

# A Comprehensive Analysis of Proprietary and Open Source Data Mining Tools

Sonia Rani Chowdhary, Vikash

Computer Application, Post Graduate Government College, Chandigarh, Chandigarh, India

## ABSTRACT

The Powerful software tools and techniques required for the development of data mining applications. With the rapid development of technologies and business interest in using electronics and latest technologies plays important role in improvement of data mining field. Data mining access the meaningful and efficient information available in worldwide which is helps in decision making. This paper described the (a) various tools and techniques used by data mining applications. (b) compared features and limitations both in Proprietary and open sources data mining tools. (c) technical analysis of proprietary and open source data mining tools. On the basis of well-designed User interface, short time analysis, statistical and mathematical analysis user can select the best tool as per their requirements. Analysis of these tools makes easy to select appropriate tool.

**Keywords :** Data Mining Tools, Open Source Tools, Proprietary tools, Technical Specifications

## I. INTRODUCTION

Today with the development of technologies the needs of data become more important in everyday life. The size of data increases day by day. Amount of information double every 20 months according to estimations [1]. Data mining is process of discovering useful pattern among large amount of data such as Big data and transformed into more informative form. It is a core step of knowledge discovery. Global IP traffic is expected to reach more than 190 exabytes per month in 2020 according to Cisco Forecast [2] and in 2025 total amount of data expected to reach over 160 zettabytes [3]. Worldwide market for Business Intelligence (BI) accounted over 17 Billion USD in 2016 according to Gartner and 2021 it expected to reach more than 26 Billion USD, where modern BI and analytics expands more rapidly than other fields [4]. Data mining is not only important in BI but also in technologies and life science sectors. It is strongly associated with data science including modifications and classifications of data by using statistical and

mathematical concepts. Data mining is also important in public opinion and marketing research [5] [6].

The purpose of this paper to not only comprehensive comparison of all proprietary and open source tools but also analysis each and every tool, which helps to select best tool for specific field or area.

## II. LITERATURE'S REVIEWS

A Comparative Study of Famous Classification Techniques and Data Mining Tools on the basis of specific parameters. Paul Y, Kumar N described various classifications techniques and data mining tools along with their advantages and disadvantages. KNN (K- Nearest Neighbor) classifier is good choice for classifications problem where the data set size is small and noise free. SVM (Support Vector Machine) are frequent and widely used classifier. Decision tree and Bayesian network discussed. They represent many combined methods but still not clear which one is best to full fill the requirements [7]. Myint Myint Than

represents a paper called "A Comparative Study of predicting teaching score by using classification algorithms" [8] this paper respectively expounds comparison between Naïve Bayesian, J48 and Random Forest Decision Tree classifiers are shown on the frequent item set is developed for data analysis. The data set analyze for better prediction. According to this paper result factors, the performance of J48 Decision tree classifier is more accurate and more precision than Naïve Bayes and Random Forest classifiers. J48 Decision tree algorithm is best for educational manager for future decision making in teacher selection by using its resulting data. Feng, Junshuai, introduced a paper called "Predicting Students' Academic Performance with Decision Tree and Neural Network" [9]. This paper explained EDM (Educational Data Mining) and Neural Networks in data mining applications. EDM and Neural Networks related to educational background and can analyze and resolving educational data problems. Both are utilized in widespread and successful data mining applications. The aim of this paper to explore the methodology such as decision tree, classifiers and neural network to predict student performance in the context of EDM. Sibel Barin Özkan, Sultan Muhammed et al. [10] explain in this paper many commercial and open source program to implement Data Mining application. This paper introduced open source data mining programs WEKA, Orange, KNIME. Aim of this study to determine the difference between these three data mining programs with Naïve Bayes classification algorithms by taking Bay iris, breast cancer, wine, monk, balance es data set from UCI Machine Learning Repository database. Hansheng Lei, Mahmoud Quweider et al. introduced paper "Mining Survey Data" [11] in this represents the strategies in mining survey data using computational methods. A Novel method for data preparation and dependent pattern mining is presented. It introduced DP (Dependent Pattern) algorithm that is not only mines compact sets of patterns, but also discovers meaningful patterns.

