A REVIEW PAPER ON DATA MINING TECHNIQUES, ALGORITHMS AND TOOLS

V.P. Muthukumar¹, Dr.S.Subbaiah², S.Srinivasan³

¹ Assistant Professor, Ph.D Part Time Research Scholar, Department of Computer Science, Vivekanandha College of Arts and Sciences for Women (Autonomous), Tiruchengode, Namakkal Dt., Tamilnadu.(India)

²Assistant Professor, Department of Computer Science, Vivekanandha College of Arts and Sciences for Women (Autonomous), Tiruchengode, Namakkal Dt., Tamilnadu.(India)

³Assistant Professor, Department of Computer Science, Vivekanandha College of Arts and Sciences for Women (Autonomous), Tiruchengode, Namakkal Dt., Tamilnadu.(India)

ABSTRACT

Data mining has made a great change in recent year but the problem of missing data has remained a great dispute for data mining algorithms and Tools. It is an activity of extracting some useful knowledge from a large data base, by using any of its techniques and Tools. Data mining is used to discover knowledge out of data and presenting it in a useful way to the human society. Data mining is the notion of all methods and techniques which allow analyzing very large data sets to extract and discover previously unknown structures and relations out of such huge heaps of details. This paper studied the classification and clustering techniques on the basis of algorithms which is used to predict previously unknown class of objects and the data mining tools used for mining and presenting the data to user also include the trends in data mining.

Keywords: - Data Mining, Classification, Clustering Algorithms, Data Mining Tools.

I.INTRODUCTION

Data mining refers to extracting or mining knowledge from large amounts of data. Data mining is synonym for another popularly used term, knowledge Discovery from Data, or KDD[1]. Data mining is the process of extraction hidden knowledge from large volumes of raw data. Data mining has been defined as the nontrivial extraction of previously unknown and potentially useful information from data. Data mining is used to discover knowledge out of data and presenting it in a form that is easily understood to humans [2]. Data mining is the notion of all methods and techniques which allow analyzing very large data sets to extract and discover previously unknown structures and relations out of such huge heaps of details. Data Mining is the process of extracting information from large data sets through the use of algorithms and techniques drawn from the field of Statistics, Machine Learning and Data Base Management Systems (Feelders, Daniels and Holsheimer, 2000). Traditional data analysis methods often involve manual work and interpretation of data that is slow, expensive an highly subjective (Fayyad, Piatsky Shapiro and Smyth, 1996). Data Mining, popularly called as knowledge

discovery in large data, enables firms and organizations to make calculated decisions by assembling, accumulating, analyzing and accessing corporate data. It uses variety of tools like query and reporting tools, analytical processing tools, and Decision Support System (DSS) tools.

1.1 Data mining as a core process in KDD

The Knowledge Discovery in Database process comprises of a few steps leading from raw data collections to some form of new knowledge. It consists of the following steps as shown in Fig 1.1.1



Fig 1.1.1: Data mining as a step in the process of KDD

- **Data cleaning:**-It is also known as the data cleansing, it is a phase in which noise data and irrelevant data removed from the collected data.
- **Data integration:**-At this time, multiple data sources, often heterogeneous, may be combined in a common source.
- **Data selection:**-At this stage, the data relevant to the analysis is decided on and retrieved from the database collection.
- **Data transformation:**-It is also known as data consolidation or summarization, it is a phase in which the selected data is transformed into forms appropriate for the mining procedure.
- **Data mining:**-It is the essential process in which clever techniques are applied to extract patterns potentially useful.
- **Pattern evaluation:**-In this phase, truly interesting patterns representing knowledge are identified based on given measures.
- **Knowledge representation:**-It is the final phase in which the discovered knowledge is visually represented to the user. This essential set uses visualization techniques to help users understand and interpret the data mining result [3]

1.2 Data Mining Techniques:-

Set of attributes. For each tuple in the training data, the value of class label attribute is known. Classification algorithm is applied on data training data to create the model. In the second step of classification, test data is used to check the accuracy of the model. If the accuracy of the model is acceptable then the model can be used to classify the unknown tuples [4].Classification techniques were developed as an important component of machine learning algorithms in order to extract rules and patterns from data that could be used for prediction. Classification techniques are used to classify data records into one among a set of predefined classes. They work by constructing a model of training dataset consisting of example records with known class labels[5].

