

A Concise Study about Data Mining Methods, Tools and Its Trendy Applications

C.Thangamalar¹, R.Ramya², R.Lavanya³

¹Head of Department,
Dept of Computer Science,
Rajagiri Dawood Batcha college of arts and science,
Papanasam, Tamilnadu, India.
ctm2015.raja@gmail.com.

²Research Scholar,
Dept of Computer Science,
Rajagiri Dawood Batcha college of arts and science,
Papanasam, Tamilnadu, India.
ramyaecs40@gmail.com.

³Research Scholar,
Dept of Computer Science,
Rajagiri Dawood Batcha college of arts and science,
Papanasam, Tamilnadu, India.
lavanyaneha.91@gmail.com.

Abstract - Data mining is an emerging topic of the computer science research in recent years, and it has extensive applications in various fields such as Bio informatics, Criminal investigation, Research analysis, Corporate surveillance, Manufacturing engineering, Web and Semantic web. It uses machine learning, statistical and visualization techniques to discovery and present knowledge in a form, which is easily and is well explicable to humans. In today's ferociously competitive business surroundings, corporations to speedily flip these zettabytes of information into vital insights into their customers and markets to guide their promoting, investment and management ways. We are concentrating many taking survey for many data mining 102 software tools because however current information contains most structured and unstructured data, spatial data.etc., that it becomes almost not possible to manually analyze them for valuable decision-making information .Automated discovery tools have the aptitude to investigate the information and gift the extracted high level info to the analyst or decision-maker. This research paper focuses on data mining techniques, software tools, and most trendy applications.

Keywords: survey, data mining, software tools, trendy applications.

1. Introduction

Data mining refers to the " semiautomatic exploration and analysis of large quantities of data in order to discover meaningful patterns and rules"^[1].It is extraction of novel, understandable and potentially useful patterns implicitly stored in large databases, data warehouses and other massive information repositories comprising textual, numerical, graphical, spatial data. As data are being accumulated continuously as well as rapidly whether it is a research field, education sector in universities , market prediction, health and emergency care unit in hospitals, E-government, Social analytics, entertainment etc. it is very tricky to get more

rapidly and proper information by traditional manual analysis which is tedious as well as very cumbersome.

Data mining is used basically i) to cut costs through proper detection and prevention of waste and fraud, ii) obtain suitable and up-to-date information and iii) increased revenues through improved marketing strategy. Data mining finds suitable patterns and subtle relationships in data which are stored in databases, data warehouses, or other information repositories providing more advanced and effective information which is augmented by utilizing models equipped with sophisticated techniques. In other way, Data Mining allows to explore massive data in such a way that its result is obtaining of profitable knowledge extraction using data mining techniques

and now days we are taking information in terabytes level using many applications. It is a multidisciplinary field, drawing work from various areas like i) Artificial intelligence ii) Neural networks, iii) Machine learning, iv) Database technology, v) statistics, vi) pattern recognition, vii) signal processing, viii) spatial data analysis, ix) Business, x) Economics, xi) Bio-informatics^[2]. The major related fields include statistics, machine learning, artificial intelligence, databases, knowledge acquisition, pattern recognition, information retrieval, visualization, intelligent agents for distributed and multimedia environments, digital libraries, and management information systems etc. Data mining can be done in different types of data but one of the purposeful facts is that it can process spatial data which is used for geographical, chip design, medical and satellite image databases. According to [Berry et al, 2000] the goal in directed data mining is to use the available data set to build a model that describes one particular variable of interest in terms of the rest of the available data. The authors also point out that directed data mining often takes the form of predictive modeling, where one knows what he want to predict^[3]. Classification, prediction and estimation are the techniques used in directed data mining. In undirected data mining, no variable is singled out as the target. The goal is to establish some relationship among all the variables. Examples of this type include clustering, association rules, description and visualization. Each year more operations are being computerized, all accumulate data on operations, activities and performance. All these data hold valuable information, e.g., trends and patterns, which could be used to improve business decisions and optimize success. It contains maturation of the technology in recent years, however has expanded its capabilities and stature as a vital tool in the many trendy applications: a) Featured Expanded analytics b) User-friendly Interface c) Powerful algorithm that allows researchers to analyze structured and unstructured data. The extensive use of handheld, wireless, and other universal devices is a developing area, since a lot of knowledge in sequence being created and transmit would be maintain and stored only on these kinds of devices. Among the other novel exploration using data mining which are being developed, investigated, and applications identified for include Subject-based data mining, Music data mining, Knowledge grid and Sensor data mining.

2. Research Domain and Related Works

Data mining is used to extract implicit and previously unknown information from data. Data mining is the process which provides a concept to attract attention of users due to high availability of huge amount of data and need to convert such data into useful information. So, many people use the term “knowledge discovery in databases” or KDD for data mining.

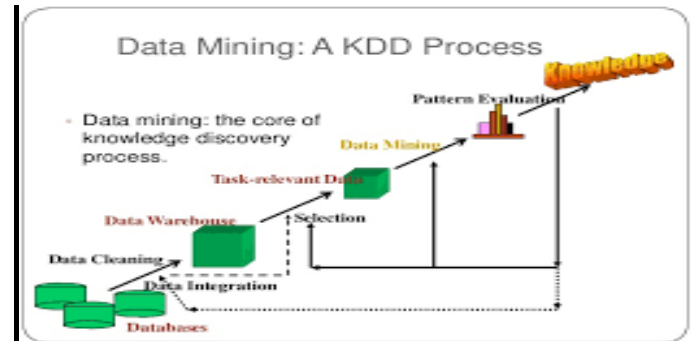


Fig 1: Schematic Diagram in KDD Process.

