Journal of the Maharaja Sayajirao University of Baroda
ISSN :0025-0422

DATA MINING TECHNIQUES AND APPLICATION

Mrs. S. Sharmila, Assistant professor, PG Department of Computer Science, Nallamuthu Gounder Mahalingam college, Pollachi.
B. Vignesh M.Sc. Computer Science, Nallamuthu Gounder Mahalingam college, Pollachi.
P. Archana M.Sc. Computer Science, Nallamuthu Gounder Mahalingam college, Pollachi.

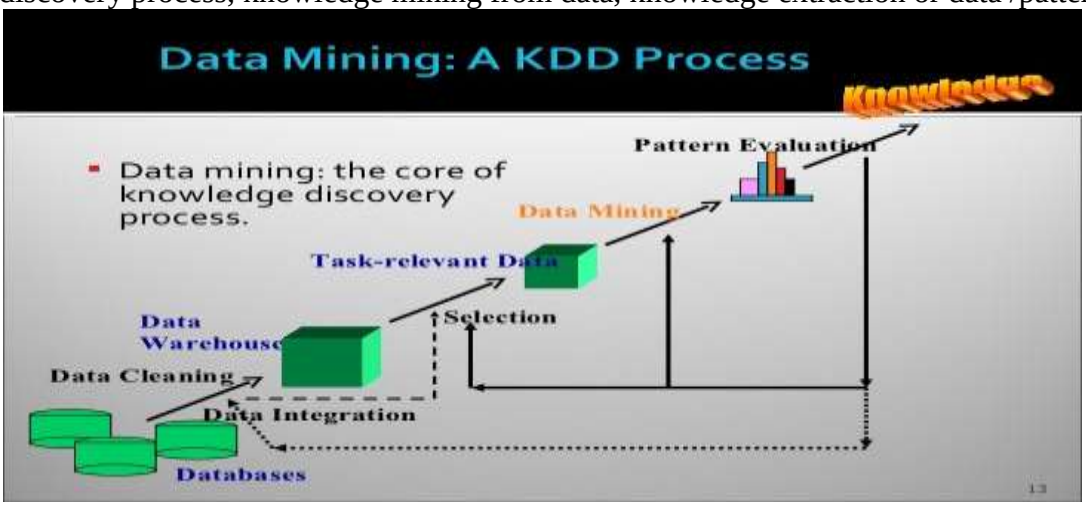
ABSTRACT

Data mining is pattern a process which finds useful from large amount of data. The paper discusses few of the data mining techniques and algorithm to improve their businesses and make excellent output.

KEYWORDS: Data mining Techniques, Data mining algorithm, Data mining application,

OVERVIEW OF DATA MINING

The progress of Information Technology has produced large amounts of records and immense data in multiple areas. The examine in databases and information technology has given growth to a tactic to store and deploy this valuable data for supplementary decision making. Data mining is a process of abstraction of useful data and patterns from huge information. It is also termed as knowledge discovery process, knowledge mining from data, knowledge extraction or data /pattern analysis.



Data mining is a reasonable process that is used to exploration through huge quantity of data in demand to find suitable facts. The goal line of this technique is to discovery patterns that were formerly strange. Once these forms are created they can supplementary be used to make convinced choices for growth of their industries.

Three phases involved are

- Exploration
- Pattern identification
- Deployment

Exploration: In the first step of data examination. Data is prepared and converted into another procedure, and imperative variable quantity and then nature of data based on the problem are determined.

Pattern Identification: Once data is discovered, advanced and well-defined for the exact variables the second step is to form pattern documentation. Recognize and choose the shapes which make the superlative calculation.