1.2.1 Clustering Approach

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. Clustering can be considered the most important unsupervised learning technique. Clustering can be considered the most important unsupervised learning technique so as every other problem of this kind. It deals with



Fig 1.2.1 Shows the Descriptive and Predictive Data Mining Techniques.

Descriptive approach includes models for overall probability distribution of the data, partitioning of whole data into groups and models describing the relationships between the variables. Predictive approach permits the value of one attribute/variable is too predicted from the known values of other attribute/variable. This paper studies the one descriptive technique i.e. clustering and one predictive technique i.e. classification and tools used for better analysis.

1.2.2 Classification Approach

Classification is a supervised learning method [3]. Data classification is two-step process. In the first step, a model is built by analyzing the data tuples from training data having a finding a structure in a collection of unlabeled data. Clustering is the process of organizing objects into groups whose members are similar in some way [9]. Cluster analysis has been widely used in many applications such as business intelligence image pattern recognition web search biology and security. In business intelligence clustering can be used to organize a large number of customers into groups where customers within a group share similar characteristics. This facilitates the development of business strategies for enhanced customer relationship management. In image recognition clustering can be used to discover cluster or subclasses in handwritten character recognition system. Suppose we

have a data set of handwritten digits where each digit is labeled as 1, 2, 3, and so on. Note that there can be a large variance in the way in which people write the same digit. Take the number 2, for example .some people may write it with a small cicle at the left bottom part, while some other may not. We can use clustering to determine sub classes for each of which represents a variation on the way in which 2 can be written. Using multiple models based on the subclasses can improve overall recognition accuracy [5].

The organization of this paper consists of following sections: Section 1 Lays the Basis of The Study, Section 2 Provides an overview of classification and clustering algorithm considered for study and Section 3 Data Mining Tools and its Comparisons, Section 4 Trends in Data Mining Section 5 Concludes the study along with scope for future work.

II.OVERVIEW OF CLASSIFICATION AND CLUSTERING ALGORITHMS

2.1 Classification Algorithms:-

A Classification Algorithm is a procedure for selecting a hypothesis from a set of alternatives that best fits a set of observations.

2.1.1 Tree based CA:-

Tree builds classification or regression models in the form of a tree structure. It breaks down a dataset into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes.

2.1.1.1) Decision Stump:-

A Decision Stump is a machine learning model consisting of a one-level decision tree.[1] That is, it is a decision tree with one internal node (the root) which is immediately connected to the terminal nodes (its leaves). A decision stump makes a prediction based on the value of just a single input feature. Sometimes they are also called 1-rules Decision stumps are often[6] used as components (called "weak learners" or "base learners") in machine learning ensemble techniques such as bagging and boosting. For example, a state-of-the-art Viola–Jones face detection algorithm employs AdaBoost with decision stumps as weak learners

2.1.1.2) J48:

J48 classifier is a simple C4.5 decision tree for classification. It creates a binary tree. The decision tree approach is most useful in classification problem. With this technique, a tree is constructed to model the classification process. Once the tree is built, it is applied to each tuple in the database and results in classification for that tuple [1][3].

2.1.2 Rules based classification algorithms

Rule based classification algorithm also known as separate-and-conquer method. This method is an iterative process consisting in first generating a rule that covers a subset of the training examples and then removing all examples covered by the rule from the training set. This process is repeated iteratively until there are no examples left to cover [7]. Rule discovery or rule extraction from data is data mining techniques aimed at

180 | Page

understanding data structures, providing comprehensible description instead of only black box prediction.

2.1.2.1) ZeroR:-

ZeroR is the simplest classification method which relies on the target and ignores all predictors. ZeroR classifier simply predicts the majority category (class). Although there is no predictability power in ZeroR, it is useful attribute, the most frequent class for each attribute value must be determined [10]. The most frequent class is simply the class that appears most often for that attribute value. A rule is simply a set of attribute values bound to their majority class. OneR selects the rule with the lowest error rate. In the event that two or more rules have the same error rate, the rule is chosen at random. The OneR algorithm creates a single rule for each attribute of training data and then picks up the rule with the least error rate.