- 1) **Data cleaning:** Data cleaning is most essential in data mining process to remove noise, irrelevant and redundancy data from collected raw data.
- 2) **Data integration:** To combining multiple data sources into single data store called target data.
- 3) **Data Selection:** After data integration, data relevant to analysis task are retrieved from database as pre-processed data.
- 4) **Data transformation:** Data transformation is the process of converting data or information from one format to another, usually from the format of a source system into the required format of a new destination system. We are using transform the useful patterns or information using summarizing, binning, regression and clustering also.
- 5) **Data Mining:** To various smart techniques and tools are applied in order to extract data pattern or rules.
- 6) **Pattern evaluation:** To strictly identify tree patterns representing knowledge.
- 7) **Knowledge representation:** This is the last stage in which, visualization and knowledge representation techniques are used to help users to understand and interpret the data mining knowledge or result. The goal of knowledge discovery and data mining process is to find the patterns that are hidden among the huge set of data and interpret useful knowledge and information.

3. Data Mining Software Tools

Data Mining is the computational process of predictive corrected patterns in large data sets involving methods using

many valuable techniques decision tree, Bayesian classifier, Artificial intelligence, with the goal to extract information from a data set and transform it into an understandable structure for further use. This paper reviews 102 tools are used how to use data mining tools, functions, techniques, platforms, supporting files, applications. There are many categorizations of data mining software tools. Classification software, Social Network Analysis, Link Analysis, and Visualization software, Statistical Analysis software, Text Analysis, Text Mining, and Information Retrieval (IR), Visualization software, Web Analytics and Social Media Analytics software, Web Usage Mining: click stream and log analysis^[5]. We are taking this survey paper for analyzing many software tools name, general characteristics, data mining tasks and its methods using applications and availability links also in particular categorization for data mining tools. Refer in the table 1,2,3,4.

Table 1: Clustering Software.

Table 2: Statistical analytics Software.

Table 3: Social and Link analysis Software.

Table 4: Web analytics and Social Software.

4. Trends In Data Mining Applications

Data mining approaches pose many tricky research issues in data mining, the development of functional and capable data mining methods and systems, the production of interactive and included data mining environments, and the applications of data mining to resolve large applications problems are significant tasks for data mining researches and data mining system and application developers. [Mannila et al, 2001] gives three general categorizes of data mining namely; Exploratory Data Analysis (EDA), Descriptive Modeling and Predictive Modeling. With EDA the goal is to explore the data without any idea of what one is looking for and typical techniques are interaction and visualization^[4]. A descriptive model presents the main features of the data. It is essentially a summary of the data permitting the study of the important aspects of the data. Clustering techniques are used in this category. In contrast, a predictive model has the specific objective of allowing one to predict the value of some target characteristic of an object on the basis of observed values of other characteristics of the object. This category includes techniques such as Classification and regression. We can then say that directed data mining, supervised learning and predictive modeling of data mining describe similar techniques that can be referred to as

supervised learning techniques. Unsupervised learning techniques, undirected data mining and descriptive modeling are techniques in the same category and will be referred to as unsupervised learning Data mining is becoming gradually more in both the private and public sectors. Industries such as banking for using transaction facilities, insurance sector, bioinformatics and retailing commonly use data mining to low costs, enrich quality in research sector, and enlarge sales. In the public sector, data mining applications primarily were used as a means to detect fraud and waste, but have grown to also be used for purposes such as measuring and improving program performance.

4.1 Bibliomining -Data Mining in Library and Information Services

In Bibliomining plays a vital role within many universities as repositories and providers of information resources. It obtains novel information resources [Ex: Books, Databases, Reference Tools, Journals, Electronic access, etc.] can be easily retrieved. Information exposed through the application of bibliomining techniques gives the major advantages in library save money, provide more appropriate programs and serve as a more extracting helpful knowledge mining for its users. Bibliomining can afford the data-based justifications for the not easy decisions and support requests library managers must make. Multidimensional analysis Tool for Online Analytical Processing in COGNOS would allow library manages to discover their traditional frequency based data in new ways by finding at statistics along easily changeable dimensions. The major tasks in bibliomining, Reference desk and Online Public Access Catalog [OPAC] serves as searching for works owned by the library. In addition, this information could be provided to the OPAC to allow users to be seeing similar works using data mining methods. Market Basket Analysis in data mining techniques can provide the same function by examining circulation histories to locate related works.

4.2 Targeted Marketing

Target marketing model helps to determine a guideline of selling methods. Selling and buying strategy for a bank ought to target those valuable customers and maintains a honest relationship so as to get larger revenue. The bank will position the target customer's precisely and implement numerous promoting methods to satisfy consumer's want identical ought to be done in corporations that area unit in retail and life

insurance quality industries. In now-a-days classification based techniques are available in handy and very extremely popular in target marketing. Example of targeted marketing is Bank direct marketing based on neural network used to data mining predictive applications like the decision tree model [C5.0] and the Multilayer Perception Neural Network [MLPNN] with back propagation are applied on the bank direct marketing. Decision tree model [C5.0] classifier it is selected because it gives more accuracy and optimizing output with comparatively high speed. Combining two methods MLPNN and C5.0 models on a real world data of bank deposit subscription.

