KNOWLEDGE REPRESENTATION: This is the final step in which the knowledge that is mined is represented in visualization methods

DATA MINING ARCHITECTURE

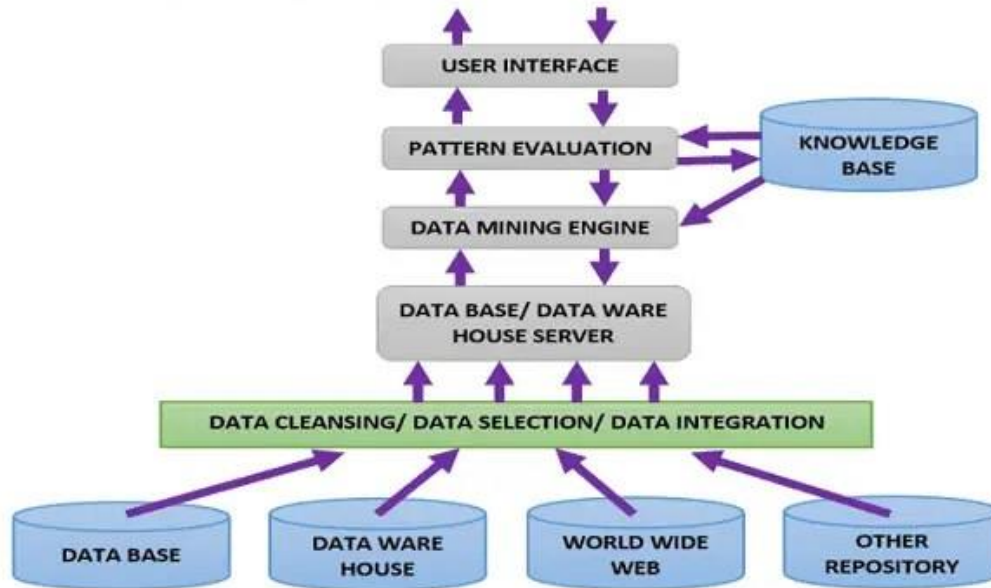


Figure 1: Data Mining Architecture

4. ROOTS OF DATA MINING

STATISTICS: Statistical data mining plays a significant role in data mining since the base and foundation of data mining relies on the statistical records. Statistics is one of the main roots of data mining too. Statistics is relevant to the subjects of regression analysis, standard deviation, standard distribution, cluster analysis, discriminate analysis, confidence intervals and standard variance. The mentioned papers help in studying the data and determining the data relationships.

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING: The use of machine learning has led to faster data mining. Artificial intelligence further helps to clean out the irrelevant data from the data sets, find the pattern of relationship amongst data which is complicated for human beings. This boosts the speed of the purpose. AI has automated the process to help mine huge datasets which contains customer database, transaction records, log files etc.,

DATABASES: The third root of data mining is databases. In data mining, huge amount of data is generated and it needs to be stored. Here, databases helps data mining in storing this data in repositories. Earlier, all these data were stored as records and fields but now with the help of databases all the data could be managed.

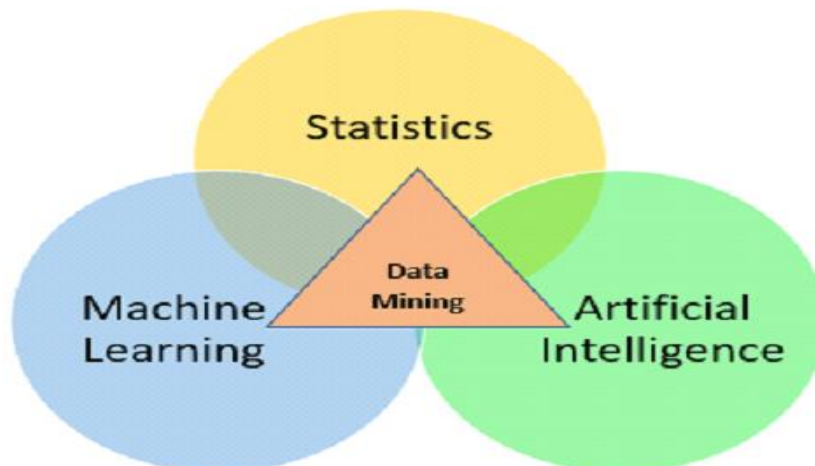


Figure 2: Roots of Data Mining

5. STEPS IN DATA MINING PROCESS

a. Business Understanding:

This Step is about understand the problem that is to be eradicated. It also involves defining the objectives of the data mining project.

b. Data Understanding:

This step is about gathering and exploring the data in order to obtain a better understanding of its structure quality and content. In this step identification of data quality issues, patterns and relationships is vital.

c. Data Preparation:

Data Preparation involves preparing the data for the purpose of analysis. In this step cleaning the data in order to remove errors, converting the data and integrating the data are done.

d. Modelling:

Using Machine Learning algorithm, a predictive model is built in this step. This includes training the model, selecting the algorithm and finally evaluating its performance

e. Evaluation:

Evaluation is using the statistical data to assess the performance of the model and checking if it is capable of predicting outcomes of new data.

f. Deployment:

In this step the deploying of the model into the production environment occurs. Integrating the model into systems and processing to make predictions in real time is involved in deployment. Deployment generates value for the organization.

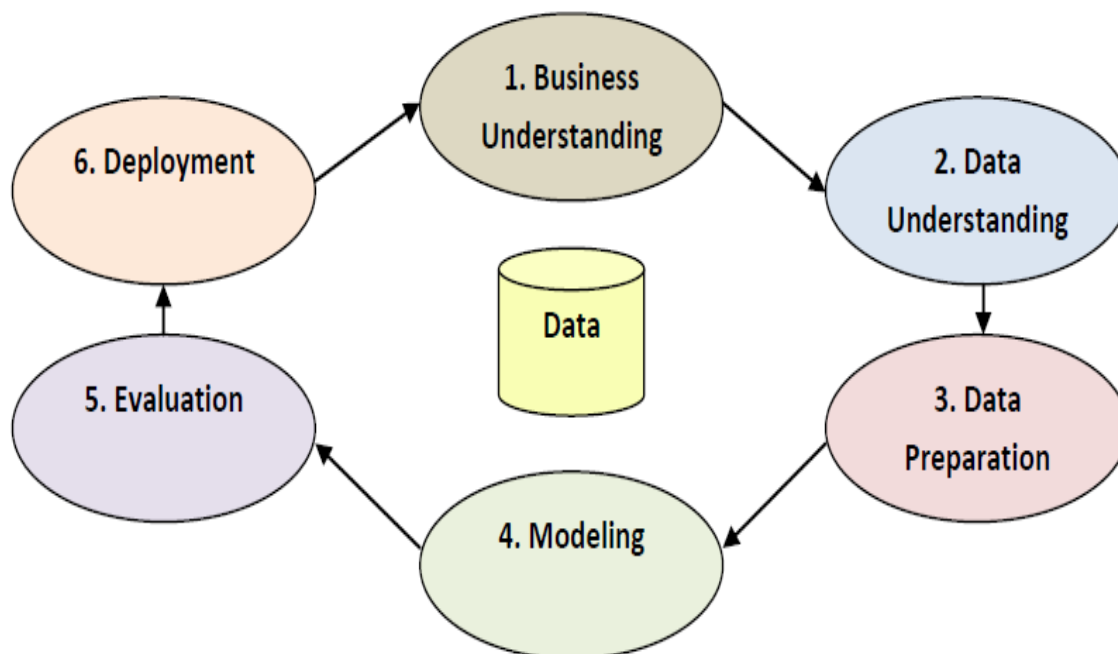


Figure 3: Steps in Data Mining process

6. DATA MINING TECHNIQUES

Data mining includes various types of tasks according to our purpose. Depending on these task types, the techniques of data mining are applied. These techniques, for our understanding could be classified as predictive and descriptive.

