

A Review Paper on Current Data Mining Applications and Techniques

M.Vignesh ^[1], Dr.V.Kathiresan ^[2]

Department of Computer Application (PG)
Dr. SNS Rajalakshmi College of Arts and Science, Coimbatore
Tamil Nadu -India

ABSTRACT

Data mining is that the technique of filtering relevant knowledge in line with one's business interests from the huge assortment of knowledge mistreatment totally different techniques. This paper conducts a proper review of the idea of data-mining, the standard Implementation process involve in data-mining, its applications in day to day field, techniques.

Keywords: KDD, Data Mining Techniques, data mining applications

I. INTRODUCTION

Data mining is that the method of sorting through massive knowledge sets to spot patterns and establish relationships to resolve issues through knowledge analysis [7].

Technically processing is that the method of extracting specific information from information and presenting relevant and usable information which can be accustomed solve problems. There are a unit completely

different forms of services within the method like text mining, web mining, audio and video mining, pictorial data processing and social network data processing.

II. DATA MINING AS A CORE PROCESS IN KDD

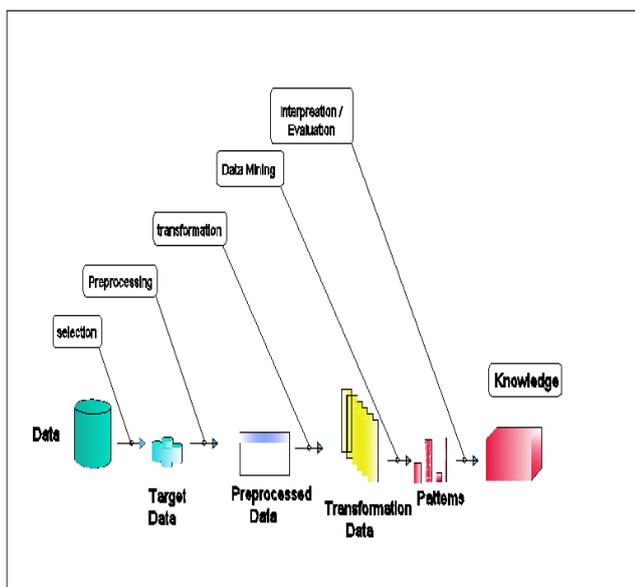


Fig.1 The list of steps involved in the knowledge discovery process

A. DATA CLEANING

The noise and inconsistent knowledge is removed.

B. DATA INTEGRATION

Multiple knowledge sources are combined.

C. DATA SELECTION

Knowledge relevant to the analysis task are retrieved from the info.

D. DATA TRANSFORMATION

Knowledge consolidated into forms applicable for mining by playing outline or aggregation operations.

E. DATA MINING

Intelligent strategies are applied so as to extract knowledge patterns.

F. PATTERN EVALUATION

Knowledge patterns are evaluated.

G. KNOWLEDGE PRESENTATION

Information is portrayed.

III. DATA MINING IMPLEMENTATION PROCESS

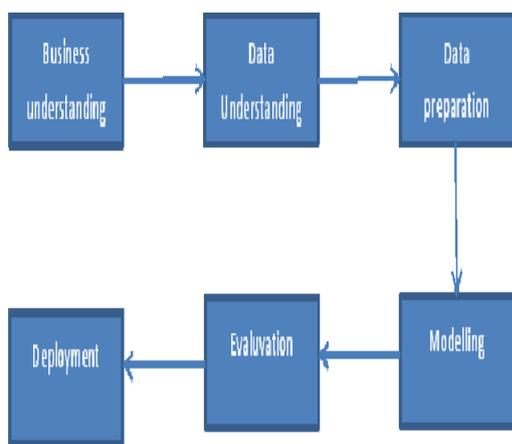


Fig.2 Data mining implementation process

A. BUSINESS UNDERSTANDING

First, it's needed to know business objectives clearly and determine what square measure the business's desires.

Next, we've got to assess the present scenario by finding the resources, assumptions, constraints and alternative necessary factors that ought to be thought of.

Then, from the business objectives and current things, we'd like to form data processing goals to attain the business objectives among the present scenario.

Finally, an honest data processing set up should be established to attain each business and data processing goals. The set up ought to be as elaborated as potential.

B. DATA UNDERSTANDING

First, the information understanding section starts with initial data assortment, that we have a tendency to collect from offered knowledge sources, to assist U.S.A. get acquainted with the info. Some necessary activities should be performed together with knowledge load and knowledge integration so as to create the info assortment with success.

Next, the "gross" or "surface" properties of no heritable knowledge have to be compelled to be examined fastidiously and reported .

Then, the info has to be explored by attempt the info mining queries, which may be addressed exploitation querying, reporting, and visualisation.

Finally, the information quality should be examined by respondent some necessary queries like "Is the no heritable

data complete?", "Is there any missing values within the no heritable data?"

C. DATA PREPARATION

The data preparation generally consumes regarding ninetieth of the time of the project. The end result of the information preparation part is that the final data set. Once obtainable information sources are known, they have to be designated, cleaned, made and typed into the specified form. The info exploration task at a larger depth is also carried throughout this part to note the patterns supported business understanding.

