

# Comparative Study of Commercial Data Mining Tools

Rupali Raut

Arti Nathe

**Abstract** -Nowadays the information technology has been growing rapidly and adopting the various applications. Due to this the progress in commercial sector is increases. Data Mining is widely used in diverse areas. There are number of commercial data mining system available today yet there are many challenges in this field. In this paper we will focus on theoretical study of various commercial data mining tools and its applications.

**Keywords:** - Commercial data mining, Data, Data Mining, Data Mining Tools.

## I. INTRODUCTION

Data mining is the process of designing to explore the data from a large amount of data. Typically it is used in business areas. Data mining helps to search consistent data or systematic relationship between entities. Data mining is the process of searching the useful related information from the large data sets. Data mining is used in the commercial business oriented application and it is also used in machine learning, artificial intelligence and statistical data mining. Data mining is the important task to discover the knowledge, future prediction and it becomes important to the today's competitive world.

## II. OVERVIEW OF COMMERCIAL DATA MINING TOOLS

In this section we discuss the theoretical approach of commercial data mining tools. Within the data mining there is many commercial tools that have been developed by many researchers for data analysis. This tool used in commercial business oriented application. Following are the some commercial data mining tools:

**A. Angoss Knowledge STUDIO:** KnowledgeSTUDIO is a predictive analysis software suite in data mining use for the development and deployment of business application. It contains data archeology, data visualization, predictive system, decision tree analysis submission. Deployment, scoring, data validation, monitoring and build out the score card [3].it is a data mining tool developed by the Angoss Software Corporation. main office of Angoss Software Corporation in Toronto, Canada, Ontario, with various branch offices situated in the United States and United Kingdom, is a supplier and service provider of predictive analytics systems software's[4]. Angoss Software Corporation was founded in 1984. Angoss' customers is a representative of insurance

sector, mutual funds, retail business, medical sciences, telecommunication and technology industries[1].

**B. Clarabridge:** Clarabridge is an American software company established in 2005 in Reston, Virginia, United States[6]. There are two products of Clarabridge. They are Clarabridge Enterprise and Clarabridge Professional, that are known as Saas.

The products of Clarabridge are used in sentiment and text analytics to automatically collect, categorize and report on structured and unstructured data[4]. These data comes from social media such as Face book, Twitter, call center notes, email, chats, and surveys[5].

**C. HP Vertica Analytics Platform:** Vertica Analytics Platform is the data mining software provided by HP. It is an analytic database management system software company. Vertica was established in 2005 by database researcher Michael Stone braker, and Andrew Palmer[7]. The assets of HP Software used by enterprise companies and the public sector areas. The cluster-oriented Vertica Analytics software suite is developed to manage large, fast-growing amount of data and provide very quick query performance when used for data warehouses and other database applications[10]. This software improves query evaluation over traditional relational database management system[11].

**D. IBM SPSS Modeler:** It is a data mining software provided by IBM. It is used to developed predictive analysis models and other analytic tasks[12]. This software has a Graphical User Interface(GUI) which allows users to easily grasp statistical and data mining algorithms without writing programs[13]. The original name of IBM SPSS Modeler is Clementine. SPSS Modeler has two separate software suit.

- **SPSS Modeler Professional** is used for structured data, such as databases, mainframe data systems, flat files.
- **SPSS Modeler Premium** contains all the features of Modeler Professional, and also Text analytics, Entity analytics, Social network analysis.

**E. KXEN Modeler:** KXEN was an American software company established in June 1998 by Roger Haddad and Michel Bera[17]. Predictive analytics software suite was developed by KXEN[16]. This software suite used by analytic professional and commercial business areas to mining the patterns from large data sets. It is also useful for variable importance, data classification, regression, segmentation, clustering, product recommendation and a social network analysis solution[15].

**F. Grapheme:** Grapheme is a data mining and visualization software provided by iChrome. iChrome is a United Kingdom. software and engineering company established in 2009[20]. iChrome has two main software programs

called as Nexus and Grapheme. Nexus is typically used for process integration. Grapheme contains the data visualization and data analysis tools[18]. Grapheme is also useful for understanding large and complex data set. It also explore the data[19]. The Grapheme offers to design interactive 2D and 3D graphics.

