Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 03, March -2018

APPLICABILITY OF DECISION TREE TECHNIQUE FOR DECISION SUPPORT TO ANALYZE STUDENTS PERFORMANCE THROUGH ORANGE AND RAPIDMINER TOOLS OF DATA MINING

Reena Hooda*

*Assistant Professor, Department of Computer Science & Applications,
Indira Gandhi University Meerpur (Rewari), Haryana, India.

Abstract - *The world's dynamics demand a magicalsway to wrestle all the unfavorablecircumstancesrapidly and develop a superbblueprintforcompetitive gain. This mandates an interactive tool to lookoutthe consequences of change in different input variables on the outcomes without causing the real effects just like computer games. Though magic and fairy tales are not real, however, the world is equipped with powerful tools to serve in almost every problem efficiently.Data mining is that resourceful developmentintegrating many techniques to extract, summarize, analyze &report data and moreprediction of unknown values through applicability of artificial intelligence, expertizeas well as availability of open source mining tools, for example; Orange, RapidMiner and SAS etc. The present paper aims atelaborating significance of decision support system in assessments by electing different inputs &criterion and selecting the best alternate based on givenmeasuresto support decision maker. Further paper highlights the importance of decision tree technique of data mining to represent the pictorial flow of information interactively and systematically.*

Keywords- *Orange, Decision Tree, Data Mining, RapidMiner.*

I. Introduction

As the market and environment fluctuate day by day, an adaptive intellectual decision making is must to face the prevailing risks like transformations in product design, quality standards, procedures, methodology change, customer's demand etc. This can causeimplementation failure of a plan and unattained objectives including system obsolescence results in short period success and survival. [1] The stress of being endurable in threatening competition is not the problem of businessperson; however, society is also facing it. Thanks to the technological achievements, many tools are in the market that can works as a faithful subordinate to help in taking decisions. Decision Support System is the amalgamation of computer and information to present complex, structured, unstructured or isolateddata in a way to assist decision maker in selecting alternates and to perform the tasks in better ways [2]. Now days, in globalized environment, the scenario is so multidimensional and shifting that deviations occur rapidly.Therefore, for various uncertain problems, Decision Support System a computer based system approaches the compliances to gripnumerouschanges by way of substitutes or path to go with elasticity and triumphing consistency. Decision Support System includes knowledge base, decision model, inputs, user friendly system interaction, combination of hardware & software backed by large dataand menu driven activities to help the user in getting the compiled data taken from diversesources to crack certain hitches [2] Decision Support System evolved from model oriented to function oriented as multifunctional entities to solve unorganized, unanticipated problems with less requirement of technical expertise; user can retrieve, search for options, get summary data to aid in taking more informed and educated decisions. [6] It is a tool to aid the decision maker,nonethelessit cannot replace him as Decision Support System does not workindependently andultimate responsibility rests with the user. Decision Support System provides the comparative analysis through different figures of various modules for separate time periods accessing useful information for simple to complex brainstorming problems. [2] Decision Support System is a gizmo contributing interactively to decision maker by opting alterations for given parameters, selecting number of nodes, levels and inspecting effectiveness of choices by amalgamating own knowledge & expertise to identify and resolve the issues. [7]

II. Historical Background of Decision Support System

Concept of Decision Support System evolved in 1950s and was implemented in 1960's. The web based Decision Making System grew during mid-1990. Various applications came into existence, interactive computer based system helped the decision maker to solve the problem on the basis of data available in data bases and model is created to improve the efficiency of the managers and various professional activities while trying to make intelligent knowledge base to facilitate critical reporting. In this line, data warehouse, OLAP supplemented the operational Decision Support System to make a system more intelligent through applicability of Artificial Intelligence and transforming from traditional method of analysis to contemporary data mining methods. Decision Support System can be categorized as passive system; active system and co-operative system. Passive simply delivers information but not any explicit proposal where as active offers the resolutions with the predictive information. Above all, as co-operative system that is responsible for maximum support through continuous interaction with the user, as a iterative process it accepts modifications, criterion from user, re-learn & evolve itself, then model is validated to present refined solutions to the user for an individual or whole organization in a networked system. The main components of the Decision Support System may include the user interaction, model, user criteria and data base comprising training as well as test data so-called knowledge base to validate the model, also choose the alternates then last one is the user interaction to give output. [4] [8] Steps or the process of Decision Support System may contain acceptability of input from the user interaction. Compare the keywords with the given equations, check the criteria for the conditions and search for the matching solutions, create a list of all possible alternates and presented to user. User can change the criteria, or accept or reject the solutions, after an iterative processing, last step is the presentation of the selected option by the user to implement it. Apart from the managerial preparations, Decision Support System application may include agricultural production, medical diagnosis, forest management or the railways, judiciary, predictive analysis, project management, mind storming etc. [2] Some of the open source Decision Support System are also available; for example Lumia, Decision Explorer, Open Rules, Paramount Decision, GRDSS etc [3]. Fraternity of University of Bath and Strathclyde initially consigned as developer of Decision Explorer and then Banxia Software. [4] Decision Support System available in variety of forms, some focus on data, other based on models or communication. Decision Support System support in both quality as well as quantity of information, quality includes relative features like good, bad, average, true or false, timeliness, relevancy, accuracy, simple , integrity & consistency etc. whereas quantity means something absolute and measurable. Decision Support System can be model oriented, knowledge centric, document focused to retrieve all type of data including web pages, supporting by new paradigms like Data Mining, Web Based Decision Support System, OLAP etc.

