Original Research Paper

Volume - 7 | Issue - 5 | May - 2017 | ISSN - 2249-555X | IF : 4.894 | IC Value : 79.96

Engineering

Predicting Learner Knowledge Level In The E-learning Environment

C.Rajendra	Research scholar, Reg.No PP Comp. Sci & Eng. 114C, Rayalaseema University, Kurnool, AP, India
Dr.B.Kavitha	Lecturer in Computer Applications, SVA Govt. Degree College(M) Srikalahasti,AP,India

ABSTRACT In the e-learning environment, learners have different knowledge levels, diverse states of mind about teaching and learning and different responses to particular e-learning situation and instructional materials. The more completely the substance designers comprehend the distinctions, the better possibility they have of meeting the differing adapting necessities of the greater part of their learners. The performance of the learners in e-learning environments is extremely affected by the way of the posted e-learning material and can be improved by providing appropriate learning content to the learners based on their knowledge level. Due to the complexity of the evolving paradigm, the approaching dynamics of learning requires the development of knowledge delivery and evaluation. Assessment of learners' knowledge level is important to adapt content presentation and to have more practical evaluation of online learners. This research tries to predict the learner knowledge level with the help of machine learning and user activity analysis. Several classification algorithms are applied for automatic prediction of Learner knowledge Level(LKL) and the corresponding results were posted.

KEYWORDS : Learner Knowledge Level, e-learning, Learning Management System, Learning Objects, Target Content, Allied Content, the performance of learners, e-learning environment, e-content developers.

1. INTRODUCTION

Much of the learning material available on World Wide Web or in the learning portals is in the form of digital content which is called as Learning Objects (LOs) and are particularly used for learning purpose in the educational environment. While designing the learning objects, care should be taken in identifying individuals' learning necessities. The individual needs were adapted in designing the learning objects. Nowadays it's commonly accepted that information technology has been developed rapidly and plays [3] a significant role in education. With the advent of technological up gradations, systems are capable of analyzing huge amounts of information within no time and also allow the user to analyze the results in a very effective way.

E-learning is characterized as a kind of learning on the premise of innovation. And it covers an assortment of learning techniques, such as virtual classroom, electronic learning, web-based learning, PC-based learning, portable learning, computerized joint effort etc[1]. Evolution of e-learning changes the role of the instructor from being an information producer to a supporter or a counselor [3]. Besides, students additionally look to assemble information by methods for self-learning environment, rehearse, audit, memory, experience, and [3] collaboration with peers, through social networks. Along these lines, learning profile of every learner can be used to distinguish the particular qualities of learners amid the learning procedure [3]. In addition, the Learning Management Systems(LMS) keep up information in log files and databases about learners' actions and interactions such as how much time spent on specific learning object, the frequency of visiting, exam performance etc. This information can be used to analyze the learners. Automatic learner knowledge level evaluation is a vital factor of online learning systems due to the nature of such systems, as there is no direct contact between tutors and learners. Moreover, the content presentation needs to be customized according to individual knowledge levels. Routine examinations, quizzes, and assignments may not yield the exact knowledge level of a learner in a situation where automatic evaluation is needed. This type of problems of online learners' evaluation attracts the attention to a more comprehensive evaluation scheme.

Learner modeling is a technique utilized as a part of intelligent mentoring frameworks to correspond to learner proficiencies and learning. Generally, in the classroom environment, teachers find out about students through years of experience [2] and acquire their understanding of students' learning through many ways, such as students' responses, questions and misconceptions, as well as their facial expressions and body language. Similarly, when it comes to e-learning environment, the system needs to be aware of learner's learning status as well. Learner modeling techniques were used to build inferences and predictions about learner knowledge level.

