Selection Criteria for Data Mining Software: A Study

Neeraj Bhargava[1] Atif Aziz[2] Rajiv Arya[3]
[1] Associate Prof and head Deptt of Computer Science School of Engineering & System Sciences
MDS University, Ajmer, Rajasthan, India.
[2] Research Scholar NIMS University
Jaipur India
[3] Prof Mechanical Engg. Deptt. Truba Institute of Engineering & Information Technology
Bhopal, India

Abstract: Nowadays software are flooded in the market, there is a need for selection criteria of software package that should be available for individuals and organizations. As the number of software continues to grow and new features added to the new software the selection of the most suitable software package is becoming more and more difficult. Wrong or immature decision would result in great loss of time and money. Research is being carried out throughout the world to evaluate the software. Researchers have not reached to consensus to generalize such selection and evaluation criteria. Improper selection of software package can lead to be a costly affair and adversely affect the business.

Keywords: *data mining, software selection, artificial intelligence, software evaluation*

1. Introduction

The data mining software market is crowded with the development and emergence of new software. Introduction of new and upgraded tools is very exciting but it creates problems in the minds of the potential buyers to take right decision. Investment in software that is most suitable for the organization can lead to improved management and customer service. Multiple categories of selection and evaluation criteria are discussed. Selection of a software package that can meet the requirements of an organization is a very time consuming and difficult task. There is a need to study the various data mining tools available in the market so that the most appropriate tool can be selected for the organization as per the requirement. [7] The purpose of this paper is to sum up the criteria which are important in software selection and evaluation process. Data mining software is costly and generally accompanied by moderately steep learning curves.[6].Selection of wrong software package can turn out to be costly and adversely affect business process.[2]

2. Literature Review

In recent years many review papers have been published outlining the selection and evaluation criteria of software package. Most papers reviewed give generalized selection criteria of various software like accounting, DSS tool, simulation tool, ERP etc. Papers discussing criteria of selection of software of any particular field like data mining are a few in numbers. There is a little work carried out in decision making frame work. However criteria

related to performance, quality, vendor, hardware and software requirements can be evaluated for any software. The task of software package selection has become more complex due to (a) difficulty in accessing applicability of software packages to the business needs of the organization due to availability of large number of number packages in the market (b) existence of incompatibilities between various hardware and software packages in the market (c) lack of technical knowledge and experience of decision makers (d) ongoing improvements in information technology [1]. A good tool will provide meaningful diagnostics to help debug problems and improve the output. [5]

3. Software Selection Criteria Framework

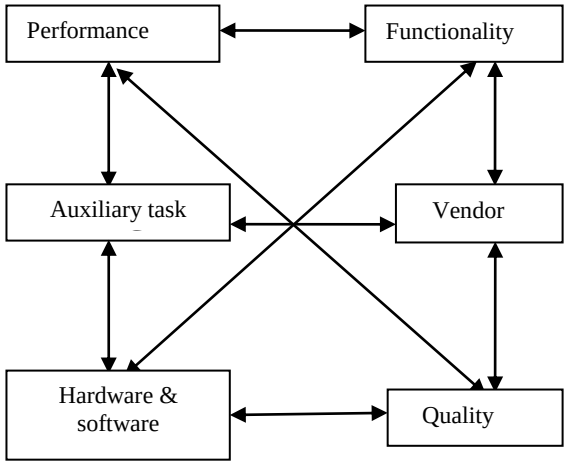

Fig.no.1

The selection of software package depends on various factors like software performance, functionality, ancillary task, and hardware and software requirements of software to run the software package efficiently, vendor responsibilities and quality or ability of software package to handle the data discrepancies. Fig no.1 depicts the relationship of various software selection criteria. All the criteria are interrelated; one cannot give high weight age to any one criteria and ignore the other criteria. For instance,

if data mining software is implemented in any engineering industry would lead to reduction of time and cost in the fields of design, manufacturing, maintenance etc. This can be possible by choosing the right data mining software analyzing and forecasting from huge amount of the data in the industry.

3.1) Performance Criteria

Sno.	Criteria	Criteria group	Criteria meaning
1	Sturdiness	Reliability	Capability: run consistently without crushing
2	Backup and recovery	Reliability	Capability: support backup and recovery feature
3	Time behavior	Efficiency	Ability: produce results in reasonable amount Relative to data size
4	Report	Output	Capability: Standard and customized report from the software package

Table No.3.1

In Table no. 1 software performance criteria is taken into consideration which depends mainly on factors like reliability, efficiency and output of the software. Any data mining software which is being implemented in any company should give consistent results for business development.

3.2) Functionality Criteria

Sno.	Criteria	Criteria meaning
1	Algorithmic variety	Availability of adequate variety of mining and algorithms available in the software and customization
2	Adaptability	Possible customization in general and for the specific company
3	Interoperability	Integration with other tools and applications
4	Security levels	Security policies supported by the software package (user identification, auditing ,data encryption)
5	Number of simultaneous users	Maximum no. of simultaneous users that can be linked and served by the system
6	Data type Flexibility	Variety of data types that are supported
7	Data sampling	Capability for random sampling of data for predictive modeling

Table No.3.2

In Table no. 2 software functionality criteria is taken into consideration which depends on adaptability and

capability of the software to customize according to the requirement of the particular organization.