III. Significance of Decision Support System

Decision Support System benefits all the three levels of management from top to bottom serving depending upon the nature of management, inputs and user requirements. It helps in performing the various organizational functions or activities, selecting the best alternate, making provisions for uncertainties etc. The positive side of the present environment is that countless tools are in hand because of expansions in communication & information technology. Higher versions of software are capable to comply with the traditional system or data formats, processing capacity enrichments, availability of unlimited information sometimes free of cost, well-connected enterprises lifting up to achieve the peaks of new developments, new fields, new business and brand new widened market to survive for a long. The worst situation is that new era brought faithless competitions, weak enterprise bonding, heterogeneous data, pressure of expertize requirement to face complexities and take quality oriented quick decisions, overloaded competitive tasks, stressed lifestyle, uncertainty in market of sudden fall down or closing of business. Moreover, hackers playing games with security are enough to blast the decisions. [9] It is an hypothesis that more information benefit the managers to take good decisions. However, if the information is processed and comprehended properly, the same can help to take better ideas and go in to new insights. Trying new ideas for this current knowledge and yes of course, subject expertise is must so both pros & cons are there. A manager or the decision maker needs information for risk assessment, uncertainty, long-time planning, trend forecasting, potential outcomes and allocation of resources in addition to making budgets with requirements of security controls, input specifications, documentation for designing of Decision Support System. [6]

IV. DECISION TREE DEVELOPMENT IN DATA MINING

Data mining is the technology and having many tools to aid managers or the decision makers to access & retrieve information, predict the unknown circumstances and take reliable decisions. Advanced data mining techniques along

with open source applications help in report generation automatically, show the comparisons diagrammatically, iteratively check the effectiveness of change in variables, draw flow chart of all possible activities plus reducing the activities to avoid overloading. Data mining aim at exploring heterogeneous data even in case of missing values, including the major popular techniques like clustering, KNN, applicability of fuzzy logic to make intelligent human like decisions similar to the real world. Apart from coding and getting expertize various open source tools are available to help decision makers like Orange, Rapid Miner etc., to provide wizard based approach for report generation, chart designing, clustering and other tasks like ranges & magnitudes of different input variables to the user free of cost. Decision is an intellectual process that a man takes in whole life depending on facts that will be proved bad or good. Sometimes taking no decision is also a decision [11] or not opting for any decision is also a decision. The trustworthiness depends upon the accuracy or reliability of inputs, methodology used, users biasness or own perceptiveness. [11] Mining in different arenas helps to take valuable decisions through digital image, data extraction, image processing, charts, raster analysis, spatial queries and vector analysis etc. [5] Decision tree is the graphical presentation of alternates to select an outcome based on predefined conditions or the criteria. Based on example data set, decision tree is trained and tested to validate the decision tree model. For testing and validating along with the example data set, test cases are also provided and checked iteratively up to the decision maker satisfaction. The results can be compared to define the clear-cut input requirements [10] or a readymade tool can be used to draw the decision making tree to get an outcome. The user has to specify the inputs and criteria of splitting each node. The decision tree will be generated based on the knowledge base [10], leaf nodes with no further splitting showing possible outcomes. Advantage of decision tree tool is that by changing the inputs or the criteria user can watch changes in traversing the tree and possible outcomes in an attractive manner without getting bored. Further advantage is that graphical presentation shows the clear picture of the flow of alternates. Therefore, it is possible to deduce the extra overloaded nodes from the tree providing a greater flexibility to the user and take decisions wisely. The current application used a dummy database having students records for the elements; Roll No, Name, MCA Marks, Extra-Curricular Activities, Graduate Marks, Soft Skills, Presentation and Sports etc. including Attendance. The application developed in Orange tool of mining using tree viewer, data table, data sampler widgets to take the data from file, creates a tree view and display the generated tree.

