Exploring Knowledge Discovery from Database using Data Mining

S. Ummul Hyrul Fathima
ME(CSE), Assistant Professor, Department of Computer Science, Thassim Beevi Abdul Kader College for Women, Kilakarai, Ramanathapuran District, India

Abstract— The need for Knowledge Discovery and Data Mining are growing exponentially in recent years as data accumulated every day, through Internet & Smart devices currently, is more than the data accumulated for the past 2000 years. Knowledge Discovery and Data Mining methods are highly required for knowledge acquisition, machine learning, database statistics, data visualization, and high-performance computing. In the field of Artificial Intelligence AI, KDD plays a vital role. Knowledge Discovery and Data Mining have been very useful in the areas of education, industry, ecommerce, digital marketing, government, and so on.
Relationship between the Knowledge Discovery and Data Mining will be discussed with greater insights in the later sections, which forms the basic understanding. In the due course, information about the exploration of knowledge in the form of data from databases through data mining techniques has been explained. Also, provide details about the real-world application of Knowledge Discovery, and directions of the future research are outlined.

Keywords—Database, Knowledge Discovery, Data Mining

I. INTRODUCTION

Knowledge Discovery in Database generally referred to as KDD, which can be defined as a way of discovering, transforming, and distilling of data together with patterns from a raw database, that will be utilized in different domains or applications. KDD is a lengthy, and sophisticated process that involves across many steps as well as iterations.

The database is an integrated collection of data maintained in one or more files organized to retrieved related information. Database to be used in Knowledge Discovery will be of type relational model. In a relational database, data are organized into tables with rows, and columns comprising records which continue to expand.

The system needed to maintain the database will be commonly referred to as RDBMS. It is a collection of procedures for retrieving, storing, manipulating, and even in few cases deleting data within a set of database tables. Knowledge Discovery and Data Mining in data processing is a methodology which will be achieved through programming, and analytical approaches to model data from a database to extract useful, and applicable knowledge. The backbone of KDD is data mining and is critical to the entire method. KDD utilizes several algorithms which are self-learning in nature to minimize useful patterns from the processed data. This process is constant feedback one with the closed-loop where a huge number of iterations will occur between the various steps as per the demand of the pattern interpretations, and algorithms.

II. STEPS INVOLVED IN KDD

Application Understanding, and Goal-Setting

Application Understanding is the first step in the process which requires awareness, and understanding of the domain, KDD will be applied. Goal-Setting determines, how the transformed data, and the patterns achieved through data mining, will be used to extract knowledge. Goal-Setting is extremely important, if set wrong, can lead to irrelevant interpretations, and negative impacts on the end-user.

Data Selection, and Integration

After the first phase mainly Goal-Setting, the data collected needs to be selected, then segregated into meaningful sets based on quality, accessibility importance, and availability. These parameters are mission-critical in data mining because they are the base of it and will affect what kind of data models are developed.

Data Preprocessing, and Cleaning

Data Preprocessing and Cleaning involves searching for missing data as well as removing noisy, deprecated, redundant, and low-grade data from the data set for improving the reliability of the data, and its effectiveness. Relevant algorithms are implemented for tracing and eliminating deplorable

data based on attributes with respect to the application.

Data Transformation

Data Transformation prepares the data that will be fed to the data mining algorithms. The data, which need to be in consolidated, and aggregate forms, will be transformed through consolidation on the basis of functions, attributes, features, etc.

Data Mining

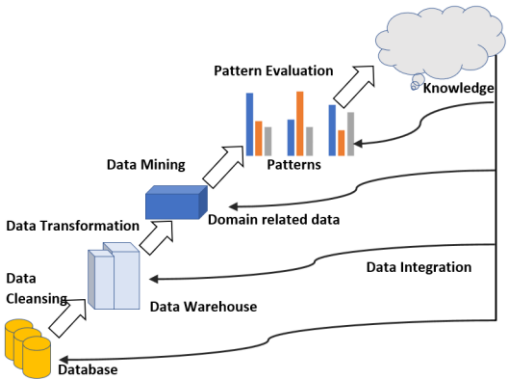
Data Mining is the backbone process of the whole KDD process. Algorithms are used to retrieve meaningful patterns from the transformed data by the process Data Mining, which helps in the prediction of models. It is an analytical tool that helps in realizing together with analyzing trends from a data set using methodologies such as artificial intelligence, advanced numerical, and statistical methods, and specialized algorithms.