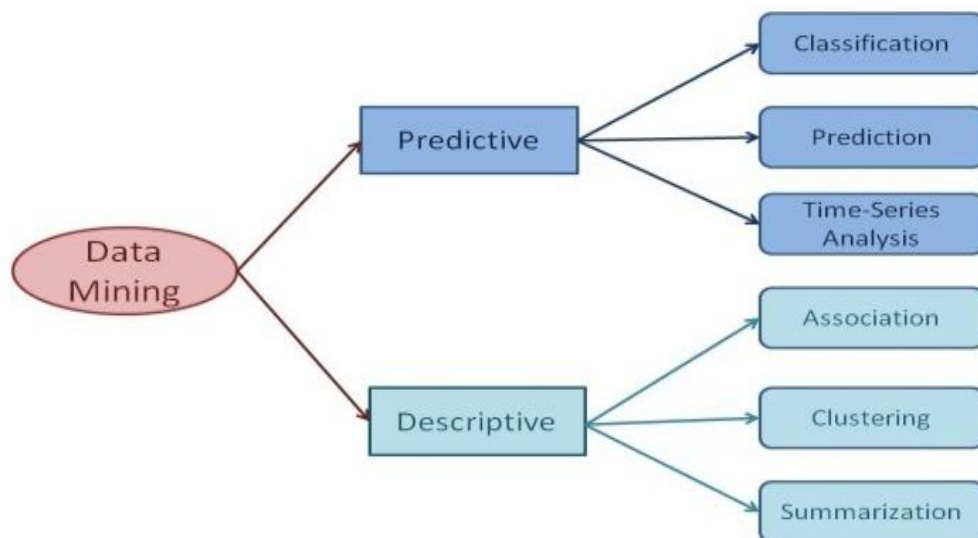


Figure 4: Data Mining Techniques

PREDICTIVE: In this, we predict results of future trends and issues priorly with the help of pre-existing data that which includes classification, regression, outlier detection and neural network etc.,

CLASSIFICATION: In a collection of data, data is classified according to some criterion which distinguishes two data. It contains labels so there is no need for training and testing data set for verifying the model created.

Eg: Classification of fabrics according to the origin and Classification of students according to age groups.

Types of classification methods:

- 1) Classification by decision tree inclusion
- 2) Bayesian classification model
- 3) Neural network
- 4) Classification by association
- 5) Support vector machines (SVM)

REGRESSION: In this regression technique, data we already have is directly used to plot the future data. Here, a relationship between one or more independent variable and dependent variable could be seen. Response variables are variable we will be predicting through the independent variables that is the known value. Regression is generally plotted on a graph. Some problems like stock prices and product failures are difficult to predict with just two variables. In this case, regression cannot be helpful.

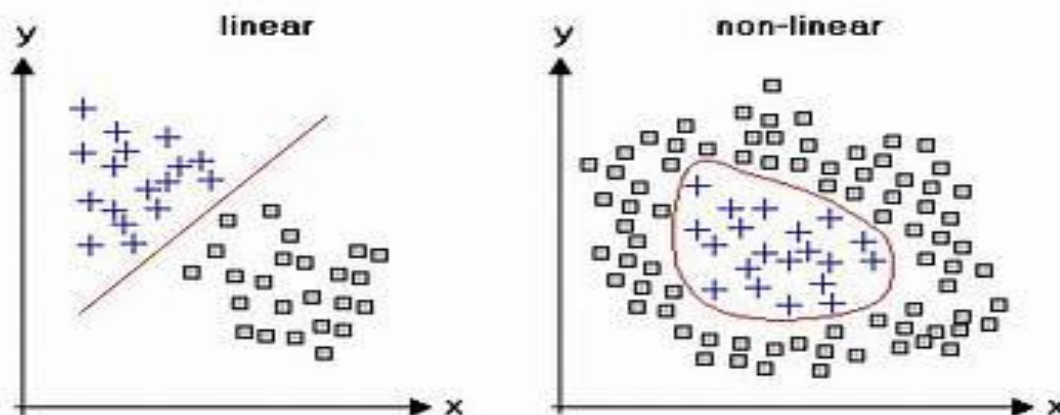


Figure 5: Regression

Types of Regression:

- 1) Linear Regression
- 2) Multivariate linear Regression
- 3) Nonlinear Regression
- 4) Multivariate non-linear Regression

ARTIFICIAL NEURAL NETWORK: Neural network is set of input or output units which contains a weight in it. It is inspired by the functioning of neural network in our brains. It is a model of classification. Neural networks are highly capable of deriving meaning from complicated set of data. It is also used to withdraw patterns and spot out trends which are too complex for humans to comprehend. Thereby, neural networks are the best for forecasting and spotting the future trends. Back propagation is a type of neural network.