Some models evaluate learners progressively and supply knowledge to other tutor modules, especially the teaching module, to allow the system to react effectively, draw the attention of learners and encourage learning. Student modeling techniques can also be used to obtain scientific insights about student learning. Student models typically produce parameter estimates after being trained on a large amount of data. Most of those parameter estimates are semantically meaningful. They may capture impacts of some student behaviors, or can show probabilities of certain actions. Hence, parameter estimates being interpretable and plausible is fundamental, through interpreting them, researchers could understand students, such as Learner Knowledge Level(LKL), Learners' learning style, learners' learning rate, interests, preferences etc. Predicting learner knowledge level is a significant task in e-learning environment.

Learner model plays an important role in classification and to decision-making in an e-learning environment. The model being used ought to have the capacity to precisely predict the learner knowledge level. This research paper focuses on predicting Learner Knowledge Level using the parameters Scale of Time spent on Target Content (TTC), Scale of Total number of Page Visits(TPV), Scale of Time spent on Allied Content (TAC), Scale of Test Performance on Target Content(PTC), and Scale of test Performance on Allied Content(PAC) obtained from the model *"Predicting Learner Knowledge Level in the e-Learning environment"*.

Following section 2 describes the role of data mining in education, section 3 presents related work, section 4 deals with data description and experimental design, section 4.1 contains a detailed discussion on results analysis and finally section 4.6 gives conclusions.

2. EDUCATIONAL DATA MINING

Many researchers have conducted comprehensive studies of educational mining on various e-learning frameworks, information resources. A study of various data mining methods utilized proficiently and precisely to record learner activities were talked about in [3].

Table 1. List of investigations to Identify learning styles

Research	Data Features	Mining Technique	Mining Purpose
Lo and Shu P[7].	E-learning systems.	Neural Networks(NN)	Kinestethic, Visual Auditory.
Chen C-H. et al[8].	E-learning systems.	Association rule(AR).	Mining learner
Mohammed Amine limam, Hamid Seghiouer, Yasyn	E-learning systems.	Decision Trees(DT).	Building profiles based on ontology for career.
Villaverde J.E in 2006 [9]	Simulated Data.	Neural Networks(NN)	detection of learning style in e learning - Felder Silverman.
Cha H. in 2006. [10]	E-learning systems.	Decision Trees.	Felder Silverman.
Garcia P. in 2007. [11]	E-learning systems.	Bayesian Networks.[2]	Felder Silverman.
Garcia E.2007. [2]	Learning Management Systems	ARM	Identifying student interactions .
Graf, S. and Kinshuk[2].	Learning Management Systems	Rule Based(RB).	Identifying student interactions on
Ozpolat E. and Akar G.B.[2]	E-learning systems	NBtrees	identifying learning styles - Felder Silverman[2].
Ahmad N. and Shamsuddin S. in 2010. [2]	Learning Management Systems.	Rough Sets.	Classification of learning style – IFS.

Table 1 tabulates a few researches that explore on identifying learners' learning style [2]. Different procedures, methodologies, and reasons for distinguishing learning styles were employed to figure out the appropriate classifier for each contextual investigation. Some of the most widely used methods are factual investigation, Neural Networks(NN), Decision Trees(DT), Bayesian Networks(BN), Genetic Algorithms(GA), Rough Sets and Association Rule Mining(ARM). [2] Recent studies on identifying learner's characteristics have concentrated primarily on understanding various factors pertaining to the learner's interaction with the e-learning systems [2]. Learning Management System(LMS) is gaining importance in creating learner oriented e-content, because of adaptability and robustness of LMS. Most of the LMSs store the interaction information of students through inbuilt log records. The information put away in the log records can without much of a stretch be utilized to examine the learner behavior in e-learning.

3. RELATED WORK

Problem-based Learning (PBL)[4], is learner centric and focuses on the learning process of the learner. PBL focuses on the learning procedure instead of the tutor who gives information to the learners. The essence of PBL becomes to build the learning object bank in the self-learning environment [4]. The recommendation with Association rule and Collaborative filtration affiliation rule was mainly followed to find out the association between the key terms which freshmen used for looking the content material.[4] And, the collaborative filtration [4] turns into carried out to automatically clear out the right keywords of each path. With such mechanisms, they may significantly gain novices on searching what substances they desired and accurately.