Pattern Evaluation/Interpretation

Post the trend, and patterns have been obtained from numerous data mining methods, and iterations, they need to be coined in precise forms of graphical representation such as bar graphs, pie charts, histograms, etc, for studying the impact of data collected and transformed during earlier steps. Pattern Evaluation/Interpretation also helps in finding the effectiveness of the data models in accordance with the domain.

Knowledge Discovery and Use

Knowledge Discovery and Use forms the final step in the KDD process and requires the 'knowledge' to be extracted from the earlier steps to be applied in the specific application or domain in a visualized format such as tables, reports, graphs, etc. Knowledge Discovery and Use drives the decision-making process for the respective application.



III. DATA MINING IN KDD PROCESS.

The past decade has seen exponential growth in generation and collection of data. Analyzing this heavily accumulated volume of data is a challenging task, because of not only for its complexity but also for its multiple numerous correlated factors.

The Data Mining process is used to extract interesting knowledge from a large amount of data stored across multiple data sources file systems, databases, data warehouses, etc. The aim of data mining is to create models for decision-making by analyzing the trend in the past and predict future behaviour. Data Mining tools can analyse the massive databases to deliver response for the queries raised by client/ server or parallel processing systems.

The data mining for the knowledge discovery process often involves the recurrent iterative application of specific data-mining methods. The overview will be addressed for the primary goals of data mining, an explication of the methods used to address goals, and a brief description of the data-mining algorithms that amalgamate these methods.

The knowledge discovery goals are described by the proposed use of the system. KDD goals can be differentiated into two types of goals viz., verification, and discovery. With the KDD goal - verification the system, which is used for KDD, is limited to verifying the user's hypothesis. With the KDD goal - discovery the system, which is used for KDD, autonomously finds new patterns.

We further categorize the discovery goal into prediction and description. Prediction is indicated as to where the KDD system discovers the patterns for predicting future comportment of some entities.

The description is nothing but where the KDD system unearths the patterns for presentation to a user in a human-understandable form We will be primarily concerned with data mining relevant to knowledge discovery.

Data mining involves fitting models to or evaluating patterns from, ascertained data. The fitted models play a vital role in inferential knowledge. The fitted model describes whether models reflect meaningful as well as useful or worthful knowledge is inclusive of the overall, interactive KDD process where

distinctive human intervention is required for decision.

The primary mathematical formalisms are used in model fitting, can be classified statistical and logical. The statistical approach focuses more on non-deterministic effects in the model, whereas a logical model will target purely deterministic. The statistical approach to data mining will be the primary focus, that tends to be widely used, and accepted basis for real-time data mining applications as the typical presence of uncertainty in data-generating processes of the real world.

Components of Data-Mining Algorithms

The next step is to start building specific algorithms for implementing the general methods we have outlined. The three primary components in any data mining algorithm will be model representation, model evaluation, and search. It is the most convenient way to express the underlying concepts of data-mining algorithms in a relatively consolidated, and compact manner.

Model representation is the methodology used to describe the discoverable patterns. If the target representation is too minimal, then no amount of training time or examples can produce a perfect model for the data. It is typically important that a data analyst should fully comprehend the representational postulates or assumptions that might be inherent in a specific method.

It is equally important that an algorithm designer clearly indicate which representational postulates or assumptions are being made by a specific algorithm. Remember that the increased representational power for models in turn increases the problem of overfitting the training data, resulting in lower prediction accuracy on unseen data.