Types of neural network:

- 1) Perception
- 2) Feed forward neural network
- 3) Multilayer perception
- 4) Convolutional neural network
- 5) Radial basis functional neural network
- 6) Recurrent neural network
- 7) LSTM
- 8) Sequence to sequence models
- 9) Modular neural network

TIME-SERIES ANALYSIS: In time series analysis, a sequence of data is collected over an interval of time. Then, the analysts record data points for the collected database over a set of periods of time. This technique therefore helps in predicting the next trend over an interval of time.

Models of time series analysis:

- 1) Classification
- 2) Curve fitting
- 3) Descriptive analysis
- 4) Explanative analysis
- 5) Exploratory analysis
- 6) Forecasting
- 7) Intervention analysis
- 8) Segmentation

PREDICTION: This technique is related with the regression technique. Here, the relationship between the dependent variable and the independent variable is studied as to bring out a pattern in which the future trend might occur. The values plotted might be in an ordered manner and it is continuous.

DESCRIPTIVE: In descriptive techniques, the database is examined to form descriptions about the pattern. The descriptions are additionally used to study about the information's. This includes clustering, sequential pattern discovery, summarization, association rule.

CLUSTERING: Clustering technique is about grouping data based on their similarities and distinguishing it from the other clusters based on their dissimilarities. To put it in technical terms, this technique focuses on maximising the intra- similarities and minimising the inter-similarities considering the data point attributes. This will result in the formation of clusters.

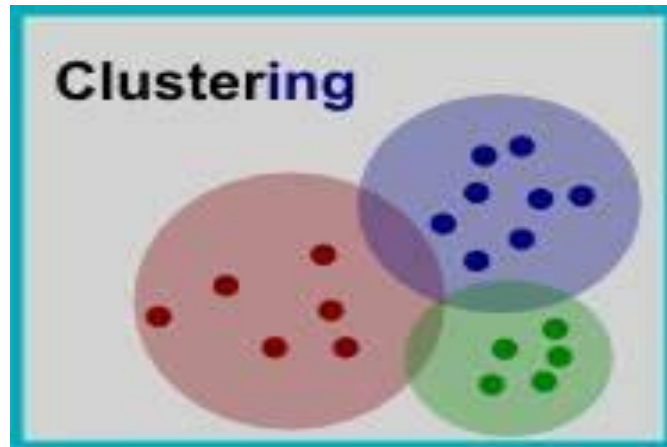


Figure 6: Clustering

Types of clustering methods are:

- 1) Partitioning method
- 2) Hierarchical method
- 3) Density based method
- 4) Grid-based method

ASSOCIATION RULE: Among large database, a relationship between two items is found to form patterns. This technique may be highly efficient in decision making process of business. For instance, this is helpful in retail management, catalogue design and cross marketing.

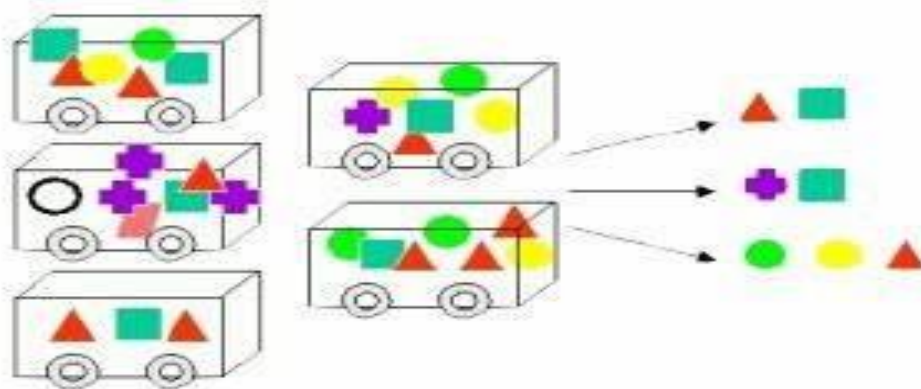


Figure 7: Association

Types of association rule:

- 1) Multilevel associational rule
- 2) Quantitative associational rule
- 3) Multidimensional association rule

SEQUENCE DISCOVERY: Sequence discovery is a data mining technique that discovers the relationship among sequential data. it is also known as sequential pattern mining. Generally, sequence discovery holds certain criterion

such as frequency, occurrence, duration in order to find the patterns. The pattern is expected to occur in a certain sequence. Eg: A person purchasing a canon camera is expected to buy a printer followed by that sequentially.

SUMMARIZATION: In summarization technique, a set of relevant tasks are gathered. Followed by that, a summary is drawn out. This technique basically gives out a short abstraction of the pre-existing data in need of analysis. The entire dataset is minimised into a report in an informative manner.

Types of summarizations:

- 1) Extractive
- 2) Abstractive summarization

7. CHALLENGES OF DATA MINING

Data mining is increasingly important considering the amount of data that needs to be managed. Data mining is also largely relied on these days because of great efficiency it shows in it. However, it too has its own challenges. In this paper, we will explore some of the main challenges.

- 1) Data Quality
- 2) Data Complexity
- 3) Data Privacy and Security
- 4) Scalability
- 5) Interpretability

8. DATAMINING APPLICATIONS

As we already saw above, data mining is about using the pre-existing data to derive knowledge from it and use it to predict the future trends. Hence, this would be highly in many sectors and fields. Data mining could apply in real life problems which relies on predication and forecasting. Thereby data mining is efficiently used in case of decision-making queries. It has greatly contributed in invention of technologies. The benefited organisation includes retail stores, hospitals, insurance companies and security associations. However, data mining is not fully developed yet. Yet, data mining helps business massively as it helps to understand about the customer behaviour, future upcoming problems and finding solutions.

BUSINESS SECTOR: In business sector, data mining is broadly used in extracting abstracts of data. Most business organisations use data mining for analyzation os their Performance and deriving customer feedback. Prediction and forecasting done by the methods od data mining help the business sector to imply models and understand threats, risks and opportunities. Data mining acts as an aid to business sector in a way that it helps profoundly in the matters of decision making.

EDUCATION: Data mining in this field is a new but useful step. Education sector could be seen in a rapid growth over the past few years. Data mining in this field concentrates on the growth of educational institutions, preferred path taken by students and the behavioural changes and adaptation of students. Descriptively, data mining could understand the models that students prefer as easy and solve the resented problems. The patterns in which students learn are examined and studied to come up with better teaching assistance.

CRIMINAL INVESTIGATION: In the field of criminal investigation, the sets of previous record are stored. from this set of data, the pattern is derived. Using these patterns, the criminal and accused are discovered. In the second case, the criminal behaviour is identified using outlier analysis due to odd data. Data mining tools used in this field are Weka, H2o, Orange etc.

HEALTHCARE: Data mining can be used to check the effectiveness of medical treatments. Healthcare can prevent or eradicate chronic diseases by identifying and tracking it with the help of data mining. Additionally, detection of inappropriate prescription by hospitals and clinics could be tracked. Application of data mining could be seen in the pharmaceutical industry where the patterns help in inventing new services. The record of patients in a hospital is used as a data to deliver ana analysis about the symptoms, causes, preventive treatments etc

E-COMMERCE: Data mining techniques specifically association rule is contributing to cross marketing. Data mining remarkably marks the patterns of searching patterns. This will help in improving the sites according to the trends.

BANKING AND FINANCE: Banks use data mining in several areas like marketing, fraud detection, risk management and investment banking. Data mining like in many fields plays a important role in decision making matters of banking. For loan approval, data mining's prediction technique is used. Stocks are also predicted with data mining though it is a complex mining process.

CLIMETAALOGY: Assessing the records of weather from a particular place in a particular period of time with data mining could predict the future meteorological cycle. Data mining play an important role in detecting the natural calamities that might occur in a place. Moreover, collecting the weather reports of a place over a long time will give us the report about the climatic condition of that place. The summarization tactic is used here.