4.3 Health Care Applications and DNA Analysis

A novel application in data processing and knowledge extraction tools may be subjected in active and innovative research. One motivation behind the fostering applications in trendy biological applications can be increasing day by day life time. In data mining health care applications and DNA analysis used to advanced predictive analysis of diabetic treatment through the regression-based data mining techniques and embracing to find the sequence finding, protein function domain detection, protein function domain detection, protein sub-cellular location prediction etc. The Oracle Data Miner (ODM) was implementing as a software mining tool for predicting diabetes. The support vector machine algorithm was used for experimental analysis. In advanced techniques large-scale RNA-seq data analysis, log-note analysis of community service program brokers, and anonymized medical record developing by researchers. In addition to Electronic Health Records (EHR), health insurance claims, medical imaging databases, disease registries, spontaneous reporting sites. In data mining techniques are illustrated as Statistical Analysis is used to Characterization of healthcare data, Text Mining is used to mining free text in electronic medical records, Visual Analysis is used to Meaningful use of healthcare data for improved patient care and cost-reduction, Data quality assessment and improvement: preprocessing and cleaning are used to missing data treatment, Pattern Detection is used to Hypothesis generation from observational data and Visualization are predict to Prescriptions drugs and interactions. Data cleaning based on mathematical morphology S.Tang^[7]. In the field of bioinformatics and medical image understanding, data noise frequently occurs and deteriorates the classification performance; therefore on effective data

cleansing mechanism in the training data is often regarded as one of the major steps in the real world inductive learning applications. In this paper the related work on dealing with data noise is firstly revived, and then based on the principle of mathematic morphology, the morphological data cleansing algorithms are proposed and two concrete morphological algorithms. Data mining has been used extensively in the medical industry. Data visualization and artificial neural networks are especially important areas of data mining applicable in the medical field.

For example, Neuro Medical Systems used neural networks to perform a Pap smear diagnostic aid. Vysis Company uses neural networks to perform protein analyses for drug development. The University of Rochester Cancer Center and the Oxford Transplant Center use Knowledge Seeker, a decision tree-based technology, to help with their research in oncology.

4.4 Educational Data Mining

Educational mining is a rising discipline involved with developing strategies for exploring the distinctive forms of knowledge that come from academic settings and exploitation those ways to higher understandable to students and also the settings during which they learn. Educational mining is taken for student use of Interactive learning Environments, Computer-supported, Co-operative learning with neural networks techniques, or administrative data from faculties and universities. It typically has multiple levels of purposeful hierarchy, which often need to be determined by properties in the data itself, rather than in advance^[7]. In generic frameworks and methods to develop tools, frameworks, and its method, algorithms, approaches and so forth. Education process mining is used to extract process related knowledge from event logs recorded by the educational systems. In these mining educational data has capable of mining assessments data, mining browsing or interaction data mining becomes the result of educational research. Educational Data Mining researchers are addressing questions of cognition, metacognition, motivation, affect, language, social discourse, etc. using data from intelligent tutoring systems, massive open online courses, educational games and simulations, and discussion forums. Examples of Educational Data Mining Tools: EPRules are used to discover prediction rules to provide feedback for courseware authors; GISMO is used to visualize what is happening in

distance learning classes, TADA-ED is used to help teachers to identify relevant patterns in students' online exercises; LOCO-Analyst is used to provide teachers with feedback on the learning process, MMT tool is used to facilitate the execution of all the steps in the data mining process of Moodle data for newcomers. SNAPP is used to visualize the evolution of participant relationships within discussions forums. AAT is used to access and analyze students' behavior data in learning systems. DRAL is used to discover relevant e-activities for learners. E-learning Web Miner is used to discover student's behavior profiles and models about how they work in virtual courses.

4.5 Earthquake Prediction

Hierarchical and non hierarchical method used include in this research. The problem of earthquake prediction is based on data extraction of precursory phenomena and it is very challenging task but data mining tools and method used. **J-S. Lai , F.Tsai** ^[8] verification and risk assessment for land slide in the Shimen Reservoir watershed of Taiwan using spatial data analysis and data mining In this study, eleven factors were considered, including elevation (Digital Elevation Model, DEM), slope, aspect, curvature, NDVI (Normalized Difference Vegetation Index), fault, geology, soil, land use, river and road and find the result accuracy and kappa coefficient. This paper presented decision tree algorithm, spatial data, and mechanism for filtering uncertain data from the data sets to improve the reliability of landslide predictions and given the accuracy and kappa coefficient in verification can reach 98.1 % and 0.8829. and given the more reliable result in the study site. In our life we are focused on the spatial data and geographical data mining such as scientific data.

5. Conclusion

We conclude that Data Mining and its current technologies tend to develop towards expert orientation, and knowledge discovery applications development is a problem-oriented domain. Finally, the capability to constantly change and obtain new thoughtful ideas is the power of data mining technologies and will be the application of future works. Future research in this area includes using data mining techniques to extract the essential task for all consecutive applications.