### III. OVERVIEW OF PROPRIETARY DATA MINING TOOLS

Sisense is best suited for business intelligence software. Non-technical persons can easily use it by its drag and drop facility. SSDT (SQL Server Data Tools) has all phases of database. It is a declarative model and has visual tool for development like intelligence, code navigation tools and programming support via c#, visual basic etc. Oracle data mining is based on relational database. It has algorithms for data classifications, prediction, regression and specialized analytics which helps to make better prediction. IBM Cognos helps to meet organization needs like Cognos connections, query studio, report studio, analysis studio, event studio and workspace advance. IBM SPSS Modeler produced by SPSS and then acquired by IBM. It has visual interface and no need of programming knowledge to interact with it. SAS data mining is analysis and data management SAS institute develop Statistical Analysis System (SAS). It helps to take timely decision. Tera data mining follow no sharing structure it means it has their own memory and processing ability. Board gather data from heterogeneous resources and analyze it. Dundas BI is a quick insight tool. It has Gap free protection of document feature which provides good accessibility across many devices [12].

### IV. OVERVIEW OF OPEN SOURCE DATA MINING TOOLS

Rapid Miner is one of the best predictive analysis. It is based on client/server model. Orange is components based and data visualization tool. It helps users to take smart decision. Weka is also known as Waikato environment. SQL database access features also available in Weka. KNIME (Konstanz Information Miner) is basically used for pharmaceutical research but also used in other areas like business intelligence

and financial data analysis. Apache Mahout has clustering, classification and collaboration filtering features. Rattle tool generate duplicate code for any activity happen in GUI by using inbuilt language code tab. Data Melt is known as DMelt. This tool is basically for engineers, scientist and students. KEEL (Knowledge Extraction based on Evolutionary Learning) used to discover large number of different knowledge tasks. R(Revolution) data mining is very

useful for complicated analysis without knowing the depth of computer system [13].

### V. COMPARATIVE ANALYSIS OF PROPRIETARY AND OPEN SOURCE DATA MINING TOOLS

The best available proprietary and open source data mining tools were chosen and analytical study was made by taking into account technical specifications and features