SECURITY: Data mining has the ability to monitor several systems. In case of any intrusion, the system will be detected with the help of data mining. Therefore, data mining helps finding out the scams.

CLOUD COMPUTING: The main source and origin of data at the present data is cloud computing. It is highly preferred for its efficient advantages such as fast servers, reliability, less cost of infrastructure etc. hence, KDD techniques are more suitable here. this will help in deriving pattern from complex set of data.

MARKETING SECTOR: In the marketing field and in retail stores, data mining is seen as a high attribute source. This is so because data mining helps in prediction of the future trends and helps the retail store in updating itself. It will show the requirements of the customers in an order. This allows the manager of the store to discount prices based on their strategy. information given by data mining are the base in which the stores develop marketing strategies.



Figure 8: Data Mining Applications

9. FINDINGS OF THE SURVEY

From the comprehensive research done on the subjects of data mining, this paper presented the centre point of the concept of data mining. The conclusions drawn from the research provides the key points of data mining's future.

Data mining, being a growing technology has numerous advantages. Few of these advantages are given below:

It helps companies and business cooperatives collect reliable information. Compared to other applications for data gathering, data mining is both effective and cost efficient. Additionally, it helps in decision making. It is beneficial for fraud detection and detecting credit risks. Out of this, product safety can be built.

Data mining as we know, is widespread too. So, it contains its own risks and challenges. Based on the research done, we could find out the obstacles faced while using data mining. Here are the few major ones:

Data is everywhere and collecting all the data is impractical. So, the accuracy might be slightly mistaken. At times, the data might contain mistakes, inconsistencies, omissions, errors and duplications and this will ultimately lead to inaccuracies in the predictions. Sometimes, the data gathered may contain personal and confidential information. This might lead to privacy related issues.

10. CONCLUSION

In our considerably digitalized world, tremendous amount of data is generated and stored every second. Yet, we strive for knowledge and understanding. Data mining answers our quest for knowledge in an efficient way. In this paper, we discussed about concepts of data mining, KDD techniques. We mainly focused about the techniques of data mining which is basically classified into predictive (classification, regression, artificial neuron network, prediction, time series analysis) and descriptive (cluster analysis, association rule, summarization, sequence discovery). Moreover, an overview about different applications of data mining is presented.