Model evaluation criteria are quantitative statements (or fit functions) of how better a particular pattern (a model, and its parameters) satisfies the goals of the knowledge discovery process. Say, for example, predictive models are mostly judged by the empirical prediction accuracy on few test sets. Descriptive models can be examined along the dimensions of predictive accuracy, novelty, utility, and understandability of the fitted model.

The Search method consists of two components viz., parameter search and then the model search. Once the model representation (or list of representations), and the model evaluation criteria are determined, then the data mining problem has been lowered due to purely an optimization task, which is to find the parameters and models from the selected list that optimize the evaluation criteria.

In parameter search, the algorithm must look for the parameters that optimize the model evaluation criteria on the given observed data, and a fixed model representation. Model search occurs as an iteration over the parameter search method: The model representation has to be changed so that a list of models is considered, in analytic methods. It comprises of different components of data modeling, data transformation, data migration, data integration, and aggregation.

IV. DATA MINING METHODS

A. Data Cleaning, and Preparation

Data cleaning and preparation forms the basis at the same time most important part of data mining. Raw data needs to be cleansed and formatted to use Data Mining means extracting knowledge in terms of knowledge from the large amount of data stored in the database.

The knowledge or information extracted can be used in analyzing by real time applications. The extracted knowledge plays a vital role in the latest research in Statistics, Machine Learning, and Artificial Intelligence. For performing research process, data mining includes various methodologies such as fundamental issues, classification, and clustering, trend, and deviation analysis, dependency modelling, integrated discovery systems, next-generation database system (like in-memory databases, column-based databases), and application case studies.

B. Tracking Patterns

Tracking patterns is one of the most basic but precious data mining methods. Together with identifying patterns within data sets, tracking patterns also monitor modification in trends over a period of time, allowing to make intelligent decisions such as regular decreases or increases in traffic during the day/ week/ festival time say Diwali/ Christmas eve or when certain products tend to sell more often, such as chat items during IPL in Swiggy, and to promote accordingly.

C. Classification

Data analysis through classification is utilized to look out for actual, and vital information. Classification is

acknowledged to be a more complex method across other data mining techniques. Information is classified into different classes, for instance, credit customers can be segregated according to risk categories such as low, medium, high, or even critical.

D. Association

Association learning is used to discover the things that tend to occur together in pairs or larger groups. The association goes beyond simple correlation, it extends beyond pairs to account for larger groupings of items.

For example, in Flipkart site, evaluating the occurrence of users buying coolers with fashion t-shirts in case of men or even to the case whenever a lady buys fashion t-shirts, she also adds matching fancy earrings, neckwear, hand bands, coolers.

E. Outlier detection.

In most cases, recognizing the pattern will not provide a clear picture of the data set. In such cases, we need to be able to identify outliers or anomalies in the data. For example, if western fashion brand site, during festivals the users are almost exclusively male, but suddenly there's a spike in female purchasers, we need to investigate the change in trend to discover the driving factor so that it can be replicated or target the female audience in next campaign with the factor.

F. Cluster

A Cluster can be defined as the collection of similar data objects. The objects that are similar to one another within the specified cluster, but they are different, or they are rather dissimilar or they are completely unrelated to the objects in other clusters. Clustering analysis is the method of identifying clusters, and groups from the data in a way that the level of association between two objects is maximum if they belong to a cluster or same group, and minimal if otherwise.

The outcome of clustering analysis can be used in the creation of customer profiling.

G. Regression

Regression is considered as the most straightforward yet simple version of predictive power. Regression analysis is generally performed when wanted to predict the value of a given feature in a continuous manner based on the values of other attributes in the data, assuming that the dependency lies in the either linear or nonlinear model.

Regression methods are highly used in data science, and logistic regression terms tend to pop up almost in every aspect of the field.

H. Prediction

Prediction is a data mining method, in turn, is a mixture of other data mining methods like classification, clusters, sequential patterns, etc. The prediction method analyzes events that occurred in the past for predicting the events in the future. The prediction method has lot of applications but is commonly used in sales for predicting profit/ loss.