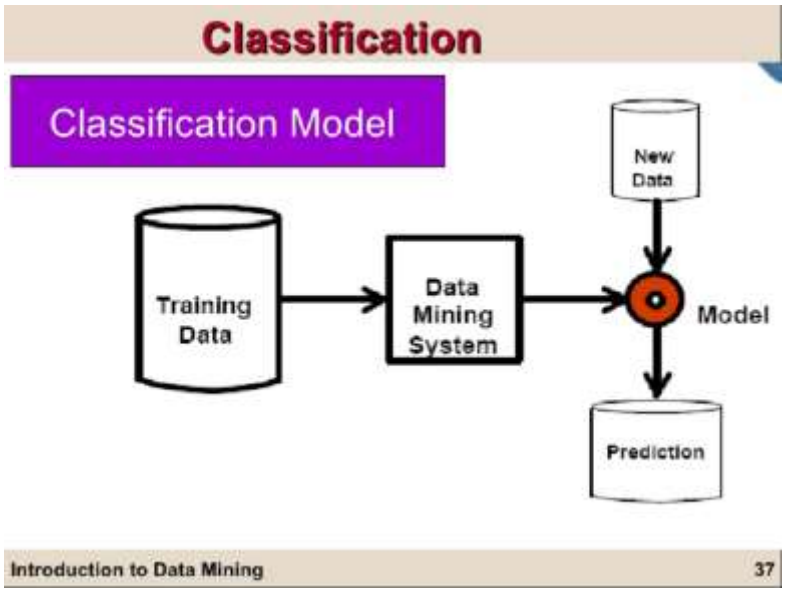
Deployment: Forms are arranged for anticipated result.

DATA MINING ALGORITHMS AND TECHNIQUES

Countless processes and procedures like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are used for data detection from files.

Classification

Classification is the utmost frequently practical data mining technique, which services a set of pre-classified examples to progress a classical that can organize the population of chronicles at huge. Scam recognition and acknowledgement danger requests are mainly well well-matched to this category of examination. This method commonly services conclusion diagram otherwise neural network-based classification algorithms. The facts classification process includes learning and classification. In learning the working out data are examined by cataloging algorithm. In classification examination of data are used to estimate the precision of the classification rules. If accurateness is suitable the guidelines can be useful to the new information tuples. For a scam detection application, this would include broad archives of mutually duplicitous and lawful actions single-minded on a record-by-record basis. The classifier-training algorithm uses these pre-classified examples to determine the set of limitations required for proper judgement. The algorithm codes these restrictions into a classical called a classifier.



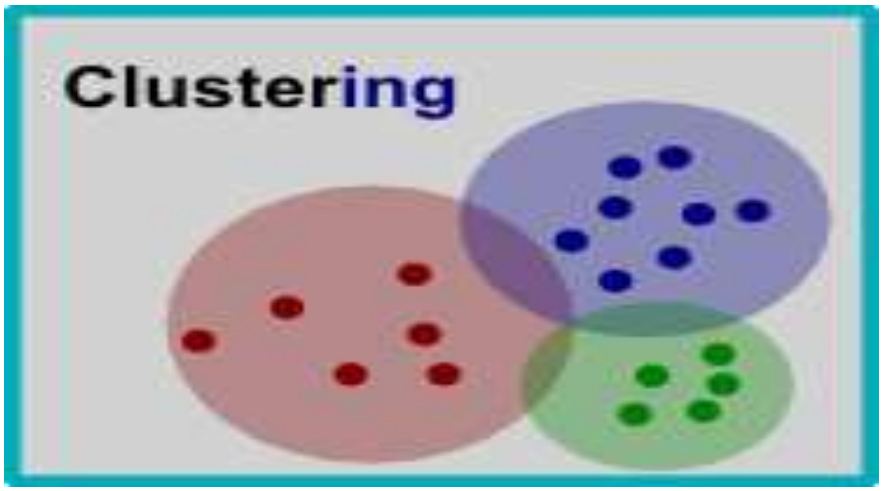
Types of classification models:

Classification by decision tree induction

- Bayesian Classification
- Neural Networks
- Support Vector Machines (SVM)
- Classification Based on Association

Clustering

Clustering can be said as credentials of similar modules of items. By using clustering techniques we can added recognize compressed and thin regions in object space and can learn overall delivery pattern and relationships among facts features. Classification approach can also be used for active means of unique groups or classes of article but it becomes expensive so clustering can be secondhand as preprocessing tactic for characteristic subgroup gathering and classification. For example, to form group of customers based on purchasing patterns, to groups genetic factor with parallel functionality.