D. MODELING

First, modelling techniques ought to be chosen to be used for the ready dataset.

Next, the check situation should be generated to validate the standard and validity of the model.

Then, one or additional models square measure created by running the modelling tool on the ready dataset.

Finally, models ought to be assessed fastidiously involving stakeholders to form positive that created models square measure met business initiatives.

E. EVALUATION

In the analysis section, the model results should be evaluated within the context of business objectives within the initial section. During this section, new business needs is also raised thanks to the new patterns that are discovered within the model results or from different factors. Gaining business understanding is associate degree reiterative method in data processing. The go or no-go call should be created during this step to makeover to the readying section.

F. DEPLOYMENT

The data, that we have a tendency to gain through data process, must be conferred in such the way that stakeholders will use it after they need it. supported the business needs, the preparation part may well be as easy as making a report or as complicated as a repeatable data process across the organization. within the preparation part, the plans for preparation, maintenance, and observance need to be created for implementation and conjointly future supports.

IV. CURRENT DATA MINING APPLICATIONS

A. BUSINESS SECTOR

In business world data processing is essentially used for analyzing performance, gain index, and client feedback analysis, and analysis of the stock values of existing organizations and their market trends to help in future business selections and Classification and agglomeration of consumers for targeted promoting and Detection of cash lamination and different money crimes.

B. BANKING & FINANCE

Assessment of individual banking records to get completely different selling ways for a target client section, authorization, stock foretelling, checking completely different forms of fraud and concealment.

C. BIO-INFORMATICS

Data Mining approaches appear ideally fitted to Bioinformatics, since it's data-rich. Applications of knowledge mining to bioinformatics embody factor finding, macromolecule operate reasoning, un wellness diagnosing, wellness prognosis, un wellness treatment improvement, macromolecule and factor interaction network reconstruction, knowledge cleansing, and macromolecule sub-cellular location .

D. FUTURE HEALTHCARE

Data mining holds nice potential to boost health systems. It uses information and analytics to spot best practices that improve care and scale back prices. Researchers use data processing approaches like multi-dimensional databases, machine learning, soft computing, information mental image and statistics. Mining will be wont to predict the degree of patients in each class. Processes are developed that certify that the patients receive acceptable care at the proper place and at the proper time. data processing may also facilitate health care insurers to discover fraud and abuse.

E. EDUCATION

There is a brand new rising field, referred to as instructional data processing, considerations with developing strategies that discover data from knowledge originating from instructional Environments. The goals of EDM are known as predicting students' future learning behavior, finding out the results of instructional support, and advancing knowledge base regarding learning. data processing will be employed by an establishment to require correct selections and additionally to predict the results of the scholar. With the results the establishment will specialize in what to show

and the way to show. Learning pattern of the scholars will be captured and wont to develop techniques to show them.

F. MANUFACTURING ENGINEERING

Knowledge is that the best quality a producing enterprise would possess. Data mining tools may be terribly helpful to find patterns in advanced producing process. Data processing may be employed in system-level planning to extract the relationships between product design, product portfolio, and client desires information. It may also be accustomed predict the merchandise development span time, cost, and dependencies among alternative tasks.

G. CRM

Customer Relationship Management is all concerning deed and retentive customers, conjointly rising clients' loyalty and implementing customer targeted ways. To keep up a correct relationship with a client a business have to be compelled to collect information and analyze the knowledge. This is often wherever data processing plays its half. With data processing technologies the collected information will be used for analysis. Rather than rambling wherever to focus to retain client, the seekers for the answer get filtered results.

H. CLIMATOLOGY

Assessment of atmospheric condition over a amount of your time thus on predict future meteoric patterns for decisive natural calamities like cyclone and additionally forecasting.

I. CRIMINAL INVESTIGATION

Criminology may be a method that aims to spot crime characteristics. Really crime analysis includes exploring and detective work crimes and their relationships with criminals. The high volume of crime information sets And conjointly the quality of relationships between these types of data have created sociology an acceptable field for applying data processing techniques. Text primarily based crime reports will be reborn into data processing files. These data will be accustomed perform crime matching method.

J. E-COMMERCE

Data mining techniques square measure utilized in ecommerce to research client search patterns to push up sale and cross sale.

K. CORPORATE SURVEILLANCE

Corporate police work is that the observation of someone or group's behavior by a company. the info collected is most frequently used for selling functions or sold to

different firms, however is additionally often shared with government agencies. It will be utilized by the business to tailor their product fascinating by their customers. The info will be used for marketing functions, like the targeted advertisements on Google and Yahoo, wherever ads area unit targeted to the user of the computer program by analyzing their search history and emails.

V. DATA MINING TECHNIQUES

A. TRACKING PATTERNS

One of the foremost basic techniques in data processing is learning to acknowledge patterns in your knowledge sets. This is often typically a recognition of some aberration in your knowledge happening at regular intervals, or associate ebb and flow of an exact variable over time. for instance, you would possibly see that your sales of an exact product appear to spike simply before the vacations, or notice that hotter weather drives additional folks to your web site.