I. Sequential Patterns

Sequential patterns are the data mining method that sought to identify and retrieve similar patterns, regular events or trends from data over a course of business period. In sales, with the past or historical transaction data, businesses can discover the set of items that customers usually buy together at specific occasions in a year. Using these patterns, businesses can recommend customers to buy it with better deals or conduct campaigns based on their purchasing trend in the past.

J. Decision Tree

The Decision Tree method refines the Classification method in a structured way to find solutions. A decision tree is a hierarchical tree like structure that enables users to understand effectively and is really simple as well as fast. In the Decision Tree, non-leaf node denotes a test on an attribute, and the branch indicates the outcome of a respective test, while the leaf node indicates the class label.

In the decision tree, attribute values of a tuple are tested right from the root all the way to the leaf node. Decision trees are quite popular as prior domain knowledge is not required, and these can represent multidimensional data. Classification rules can be derived easily from the decision trees.
The application of decision trees is manufacturing, production, medicine, astronomy, etc.

***K.** Statistical Approaches*

Earlier, statisticians used to develop methods that are used for evaluating hypotheses as well as determining whether differences can be attained to random chance. But those Statistical theories support prediction methods, and data models.

Generally classical statistical models, mainly linear models, assume data is clean as well as smaller in size, and they tend to break down when facing massive data sets. To avoid this, Bayesian inference is the most used statistical approach for knowledge discovery in databases. There are three Bayesian methods available for data mining namely Naive Boyes classifier, Autoclass, and Bayesian networks

L. Data Visualization

Data visualization is useful for data mining as the output is presented nicely. Analysts can attain a better understanding of the data through using visual tools because they can focus their attention on the patterns that are the outcome of other methods. With the variations in dimensions, colour, and depth, it is possible to discover new associations, and also helps in improving the differentiation between them.

Data visualization is a very handy method for the discovery of relationships, patterns, missing, and exceptional values. However, Data visualization's greatest limitation is that visualization must collapse many different dimensions into a two- or three-dimensional screen. Moreover, tools developed for data visualization usually need considerable training, and will not be suitable for people who have difficulty with spatial analysis or who are colour blind.

M. Neural networks

Neural networks are mainly utilized in deep learning algorithms. Neural networks parse training data like how the human brain is interconnected by the nodes. Each node comprises bias or threshold, inputs, weights, and an output. If the threshold has been exceeded, then Neural networks activate or fire node, which passes the data to the proceeding layer of the network.

This mapping function has been learned by the Neural networks through supervised learning, adjusting on basis of the loss function, the process from gradient descent. The Neural Network model's accuracy to provide the correct answer when the cost function is reaches or closer to zero. Three pillars of the neural network are Model, Learning Algorithm (supervised or unsupervised), and applications

N. Data warehouses

A Data Warehouse is the methodology that accumulates the data from different kinds of sources within the same organization to provide meaningful business insights. The extremely large amount of data comes from multiple places such as Sales, Marketing, and Finance.

The data warehouse is developed for data analysis and supports decision-making to the management in the business organization. Warehousing is an integral aspect of data mining processes. In certain cases, analysts may start from the scratch to be precise by selecting the data they want and build a data warehouse based on their specifications.

O. Long-term Memory Processing

Long-term memory processing is designed in such a way that it helps to scale data in memory and facilitates higher weight to the input in the sequence. The long-term memory processing method avoids overfitting of memory by scaling the cell state after attaining the optimal results.

The long-term memory network (LTM) is primarily used in remembering the huge sequences in turn to prevent the suffering of the learning model from the vanishing gradient problem. The ultimate feature of long-term memory processing is that it does not flush out the past sequence, but incorporates the same along with the current inputs, then generalizes the past sequences, and resulting in a higher emphasis on the new inputs.