**G. LION solver:** It is a collection of software application used in a data mining, business intelligence, and modeling that helps to implement the Learning and Intelligent Optimization (LION) approach[18]. LIONsolver helps to develop a model and visualize them. It also helps to improve business and engineering applications. This software is helpful for decision making [18]. LIONsolver develops from research Principles in Reactive Search in 2011.

**H. Microsoft Analysis Services:** It is a data mining software suite developed by Microsoft. It provides SQL Server Analysis Services, online analytical processing (OLAP), data mining and report generation tool [21]. It contains the number of services in SQL Server related to commercial business intelligence and data warehousing [22]. The services are Integration Services and Analysis Services. Microsoft releases this software in 2010.

**I. NetOwl:** NetOwl is a software suit developed by private (a division of SRA international). The first NetOwl product was released in 1996[25]. This software suit offer to analyze the Big data and convert it into the textual data formats – reports, web, data on social media, as well as structured patterns information about people, organizations, places, and other things[24].

**J. Oracle Data Mining:** Oracle Data Mining (ODM) is a data mining software developed by Oracle. It is an alternative for Relational Database Management System (RDBMS) Enterprise Edition[27]. It creates many data mining and data analysis algorithms for data classification, feature prediction, regression, relation between the entities, feature selection, outlier data detection, feature mining, and specialized analytics[28]. It helps to create, manage, operate and implementation of data mining application in business. Oracle Data Mining software provides Oracle SQL functions which helps to the data into the database properly[28].

**K. SAS Enterprise Miner:** it is mining software suit developed by the SAS Institute. SAS means Statistical Analysis System. It offers a predictive analytics, business intelligence, and data management [29]. This software suit was developed at North Carolina State University from 1966 to 1976. SAS is useful to extract, alter, manage and access data from a various data sources and perform numerical analysis on it. it provides a Graphical User Interface (GUI) environment to user[30].

**L. STATISTICA Data Miner:** it is data mining software suit provided by StatSoft released in 2013[33]. STATISTICA contain a statistics, data analysis, data management, , data mining, and data visualization tools which is help to analyze the data from large data set[31]. It

provides the clustering, classification and predictive analysis [31].

**M. Qlucore Omics Explorer:** It is a data mining software package provided by Qlucore. Qlucore is a software firm situated in Lund, Sweden [33]. It provides bioinformatics software products for life science which is helpful for biotech industrial areas. Qlucore provides the Graphical User Interface (GUI) environment to users which is helps to visualize the real time data and easier for biomedical scientists to analyze the data[34]. It provides the 2D and 3D environment that helps to visualize the data in 2D and 3D format.

### III. COMPARATIVE STUDY OF TOOLS

There are number of commercial data mining tools available some of them were chosen and analytical study was made by taking into account technical specifications and feature.