2.1.2.2) OneR:

OneR or "One Rule" is a simple algorithm proposed by Holt. The OneR builds one rule for each attribute in the training data and then selects the rule with the smallest error rate as its one rule. The algorithm is based on ranking all the attributes based on the error rate [9]. To create a rule for an attribute, the most frequent class for each attribute value must be determined. The most frequent class is simply the class that appears most often for that attribute value. A rule is simply a set of attribute values bound to their majority class. OneR selects the rule with the lowest error rate. In the event that two or more rules have the same error rate, the rule is chosen at random. The OneR algorithm creates a single rule for each attribute of training data and then picks up the rule with the least error rate.

2.1.2.3) PART:-

PART is a Partial Decision Tree Algorithm, which is the developed version of C4.5 and RIPPER algorithms. PART is a Separate-and-Conquer rule learner proposed by Eibe and Witten [54]. The algorithm producing sets of rules called decision lists which are ordered set of rules. A new data is compared to each rule in the list in turn, and the item is assigned the category of the first matching rule (a default is applied if no rule successfully matches). PART builds a partial C4.5 decision tree in its each iteration and makes the best leaf into a rule. The algorithm is a combination of C4.5 and RIPPER rule learning [13].

2.2 Clustering Algorithms:-

Clustering is a data mining technique to group the similar data into a cluster and dissimilar data into different clusters.

2.2.1 K-Mean algorithm:-

K-mean is an iterative clustering algorithm in which items are moved among sets of clusters unit the desired set is reached. As such, it may be viewed as a type of squared error algorithm, although the convergence criteria need not be defined based on the squared error. A high degree of similarity among elements in clusters is obtained, while a high degree of similarity among elements in clusters is obtained while a high degree of dissimilarity among elements in different clusters is achieved simultaneously[6].

2.2.1.1 Sets of algorithm:-

- a) First it selects the initial k prototypes arbitrarily.
- b) The squared error criterion is used to determine the clustering quality.
- c) In each iteration the prototype of each cluster is re-computed to be the cluster mean.
- d) The basic version of k- means does not include any sampling techniques to scale to huge databases.
- e) For determining a baseline performance as a benchmark for other classification methods [8]. It is the simplest method which algorithms actually creates sets of clusters.

2.2.2 Hierarchical Algorithm:-

Hierarchical clustering Hierarchical algorithm differs in how the sets are created. A tree data structure called a dendrogram can be used to illustrate the hierarchical clustering technique and the sets of different clusters. The root is a dendrogram tree contains one cluster where all elements are together. The leaves in the dendrogram each consist of a single element cluster. Internal nodes in the dendrogram represent new clusters formed by merging the clusters that appear as its children in the tree. Each level in the tree is associated with the distance measure that was used to merge the clusters. All clusters created at a particular level were combined because the children clusters had a distance between them less than the distance value associated with this level in the tree [6].

III.DATA MINING TOOLS

Today's various data mining tools that are available to handle or manage the large number of datasets and also to improve the quality of data, such tools are RapidMiner, Weka, R, scikit-learn, KNIME, Orange, KEEL, Tanagra etc.These data mining tools makes easy for analyst to get the knowledgeable information.Data mining tools are used to predict future trends, behaviours, allowing business to make proactive, knowledge driven decissions[16]. The various Data mining techniques and algorithms have been implemented on these tools to extract the information and also to check their efficiency and accuracy. In this paper, we are going to discuss and compare only three tools among of these that are; RapidMiner, WEKA, and KNIME which are using the same platform(Java).The description of these tools are as follows:

3.1 Weka

The WEKA(Waikato Environment for Knowledge Analysis) is a Data Mining tool, developed at the University of Waikato, New Zealand that is suitable for machine learning algorithm for data mining tasks and well suited for developing new machine leaning schemes [17]. These algorithms can be applied directly to dataset or can be called from your own java code [16].WEKA is an open source software issued under the GNU(General Public License) agreement. WEKA provide four application interface : Explorer, Experimenter, Knowledge flow, and

Simple Command line [18]. But Explorer is the main interface of WEKA. WEKA is Java based software and can run in different platforms. With the Java based version, the tool is so revolutionary and used in various application including visualization and algorithm for data analysis and predictive modeling [17]. It is freely available for download and offers many powerful features.

3.1.1 Features

- WEKA is Java based open source data mining tool.
- It is easy to use for beginners and has the ability of running several learning algorithms and comparing.
- It is platform independent.
- It performs various data mining tasks including: Data preprocessing, Classification rules, regression, Clustering, association rules, visualization, feature selection and improving the knowledge discovery.
- WEKA has 49 Data preprocessing tools, 76 Classification/regression algorithm, 8 Clustering algorithm, 3 algorithm for finding association rules, 15 attribute/subset evaluator plus 10 search algorithm for feature selection [19].
- There are various built in features.
- There is no programming and coding language required.