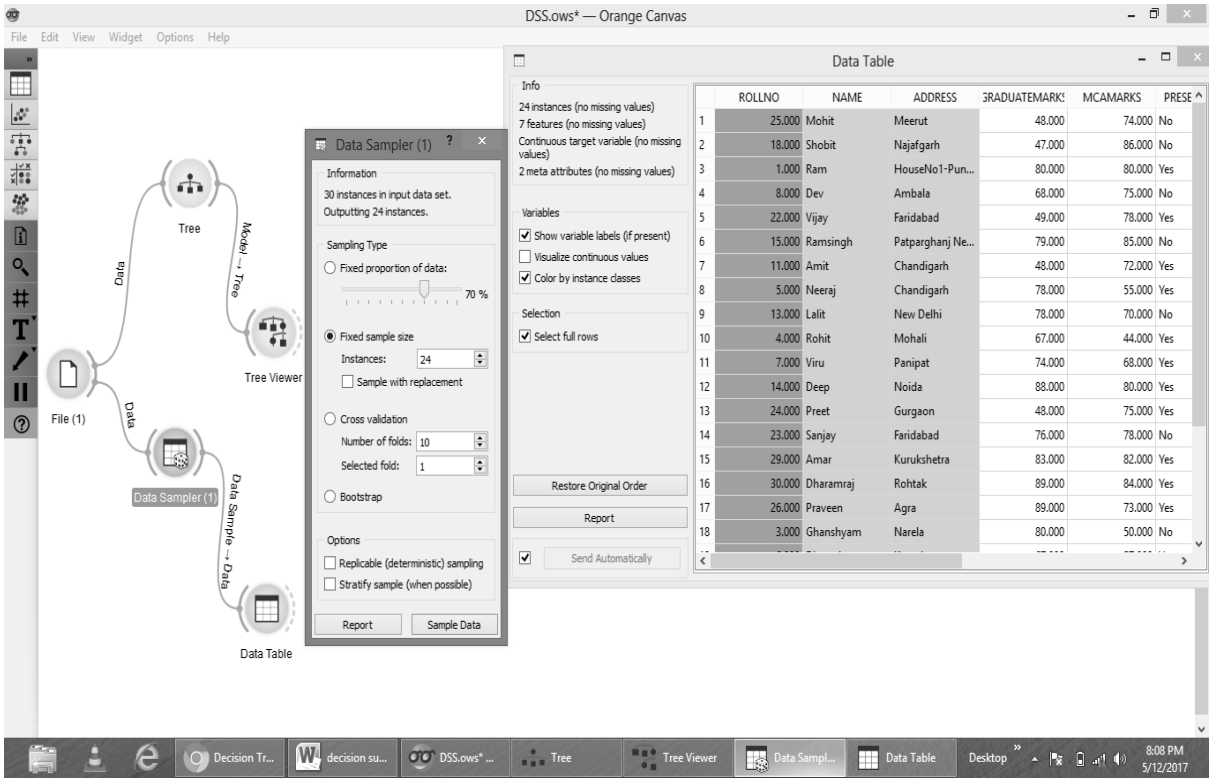
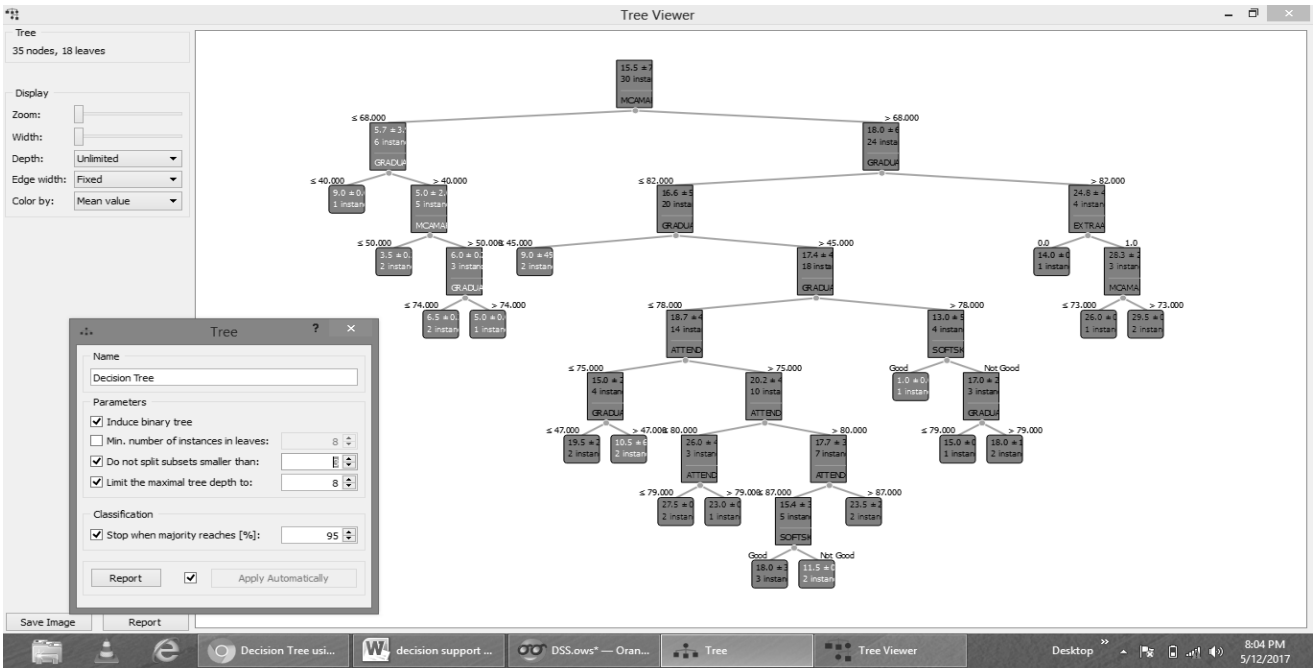