References

- Michael J. A. Berry, Gordon S. Linoff, Data mining Techniques: For marketing, sales and customer support, Wiley Publications, 1997.
- Amit Kumar Patnaik, Biswaranjan Nayak, Dr. Srinivas Prasad," Data Mining and Its Current Research Directions". Information Sciences, Volume181- 7, 1264-1284, 2011.
- Berry and Gordon S. Linoff , Michael J.A, Mastering Data Mining: The Art and Science of Customer Relationship Management , Wiley Computer,USA,2000.
- [4] Mannila , David Hand, Heikki Mannila and Padhraic Smyth,Principles of data mining, Cambridge Massachusetts, MIT Press, 2001.
- [5] KDNuggets, Software Suites for Data Mining and Knowledge Discovery,2002 (<http://www.kdnuggets.com>)
- [6] Sheng TANG and Si-ping CHEN "Data Cleansing Based on Mathematic Morphology" Bioinformatics and Biomedical Engineering, 2008.ICBBE 2008. The 2nd International Conference on May(IEEE-EXplore) 2008 755 - 758
- [7] Kenneth R. Koedinger,Sidney D'Mello, Elizabeth A. McLaughlin,Zachary A. Pardos, Carolyn P. Rose,"Data mining and education,"Wiley Interdisciplinary Reviews: Cognitive Science, Volume 6, Issue 4, pp. 333-353,2015.
- [8] J-S. Lai, F. Tsai ,” Verification and risk assessment for landslides in the shimen reservoir watershed of Taiwan using spatial analysis and data mining”, International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XXXIX-B2, 2012.
- [9] Navita Kumari , “Business Intelligence In A Nutshell”, International Journal of Innovative Research in Computer and Communication Engineering Volume 1, Issue 4, June 2013, pp.969-975.
- [10] Abdul-Aziz Rashid Al-Azmi, “Data, Text, And Web Mining forBusiness Intelligence: A Survey”, International Journal ofData Mining & Knowledge Management Process Volume 3,Issue.2, March 2013, pp.1-21.
- [11] Jayanthi Ranjan, “Business Intelligence: Concepts, Components,Techniques and Benefits”, Journal of Theoretical and AppliedInformation Volume 9, Issue 1, November 2009, pp. 60 - 70.
- [12] Vishal Gupta, Gurpreet S. Lehal, “A Survey of text miningtechniques and applications”, Journal Of Emerging Technologies In Web Intelligence, Volume 1, Issue 1, August 2009, pp. 60-76.
- [13] Divya Nasa , “ Text Mining Techniques - A Survey ”, International Journal of Advanced Research in Computer Science and Software Engineering , Volume 2, Issue 4, April 2012, pp.50-54.

- [14] Muhammad Obeidat, Sarah North, Vebol Rattanak, "BusinessIntelligence Domain and Beyond", Universal Journal of Industrial and Business Management, Volume 2, Issue 6, August 2014, pp. 127-134.
- [15] S. T. Dumais, "Latent Semantic Analysis", Annual Review of Information Science and Technology, Volume 38, Issue 4, 2004, pp. 189-230.
- [16] K.L.Sumathy, M.Chidambaram, "Text Mining: Concepts, Applications, Tools and Issues – An Overview", International Journal of Computer Applications, Volume 80, Issue 4, October 2013, pp.29-32.
- [17] Rajender Singh Chhillar "Extraction Transformation Loading – A Road to Data warehouse", 2nd National Conference Mathematical Techniques: Emerging Paradigms for Electronics and IT Industries, India, pp. 384-388.
- [18] Palak Gupta, Barkha Narang, "Role of Text Mining in Business Intelligence", Gian Jyoti E-Journal, Volume 1, Issue 2, March 2012.
- [19] Li Gao, Elizabeth Chang, and Song Han, "Powerful Tool to Expand Business Intelligence: Text Mining", Proceedings Of World Academy Of Science, Engineering And Technology, Volume 8, October 2005.
- [20] Rafi Ahmad Khan, Dr. S.M. K. Quadri, "Business Intelligence: An Integrated Approach", Business Intelligence Journal, Volume 5, Issue 1, January 2012, pp.64-70.
- [21] Knowledge Storm, Data Mining, 2002. (<http://www.knowledgestorm.com>)
- [22] Two Crows Corporation, Introduction to Data Mining and Knowledge Discovery, 1999. (<http://www.twocrows.com>)
- [23] John F. Elder IV & Dean W. Abbott, A Comparison of Leading DataMining Tools, 1998. (<http://www.datamininglab.com>)
- [24] Margaret H. Dunham, Data Mining Introductory and Advanced Topics, Pearson Education, New Delhi, 2009.

| CLUSTERING SOFTWARE | | | | | | | |
|---------------------|---|--------------------------|---------------|----------------------------|---|--|---|
| SNO | DATA MINING TOOLS | GENERAL CHARACTERISITICS | | DATA MINING TASKS &METHODS | | USING APPLICATIONS | LINK |
| | | TARGETED PLATFORMS | LEGAL SOURCES | DISCOVERY TASKS | METHODS | | |
| 1 | Bayesia Lab | MultiPlatform | Commerical | Clustering , Segmentation | Bayesian Networks | Epidemiology, Financial Analytics | http://www.bayesia.com/ |
| 2 | Clustan Graphics3 | MultiPlatform | Commerical | Clustering | Clustering Analysis | Market analysis or Administrative purposes | http://www.clustan.com/ |
| 3 | CMSR Data Miner | MultiPlatform | Commerical | Clustering | Decision Tree Classification and Segmentation. | Time-series Forecasting and Trend Analysis | http://www.roselladb.com/starprobe.com/ |
| 4 | Cviz | MultiPlatform | | Clustering | Multiple Regression | Data Visualization and Scientific Processing | www.sai.msu.su/sal/D/1/CVIZ.com/ |
| 5 | IBM SPSS Modeller | MultiPlatform | Commerical | Clustering | Analytic Tasks | Healthcare Quality Improvement | http://www.ibm.com/analytics/us/en/technology/spss.com/ |
| 6 | NeuroXL Clusterizer | Single Platform | Commerical | Clustering | Netural networks | Stock price prediction,Sales prediction,Sports prediction | http://neuroxl.com/download.com/ |
| 7 | Persimplex | MultiPlatform | | Clustering | Fuzzy Logic | Numeric data Processing | www.perSimplex.biz |
| 8 | PolyAnalyst | MultiPlatform | Commerical | Clustering | Cluster Discovery, Classification Discovery | Text Categorization Applications | http://www.megaputer.com |
| 9 | Viscovery Explorative data mining modules | MultiPlatform | Commerical | Clustering | SOM Visualization | Applications in Healthcare and Scientific Research | https://www.viscovery.net/news |
| 10 | visipoint | MultiPlatform | Commerical | Clustering | Fuzzy Modelling, Summon Mapping | Pest Control Applications and Health and Nutrition related Applications | www.visipoint.fi |
| 11 | Autoclass C | MultiPlatform | Free | Clustering | Unsupervised Bayesian Classification | Predictive Analysis | https://ti.arc.nasa.gov/icmp/projects/bayes-group/group/autoclass/autoclass |
| 12 | Cluto | Single Platform | Free | Clustering | Graph Partitioning | Information Retrieval, Customer Purchasing Transactions | glaros.dtc.umn.edu/gkhome/cluto/cluto/overview |
| 13 | Databionic ESOM Tools | MultiPlatform | Free | Clustering | High dimensional with U-matrix ,P-matrix component planes | Scientific Research,Life Science Applications | databionic-esom.sourceforge.net |
| 14 | David DOWE Mixture Modeling Page | Single Platform | Free | Clustering | Knowledge Discovery | Artificial Intelligence applications,Machine learning | www.csse.monash.edu.au/~dld/cluster.html |
| 15 | ELKI | MultiPlatform | Free | Clustering | Knowledge Discovery | High Level Visualization to Predict Diseases | elki.dbs.ifi.lmu.de |
| 16 | MCLUST/EMCLUS | MultiPlatform | Free | Clustering | Knowledge Discovery | Character Recognition,Tissue Prediction | www.stat.washington.edu/mclust |
| 17 | MDL clustering | MultiPlatform | Free | Clustering | Unsupervised attribute ranking, Discretization and clustering | Data Exoploration applications ,Information Retrieval | www.math.uwaterloo.ca/~rwardfor/students/SteveShi/essay . |
| 18 | Snob | MultiPlatform | Free | Clustering | Classification Analysis | Healthcare mining,Sensor data mining for emergency detection, Sensor data fusion for decision support, Gene expressions and MRI Images | www.csse.monash.edu.au/~dld/Snob.html |
| 19 | Starprobe | MultiPlatform | Free | Clustering | Linear Logistic Exponential | Pattern Recognition | www.roselladb.com/starprobe1.htm |
| 20 | Grachus | MultiPlatform | Free | Clustering | Clustering Tools | Image Processing ,Pattern Recognition | www.pqquest.com/grachus-a-fast-graph-clustering-tool |