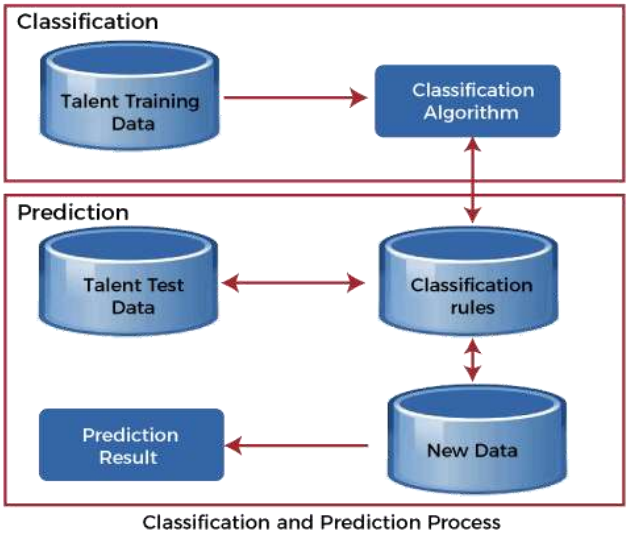


Types of clustering methods
Partitioning Methods

- Hierarchical Agglomerative (divisive) methods
- Density based methods
- Grid-based methods
- Model-based method

Predication

Regression performance can be modified for predication. Regression examination can be used to classical the connection among one or more self-determining variables and dependent variables. In data mining self-governing variable quantity are attributes previously known and reply variables are what we want to forecast. Tactlessly, several real-world problems are not simply prediction. For occurrence, sales volumes, stock prices, and product failure rates are all very difficult to expect because they may hang on on intricate relations of multiple predictor variables. Therefore, more complex techniques (e.g., logistic regression, decision trees, or neural nets) may be necessary to estimate future values. The same model types can often be used for both regression and classification. For sample, the CART (Classification and Regression Trees) decision tree algorithm can be used to build both classification trees (to classify categorical response variables) and regression trees (to forecast continuous response variables). Neural networks too can create both classification and regression models.



Classification and Prediction Process

Types of regression methods
Linear Regression

- Multivariate Linear Regression
- Nonlinear Regression
- Multivariate Nonlinear Regression

Association rule

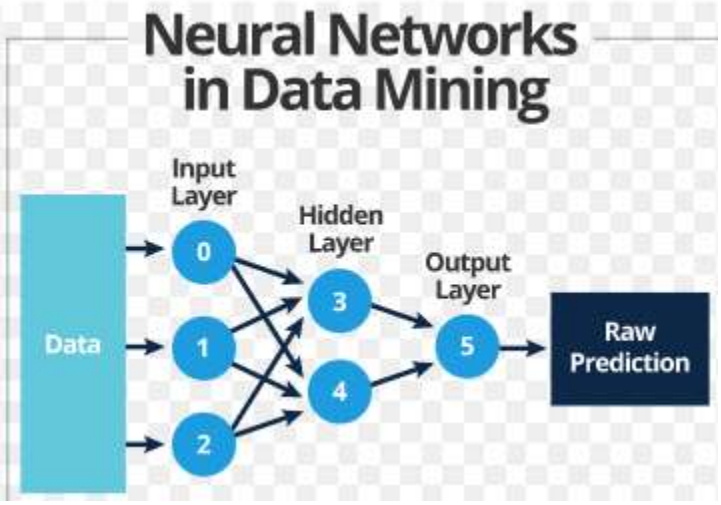
Association and correlation is regularly towards find recurrent item set discoveries among huge information sets. This type of outcome helps businesses to make confident decisions, such as directory design, cross marketing and customer shopping behavior examination. Association Rule algorithms need to be able to generate rules with self-confidence values less than one. However the number of possible Association Rules for a given dataset is generally very large and a high quantity of the rules are usually of little value.

Types of association rule
Multilevel association rule

- Multidimensional association rule
- Quantitative association rule

Neural networks

Neural network is a set of connected input/output units and each construction has a weight present with it. During the learning phase, network learns by adjusting weights so as to be able to predict the correct class labels of the input tuples. Neural networks have the extraordinary capability to originate denotation from complex or inaccurate information and can be used to abstract patterns and perceive movements that are too multipart to be perceived by either humans or other computer techniques. These are suited for persistent valued inputs and outputs. For example handwritten character shakeup, for training a computer to articulate English text and many real world business problems and have already been successfully practical in many industries. Neural networks are best at recognizing shapes or leanings in information and well well-matched for calculation or prediction needs.



Data Mining Application

Data mining is a moderately new skill that has not fully ripened. Despite this, there are a number of industries that are previously using it on a steady basis. Some of these establishments include trade goods, hospitals, banks, and insurance companies. Many of these establishments are joining data mining with such things as figures, pattern recognition, and other significant tools. Data mining can be used to find decorations and associates that would otherwise be tough to find. This technology is popular with many businesses because it permits them to learn more about their clients and make clever marketing decisions. Here is overview of business difficulties and resolutions found using data mining technology.

