

Basic Review on Data Mining

Bhagyashree Kambhare¹, Yamini Laxane², Achal Kharabe³, Tushar Mankar⁴

¹HOD and Professor, Department of MCA, smt Radhikatai pandav college of engineering Nagpur, Maharashtra, India.

²Professor, Department of MCA, smt Radhikatai pandav college of engineering Nagpur, Maharashtra, India.

^{3,4}Department of MCA, smt Radhikatai pandav college of engineering Nagpur, Maharashtra, India.

To Cite this Article: Bhagyashree Kambhare¹, Yamini Laxane², Achal Kharabe³, Tushar Mankar⁴, “Basic Review on Data Mining”, Indian Journal of Computer Science and Technology, Volume 04, Issue 02 (May-August 2025), PP: 210-213.

Abstract: Data excavation is a process that extracts valuable and unique embellished dossiers from enormous collections of dossiers. Many data excavating habits, algorithms, and friendships were employed to acclimate dossier excavating technologies in order to improve administrative progress and establish superior more effective habits. There is overwhelming development in data group as a result of the ever-increasing employment of calculators and stereos bias, as well as the extreme rise in canny capacity and storage ability. The storage of data in a dossier depository allows for total resourcefulness in piercing and systematizing a trustworthy current table meets that require the particular finishes known as data excavation forms.

Key Words: Technology, Pre-processing, Identification, Mining, Pattern.

1. INTRODUCTION

Information Mining is as prepare created & outlined to investigate expansive amounts of information show in bulk organize to find significant designs. The thing of information mining is filtration of information from expansive database, which can be utilized in major decision-making trade operations with the utilize of information mining instruments. Information mining, famously known Information Mining As Information Revelation in Databases (KDD), it's the nontrivial birth of certain, to begin with obscure and possibly valuable data from information in databases. Information mining is the birth of resigned prophetic examination from huge databases; it's a vital innovation with expansive outcome to assist associations concentrate on the foremost vital data in their information stockpiles. Information booby-trapping tools predicts unborn patterns and practices, makes a difference affiliations to create visionary and prophetic information- driven suppositions.



Fig 1: Data Mining

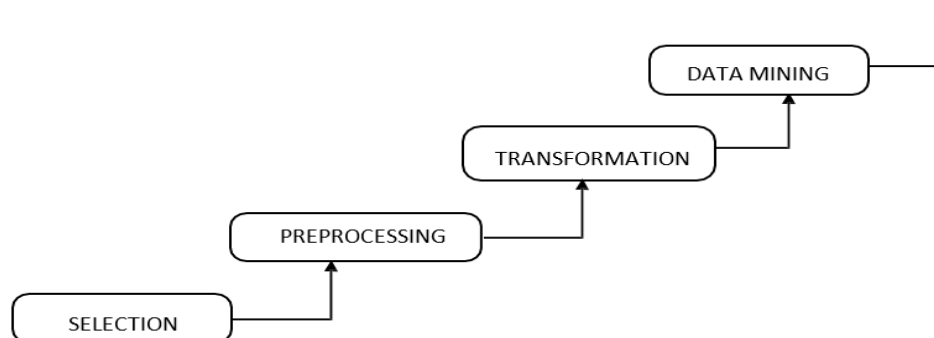


Fig 2: Knowledge Discovery Process

II. DATA MINING TECHNIQUES

Data mining techniques can yield the benefits of automation on existing software and hardware platforms to increase the value of existing information resources which is in data format, and can be implemented on a new products and systems. Various algorithms and techniques like Classification, Decision Trees, Clustering, Artificial Intelligence, Regression, and Neural Networks, Genetic Algorithm, Association Rules, Nearest Neighbour method etc., are used for knowledge discovery from data mining processes.

Classification

Classification is the most commonly applied data mining technique, which employs a set of pre-classified examples to develop a model that can classify the population of records at large. The data classification process involves learning and classification. In Learning the training data are analysed by classification algorithm.

Clustering

Clustering can be defined as identification of same classes of objects. By using clustering techniques we can further identify and segregate dense and sparse regions in object space and can find out overall distribution pattern and correlations among data attributes.

Prediction/Regression

Regression technique can be adapted for prediction purpose. Regression analysis can be used to model the relationship between one or more distinct variables such as independent variables and dependent variables. In data mining independent variables are attributes already known and dependent or response variables are what we want to predict.