P. Machine learning, and artificial intelligence

Machine learning and artificial intelligence represent the highly advanced methods in the data mining process. When working with data at a large scale, highly accurate predictions can be attained methods such as deep learning which are part of the advanced forms in machine learning.

Machine learning, and artificial intelligence data mining methods are very much helpful for determining results from semi-structured as well as unstructured data. These methods are absolutely handy for processing data in AI applications such as speech recognition, computer vision, or complicated text analytics with the help of Natural Language Processing.

V. DATA MINING TOOLS

The requirement is to perform different algorithms such as clustering or classification on the data set selected, processed, and transformed for visualizing the results themselves. The framework which

provides us a phenomenon that data represent and better insights for data is referred to as a data mining tool.

Data Mining tools provide algorithms as well as procedures for the objective of extracting patterns, groupings, trends from extremely large sets of data, and translating data into refined information.

A. Orange Data Mining

Orange is a perfect combination of data mining software, and a machine learning suite, which supports the visualization, and is software based on components written in the Python programming language. The components of Orange are called widgets, which range from pre-processing, and data visualization to the value judgment of algorithms, and predictive modeling.

Widgets provide significant functionalities such as

a) Displaying data table, and providing a way to select features
b) Data reading
c) Training predictors and comparison of learning algorithms
d) Data element visualization

Orange provides a more interactive, and enjoyable atmosphere compared to other dull analytical tools, which enables the data mining experts to operate in a quite exciting way.

In Orange, the data source is formatted quickly to the desired pattern, and widgets can be transferred easily without any fuss. Orange allows the data mining experts to take smarter decisions in a short span of time by comparing and analyzing the data in a quick fashion. Python scripts can keep running in a terminal window, an integrated development environment (IDE) like PyCharm, and PythonWin, power shells like iPython.

B. SAS Data Mining

SAS refers to Statistical Analysis System, a product of the SAS Institute created dedicatedly for analytics, and data management. SAS can mine data, change it, manage information from heterogenous sources, and analyze statistics that too in a graphical UI for non-technical users.

SAS facilitates data mining experts to analyze big data and also provides accurate insight for decision-making purposes in a timely manner. The memory processing architecture of SAS is distributed, that is highly scalable, which is suitable for data mining, optimization, and text mining.

C. DataMelt Data Mining

DataMelt is a visualization, and computation environment, DataMelt provides an interactive structure for data analysis as well as visualization. DataMelt is also known as DMelt, which is primarily designed for students, engineers, and scientists.

DMelt has been developed using JAVA with support to install on any operating system which has Java Virtual Machine, JVM. DMelt consists of science, and mathematics libraries. Scientific libraries are used for drawing the 2D/3D plots while Mathematical libraries are used for algorithms, curve fitting, random number generation, etc. DMelt can be utilized for the analysis of extremely large data sets, statistical reporting, and data mining. The applications of DMelt are natural sciences, financial markets, and engineering.

D. Rattle

Rattle is another data mining tool based on GUI which uses the R programming language. Rattle exposes the statistical power of the R programming language by providing significant data mining capabilities. While Rattle has a well-developed, and comprehensive graphical user interface, Rattle also has an integrated log code window that will produce duplicate code for any specific GUI operation.

The data set produced by Rattle can be viewed and edited, also Rattle gives additional facility to review code, use it for many purposes, and extend the code without any limitations.

E. Rapid Miner

Rapid Miner is one of the most popular predictive analysis systems used in data mining, created by the company Rapid Miner. Rapid Miner is developed using the JAVA programming language, and Rapid Miner offers an integrated development environment for text mining, deep learning, machine learning, and predictive analysis.
Rapid Miner can be used in a wide range of applications, including but not limited to commercial applications, application development, education, training, machine learning, and research.

VI. CONCLUSION AND FURTHER STEPS

Data mining is a very powerful as well as a useful methodology for generating vital information for decision making. Currently, data mining is done mainly on simple numeric as well as categorical data. Data mining can be utilized in more complex data types. For any model that has been designed and selected, further refinement is possible by exploring other attributes as well as their relationships.