**Table-1 : Technical Overview proprietary data mining tools**

Sno.	Tool Name	Technologies/Algorithms	Language	Features	Limitations
1.	Sisense [19]	<ul style="list-style-type: none"> <li>In-Chip analytics (In Memory Technology)</li> <li>Prism 10X</li> <li>Crowd Accelerated BI</li> <li>Sisense Pulse</li> </ul>	<ul style="list-style-type: none"> <li>ETL</li> <li>Python</li> <li>R</li> </ul>	<ul style="list-style-type: none"> <li>Data Size limitations removed</li> <li>Unifying Data</li> <li>Advanced analytics</li> <li>Wide range of widgets such as gauges, charts and graphs</li> <li>Complex business queries without programming or SQL writing</li> </ul>	<ul style="list-style-type: none"> <li>Size of Sisense application is very heavy</li> <li>Sisense’s dashboards only interact on the web</li> <li>Navigation and filtering on mobile platform should be improved.</li> <li>Most of researchers are not familiar with it.s</li> </ul>
2.	SSDT (SQL Server Data Tool) [20]	<ul style="list-style-type: none"> <li>SQL Server database technology</li> <li>Azure SQL databases</li> <li>Analysis Services (AS) data models</li> <li>Integration Services (IS) packages</li> <li>Reporting Services (RS)</li> </ul>	<ul style="list-style-type: none"> <li>SQL</li> </ul>	<ul style="list-style-type: none"> <li>Can create offline Database</li> <li>Analysis Services (AS) data models</li> <li>Integration Services (IS) packages</li> <li>Reporting Services (RS) reports</li> </ul>	<ul style="list-style-type: none"> <li>Required knowledge of database system.</li> <li>DAC (distributed client-server architecture) framework compatibility issues</li> </ul>
3.	Oracle Data Mining [21]	<ul style="list-style-type: none"> <li>Naive Bayes</li> <li>Support Vector Machine</li> <li>Decision Tree</li> <li>Enhanced K-Means</li> <li>Orthogonal Partitioning Clustering</li> <li>Expectation Maximization</li> <li>Apriori</li> <li>Non-negative Matrix Factorization</li> <li>Principal Components Analysis (PCA)</li> <li>Singular Vector Decomposition</li> </ul>	<ul style="list-style-type: none"> <li>PL/SQL</li> <li>Java APIs</li> <li>SQL</li> </ul>	<ul style="list-style-type: none"> <li>Association Model Aggregation Metrics</li> <li>Support for Explicit Semantic Analysis Algorithm</li> <li>Support for Partitioned Models</li> </ul>	<ul style="list-style-type: none"> <li>Required knowledge of database system.</li> <li>Navigation system should be improved</li> </ul>
4.	IBM Cognos [22]	<ul style="list-style-type: none"> <li>Correlation</li> <li>decision-tree algorithms</li> <li>predictive algorithms</li> <li>machine learning</li> <li>natural language processing</li> </ul>	<ul style="list-style-type: none"> <li>Java</li> </ul>	<ul style="list-style-type: none"> <li>Visualize organizational performance</li> <li>Protect data</li> <li>Share critical insights easily</li> <li>Automated data preparations</li> </ul>	<ul style="list-style-type: none"> <li>Its file become very large in size</li> <li>Limited data size used</li> </ul>
5.	IBM SPSS Modeler [23]	<ul style="list-style-type: none"> <li>decision trees</li> <li>neural networks and regression models</li> <li>Generative adversarial networks (GANs)</li> </ul>	<ul style="list-style-type: none"> <li>Python</li> </ul>	<ul style="list-style-type: none"> <li>Support for many data sources</li> <li>Easy model deployment</li> <li>Powerful graphics engine</li> <li>Automatic data preparation</li> </ul>	<ul style="list-style-type: none"> <li>Text Analytics node restrictions</li> <li>job will fail if the firewall blocks the connection between the two servers.</li> </ul>