3.3) Auxiliary task support

Sno.	Criteria	Criteria meaning
1	Data Cleansing	Capability to modify spurious values in the data
2	Data filtering	Capability for selection of subsets based on user defined selection criteria
3	Binning	Availability of bin of continuous data for improving efficiency
4	Record deletion	Capability to delete record which may be biased from entire population of records
5	Handling blanks	Capability to handle blanks in the entries

Table No.3

In Table no. 3 Ancillary task support criteria is taken into consideration to find software capability to handle data with discrepancies and faults like blanks .

3.4) Software Quality Characteristics

Sno	Criteria	Criteria Group	Criteria meaning
1	Vertical solution	Personalization	Customized version of software package accommodating requirements of specific industry
2	Interface type	Personalization	User Interface type of package
3	DBMS standard	Portability	Breath of database management systems that can be accessed by software package (SQL server, oracle)
4	Error reporting	Usability	Error reporting and messaging ability
5	User interface	Usability	Ease of use of user interface
6	Platform variety	Portability	Variety of platforms on which software package can be used
7	Action history	Usability	Availability modification of history of actions in data mining process
8	Technical Source	Opinion	Opinion about software package potential vendors, in house experts, external consultant

Table No.3.4

In Table no. 4 quality criteria is taken into consideration which includes software capability to modify history of actions being carried out, portability andeasily understandable GUI.

3.5) Criteria related to vendor

SNo.	Criteria	Criteria group	Criteria meaning
1	User manual & tutorial/training	Vendor	Availability of User manual for important information, tutorial for learning and troubleshooting guide
2	Maintenance and upgrading	Vendor	Up gradation and annual maintenance contract
3	Consultancy	Vendor	Availability of technical support
4	Indirect benefits	Benefits	Improvement in customer service

Table No.3.5

In Table no. 5 vender criteria is taken into consideration which includes software upgrading, training and technical support from the vendor for improved customer service.

3.6) Criteria related to hardware and Software

SNo.	Criteria	Criteria group	Criteria Meaning
1	Internal and external memory	Hardware	Primary storage and Secondary storage needed to run the package
2	Source code	Software	Availability of source code

Table No.3.6

System requirements, including supported computer platforms are often particular to a company or project.[4] In Table no. 6 hardware and software criteria is taken into consideration which includes the resources that are available with the organization and the investment required for software implementation and smooth functioning of the software package.

4. Discussion and Conclusion

Installation and Implementation of a software package in an organization to effectively increase the decision making and forecasting capabilities involves huge investment. Data mining Software package if wrongly chosen would lead to loss of money and time. This makes a requirement for companies to be aware of best suited data mining software package in order to extract knowledge as possible from databases and able to improve. The software selection criteria discussed above can be given weights by software selection experts according to the requirement of the organization. Different organizations might opt for

weights for different criteria. Secondly, during the selection of software generally the evaluator keeps a particular as his/her favorite software. In this case that particular software should be as a reference tool. In a score of 5 the reference tool is receives a score of 3 for criterion and other tools are rated against the reference tool for each criterion. WEKA toolkit is widely used toolkit for machine learning and data mining that was originally developed at the university of Waikato in New Zealand.[3][8] For further analysis WEKA software will be opted for as a reference tool.

The advantage of having generic software selection criteria would lead to reduction in time and money wastage of the experts and company as a whole. The limitations would bethat it will lead to eradication of brain storming sessions by experts which becomes the basis of new ideas and new factors of software selection criteria. The applications of this can be any industry like medical, engineering, manufacturing, pharmaceutical etc for data analysis and forecasting i.e. knowledge discovery for knowledge management.

5.References

[1] Jadhav A.S., Sonar R.M., "Framework for evaluating and selection of the software packages: A hybrid knowledge based system approach," the journal of system and software 84 2011, pp1394-1407.

[2] Jadhav A.S., Sonar R.M., "Evaluating and selecting software package: A review," Information and software technology 51, 2009, pp555-563

[3]Wahbeh A.H., Al-Radaideh Q.A., Al-Kabi M.N., Al-Shawakfa E.M., "A Comparison study between data mining tools over some classification methods," International journal of advanced computer science and applications special issue on artificial intelligence.

[4] Abbot D.W., Matkovaky I.P., Elder J. IV, "An evaluation of High end Data mining tools for fraud detection"

[5] Ogwueleka, Nonylum F. and Okeke, Knolika G. N., "Methodology and tool selection criteria in data mining" Global Advanced Research Journal of Engineering vol 16, 2012,131-137

[6] Collier K., Carey B., sautter D., Marjaneini C., A methodology for evaluating and selecting data mining software, proceedings of the 32nd Hawaii international conference on system sciences- 1999

[7] Bhargava N. , Arya R., Rajput A., Aziz A., Data Mining: A tool for knowledge management and eco-friendly sustainable development, journal of computer and information technology vol.3 pp13-18 2012

[8] Tiwari M., Jha M.B., Yadav O.P., Performance analysis of Data mining algorithms in Weka, IOSR journal of computer engineering Vol 6, no. 3, pp 32-41,2012.

First Author Dr. Neeraj Bhargava is presently working as Associate Prof and head Department of Computer Science School of Engineering & System Sciences MDS University, Ajmer, Rajasthan, India.

Second Author Mr. .Atif Aziz is presently pursuing PhD . He has authored many research articles in various reputed journals.

Third Author Dr. Rajeev Arya has been involved in various research projects.