Table 1: Comparative study of commercial data mining tools

Sr. No.	Name	Company Name	Released in	License Type	Data mining Approach	Platform	Web sites
1	Angoss Knowledge STUDIO	ANGOSS	1996	Commercial	Classification Discovery, Cluster Discovery, Data Visualization,	Windows XP/7(32/64 bit), Vista , Windows Server 2003-08, Unix and Linux	<a href="http://www.angoss.com/analyticssoftware/KnowledgeSTUDIO.php">http://www.angoss.com/analyticssoftware/KnowledgeSTUDIO.php</a>
2	Clarabridge Enterprise	Clarabridge	2008	Commercial	text analysis	Windows XP/7(32/64 bit), Vista , Windows Server 2003-08,	<a href="http://www.clarabridge.com">http://www.clarabridge.com</a>
3	HP Vertica Analytics Platform	HP Vertica	2011	Commercial	Clustering, Pattern Matching, statistical analysis , geospatial analysis.	Windows XP/7(32/64 bit),	<a href="http://www.vertica.com">www.vertica.com</a>
4	IBM SPSS Modeler	IBM	2010	Proprietary software	Text Analytics, Social Network Analysis, prediction analysis, statistical analysis	Windows, Linux, UNIX	<a href="http://www1.ibm.com/software/analytics/spss/products/modeler">http://www1.ibm.com/software/analytics/spss/products/modeler</a>
5	KXEN Modeler	KXEN	2009	Proprietary software	classification, regression, segmentation, time series, social network analysis	Windows, Linux, UNIX	<a href="http://www.kxen.com">www.kxen.com</a>
6	Grapheme	iChrome	2011	Proprietary commercial software	data visualization, statistical analysis,	Linux, Windows	<a href="http://www.ichrome.com/grapheme">www.ichrome.com/grapheme</a>
7	LIONsolver	LIONsolver	2011	Proprietary commercial software, free for academic use	data visualization, statistical analysis	Unix and Mac OS X operating systems, windows	<a href="http://lionoso.com">http://lionoso.com</a>
8	Microsoft Analysis Services	Microsoft SQL Server Analysis Services	2010	Microsoft EULA	Data analysis and aggregations	Windows XP/7(32/64 bit), Vista , Windows Server 2003-08, Unix and Linux	<a href="http://technet.microsoft.com/en-us/sqlserver/cc10300.aspx">http://technet.microsoft.com/en-us/sqlserver/cc10300.aspx</a>
9	NetOwl	NetOwl	1996	Proprietary commercial software	Text analytics and entity analytics	Windows XP/7(32/64 bit), Vista , Windows Server 2003-08	<a href="http://www.netowl.com">http://www.netowl.com</a>
10	Oracle	Oracle	2009	Commercial	classification,	Windows	<a href="http://www.oracle.com">http://www.oracle.com</a>

	Data Mining	Corporation			prediction, regression, associations, feature selection, anomaly detection, feature extraction	XP/7(32/64 bit), Vista , Windows Server 2003-08, Unix and Linux	
11	SAS Enterprise Miner	SAS Institute	2004	Proprietary	predictive analytics, statistical analysis	Windows, IBM mainframe , Unix/Linux, OpenVMS Alpha	www.sas.com
12	STATISTICA Data Miner	StatSoft	1995	Proprietary software	statistical analysis, data visualization, clustering, classification and predictive analysis.	Windows XP/7(32/64 bit), Vista , Windows Server 2003-08	www.statsoft.com
13	Qlucore Omics Explorer	Qlucore	2007	Proprietary software	data visualization	Windows XP/7(32/64 bit), Vista , Windows Server 2003-08	www.qlucore.com

#### IV. CONCLUSION

Data mining is the process of extracting the meaningful pattern from large data set. Today's there is many commercial data mining tools for extracting the information. Many Researcher work on these tools to enhanced environment along with more technical improvements. These tools provide the interactive Graphical environment and flexibility for visual programming. In this paper, we will discuss the specific details along with description of Data mining is the process of extracting the meaningful pattern from large data set. Today's there is many commercial data mining tools for extracting the information. Many Researcher work on these tools to enhanced environment along with more technical improvements. These tools provide the interactive Graphical environment and flexibility for visual programming.

#### V. REFERENCES

1. By Dan Primack, CNN Money. "Private equity deals." June 12, 2013. Retrieved June 17, 2013.
2. COMSOL ONLINE - ANGOSS - Knowledge Engineering
3. <http://www.angoss.com/wpcontent/uploads/2011/08/KnowledgeSTUDIO.pdf>
4. <http://www.angoss.com/>
5. "clarabridge - United States Patent and Trademark Office Search Results". Search.usa.gov. Retrieved 8 December 2014.
6. "Start-Up". Washingtonpost.com. Retrieved 8 December 2014.
7. <http://www.clarabridge.com/>
8. www.vertica.com
9. Network World staff: "New database company raises funds, nabs ex-Oracle bigwigs", LinuxWorld, February 14, 2007
10. Brodtkin, J: "10 enterprise software companies to watch", Network World, April 11, 2007
11. HP News Release: "HP to Acquire Vertica: Customers Can Analyze Massive Amounts of Big Data at Speed and Scale" Feb. 2011
12. HP News Release: "HP Completes Acquisition of Vertica Systems, Inc." March 22, 2011.
13. Delen, Dursun (2009); Predicting Movie Box-Office Receipts Using SPSS Clementine Data Mining Software, in Nisbet, Robert; Elder, John; Miner, Gary (2009). Handbook of Statistical Analysis and Data Mining Applications. Elsevier. pp. 391–415. ISBN 978-0-12-374765-5.
14. <http://www01.ibm.com/software/analytics/spss/products/modeler>
15. "SAP Extends the Power of Predictive Analytics to Unlock Big Data With Acquisition of KXEN". September 10, 2013
16. Chris Kanaracus, September 10, 2013, PC World. "SAP buying KXEN for predictive analytics".
17. www.kxen.com