Personalized Recommendation: Due to the fast expansion of the internet, resource advice mechanism is extensively employed in e-business. The same techniques were also adapted to e-Learning [1]. In personalized recommendation, three aspects of personalization can be [1] revealed the presentation sequence of learning objects, personalized presentation and gaining knowledge of an individual. Primarily by comparing the learner model and the classification of resources, resource recommendation mechanism pushes the correct resources to each learner after comparing the learner model and classification of learning objects Furthermore, the learners' preferences will be varying over a period [1]. So the recommendation strategies could take into consideration, the time of historical records. Owing to the repeatability and periodicity of learning process, there are some dependence relationships among learners' historical access records which can re- flect resource access patterns and learner's latent preference information. Unfortunately, existing content- filtering-based and collaborative-filtering-based recommendation systems always neglect these useful information [2]

4. DATA DESCRIPTION AND EXPERIMENTAL DESIGN

To predict the learner knowledge level, a sample of 500 students was taken under study. These students have gone through the learning objects placed in the Learning management system *"Predicting Knowledge level of learner in the e-learning environment"* as shown in figure 1.

Figure 1. Tool to get learner activity data

The LMS *"Predicting Knowledge level of learner in the e-learning environment"* was designed for data collection of this research work. Data collected from the LMS weblog was used for experimentation and analysis. Mentor as admin has the right to provide logins to learners and only authorized learners can use LMS. Target content, allied content and also question bank and all Learning Objects were linked to the LMS. Before taking the test, the learner must go through the target content and allied content. The input parameters, the amount of time spent on the target content, time spent on the allied content, the total number of page visits were recorded into the log file. From the tests taken by the learners, test performance on target content and on allied content was evaluated and stored in the log file. The data values from the log file were given to human domain experts to analyze the knowledge level of learners as {Beginner, Intermediate, Expert, Advanced}. These values were populated to output parameter Learner Knowledge Level(LKL). The description of various input and out parameters were given below in table 2

Table 2. Description of Input Variables

Variable	Description
TTC -	Scale of Time spent on Target Content
TPV -	Scale of Total number of Page Visits
TAC -	Scale of Time spent on Allied Content
PTC -	Scale of test Performance on Target Content
PAC -	Scale of test Performance on Allied Content
LKL -	Learner Knowledge Level

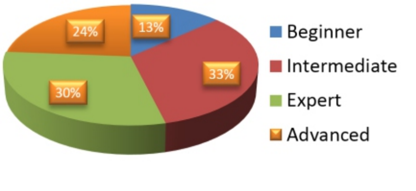


Figure 2. Distribution of Training data set

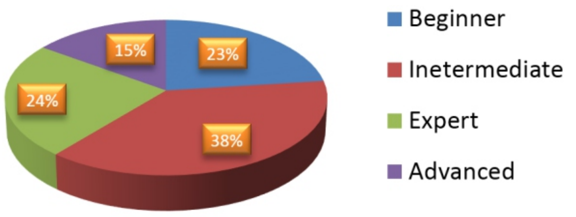


Figure 3. Distribution of Test data set

Input parameter values are then scaled to the normalized range using the standard technique.

$$\text{normalized value} = \left(\frac{\text{input value} - \text{initialLow}}{\text{initialHigh} - \text{initialLow}}\right) * (\text{finalHigh} - \text{finalLow}) + \text{finalLow}$$

From the 500 instances, 375 were considered as training dataset and the remaining 125 as test data set. According to the human domain experts analysis, the distribution ratio of Learner Knowledge Levels (LKL) in training data set and test dataset were shown in the Figure 2 and Figure 3

Table 3. Performance of various Classification Algorithms using training data set