Research in data mining will result in supplement to determine the most interesting characteristics in the data. Given the advantages of data mining, there is no second thought that it will become increasingly more essential to organizations, both commercial as well as not commercial ones.

Data mining may be handy to anticipate the future directions that data mining, as well as its related areas, may take, which can be classified under the major categories of data, applications, software, and hardware.

Natural language makes the path for mining in free-form text, specifically for automated annotation, and indexing prior to classification of data corpora. Unlimited parsing capabilities in Natural language can help significantly in the process of deciding what an article actually refers to.

Thus, the spectrum from simple natural language processing methodologies, all the way to language mastering can help predominantly. Together with these capabilities, natural language processing can contribute principally as an effective interface for describing hints to mining algorithms, conceptualizing, and detailing knowledge discovered by a KDD system.

Intelligent agents can be propelled to accumulate necessary information from a wide range of sources. Along with this, information agents can be triggered remotely over intranet / internet or can be activated on the occurrence of specific events, and then launch an analysis operation. To sum up, agents also can assist to navigate, and model the World-Wide Web, the competitive area growing in importance.

Uncertainty in AI includes but is not limited to problems for maintaining uncertainty, proper interpretation mechanisms in the essence of uncertainty, and the reasoning about the root cause, all underlying concepts to KDD theory, and practice.

Knowledge representation includes ontologies, a newly evolved methodology for representing, storing, and accessing knowledge. In addition to that, Knowledge representation includes the schemes for representing knowledge and empowering the use of prior human knowledge about the underlying process by the KDD system.

REFERENCE

Books and Journals

[1] Adela Tudor, Adela Bara, Iuliana Botha. The Bucharest Academy of Economic Studies Bucharest ROMANIA. Solutions for analyzing CRM systems - data mining algorithms. International Journal of Computers Issue 4, Volume 5, 2011.

[2] Arun K Pujari. Data Mining Techniques. The Orient Blackswan, 2016.

[3] Berndt, D., and Clifford, J. Finding Patterns in Time Series: A Dynamic Programming Approach. 1996.

[4] Berry, J.. Database Marketing. Business Week, September 5, 56–62. 1994.

[5] Buntine, W. Graphical Models for Discovering Knowledge. 1996.

[6] C. Bash, C. Patel, A. Shah, R. Sharma, "The Sustainable Information Technology Ecosystem", ITHERM'08, June 2008, Orlando, FL.

[7] Chien-Hua Wang and Chin-Tzong Pang. Applying Fuzzy Data Mining for an Application CRM H. -J. Zimmermann, Fuzzy sets, Decision Making, and Expert Systems, Kluwer, Boston. 1991.

[8] Fielding, Nigel G. and Raymond M. Lee. Using Computers in Qualitative Research. SAGE Publications, 1991.

[9] Friedman, J. H. Multivariate Adaptive Regres-sion Splines. Annals of Statistics 19:1–141. 1989.

[10] Geman, S.; Bienenstock, E.; and Doursat, R. Neural Networks and the Bias/Variance Dilemma. Neural Computation 4:1–58. 1992.

[11] Hall, J.; Mani, G.; and Barr, D. Applying Computational Intelligence to the Investment Process. In Proceedings of CIFER-96: Computational Intelligence in Financial Engineering. Washington, D.C.: IEEE Computer Society, 1996.