Figure 1: Shows the information for the variables, selection, number of instances, features and Meta attributes plus the data sampler information.

Different modules in Orange are the Data, Visualize, Evaluate, Data sampler, model and unsupervised widgets to solve the problems. Data sampler contain the information regarding number of input and output instances, sampling size either fixed proportion of the data or user defined instances in fixed sampling, cross validation like number of the folds and selected folds or the bootstrap plus the sampling option for deterministic sampling or the stratified sampling. Data table is

used to display the database elements and records of the students including the variable labels, if present, as shown in figure1. Figure 2 shows tree viewer to represent the flow of information through tree, having a root and leaf nodes. Flexibility is available in the Orange for the user in a form that shows the number of nodes and the leaf, width of the tree, user defined depth like number of levels user wants, edge size, color of the nodes on the bases of mean value, number of



the instances, or variances etc. plus the label is user defined.

Figure 2: Shows the tree view of the data.

As compare to Orange, decision tree development in the RapidMiner is somewhat complex and user has to take care of certain parameters. Strict input requirements to avoid the errors however great flexibility is provided in RapidMiner. The user can specify the label, can change the label, criterion, minimum leaf size, minimum size for split, maximum depth, confidence and pruning etc. during run time, tree can viewed in different ways and can be saved. User can see text view plus graphical view at run time. So run time RapidMiner provides a great flexibility to view data in circular form, tree form, tree (tight form), balloon shape or others as shown in figure 3.