					<ul style="list-style-type: none"> <li>Navigation system should be improved</li> </ul>
6.	SAS Data Mining [24]	<ul style="list-style-type: none"> <li>cutting-edge algorithms</li> <li>Clustering</li> <li>Anomaly detection</li> <li>Association rule learning</li> <li>Principal component analysis</li> <li>Affinity grouping</li> <li>Regression</li> <li>Neural networks</li> <li>Decision trees</li> <li>Support vector machines</li> </ul>	<ul style="list-style-type: none"> <li>C</li> <li>Python</li> <li>R</li> </ul>	<ul style="list-style-type: none"> <li>Easy-to-use GUI and batch processing</li> <li>Model comparisons, reporting and management.</li> <li>Sophisticated data preparation, summarization and exploration.</li> <li>Automated scoring</li> </ul>	<ul style="list-style-type: none"> <li>Lack of graphic representation</li> <li>Difficult Text Mining</li> <li>Difficult than R</li> <li>Costly</li> </ul>
7.	Tera Data [25]	<ul style="list-style-type: none"> <li>Cutting-edge technology</li> <li>Machine Learning Algorithms</li> </ul>	<ul style="list-style-type: none"> <li>SQL</li> <li>R</li> <li>Python with Scala</li> <li>Go</li> <li>JavaScript</li> </ul>	<ul style="list-style-type: none"> <li>Unlimited Parallelism</li> <li>Shared Nothing Architecture</li> <li>Linear Scalability</li> <li>Robust Utilities</li> </ul>	<ul style="list-style-type: none"> <li>Multiloading utility limitation</li> <li>Lack of graphic representation</li> </ul>
8.	Board [26]	<ul style="list-style-type: none"> <li>cutting-edge technology</li> <li>in-memory technology called HBMP (Hybrid Bitwise Memory Pattern)</li> <li>combination of HTML5 technology with the "Toolkit"</li> </ul>	<ul style="list-style-type: none"> <li>Python</li> </ul>	<ul style="list-style-type: none"> <li>Multi-language</li> <li>Single logical view of corporate data</li> <li>Data Connectivity</li> <li>Granular security</li> <li>Server Clustering</li> </ul>	<ul style="list-style-type: none"> <li>Limited reporting capability</li> <li>Big size</li> <li>Data limitation</li> </ul>
9.	Dundas BI [27]	<ul style="list-style-type: none"> <li>HTML5 web technology</li> <li>server-based technology</li> <li>Microsoft Technology Integration</li> </ul>	<ul style="list-style-type: none"> <li>ETL</li> <li>Python</li> <li>R</li> <li>.Net</li> </ul>	<ul style="list-style-type: none"> <li>Natural Language User Input</li> <li>Table Relationships</li> <li>Multi-Tenancy Improvements</li> <li>Heat Map Chart</li> </ul>	<ul style="list-style-type: none"> <li>Not flexible with 3D charts</li> <li>Require large server RAM for memory processing</li> <li>Does not support text analytics</li> <li>.</li> </ul>
10.	Rapid Miner [28]	<ul style="list-style-type: none"> <li>machine learning algorithms</li> <li>classification and clustering algorithms</li> <li>k-Means Clustering</li> </ul>	<ul style="list-style-type: none"> <li>Language Independent</li> </ul>	<ul style="list-style-type: none"> <li>More than 20 new functions for analysis and data handling, including multiple new aggregation functions</li> <li>File operators to operate directly from Rapid Miner</li> <li>A macro viewer that shows macros and their values in real time during process execution</li> <li>Intuitive GUI</li> </ul>	<ul style="list-style-type: none"> <li>Requires best knowledge of database management.</li> </ul>
11.	Orange [29]	<ul style="list-style-type: none"> <li>machine learning algorithms</li> <li>classification and regression algorithms</li> <li>K-nearest neighbor (KNN)</li> <li>Random Forest</li> <li>Neural Network</li> <li>Naïve Bayes and CN2 Rule Inducer</li> </ul>	<ul style="list-style-type: none"> <li>Python</li> <li>C++</li> <li>C</li> </ul>	<ul style="list-style-type: none"> <li>Visual Programming, Visualization, Interaction and Data Analytics</li> <li>Large toolbox, Scripting interface</li> <li>Extendable Documentation</li> </ul>	<ul style="list-style-type: none"> <li>Limited reporting capability</li> <li>Very big installation.</li> </ul>
12.	Weka [30]	<ul style="list-style-type: none"> <li>machine learning algorithms</li> <li>classification/regression</li> <li>Clustering</li> <li>searching Algorithms</li> </ul>	<ul style="list-style-type: none"> <li>Java</li> </ul>	<ul style="list-style-type: none"> <li>Forty-nine data preprocessing tools, seventy-six classification/regression algorithms, eight clustering algorithms, fifteen attribute/subset evaluators, ten search algorithms for feature selection.</li> <li>three algorithms for finding association rules</li> <li>Good documentations- a whole textbook being devoted to it.</li> </ul>	<ul style="list-style-type: none"> <li>Weak classical statics</li> <li>Weak CSV reader.</li> </ul>
13.	KNIME [31]	<ul style="list-style-type: none"> <li>machine learning algorithms</li> <li>Tree Ensemble</li> <li>Decision Tree</li> </ul>	<ul style="list-style-type: none"> <li>Java</li> </ul>	<ul style="list-style-type: none"> <li>Scalability, Intuitive user interface, High extensibility</li> <li>well-defined API for plugin extensions</li> </ul>	<ul style="list-style-type: none"> <li>Poor parameter optimization</li> <li>Limited error measurements</li> </ul>