Table 1: clustering software

| STATISTICAL ANALYTICS SOFTWARE | | | | | | | |
|--------------------------------|----------------------------|-------------------------|---------------|----------------------------|---|--|--|
| SNO | DATA MINING TOOLS | GENERAL CHARACTERISTICS | | DATA MINING TASKS &METHODS | | USING APPLICATIONS | LINK |
| | | TARGETED PLATFORMS | LEGAL SOURCES | DISCOVERY TASKS | METHODS | | |
| 1 | Analyse-it | Multi Platform | Commerical | Statistical Analysis | Data Visualization | Applictions in Industry and Education Purpose | https://analyse-it.com |
| 2 | Appricon's Analysis Studio | Single Platform | Commerical | Statistical Analysis | Advanced Logistic Regression Modeling, Regression Modeling | Applications in Healthcare and Scientific Research | www.appricon.com |
| 3 | Data Desk | Multi Platform | Commerical | Statistical Analysis | Data Analysis | Academic Purposes | https://datadesk.com |
| 4 | DM STATware | Single Platform | Commerical | Statistical Analysis | Analysis and Mining and Modeling | Marketing Analysis | www.dmstat.com |
| 5 | JMP | Single Platform | Commerical | Statistical Analysis | Data Exploration and Analysis | Genomics,Clinical Data Analysis | http://www.jmp.com/en_us/software.html |
| 6 | Lumenaut | Single Platform | Commerical | Statistical Analysis | Monte Carlo Simulation, Decision Tree and | Finance, Operations Management, Marketing | http://www.lumenaut.com/ |
| 7 | MODSTAT | Single Platform | Commerical | Statistical Analysis | very cheap DOS-based program | Academic Purposes | http://lifestream.aol.com/ |
| 8 | PermuteIt™ | Single Platform | Commerical | Statistical Analysis | Data Exploration and Analysis | Manufacturing,Pharmaceuticals ,Medical,Healthcare | http://www.datamineit.com/DM_I_software.html |
| 9 | Palisade | Single Platform | Commerical | Statistical Analysis | New and Improved Tornado Graphs ,Faster Optimization with | Applications in Finance and Securities | http://www.palisade.com/risk/default.asp |
| 10 | R-PLUS Professional | Single Platform | Commerical | Statistical Analysis | Customized Analytics | Big Data Analytics | |
| 11 | Simsat for Windows | Single Platform | Commerical | Statistical Analysis | GLM ANOVA/ANCOVA, BOOTSTRAP RESAMPLING analysis | Educational and Multimedia Applications | https://provalisresearch.com/products/simstat/ |
| 12 | StarProbe | Single Platform | Commerical | Statistical Analysis | clustering,Radial Basis Function | Predictive Applications | www.roselladb.com/starprobe.html |
| 13 | StatCrunch | Single Platform | Commerical | Statistical Analysis | Drill Down Statistics | To analyzing tweets applications | www.statcrunch.com |
| 14 | StatEL | Single Platform | Commerical | Statistical Analysis | Data Management and Data Analysis Tool | Statistical Testing,Medical Statistical Testing and Data | www.adscience.eu |
| 15 | STATISTICA | Single Platform | Commerical | Statistical Analysis | Sentimental Analysis, Warranty Analysis | Risk Management, Churn Analysis,Text Mining | www.statsoft.com |
| 16 | Statit | Single Platform | Commerical | Statistical Analysis | Data Access,Automation,Better | Health Care, Manufacturing | www.statit.com/index.shtml |
| 17 | Educational Statistics | Single Platform | Commerical | Statistical Analysis | Data Management and Data Analysis Tool | Educational and Multimedia Applications | symynet.com/educational_software/ |
| 18 | TIBCO Spotfire S-PLUS | Single Platform | Commerical | Statistical Analysis | Database Marketing Analysis and Mining and Modeling | Big Data Analytics,Content Analytics,Predictive Analytics,Location | spotfire.tibco.com/products/s-plus/statistical_analysis_software |
| 19 | Vitalnet | Single Platform | Commerical | Statistical Analysis | Data Warehouse / Data Analysis Software | Health Datawarehouse, Health Planning, Health Statistics, Healthcare Benchmarking, Healthcare Data Analysis software, Health Data Query, | www.ehdp.com/vitalnet/ |
| 20 | WINKS SDA | Single Platform | Commerical | Statistical Analysis | T-tests (paired t test, Independent group t test/unpaired t test), ANOVA, Regression, correlation, Repeated measures, Logistic Regression, Times Series Analysis, Chi-Square, Bland-Altman, | Survival Analysis Management | www.texasoft.com |
| 21 | WIZARD | Single Platform | Commerical | Statistical Analysis | Elegant graphical results, Instantaneous statistical tests, Advanced multivariate modeling | Manufacturing,Pharmaceuticals ,Medical,Healthcare | www.wizardmac.com |
| 22 | XLSTAT | Single Platform | Commerical | Statistical Analysis | DataFlagger, MinMaxSearch, Remove | XLSTAT-Sensory,XLSTAT-Marketing | www.xlstat.com/en |
| 23 | Develve | Single Platform | Free | Statistical Analysis | Repeated measures, Logistic Regression, Times Series Analysis, Chi- | Predictive Applications | https://develve.net/ |
| 24 | Microsirris | Multi Platform | Free | Statistical Analysis | Statistical and Data Management Package | Bioinformatics Research in Universities, | http://www.microsirris.com/ |