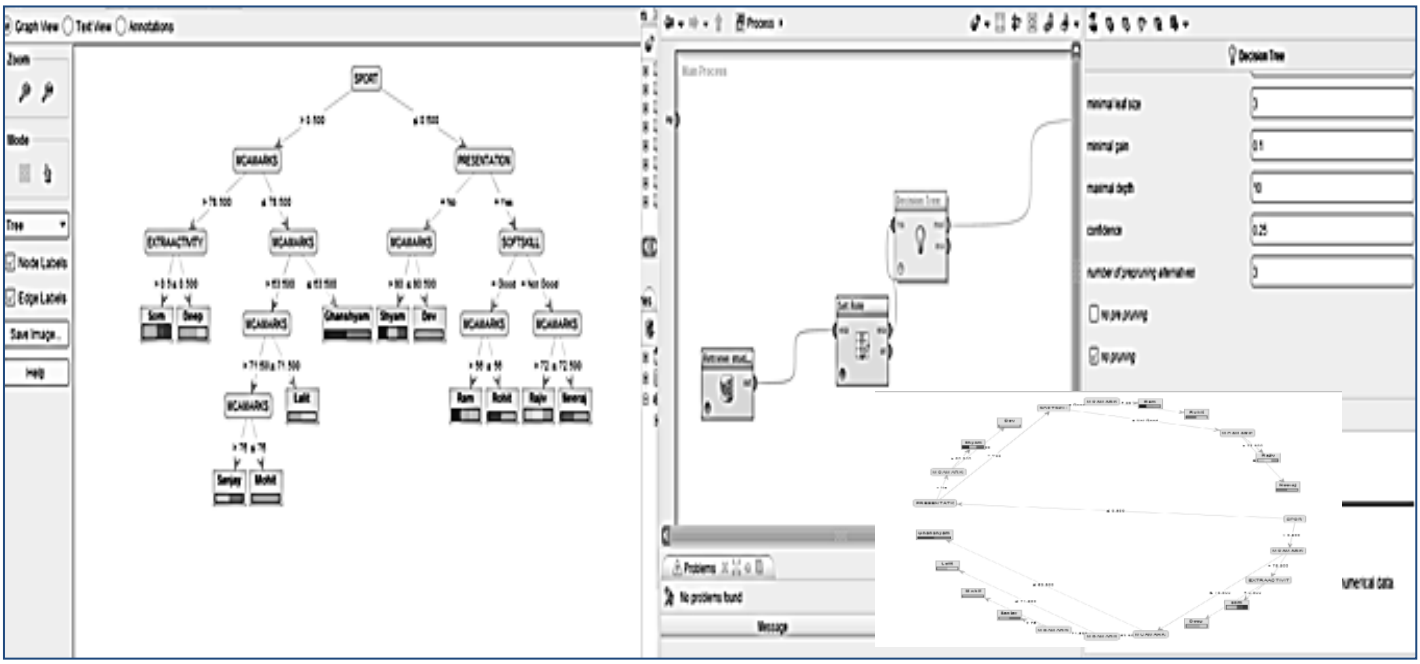


Figure 3: Shows the decision tree in RapidMiner in different formats.

Orange provides artistic effects at run time; colors of the nodes whereas Miner provides data view, meta data view, plot view, graph view and also flexibility to set the level of confidence and number of leaf nodes. Overall each software has its own unique features thus if the user has some expertize in the mining techniques for instance; if user has an idea of confidence and support in association rules of mining, pruning, clustering and prediction, distance measure, database constraints, input formats, meta data or other methodologies in addition to their functioning or the requirements, it will be more easy to work with the mining software to analyze and display the outputs. Outcomes can easily be inferred and examined in different forms. It will certainly benefit a lot in captivating speedy results supported by robust exploration through Mining tools.

V. CONCLUSION

Technology helps a lot, if used with enthusiasm. Same way, data mining is a great pool of techniques and if practiced, can helps a lot in data analysis. Moreover, it is not mandatory to develop or write codes each time. Different open source tools of data mining are available which can be installed on system and could be employed for the data analysis & reporting rapidly. Almost every type of statistical, graphical representation, model building, and predictions as well can be prepared easily. Thus data can be observed from different perceptions with extensive documentation to acquire additional tools for imminent expertize improvements. The present paper underlined the decision tree methodology to yield choices out of many alternatives and exhibit the tree using Data mining open source tools.

REFERENCES

[1] Adela Bâra and Ion Lungu []. "Improving Decision Support Systems with Data Mining Techniques". Retrieved from: https://www.intechopen.com/books/advances-in-data-mining-knowledge-discovery-and-applications/improving-decision-support-systems-with-data-mining-techniques-
[2] https://en.wikipedia.org/wiki/Decision_support_system
[3] http://www.butleranalytics.com/5-decision-support-software-free-and-open-source/
[4] http://banxia.com/dexplore/
[5] http://wgbis.ces.iisc.ernet.in/foss/index.php?option=com_content&task=view&id=17&Itemid=40
[6] Randall E. Louw [2002]. "*Decision Support Systems*". Retrieved from: http://www.umsl.edu/~sauterv/analysis/488_f02_papers/dss.html.
[7] DINESH THAKUR [ ]. "What is DSS? What is the purpose of Decision Support System in MIS". Retrieved from http://ecomputernotes.com/mis/structure-and-classification/what-is-dss-what-is-the-purpose-of-decision-support-system-in-mis
[8] http://dsssystem.blogspot.in/2010/01/components-of-decision-support-systems.html
[9] L. G. Kabari1* and E. O. Nwachukwu2 [2013]. "Decision Support System Using Decision Tree and Neural Networks ". Computer Engineering and Intelligent Systems . Retrieved from: http://www.iiste.org/Journals/index.php/CEIS/article/view/6107/6241
[10] GarimaSahu and Rakesh Kumar Khare [2014]. "Decision Tree Classification based Decision Support System for Derma Disease". Retrieved from: http://research.ijcaonline.org/volume94/number17/pxc3896171.pdf
[11] http://www.decision-making-confidence.com.