Neural Network

Neural network is a set of connected input/ affair units and each connection has a weight present with it. During the literacy phase, network learns by conforming weights so as to be suitable to prognosticate the correct class markers of the input tuples. Neural networks have the capability to decide meaning from complicated, precise or squishy data and can be used to separate different patterns and descry colorful trends that are too complex to be noticed by either humans or other artificial computer ways. Neural networks are best at relating patterns or trends in data and well suited for vaticination or rainfall soothsaying needs.

III. DATA MINING TASK

The data mining tasks are of different types depending on the use of data mining result the data mining tasks are classified as follows:

Exploratory Data Analysis

It is simply exploring the data without having any clear ideas of requirements needed for. These techniques are interactive and visually effective.

Descriptive Modeling

It describe all the data, this includes models for overall probability distribution of the data, classification and partitioning of the p-dimensional space into groups and models describing the relationships between the variables which represents further model.

Predictive Modeling

This model permits the value of one variable to be prognosticated by the known values of other variables.

Discovering Patterns and Rules

It concern with pattern detection, the aim is spotting fraud detection for regions of the space defining the different types of transactions where the data points significantly different from the rest.

Retrieval by Content

It's chancing pattern analogous to the pattern of interest in the data set. This task is most generally used for textbook and image data sets.

IV. DRAWBACK OF DATA MINING

As Data Mining has advantages like briskly data analysis, Data Restoration, briskly Decision and new Invention like this Data Mining have several debit over this are as follows:

Complexity

Data mining can be complex and require technical skills and software tools. Smaller companies may find this a barrier to entry.

Scalability

Data mining can be difficult to scale, as larger datasets needs more powerful hardware infrastructure.

Data Quality

Data mining depends on the assumption that data is reliable. Inaccurate or biased data can skew outcomes and wrong decision-making.

Privacy Concern

Data mining can override data safety and can raise privacy concerns, as companies may sell customer data to other third-party businesses.

Cost

Data mining can be costly, with costs for data managing tools, data mining tools, data acquisition, and IT infrastructure.

V.FUTURE SCOPE OF DATA MINING

Data mining have various features with respect to the present time. Some of future scope of data mining are as follows:

Customer Relationship Management (CRM)

Enhancing customer interactions and loyalty through personalized marketing strategies, customer segmentation, and helps to improve customer service strategies.

Fraud Detection and Security

Identifying distinct and unusual patterns and anomalies that could help to detect fraud, also helps significantly improving security measures in financial transactions and information systems with the help of data mining.

Market and Sales Analysis

Data mining algorithms helps analysing market trends and customer purchasing patterns to optimize product placement, inventory management, and sales strategies.

Healthcare and Medical Analysis

Improving patient healthcare and outcomes by analysing medical records for trends, effective treatments, and early detection of diseases by data collection methodologies.

Manufacturing and Production

Optimizing production processes, quality control, and supply chain management through predictive maintenance forecasting and product demand with the help of predictive analysis.

E-commerce and Web Optimization

Enhancing user experiences on websites through personalized content, product recommendations, navigation and optimized site by making web user-friendly.

Social Media and Sentiment Analysis

Collecting public opinion, brand perception, and customer satisfaction by analyzing sentimental behavior expressed on social media platforms such as Facebook, twitter and Instagram.

Financial Analysis

Assessing financial risk, predicting stock market trends and market crisis, and optimizing investment portfolios by analysing financial data and market conditions as per requirement.

Research and Academic Applications

Advancing knowledge in various academic fields by uncovering new insights from data, supporting hypothesis testing, and facilitating data-driven research methodologies.

VI.CONCLUSION

The report shows the result that data mining is a growing technology nowadays, it has been growing fluently, which leads to high amount of data generation. This has many benefits like this is very time-saving methods also easy to use and easy to learn to anyone. Data mining techniques and algorithms such as classification, clustering etc., helps in identifying the patterns and to predict the trends in any domain. It has many disadvantages such as data mining never be easily available in cheap cost also it may lead to data safety risk. Data mining have drawbacks like complexity, scalability and privacy concerns. Whereas, data mining task includes descriptive modelling and predictive modelling.

Acknowledgements

We would like to thank our guide Prof. Bhagyashree Kumbhare who provided expertise and insight that greatly helped in our research and for providing this platform.

References

- [1]. https://www.researchgate.net/publication/49616224_Data_mining_techniques_and_applications
- [2]. <https://www.irjet.net/archives/V4/i11/IRJET-V4I11345.pdf>
- [3]. https://www.vssut.ac.in/lecture_notes/lecture1422914558.pdf
- [4]. https://www.academia.edu/22412092/DATA_MINING_CONCEPTS_AND_TECHNIQUES_3RD_EDIT
- [5]. <https://datamineaz.org/textbooks/hanDataMiningConceptual.pdf>