18. Battiti, Roberto; Mauro Brunato (2014). The LION way. Machine Learning plus Intelligent Optimization. Trento, Italy: LIONlab, University of Trento. ISBN 978-14-960340-2-1.
19. Battiti, Roberto; Mauro Brunato; Franco Mascia (2008). Reactive Search and Intelligent Optimization. Springer Verlag. ISBN 978-0-387-09623-0.
20. <http://ichrome.com/solutions/nexus>
21. "MS SQL Server 7.0 OLAP Services".
22. <http://technet.microsoft.com/enus/sqlserver/cc510300.aspx>
23. Indurkha, Nitin et al. (2005). Text Mining: Predictive Methods for Analyzing Unstructured Information, New York: Springer, p. 154-155, ISBN 0-387-95433-3. Retrieved 2013-07-02.
24. Jackson, Peter; Moulinier, Isabelle (2002). Natural Language Processing for Online Applications: Text Retrieval, Extraction, and Categorization, Philadelphia: John Benjamins B.V., p. 117, ISBN 90-272-4989-X. Retrieved 2013-07-02.
25. <http://www.netowl.com>
26. <https://www.netowl.com/entity-extraction/>
27. <http://www.Oracle Data Mining>
28. T. H. Davenport, Competing on Analytics, Harvard Business Review, January 2006
29. Arthur Li (10 April 2013). Handbook of SAS DATA Step Programming. CRC Press. p. 149. ISBN 978-1-4665-5238-8.
30. SAS Institute Inc. and World Programming Limited (England and Wales High Court (Chancery Division) July 23, 2010).
31. Christian H. Weiss "Commercial meets Open Source: Tuning STATISTICA with R" R-Project. March 2008.
32. [www.statsoft.com](http://www.statsoft.com)
33. BBC: Read about how Qlucore Omics Explorer is used at King's College in London to analyze gene expression data from worms. <http://news.bbc.co.uk/2/hi/technology/8386807.stm>
34. SP2 Magazine: Read about how Qlucore solution enables new working models for biologists and statisticians. Page 33 and onwards. [http://www.labnews.co.uk/laboratory\\_article.php/4795/5/getting-through-the-data-maze](http://www.labnews.co.uk/laboratory_article.php/4795/5/getting-through-the-data-maze).
35. [www.qlucore.com](http://www.qlucore.com)
36. Rabee M. Reffat, John S. Gero and Wei Peng, "Using data mining on building maintenance during the building life cycle" Proceedings of the 38th Australian & New Zealand Architectural Science Association (ANZASCA) Conference, School of Architecture, University of Tasmania, Tasmania, Australia, pp. 91-97.
37. Mihai ANDRONIE, Daniel CRI AN, "Commercially Available Data Mining Tools Used In The Economic Environment", Database Systems Journal vol. I, no. 2/2010

## AUTHOR'S PROFILE



**Rupali Raut** received the Master in Computer Application(MCA) from P. G. Dept. of Comp. Sci. and technology, DCPE, HVPM Amravati. Currently she is working as a Assit. Professor at MCA dept. at P. G. Dept. of Comp. Sci. and Technology, DCPE, Hanuman Vyayam Prasarak Mandal Amravati. She is having 2 years teaching experience and specialization is in software development as well as in data mining. She has developed a software 'Online Feedback System'. She get awarded by the 'Gold Medal' and 'Rajguru Award' for getting first merit in MCA 2012 at P. G. Dept. of Comp. Sci. and Tech. D.C.P.E. HVPM Amravati.



**Arti Nathe** received the Master in Computer Application(MCA) from sant Gadge Baba Amravati university Amravati. She is currently working as a Assit. Professor at MCA dept. at P. G. Dept. of Comp. Sci. and Technology, DCPE, Hanuman Vyayam Prasarak Mandal Amravati. She has 2.5 teaching experience and field of specialization is data mining.