3.1.2 Advantages

- Easy to manipulate the data.
- Provide access to SQL databases.
- It provides two options for the user to interact through Explorer and Command line [20].
- Specially used for data mining.
- It provides various machine learning algorithms for data mining tasks.
- It supports various standard Data mining tasks that include: Data preprocessing, Clustering and Classification, Regression, Visualization and Feature selection [21].

3.1.3 Disadvantages

- Memory is limited and has lesser performance [15].
- Data visualization and data survey is limited.
- Not better suitable option for the large datasets as they roughly handled.
- Lacking in the representation to the result of processing.
- Limited ability to partition dataset to training and test set [22].
- It doesn't accept data in every format (data format constraints).
- Not good in interfacing with other software [15].

3.2 Rapid Miner

RapidMiner, previously YALE (Yet Another Learning Environment) was developed at the Technical University of Dortmund in 2001 by Ralf Klinkenberg, Ingo Mierswa and Simon Fischer.After, this software name was changed in 2007 from YALE to RapidMiner and is developed by the company RapidMiner, Germany. RapidMiner is an open source java based system for data mining and provides an integrated environment for machine learning, data mining, text mining ,predictive analysis and business analytics and is mainly used for business and industrial application[16].RapidMiner is the most powerful, easy to use and intuitive Graphical User Interface for the design of analytic process, that contain several "operators".The operator functions as a single task in their process in which the input is produced by the existing output of the operator[15].

3.2.1 Features

- It is platform independent.
- It has compatibility with various databases like oracle, MySQL, Excel, SPSS, Microsoft SQL server etc.
- It provides Drag and Drop interface to design the analytics process.
- It supports and accepts new data drivers.
- It provides more than 500 operators for all machine learning procedures, and also combines learning schemes and attributes evaluators of the WEKA learning environment [23].
- It allow user to work with different sizes and types of data sources.

3.2.2 Advantages

- It has enormous flexibility.
- It provides the integration of maximum algorithm of such tools.
- Easy to debug the errors.

3.2.3 Disadvantages

• Limited partitioning abilities for dataset to training and testing sets.

3.3 KNIME

Konstanz Information Miner is an open source general data mining tool that is based on the Eclipse platform, developed and supported by KNIME.com.AG. In 2004, the KNIME initially developed by the team of software engineer at the University of Konstanz, Germany and in 2006, the initial version of KNIME was released [14]. KNIME is very powerful tool for analytical task, extracting data and knowledge from the web communities. The KNIME base version already incorporates hundreds of processing nodes for data I/O, preprocessing and

184 | Page

cleansing, modeling, analysis and data mining as well as various interactive views, such as scatter plots, parallel coordinates and others[24].In KNIME, representation of data sources and sinks, mining algorithm, transformations, visualizations, etc defined by set of nodes called "workflow" and each node has its specific input and output ports that depends on the functionality of the node [25]. For both simple and complex data types, KNIME allows revolutionary analysis to discover trends and predict future results. KNIME uses for teaching as well as research which allows integrating the new algorithm and tools in a simpler manner.

3.3.1 Features

- Available to everyone i.e., allow users to use the well- defined node API to add proprietary extensions.
- Intuitive user interface.
- It is easy to use and handle different functions.
- KNIME modules cover a wide variety of functionalities like, I/O, data manipulation, views, hilting etc to better understand your data.
- It provides the users to create data flows or pipeline visually, users can selectively execute some or all analysis steps, study the results, prototypes, and collaborative interpretations [20].
- For cross validation and independent validation, it provides functionality to save parameters.

3.3.2 Advantages

- The major benefit of this is easy to use plug-in [25].
- It based on the node work which includes more than 100 nodes to examine the data [26].
- It provides data flow process by dragging and dropping new nodes.

3.3.3 Disadvantages

- Less suitable option for large complex workflows.
- Partitioning ability is limited for dataset.

3.4 R-Programming Tool:

This is written in C and FORTRAN [28], and allows the data miners to write scripts just like a programming language/platform. Hence, it is used to make statistical and analytical software for data mining. It supports graphical analysis, both linear and nonlinear modeling, classification, clustering and time-based data analysis.