B. CLASSIFICATION

Classification could be a additional advanced data processing technique that forces you to gather varied attributes along into discernible classes, that you'll then use to draw additional conclusions, or serve some perform. for instance, if you're evaluating information on individual customers' monetary backgrounds and buy histories, you would possibly be able to classify them as "low," "medium," or "high" credit risks. you'll then use these classifications to be told even additional concerning those customers.

C. ASSOCIATION

Association is expounded to following patterns, however is a lot of specific to dependently coupled variables. during this case, you'll explore for specific events or attributes that area unit extremely related with another event or attribute; for instance, you would possibly notice that once your customers get a particular item, they additionally typically get a second, connected item. this is often typically what's accustomed populate "people additionally bought" sections of on-line stores.

D. OUTLIER DETECTION

In several cases, merely recognizing the overarching pattern can't offer you a transparent understanding of your knowledge set. you furthermore might have to be compelled to be ready to establish anomalies, or outliers in your knowledge. for instance, if your purchasers square measure virtually solely male, however throughout one strange week in Gregorian calendar month, there's an enormous spike in feminine purchasers, you'll need to research the spike and

see what drove it, therefore you'll either replicate it or higher perceive your audience within the method.

E. CLUSTERING

Clustering is incredibly kind of like classification, however involves grouping chunks of information along supported their similarities. for instance, you may value more highly to cluster totally different demographics of your audience into different packets supported what quantity income they need, or however typically they have a tendency to buy at your store.

F. REGRESSION

Regression, used primarily as a variety of designing and modelling, is employed to spot the probability of a particular variable, given the presence of different variables. for instance, you may use it to project a particular value, supported different factors like accessibility, client demand, and competition. a lot of specifically, regression's main focus is to assist you uncover the precise relationship between 2 (or more) variables in a very given information set.

G. PREDICTION

Prediction is one in every of the foremost valuable data processing techniques, since it's accustomed project the categories of knowledge you'll see within the future. In several cases, simply recognizing and understanding historical trends is enough to chart a somewhat correct prediction of what is going to happen within the future.

VI. CONCLUSION

In this paper we have mentioned the idea of data Mining Implementation process ,data mining applications. we tend to have additionally lined the data mining techniques. data processing brings heaps of advantages to businesses, society, governments in addition as individual. data processing involves the utilization of knowledge analysis tools to find antecedently unknown, valid patterns and relationships in massive knowledge sets. unremarkably use data processing to scale back prices, enhance analysis, and increase sales. data mining applications are initially were used as a way to sight fraud and waste. triple-crown data processing still needs experienced technical and analytical specialists United Nations agency will structure the analysis and interpret the output.

REFERENCES

- [1] Venkatadri.M ,Dr. Lokanatha C. Reddy "A Review on Data mining from Past to the Future" in International Journal of Computer Applications (0975 – 8887) Volume 15– No.7, February 2011

- [2] Tipawan Silwattananusarn and Assoc.Prof. Dr. KulthidaTuamsuk “Data Mining and Its Applications for Knowledge Management : A Literature Review from 2007 to 2012 ” in International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.5, September 2012
- [3] Samiddha Mukherjee, Ravi Shaw, Nilanjan Haldar, Satyasan Changdar “A Survey of Data Mining Applications and Techniques” inSamiddha Mukherjee et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (5) , 2015, 4663-4666
- [4] Mansi Gera, Shivani Goel“Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity” in International Journal of Computer Applications (0975 – 8887) Volume 113 – No. 18, March 2015
- [5] Vikas Gupta, Prof. Devanand “A survey on Data Mining: Tools, Techniques, Applications, Trends and Issues” inInternational Journal of Scientific & Engineering Research Volume 4, Issue3, March-2013 ISSN 2229-5518
- [6] Bhavani Thuraisingham, Latifur Khan, Mohammad M. Masud, Kevin W. Hamlen “Data Mining for Security Applications” in International Conference on Embedded and Ubiquitous Computing
- [7] Shakir Khan, Dr. Arun Sharma, Abu Sarwar Zamani, Ali Akhtar “Data Mining For Security Purpose & Its Solitude Suggestions” in INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH VOLUME 1, ISSUE 7, AUGUST 2012 ISSN 2277-8616
- [8] Dileep Kumar Singh, Vishnu Swaroop “Data Security and Privacy in Data Mining: Research Issues & Preparation” in International Journal of Computer Trends and Technology- volume4Issue2- 2013
- [9] Kenji YAMANISHI, Jun-ichi TAKEUCHI and Yuko MARUYAMA “Data Mining for Security” in Journal of Advanced Technology, Vol. 2, No. 1
- [10] Thitimanan Damrongsakmethee and Victor-Emil Neagoe “Data Mining and Machine Learning for Financial Analysis” inIndian Journal of Science and Technology, Vol 10(39), DOI: 10.17485/ijst/2017/v10i39/119861, October 2017