[12] Hernandez, M., and Stolfo, S. The MERGE-PURGE Problem for Large Databases. In Proceedings of the 1995 ACM-SIGMOD Conference, 127–138. New York: Association for Computing Machinery, 1995.
[13] Horvitz, E., and Jensen, F. Proceedings of the Twelfth Conference of Uncertainty in Artificial Intelligence. San Mateo, Calif.: Morgan Kaufmann, 1996.
[14] Jain, A. K., and Dubes, R. C. Algorithms for Clustering Data. Englewood Cliffs, N.J.: Prentice-Hall, 1988.
[15] Jiawei Han , Micheline Kamber. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011.
[16] Jing He. Advances in Data Mining: History and Future, Third international Symposium on Information Technology, 2009.
[17] Karl Rexer, Heather Allen, & Paul Gearan. Data Miner Survey Summary, presented at Predictive Analytics World, Oct. 2010.
[18] Kotsiantis, S., Kanellopoulos, D., Pintelas, P. Multimedia Mining. WSEAS Transactions on Systems, No 3, s. 3263-3268. 2004.
[19] Langley, P., and Simon, H. A. Applications of Machine Learning and Rule Induction. Communications of the ACM 38:55–64. 1995.
[20] M. Marwah. "Data analysis, Visualization and Knowledge Discovery in Sustainable Data Centers", Proc. of ACM Compute, Jan 9-10, 2009, Bangalore, India.
[21] Ma, Y.; Guo, Y.; Tian, X.; Ghanem, M. Distributed Clustering-Based Aggregation Algorithm for Spatial Correlated Sensor Networks. IEEE Sensors Journal 11 (3): 641. 2011.
[22] Mani and Inderjeet. Automatic Summarization. Amsterdam/Philadelphia: John Benjamins, 2002.
[23] Margaret H. Dunham. Data Mining: Introductory and Advanced Topics. Wiley, 2016.
[24] Mehmed Kantardzic. Data Mining: Concepts, Models, Methods and Algorithms. Wiley, 2017.
[25] MUC 6. Proceedings of the Sixth Message Understanding Conference (MUC-6). Morgan Kaufmann, 1996.
[26] OscarMarbán, Gonzalo Mariscal and Javier Segovia. A Data Mining & Knowledge Discovery Process Model. In Data Mining and Knowledge Discovery in Real Life Applications 438-453, Austria, 2009.
[27] Pang-Ning Tan, Michael Steinbach, Vipin Kumar. Introduction to Data Mining. Pearson, 2016.
[28] Parteek Bhatia. Data Mining and Data Warehousing: Principles and Practical Techniques. Cambridge, 2019.
[29] Pearl, J. Probabilistic Reasoning in Intelligent Systems. San Francisco, Calif.: Morgan Kaufmann, 1988.
[30] Pieter Adriaans and Dolf Zantinge. Data Mining. Pearson, 2002.
[31] Quinlan, J. C4.5: Programs for Machine Learning. San Francisco, Calif.: Morgan Kaufmann, 1992.
[32] Ralph Kimball and Joe Caserta. The Data WarehouseETL Toolkit: Practical Techniques for Extracting, Cleaning, Conforming, and Delivering Data. Wiley, 2007.
[33] Ralph Kimball and Margy Ross. The Kimball Group Reader: Relentlessly Practical Tools for Data Warehousing and Business Intelligence Remastered Collection. Wiley, 2016.
[34] S.K. Shinde and Uddagiri Chandrasekhar. Data Mining and Business Intelligence. Dreamtech Press, 2015.
[35] Silverman, B. Density Estimation for Statistics and Data Analysis. New York: Chapman and Hall, 1986.
[36] van Kuppevelt, Jan, Ulrich Heid, and Hans Kamp. Best practice in spoken language dialogue systems engineering - introduction to the special issue. Natural Language Engineering, 6(3 & 4):205–212, 2000.
[37] Vikram Pudi and P. Radha Krishna. Data Mining: Concepts and Techniques. Oxford, 2012.

Websites

[1] www.kdnuggets.com
[2] www.loginworks.com
[3] www.mozenda.com
[4] www.smartertools.com
[5] www.javatpoint.com
[6] www.scalar.com
[7] www.datacamp.com
[8] www.infogix.com
[9] www.guru99.com
[10] www.onix-systems.com
[11] www.analyticsindiamag.com
[12] www.upgrad.com
[13] www.frontiersin.org