11. REFERENCES

- [1]. K. Vanishree and K. Jayanthi, "Survey on data mining," *Int. J. Mech. Eng. Technol.*, vol. 8, no. 11, pp. 10–20, 2017.
- [2]. N. Padhy, P. Mishra, and R. Panigrahi, "The Survey of Data Mining Applications and Feature Scope Keywords Data mining task, Data mining life cycle, Visualization of the data mining model, Data mining Methods, Data mining applications," *Int. J. Comput. Sci. Eng. Inf. Technol.*, vol. 2, no. 3, pp. 43–58, 2012.
- [3]. A. Anshu, "Review Paper on Data Mining Techniques and Applications," *Int. J. Innov. Res. Comput. Sci. Technol.*, vol. 7, no. 2, pp. 22–26, 2019.
- [4]. L. Ren and Y. Chen, "Research on the Application of Data Mining Technology in Military Audit," *Proc. - 2021 Int. Conf. Educ. Inf. Manag. Serv. Sci. EIMSS 2021*, vol. 2021, pp. 277–283, 2021.
- [5]. Q. Chang and J. Hu, "Research and Application of the Data Mining Technology in Economic Intelligence System," *Comput. Intell. Neurosci.*, vol. 2022.
- [6]. T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using Educational Data Mining," *Proc. 2016 Int. Conf. Data Min. Adv. Comput. SAPIENCE 2016*, pp. 91–95, 2016.
- [7]. S. S. Al-Rifai, A. M. Shaban, M. S. M. Shihab, A. S. Mustafa, H. A. Alhalboosi, and A. M. Shantaf, "Paper Review on Data Mining, components, and Big Data," *HORA 2020 - 2nd Int. Congr. Human-Computer Interact. Optim. Robot. Appl. Proc.*, 2020.
- [8]. E. S. Priya and K. Anandhan, "An Overview of Data Mining - A Survey Paper," *Int. J. Mod. Comput. Sci.*, vol. 6, no. 1, pp. 2320–7868, 2018.
- [9]. S. Lakshmi Prabha and A. R. M. Shanavas, "Educational Data Mining Applications," *Oper. Res. Appl. An Int. J.*, vol. 1, no. 1, pp. 23–29, 2014.
- [10]. A. Iyer, J. S, and R. Sumbaly, "Diagnosis of Diabetes Using Classification Mining Techniques," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 1, pp. 01–14, 2015.
- [11]. M. Durairaj and V. Ranjani, "Data Mining Applications In Healthcare Sector : A Study," *Int. J. Sci. Technol. Res.*, vol. 2, no. 10, pp. 29–35, 2013.
- [12]. A. S. Ashour, N. Dey, and D. N. Le, "Biological data mining: Techniques and applications," *Min. Multimed. Doc.*, vol. 1, no. 4, pp. 161–172, 2017.
- [13]. H. A. Madni, Z. Anwar, and M. A. Shah, "Data mining techniques and applications - A decade review," *ICAC 2017 - 2017 23rd IEEE Int. Conf. Autom. Comput. Addressing Glob. Challenges through Autom. Comput.*, no. September, 2017.
- [14]. S. D. GHEWARE, A. S. KEJKAR, and S. M. TONDARE, "Data Mining Task Tools Techniques and Applications," *Ijarcce*, no. November 2015, pp. 8095–8098, 2014.
- [15]. D. Papakyriakou and I. S. Barbounakis, "Data Mining Methods: A Review," *Int. J. Comput. Appl.*, vol. 183, no. 48, pp. 5–19, 2022.
- [16]. A. N. Paidi, "Data Mining : Future Trends and Applications," *Int. J. Mod. Eng. Res.*, vol. 2, no. 6, pp. 4657–4663, 2012.
- [17]. G. Schuh et al., "Data mining definitions and applications for the management of production complexity," *Procedia CIRP*, vol. 81, no. March, pp. 874–879, 2019.

- [18]. P. G. Subbarao, P. S. Khan, and K. V. Kumar, "Case Study on Data Mining Application in Health Care Monitoring Systems," *Res. Inven. Int. J. Eng. Sci.*, vol. 6, no. 5, pp. 79–82, 2016.
- [19]. C. Mehta, "Basics of Data Mining: A Survey Paper," *Int. J. Trend Res. Dev.*, vol. 4, no. 2, pp. 2394–9333, 2017.
- [20]. M. K. Najafabadi, A. H. Mohamed, and M. N. Mahrin, "A survey on data mining techniques in recommender systems," *Soft Comput.*, vol. 23, no. 2, pp. 627–654, 2019.
- [21]. K. Neha and M. Y. Reddy, "A study on applications of data mining," *Int. J. Sci. Technol. Res.*, vol. 9, no. 2, pp. 3385–3388, 2020.
- [22]. A. B. Devale, "applications of Data Mining Techniques in," vol. 2, no. 4, pp. 31–40, 2012.
- [23]. S. Anwar Lashari, R. Ibrahim, N. Senan, and N. S. A. M. Taujuddin, "Application of Data Mining Techniques for Medical Data Classification: A Review," *MATEC Web Conf.*, vol. 150, pp. 1–6, 2018.
- [24]. S. Mukherjee, R. Shaw, N. Haldar, and S. Changdar, "A Survey of Data Mining Applications and Techniques," *Int. J. Comput. Sci. Inf. Technol.*, vol. 6, no. 5, pp. 4663–4666, 2015.
- [25]. O. A. Santana, "Teachers in Brazil: Social Mobility and Elements for Federalism Strategies," *Creat. Educ.*, vol. 06, no. 15, pp. 1629–1637, 2015.
- [26]. H. Hassani, X. Huang, E. S. Silva, and M. Ghodsi, "A review of data mining applications in crime," vol. 9, no. 3, 2016.

DOI: <https://doi.org/10.15379/ijmst.v10i3.2736>

This is an open access article licensed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted, non-commercial use, distribution and reproduction in any medium, provided the work is properly cited.