3.4.1 Features of R Tool

• R is more than just a domain-specific programming language aimed at data analysis.

- It performs multiple calculations with vectors.
- It is processing more than just statistics.
- It can running code without a compiler

3.4.2 Advantages

- R is a programming language and environment developed for statistical analysis by practicing statisticians and researchers.
- R is free and open source software, allowing anyone to use and, importantly, to modify it.
- R has no license restrictions [27].
- R has over 4800 packages available from multiple repositories specializing in topics like econometrics, data mining, spatial analysis, and bio-informatics.
- R is cross-platform. R runs on many operating systems and different hardware. It is popularly used on GNU/Linux, Macintosh, and Microsoft Windows, running on both 32 and 64 bit processors [27].

3.4.3 Disadvantages

- In R, quality of some packages is less than perfect [27].
- In R, no one to complain, if something doesn't work [27].
- R is a software Application that many people devote their own time to developing.
- R commands give little thought to memory management, and so R can consume all available memory.

3.5 Comparison of Data Mining Tools

This comparative study of Data Mining Tools based on their parameters is listed in TABLE 1. The main motive of this comparison is to make the best selection of tool with respect to their areas.

The tools are used to produce the best data mining results to the researchers as well as the users. Here we have compared are WEKA, Rapid Miner, KNIME and R Tool.

Parameters	WEKA	RapidMiner	KNIME	R Tools
				University of Auckland
	University of	D INC	Swiss company	New
Developer	Waikato,	RapidMiner,	Knime.com	Zealand. John
	New Zealand.	Germany.	AG, Switzerland.	Chambers and colleagues developed
				R at Bell Laboratories.
Programming	Iava	Lava	Lovo	Primarily in R, and sometimes in Java,
Language	Java	Java	Java	C, C++, and Fortran.
Released date	1993	2006	2004	2000

 Table1. General Parameters of these tools

186 | Page

GNU GNU GNU AGPL License General Public General Public Proprietary General Public License License License Availability **Open Source Open Source Open Source Open Source** Current 3.7.13 6.5 2.10 3.4.2 Version Financial forecasting, Machine Customer Targeted learning, Data intelligence, marketing, visualization, Finance, Medical time series and Manufacturing, Computational statistics, diagnosis, Areas analysis, text Chemical library, perception, and data science. credit card mining, fraud enumeration, fraud detection. detection Retail, Cross text mining, Etc industry etc. weather forecasting etc Linux, Windows, GNU/Linux, Macintosh, and Cross Platform Cross Platform Portability OS X Microsoft Windows Usability Easy Easy Easy Easy Oracle, IBM MySQL, Postgre DB2, SQL, MSQL MySQL, SQLite, Microsoft SQL Oracle, IBM DB2, Compatibility Server, Oracle, Postgre Server. Microsoft SQL Server, with database ODBC, Sqlite SQL, Hive Mysql, Excel, Mysql, Excel, Access, SPSS etc 3.x, HSQLDB connector etc. Access, SPSS etc. etc. Platform Platform Platform Platform Platform independent independent independent independent Supporting Easy to use but Flexibility Flexible Same as WEKA Flexible not enough flexible Limited Better Better Visualization High End Graphics Support visualization visualization visualization GUI Has good but It has better It has better GUI It has better GUI

not better as	GUI	
much as other		
mentioned tools		

IV. DATA MINING TRENDS

Here is the list of trends in data mining that reflects pursuit of the challenges such as construction of integrated and interactive data mining environments, design of data mining languages [1]:

- Application Exploration
- Scalable and Interactive data mining methods
- Integration of data mining with database systems, data warehouse systems and web database systems.
- Standardization of data mining query language
- Visual Data Mining
- New methods for mining complex types of data
- Biological data mining
- Data mining and software engineering
- Web mining
- Distributed Data mining
- Real time data mining
- Multi Database data mining
- privacy protection and Information Security in data mining

V. CONCLUSIONS AND FUTURE WORK

This paper presents a detailed description of data mining techniques, algorithms and Tools. It list the trends in Data mining. Therefore, Data Mining is the process of discovering interesting knowledge from large amounts of data stored either in databases, data warehouses, or other information repositories. The various algorithms and tools used for the mining of data are specified in detail. The future scope provides enhancement and efficiency of data in the system. They could lead to better, faster and qualitative exaction of data with better tools and techniques.