		<ul style="list-style-type: none"> <li>• Random Forest, and Naïve Bayes learner and predictor nodes</li> </ul>		<ul style="list-style-type: none"> <li>• sophisticated data handling, intelligent automatic caching of data, Data visualization</li> <li>• Import/export of workflows, Parallel execution on multi-core systems</li> <li>• Command line version for "headless", "batch executions", Hitting,</li> </ul>	<ul style="list-style-type: none"> <li>• no wrapper methods for descriptor selection</li> </ul>
14.	Apache Mahout[32]	<ul style="list-style-type: none"> <li>• Clustering, classification</li> <li>• batch based collaborative filtering,</li> </ul>	<ul style="list-style-type: none"> <li>• Java</li> <li>• Scala</li> </ul>	<ul style="list-style-type: none"> <li>• Extensive programming environment</li> <li>• Pre-made algorithms</li> <li>• Math experimentation environment</li> <li>•</li> </ul>	<ul style="list-style-type: none"> <li>• Limited Algorithms</li> <li>• User interface should be improved</li> </ul>
15.	Rattle[33]	<ul style="list-style-type: none"> <li>• Decision trees, boosting</li> <li>• random forests</li> <li>• support vector machines</li> <li>• generalised linear models</li> <li>• neural networks</li> </ul>	<ul style="list-style-type: none"> <li>• R</li> </ul>	<ul style="list-style-type: none"> <li>• Graphical user interface</li> <li>• Data Visualization</li> <li>• Statistical power of R</li> </ul>	<ul style="list-style-type: none"> <li>• Limited data sources used</li> <li>• Generate duplicate code for any activity in GUI</li> </ul>
16.	Data Melt [34]	<ul style="list-style-type: none"> <li>• Mathematical and statistical algorithms</li> <li>• Linear regression, curve fitting</li> <li>• cluster analysis, neural networks</li> <li>• fuzzy algorithms</li> <li>• analytic calculations and interactive visualisations using 2D/3D plots</li> <li>• histograms</li> </ul>	<ul style="list-style-type: none"> <li>• Python</li> <li>• Ruby</li> <li>• Groovy</li> <li>• java</li> </ul>	<ul style="list-style-type: none"> <li>• Numeric computation, statistics, analysis of large data volumes ("big data") and scientific visualization.</li> <li>• computational platform</li> <li>• creates high-quality vector-graphics images</li> </ul>	<ul style="list-style-type: none"> <li>• Data limitations</li> <li>• Limited algorithms and formulas used</li> </ul>
17.	KEEL [35]	<ul style="list-style-type: none"> <li>• Evolutionary algorithms</li> <li>• extraction algorithms</li> <li>• Computational Intelligence based learning algorithms,</li> </ul>	<ul style="list-style-type: none"> <li>• Java</li> </ul>	<ul style="list-style-type: none"> <li>• Classification Discovery</li> <li>• Cluster Discovery</li> <li>• Regression Discovery</li> <li>• Association Discovery</li> <li>• Data Visualization</li> <li>• Discovery Visualization</li> <li>• a user-friendly graphical interface</li> <li>• evolutionary learning</li> </ul>	<ul style="list-style-type: none"> <li>• Limited Algorithms</li> </ul>
18.	R[36]	<ul style="list-style-type: none"> <li>• Machine learning algorithms, Naïve Bayes algorithm, confusion matrix</li> </ul>	<ul style="list-style-type: none"> <li>• C</li> <li>• Fortran</li> <li>• R</li> </ul>	<ul style="list-style-type: none"> <li>• Data Exploration</li> <li>• Outlier detection</li> <li>• Clustering</li> <li>• Text Mining</li> <li>• Time Series Analysis</li> <li>• Social Network Analysis</li> <li>• Parallel Computing, Graphics</li> <li>• Visualization of geo spatial data</li> <li>• Web Application Big data</li> <li>• Data and error handling</li> </ul>	<ul style="list-style-type: none"> <li>• Requires Knowledge of array language</li> <li>• Less specialized for Data Mining.</li> </ul>

Table-1 represents the technical overview of the proprietary and open source data mining tools which includes name of tool and description of technologies and algorithms, language.

## VI. ANALYSIS

This paper examined some best proprietary and open source data mining tools. Proprietary tools are marketed and protected by registered trade and open source tools are freely accessed by the user. Every tool has their own features and limitations.