Table 2: Statistical analytics software

| SOCIAL AND LINK ANALYSIS SOFTWARE | | | | | | | |
|-----------------------------------|---|-------------------------|---------------|--|---|---|---|
| SNO | DATA MINING TOOLS | GENERAL CHARACTERISTICS | | DATA MINING TASKS & METHODS | | USING APPLICATIONS | LINK |
| | | TARGETED PLATFORMS | LEGAL SOURCES | DISCOVERY TASKS | METHODS | | |
| 1 | Analyst's Notebook | Multi Platform | Commerical | Social Network Analysis | i2 Intelligence-Led Operations Platform, delivers rich assisted Data mining analysis | Fraud detection, Risk Management, visualization capabilities to support analysts to prevent crime and terrorism. | http://www.ibm.com/analytics/us/en/safer-planet/ |
| 2 | AdvancedMiner Social Network Analysis (SNA) | Multi Platform | Commerical | Social Network Analysis | General Data Analysis and Data Mining | Increasing churn prediction efficiency, Detecting potentially fraudulent situations | http://algolytics.com/products/advancedminer-sna/ |
| 3 | Centrifuge | Multi Platform | Commerical | Social Network Analysis | Predictive Mining | Analysts and investigators an integrated suite of capabilities that can help them rapidly understand and glean insight from new data sources. | https://www.crunchbase.com/organization/centrifuge-systems |
| 4 | IBM SPSS Clementine | Multi Platform | Commerical | Social Network Analysis | Predictive Modeling and Data Mining | Statistical analysis and reporting, Decision management and deployment, Big data analytics | http://www.ibm.com/analytics/us/en/technology/spss/ |
| 5 | Hypersoft OmniContext | Multi Platform | Commerical | Social Network Analysis and collaborative analysis | Operational and Structural analysis | The source data used to gather all of the relevant information includes all interaction and communication channels | https://hypersoft.com/operational_intelligence/Operational_intelligence.aspx |
| 6 | Idiro SNA Plus | Multi Platform | Commerical | Social Network Analysis | Attribution Modelling | Idiro uses advanced predictive modelling to help you predict the future behaviour of your customers. Use this information to improve marketing targeting. | http://idiro.com/ |
| 7 | InFlow | Single Platform | Commerical | Social Network Analysis | Information/knowledge Extraction | Services to map and measure networks in organizations, communities, and other complex | http://www.orgnet.com/ |
| 8 | KXEN Social Network Analysis (KSN) | Multi Platform | Commerical | Social Network Analysis | Predictive Modeling and Data mining | Business Intelligence | http://www.kxen.com/index.php?option=com_content&task=view&id=564&Itemid=10 |
| 9 | Linkurious | Multi Platform | Commerical | visualization software with social network anlysis | Advanced Data Analysis and Collaborative Analysis | Fraud Detection, Network Management, Healthcare , Knowledge Management Cyber security, Intelligence Analysis, Telecommunications | http://linkurio.us/ |
| 11 | MarketVisual | Multi Platform | Commerical | Social Network Analysis | Online Visual Relationship Mapping | Market Analysis | http://www.marketvisual.com/ |
| 12 | NETMAP analytics | Multi Platform | Commerical | Social Network Analysis | Network Data Mining (NDM) | Retail, Government, Commercial, Insurance, Criminal and Marketing | http://netmap.com.au/ |
| 13 | NetMiner | Multi Platform | Commerical | Social Network Analysis | Mining Module based on Machine Learning with enhanced 3D Network Map | Knowledge Management applications | http://www.netminer.com/main/main-read.do |
| 14 | Networked Insights | Multi Platform | Commerical | Social Network Analysis | Collective Intelligence and Search | Customer Sastification | http://www.networkedinsights.com/ |
| 15 | NetworkX | Multi Platform | Commerical | Social Network Analysis | Creation, Manipulation, and study of the structure, Dynamics, and | Retail, Government, Commercial, Insurance, Criminal and Marketing organisations | http://networkx.lanl.gov/ |
| 16 | NodeXL | Multi Platform | Commerical | Social Network Analysis | Social media network data streams, Advanced network metrics, and Text and Sentiment analysis, and Powerful Report | Explore Network Graphs | http://nodexl.codeplex.com/ |
| 17 | OmniAnalyser | Multi Platform | Commerical | Social Network Analysis | Analytics, Online Analytics | Business Intelligence | https://hypersoft.com/omnia_nalyser.htm |
| 18 | Sentinel Visualizer | Multi Platform | Commerical | Link analysis | Data Integration ,Data Analysis, Advanced Link Analysis, Data Visualization, Geospatial Mapping | Anti-Corruption and Bribery Gang Activity, Telephone Call Data Records. | http://www.fmsag.com/ |
| 19 | Sonamine | Multi Platform | Commerical | Link analysis | Massive Social Networks and Graphs. | Criminal and Marketing organisations etc., | http://www.sonamine.com/home/ |
| 20 | SVAT | Multi Platform | Commerical | Link analysis | Text mining with Visual Search | Loan Fraud Prevention, Organized Crime Investigation, Bitcoin Fraud | http://www.nfgsvat.com/ |