REFERENCES

- [1] Jiawei Han and Micheline Kamber, "Data Mining Concepts and Techniques", Second Edition 2006.
- [2] Prajapati. D, Prajapat. J, "Handling missing values: Application to University Data Set", August, 2011.

- [3] Han. J, Kamber. M, Pei. J, " Data Mining Concepts and Techniques", Third edition The Morgan Kaufmann Series in Data Management Systems Morgan Kaufmann Publishers, July 2011
- [4] Kabra. R, Bichkar. R, "Performance Prediction of Engineering Students using Decision Tree", International Journal of computer Application s, December ,2011
- [5] VikramPudi, PRadha Krishna "Data Mining", Oxford University Press, First Edition, 2009
- [6] Margaret H. Dunham, "Data Mining Introductory and Advanced Topics", Dorling Kindersley Pvt.Ltd.India, Sixth Edition, 2013.
- [7] Phyu, Thair Nu. "Survey of classification techniques in data mining."International MultiConference of Engineers and Computer Scientists, 2009.
- [8] http://www.saedsayad.com/ zeror.htm
- [9] Tayel, Salma, et al. "Rule-based Complaint Detection using RapidMiner", Conference: RCOMM 2013, At Porto, Portugal, Volume: 141-149,2014
- [10] http://mydatamining.wordpress.com/2008/04/14/rule-learner-or-rule-induction/
- [11] Vijayaran S, Sudha. "An Effective Classification Rule Technique for Heart Disease Prediction". International Journal of Engineering Associates, February 2013.
- [12] Buddhinath, Gaya, and Damien Derry."A simple enhancement to One Rule Classification." Department of Computer Science & Software Engineering. University of Melbourne, Australia (2006).
- [13] Ali, Shawkat, and Kate A. Smith."On learning algorithm selection for classification." Applied Soft Computing, 2006.
- [14] A. Jovic, K. Brikic, N. Bogunovic, An Overview of free software tools for general Data mining.
- [15] S. Sarumathi, N. Shanthi, S. Vidhya, M. Sharmila, A Review: Comparative Study of Diverse Collection
- [16] K. Rangra, K.L. Bansal, Comparative Study of Data Mining Tools, International Journal of Advanced Research in Computer Science and Software Engineering, 4(6), June 2014.
- [17] P.S. Patel, S.G. Desai, Comparative Study on data Mining tools, International Journal of Advanced Trends in Computer Science and Engineering, 4(2), April 2015.
- [18] S. Srivastava, WEKA: A Tool for Data Preprocessing, Classification, Ensemble, Clustering and Association Rule mining, International Journal of Computer Applications, 88(10), February 2014.
- [19] S.K. David, Amr T.M. Saeb, K.A. Rubeaan, Comparative Analysis of Data Mining Tools and Classification Techniques using WEKA in Medical Bioinformatics, Computer Engineering and Intelligent System, 4(13), 2013.
- [20] K. Saravanapriya, A Study on Free Open Source Data Mining Tools, International Journal of Engineering and Computer Science, 3(12), December 2014.
- [21] S. Singhal, M. Jena, A study on WEKA tool for Data Preprocessing, Classification and Clustering, International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2(6), May 2013.

- [22] H. Solanki, Comparative Study of Data Mining Tools and Analysis with Unified Data Mining Theory, International Journal of Computer Applications, 75(16), August 2013.
- [23] M. Vijayakamal, M. Narendhar, A Novel Approach for WEKA & Study On Data Mining Tools, International Journal of Engineering and Innovative Technology (IJEIT), 2(2), August 2012.
- [24] L. Kataria, Implementation of Knime-Data Mining Tool, International Journal of Advanced Research in Computer Science and Software Engineering, 3(11), November 2013.
- [25] S. Gunnemann, H. Kremer, R. Musiol, R.Haag, T. Seidl, A Subspace Clustering Extension For the
- [26] KNIME Data Mining Framework, 2012 IEEE 12th International Conference on Data Mining Workshops.
- [27] P. Subathra, R. Deepika, K. Yamini, P. Arunprasad, S.k Vasudevan, A Study of Open Source Data Mining Tools and its Applications, 10(10), 2015
- [28] http://analyticstrainings.com/?p=101
- [29] https://en.wikipedia.org/wiki/R_(programming_language)