Healthcare

Healthcare is one of the ever-growing segments. There are permanently people out there with sicknesses who are looking for well-being and that's why data is always graceful at the health centers. With the appropriate analyzing of information, we can find out which is the greatest course of treatment for a patient. A lot of variances happen during treatment, to reduce the errors data mining is used as a tool. It also helps in avoiding mistaken medicine to a patient or to evade contact with a harmful drug.

By using analysis, we can find out which physician is best suitable for a patient looking into their preceding records where they have solved a remedial case. We can use data mining in the healthcare industry to find out the acquiring behaviors of patients. It was not like that people didn't work earlier, data mining has just made it easy and has stretched the limitations.

• Customer Relationship Management

Every company wants to lift themselves economically and decently and that's why we have jobs like association managers. Client relationship administration and buyer satisfaction are very important for a firm. When we don't know about the client or their favorites, we waste time frustrating to encourage them with the incorrect picks. That is where data mining chips in, it gives us the essential details about the behaviors of a customer in any wanted field. With the help of this information, one can focus on the required topics and help the client in a better method. Data mining finds out what is best suitable for a client and helps in building better client relations.

• Manufacturing Engineering

Manufacturing engineering instructs all the fields of engineering in it. It is basically the alteration of raw materials into a equipped commodity in the most economical way. The quality of the product is also kept in mind. Data mining delivers information about customer satisfaction concerning any particular product. One can analyze the client's necessities and fulfillment and can find a way to progress their services. Data mining stretches us the statement of the dependability of any exact product.

Analyses and inspections help in finding main worries and supplies of regulars and according to that condition requirement one can judge whether his work is successful in the right way or not. Shakings are some common pressures from which a building should be innocent if it occurs. By using data mining, we can find out the particular number of disasters that happened in any exact area and the harm detected. Raw materials are looked systematically to distribute best permanency and endurance to a building. The greatest course of action is decided and duplication of faults in business is concentrated.

CONCLUSION

Data mining has rank concerning discovery the patterns, forecasting, discovery of knowledge etc., in different business fields. Data mining methods and procedures such as classification, clustering etc., helps in finding the decorations to decide upon the upcoming trends in businesses to raise. Data mining has extensive application field almost in every industry where the data is produced that's why data mining is measured one of the most important boundaries in database and information systems and one of the most talented inter corrective growths in Information Technology.

REFERENCES

[1]. Venkatadri, M., and Lokanatha C. Reddy. "A review on data mining from past to the future." International Journal of Computer Applications 15.7 (2011): 19-22.
[2]. Tomar, Divya, and Sonali Agarwal. "A survey on Data Mining approaches for Healthcare." International Journal of Bio-Science and Bio-Technology 5.5 (2013): 241-266.
[3]. Saurkar, Anand V., et al. "A Review Paper on Various Data Mining Techniques." International Journal of Advanced Research in Computer Science and Software Engineering 4.4 (2014).
[4]. Soni, Shraddha. "A Literature Review on Data Mining and its Techniques." Indian Journal of Applied Research 5.6 (2016).
[5]. Fayyad, Usama, Gregory Piatetsky-Shapiro, and Padhraic Smyth. "The KDD process for extracting useful knowledge from volumes of data."
[6]. Gorunescu, Florin. Data Mining: Concepts, models and techniques. Vol. 12. Springer Science & Business Media, 2011.
[7]. Chen, Ming-Syan, Jiawei Han, and Philip S. Yu. "Data mining: an overview from a database perspective." IEEE Transactions on Knowledge and data Engineering 8.6 (1996): 866-883.
[8]. Parvathi, I., and Siddharth Rautaray. "Survey on data mining techniques for the diagnosis of diseases in medical domain." International Journal of Computer Science and Information Technologies 5.1 (2014): 838-846.
[9]. Romero, Cristobal, and Sebastian Ventura. "Educational data mining: A survey from 1995 to 2005." Expert systems with applications 33.1 (2007): 135-146.
[10]. Ngai, Eric WT, Li Xiu, and Dorothy CK Chau. "Application of data mining techniques in customer relationship management: A literature review and classification." Expert systems with applications 36.2 (2009): 2592-2602.
[11]. Raeder, Troy, and Nitesh V. Chawla. "Market basket analysis with networks." Social network analysis and mining 1.2 (2011): 97-113.