Sisense is a business intelligence tool. It processes data small to large organizations. Non-technical user easily operates it. But very less number of users used it. SSDT (SQL Server Data tool) is declarative model that is used for expand the database phases. It has capability to maintain, debug and refactor database. Basic database management knowledge required for SSDT. Oracle data mining is also based on database. It is used for better prediction, identify cross-selling. User can easily interact with it just because of it has drag and drop feature. IBM Cognos is Intelligence suite for reporting and data analysis. IBM SPSS Modeler is data mining and text analysis tool for built predictive model. It has a visual interface that allow user to work with data mining. SAS Data Mining is for statistical analysis. It helps user to make timely decision. Tera Data is used for business analysis. Its working on share nothing architecture its server node has their own memory and processing ability. Board is best tool for companies looking for improvement in decision making. Its interface is more attractive and comprehensive interface as compare to all BI software. Dundas BI is reporting and analytics tool. It can create attractive tables and charts. It has gap free protection of data.

Rapid Miner and Orange tools are considered appropriate for advance users with knowledge of science because additional skills are required. There are very limited visualizations support present in these tools. Weka is having many inbuilt features and there is no programming knowledge required. It is auto parameterized and have very good documentations. KNIME is highly recommendable for the users who are highly skilled. It is very robust with inbuilt features and can obtain third party library functions. Apache Mahout is for machine learning algorithms. It can perform mathematical functions. Rattle is very user friendly it has well developed UI. Data Melt has interactive framework for data analysis. It is a multiprogramming utility. KEEL is software tool to access evolutionary algorithms for data mining

problems. It is restricted by the number of algorithms. R is well-designed statistical data mining tool. It is less specialized for data mining.

## VII. CONCLUSION

There are many proprietary and open source data mining tools available. In this paper different types of data mining tools are explained. They all have their features and limitations. Most of the tools provides nice graphical user interface, focus on usability, interactivity and security of data. Data visualization is great feature of software. Statistical and mathematical functions are key features of some tools. User can use the software as per their requirement. There are many proprietary and open source tools available. According to user needs and requirements they can used them. Increase of data mining tools is still challenges for select best tool for specific use. For future work makes data mining tools more efficient, interactive, user friendly, consume less space and complete work in short time.

## VIII. REFERENCES

- [1]. Kalpana Rangra, Dr. K. L. Bansal "Comparative Study of Data Mining Tools", International Journal of Advanced Research in Computer Science and Software Engineering, (June 2014) ; Volume 4, Issue 6
- [2]. San Jose, CA Cisco "The zettabyte era: trends and analysis." Cisco., Visual Networking Index. (2017).
- [3]. Reinsel, D., Gantz, J., & Rydning, J. "Data age 2025: The evolution of data to life-critical". Framingham, MA: International Data Corporation, (2017).
- [4]. Cheung, M., Dharmasthira, Y., Eschinger, C., Anderson, R. P., Granetto, B. F., Adams, A., ... Roth, C. "Forecast: Enterprise software markets, worldwide", 2013–2020. Stamford, CT: Gartner RAS Core Research Notes Gartner, (2016).
- [5]. Ravi, K., & Ravi "A survey on opinion mining and sentiment analysis: Tasks, approaches and