| | | | | | | | |
|----|--|-----------------|------------|--|---|--|---|
| 20 | SVAT | Multi Platform | Commerical | Link analysis | Text mining with Visual Search | Loan Fraud Prevention, Organized Crime Investigation, Bitcoin Fraud Investigation. | http://www.nfgsvat.com/ |
| 21 | VisualLinks(TM) suite | Multi Platform | Commerical | Link analysis | Data Clarity | Insurance Fraud Detection System | http://www.visualanalytics.com/ |
| 22 | GraphChi | Multi Platform | Free | Graph Visualization | Disk-based large-scale graph computation system | Social networks, Web graphs, and Protein prediction in bioinformatics interaction graphs | http://select.cs.cmu.edu/publications/paperdir/osdi2012-kyrolab-lleloch-guestrin.pdf |
| 23 | Graphviz | Single Platform | Free | Graph Visualization | Representing Structural Information | software engineering, database and web design, machine learning, and in visual interfaces | http://www.graphviz.org/ |
| 24 | JUNG(the Java Universal Network/Graph) | Multi Platform | Free | Framework | Clustering, Decomposition, Optimization | random graph generation, statistical analysis, and calculation of network distances, flows, and importance measures. | http://jung.sourceforge.net/ |
| 25 | libsna (library for Social Network Analysis) | Multi Platform | Free | Social Network Analysis | Fast Processing time with graph algorithms | Modeling, analysis, and Visualization of network data | https://www.gmw.rug.nl/~huisman/sna/software.html |
| 26 | Pajek | Single Platform | Free | Graph Visualization | Analysis and Visualization of large | Community Detection | http://mrvar.fdv.uni-lj.si/pajek/ |
| 27 | R | Single Platform | Free | Graph Visualization | Statistical Computing and Graphics | Outlier Detection, Time Series Decomposition, Forecasting Time Series Clustering and Classification | https://www.r-project.org/ |
| 28 | Social Networks Visualiser (SocNetV) | Multi Platform | Free | Social network analysis and visualization | Structural Distance and Covariance methods | random graph generation, statistical analysis, and calculation of network distances | http://socnetv.sourceforge.net |
| 29 | Subdue | Multi Platform | Free | Graph Visualization | Graph Based Hierarchical Clustering | Graph Based Knowledge discovery Applications | http://www.subdue.com |
| 30 | SNA: R Tools for Social Network Analysis | Multi Platform | Free | Social network analysis | Central Markov Chain Monte Carlo (MCMC) algorithm | Graph-level indices, , Structural Equivalence Detection, Network Regression, Random Graph Generation, and 2D/3D Network Visualization. | http://cran.us.r-project.org/web/packages/sna/index.html |
| 31 | statnet | Multi Platform | Free | Social network analysis | Random Graph Models (ergm) | Analysis, Simulation and Visualization of network data. | http://statnetproject.org/ |
| 32 | Tulip | Multi Platform | Free | visualization software with social network anlysis | Information Visualization Framework | Research Purposes | http://www.tulip-software.org/ |

Table 3: Social and Link analysis software

| WEB ANALYTICS AND SOCIAL SOFTWARE | | | | | | | |
|-----------------------------------|-------------------------------------|-------------------------|---------------|--------------------------------|---|--|---|
| SNO | DATA MINING TOOLS | GENERAL CHARACTERISTICS | | DATA MINING TASKS & METHODS | | USING APPLICATIONS | LINK |
| | | TARGETED PLATFORMS | LEGAL SOURCES | DISCOVERY TASKS | METHODS | | |
| 1 | AlterWind Log Analyzer Professional | Multi Platform | Commerical | Web log analytics | Search engine optimization (SEO) | Marketing Applications , Revelant analysis | http://www.alterwind.com/loganalyzer/log-analyzer-professional.html |
| 2 | ClickTracks | Multi Platform | Commerical | Web log analytics | Search engine optimization (SEO) | Revenue and ROI Tracking | www.clicktracks.com/downloads/ClickTracks_Profession |
| 3 | CoreMetrics | Multi Platform | Commerical | Web log analytics | Digital Analytics Method and Developing Websites via cloud Infrastructre and Big Data | IOT, security ,Data Analytics | https://www-01.ibm.com/software/info/coremetrics-software/ |
| 4 | Download Analyzer | Multi Platform | Commerical | web log analysis | Pay-Per-Click Calculator in search engine marketing in web mining | Relevant Analysis | http://www.downloadanalyzer.com/ |
| 5 | Google Analytics | Multi Platform | Commerical | web log analysis | Data Collection & Management ,Data Consolidation,Data Analytics & Reporting Cohort Analysis | Marketing Applications , Revelant Analysis | https://analytics.google.com |
| 6 | Inspectlet | Multi Platform | Commerical | web log analysis | Eye-tracking Heatmaps,Scroll Heatmaps in web mining | Record and watch everything | http://www.inspectlet.com/ |
| 7 | iWebTrack | Multi Platform | Commerical | web analytics service | Track and monitor the activities of every individual visitor to your website, web statistics, visitor system information, advertising campaigns, conversion tracking, and | Track Campaign Effectiveness, | http://www.iwebtrack.com/ |
| 8 | Kampyle | Multi Platform | Commerical | web log analysis | Feedback Analytics | Integration with CRM /analytics/Ecommerce/Ticketing | http://www.kampyle.com/ |
| 9 | Lamplight Analytics | Multi Platform | Commerical | web log analysis | Asia premier social media intelligence platform | Marketing/PR Effectiveness ,Crisis Management | https://www.lamplight.me/ |
| 10 | MetaSun | Multi Platform | Commerical | web analytics software | MetaTraffic supports the following tracking methods: JavaScript for tracking any type of HTML page.ASP Includes for ASP pages only. | Drupal Upgrades or Migrations, Cloud Hosting, Custom Drupal Module Development | http://metasun.com/services |
| 11 | NedStat | Multi Platform | Commerical | Web Log Analyzer | Website monitoring service | Crisis Management | http://www.nedstat.com/web-analytics-solutions |
| 12 | Nihuo Web Log Analyzer | Multi Platform | Commerical | Web Log Analyzer | Searched File Analysis, Spider Analysis,Bounce Rate Analysis, File time taken analysis (only available for IIS) | Increment analysis support in Marketing | http://www.loganalyzer.net/features.html |
| 13 | OneStat | Multi Platform | Commerical | Real-Time Web Analytics | Website monitoring service,Professional Web Analytics, Enterprises Web Analytics | Improve visitor experience, Online Sales and ROI of Online Marketing Campaigns | http://www.onestat.com/ |
| 14 | Omniure | Multi Platform | Commerical | Real-Time Web Analytics | Crisis Management system | Email Marketing,Social Media Marketing Applications | https://my.omniure.com |
| 15 | OpenTracker | Multi Platform | Commerical | Website & Mobile app analytics | CRM & Integration API | Real-time Reporting,IP address tracking | www.opentracker.net/ |
| 16 | PostRank | Multi Platform | Commerical | Social Media Analytics | PostRank Analytics, PostRank Data Services ,PostRank Connect | Measure Marketing, and for Strategic Planning | https://en.wikipedia.org/wiki/PostRank |
| 17 | SAS Web Analytics Solution | Multi Platform | Commerical | Real-Time Web Analytics | Optimize your website's performance | Forecasting and Trending Capabilities ,Analytical Reporting,Enhanced Online Analytics Data Model | https://support.sas.com/resources/papers/tnote/webanalytics.html |
| 18 | ShinyStat | Multi Platform | Commerical | Web Statistics Analyzer | analysis,Monitoring and ROI of advertising campaigns SERD (Search Engine Ranking Detecting) Benchmarking | Internet Audience Analysis and Website Statistics | http://www.shinystat.com/ |
| 19 | StatCounter | Multi Platform | Commerical | Web Statistics Analyzer | Browser Stats,O.S. Stats,Resolution Stats,Multiple Site Management,Blocking Cookie | Use StatCounter to monitor actual human activity in real-time ,Keyword Analysis | https://statcounter.com/ |

| | | | | | | | |
|----|-------------------------------|-----------------|------------|--|---------------------------------|--|---|
| 20 | Surf Pattern Visual Analyzer | Multi Platform | Commerical | Web Search and Web Navigation Pattern Analyzer | Web Mining | Direct Marketing, Geographic Segmentation , Healthcare Fraud detection | http://www.roselladb.com/surf-pattern-analyzer.htm |
| 21 | VisiStat | Multi Platform | Commerical | Cloud-Based Web Analytics | Web Mining | CRM Integration,Geographic listings of website traffic, Discover influencers and trends in Social Media | https://www.publicapis.com/visistat/visistat |
| 22 | WebLog Expert 2.0 for Windows | Single Platform | Commerical | WebLog Expert Analytics | Web Mining | Activity statistics: daily, by hours of the day, by days of the week, by weeks and by months. Access statistics: statistics for pages, files, images, directories, queries, view time, entry pages, exit pages, bounces, paths through the site, file types, virtual domains and load balanced servers | https://www.weblogexpert.com/ |
| 23 | WebTrends | Multi Platform | Commerical | Website Measurement & Analytics | Web Optimization | Website Analytics, Measurement, Testing | https://www.webtrends.com/ |
| 24 | 123LogAnalyzer | Multi Platform | Commerical | Website Measurement & Analytics | Analyses in Web Server logfiles | General Statitics in Web Analysis | 123loganalyzer.soft112.com |
| 25 | Analog | Multi Platform | Free | Website Measurement & Analytics | IIs Web Log Analyzer | Activity Statistics available | http://www.analog.cx |
| 26 | Htminer | Multi Platform | Free | Website Measurement & Analytics | KDD | Support analysis of Web Logs (including unique visitors, sessions, transactions) | htminer.po-net.prato.it/en |

Table : 4 Web analytics and social software