- applications". Knowledge-Based Systems, (2015);89, 14–46.
- [6]. Scott, J." Social network analysis. Thousand Oaks", CA: SAGE,(2017).
- [7]. Paul Y., Kumar N. (2020)" A Comparative Study of Famous Classification Techniques and Data Mining Tools." In: Singh P., Kar A., Singh Y., Kolekar M., Tanwar S. (eds) Proceedings of ICRIC (2019);Lecture Notes in Electrical Engineering, vol 597. Springer, Cham
- [8]. Myint Myint Than" A COMPARATIVE STUDY OF PREDICTING TEACHING SCORE BY USING CLASSIFICATION ALGORITHMS" Lecturer, University of Computer Studies, Kalay, Burma , (2019); IJARIE-ISSN(O)-2395-4396, Vol-5 Issue-4
- [9]. Feng, Junshuai, "Predicting Students' Academic Performance with Decision Tree and Neural Network" (2019); Electronic Theses and Dissertations. 6301. Spring Term 2019
- [10]. Sibel Barin Özkan, Sultan Muhammed Fatih Apaydin, Yasin Özkan, Irem Düzdar" Comparison of Open Source Data Mining Tools: Naive Bayes Algorithm" , Scientific Meeting on Electrical-Electronics & Biomedical Engineering and Computer Science (EBBT), IEEE (2019).
- [11]. Hansheng Lei , Mahmoud Quweider et al. "Mining Survey Data" 2nd International Conference on Data Intelligence and Security (ICDIS), IEEE (2019)
- [12]. <https://www.softwaretestinghelp.com/data-mining-tools/>
- [13]. Andreas Bartschat ,Markus Reischl, Ralf Mikut. " Data mining tools", Institute for Automation and Applied Informatics, Karlsruhe Institute of Technology, EggensteinLeopoldshafen, Germany, (4 January 2019)
- [14]. Jovic, K. Brkic and N. Bogunovic. "An overview of free software tools for general data mining", Faculty of Electrical Engineering and Computing, University of Zagreb / Department of Electronics, Microelectronics, Computer and Intelligent Systems, Unska 3, 10 000 Zagreb, Croatia, (May 2014); 26-30
- [15]. Mrs. G. SangeethaLakshmi1, Ms. M. Jayashree2, " Comparative Analysis of Various Tools for Data Mining and Big Data Mining" , International Research Journal of Engineering and Technology (IRJET), (Apr 2019); e-ISSN: 2395-0056, p-ISSN: 2395-0072, Volume: 06 Issue: 04
- [16]. Aderibigbe Israel Adekitan\*, Jeremiah Abolade and Olamilekan Shobayo "Data mining approach for predicting the daily Internet data traffic of a smart university", Adekitan et al. J Big Data (2019); 6:11
- [17]. Graham J Williams "Rattle: A Data Mining GUI for R" The Research Journal · (December 2009)
- [18]. Maqsud S. Kukasvadiya, Dr.Nidhi H. Divecha,"Analysis of Data Using Data Mining tool", IJEDR (2017); Volume 5, Issue 2 | ISSN: 2321-9939.
- [19]. <https://www.sisense.com>
- [20]. <https://docs.microsoft.com/en-us/sql/ssdt/sql-server-data-tools?view=sql-server-ver15>
- [21]. <https://www.oracle.com/database/technologies/advanced-analytics/odm.html>
- [22]. <https://www.ibm.com/products/software>
- [23]. <https://www.ibm.com/products/spss-modeler>
- [24]. [https://www.sas.com/en\\_in/software/enterprise-miner.html](https://www.sas.com/en_in/software/enterprise-miner.html)
- [25]. [www.teradata.com](http://www.teradata.com)
- [26]. [www.board.com](http://www.board.com)
- [27]. <https://www.dundas.com/dundas-bi/>
- [28]. <https://rapidminer.com/>
- [29]. <https://orange.biolab.si/>
- [30]. [www.cs.waikato.ac.nz/~ml/weka](http://www.cs.waikato.ac.nz/~ml/weka)
- [31]. [www.knime.org](http://www.knime.org)
- [32]. [mahout.apache.org](http://mahout.apache.org)
- [33]. [rattle.togaware.com](http://rattle.togaware.com)
- [34]. <https://jwork.org/dmelt/>
- [35]. <https://sci2s.ugr.es/keel/download.php>
- [36]. <https://www.r-project.org>

**Cite this article as :**

Sonia Rani Chowdhary, Vikash, "A Comprehensive Analysis of Proprietary and Open Source Data Mining Tools", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6 Issue 2, pp. 414-420, March-April 2020. Available at doi : <https://doi.org/10.32628/CSEIT206210>  
Journal URL : <http://ijsrcseit.com/CSEIT206210>