	Name of the classifier	Correctly classified	inCorrectly classified	Kappa statistic	MAE	RMS	TP Rate F-Area	FP Rate	Precision	Recall	F-MEASURE	Roc area
Functions	SMO	359	16	0.941	0.083	0.156	0.957	0.019	0.959	0.957	0.957	0.995
	SL	364	11	0.960	0.046	0.125	0.971	0.011	0.971	0.971	0.971	0.993
	logistic	361	14	0.949	0.032	0.131	0.963	0.014	0.963	0.963	0.963	0.981
	MLP	361	14	0.949	0.032	0.131	0.963	0.014	0.963	0.963	0.963	0.981
Lazy	Kstar	340	35	0.872	0.055	0.186	0.907	0.035	0.908	0.907	0.885	0.949
	IBK	345	30	0.890	0.043	0.199	0.920	0.029	0.921	0.920	0.920	0.920
	IB1	345	30	0.890	0.040	0.200	0.920	0.029	0.921	0.920	0.920	0.945
Trees	J48	367	8	0.971	0.013	0.100	0.979	0.008	0.979	0.979	0.979	0.994
	RF	369	6	0.978	0.040	0.116	0.984	0.006	0.984	0.984	0.984	0.992
	RT	369	6	0.978	0.040	0.116	0.984	0.006	0.984	0.984	0.984	0.992

Table 4. Performance various Classification Algorithms using test data set

	Name of the classifier	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area
Functions	SMO	113	12	0.865	0.258	0.324	0.904	0.058	0.924	0.904	0.904	0.954
	S L	106	19	0.789	0.076	0.276	0.848	0.070	0.847	0.848	0.846	0.889
	logistic	106	19	0.789	0.076	0.276	0.848	0.070	0.847	0.848	0.846	0.889
	MLP	116	9	0.901	0.039	0.156	0.928	0.026	0.931	0.928	0.928	0.996
Lazy	Kstar	98	27	0.700	0.126	0.290	0.784	0.100	0.790	0.784	0.786	0.931
	IBK	106	19	0.789	0.076	0.276	0.848	0.070	0.847	0.848	0.846	0.889
	IB1	99	26	0.712	0.104	0.323	0.792	0.093	0.796	0.792	0.791	0.850
Trees	J48	122	3	0.967	0.036	0.097	0.976	0.006	0.977	0.976	0.976	1.000
	RF	122	3	0.967	0.026	0.106	0.976	0.006	0.977	0.976	0.976	0.999
	RT	116	9	0.901	0.039	0.156	0.928	0.026	0.931	0.928	0.928	0.996

Experimentation was carried out using Weka through various standardized classifiers like Meta, Lazy, Rules, Trees and Functions. Various performance parameters like Kappa statistic, Mean Absolute Error, Root Mean Squared Error and Relative Absolute Error for training data and test data were tabulated in table 3 and table 4 respectively.

The most effective measure of prediction accuracy is known as Mean Absolute Error (MAE). The absolute error is [11] the absolute value of the difference between the forecasted value and the actual value. It is a measure used to find how close the forecasts or predictions are to the eventual consequences. The mean absolute error is given by

$$MAE = \sum(|f(x_i) - y_i|)/N$$

A quadratizc scoring rule which measures the average magnitude [15] of the error is called Root Mean Squared Error(RMSE). The equation for the RMSE is the difference among forecast and corresponding located values are each squared and then averaged over the sample [6]. Finally, the square root of the average is taken. Since the mistakes are squared before they're averaged, the RMSE offers a distinctly excessive weight to massive errors [6]. This method the RMSE is most useful while huge mistakes are especially unwanted. The Root mean squared error is given by

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i - \hat{y}_i)^2}{n}}$$

Kappa static: Interpretation of Kappa

Kappa measures the conformity between any two raters. M items are classified int C mutually exclusive categories [10].The equation for κ is:

$$\kappa = \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e}$$

Here *po* is the relative watched understanding among raters. *pe* is the theoretical likelihood of conformity. The watched information is used to figure the probabilities [5] of every viewer randomly saying each group. K=1, if the raters are in complete agreement. κ ≤ 0 otherwise.

Table 5 Prediction accuracy of Kappa Static

Agreement	Poor	Slight	Fair	Moderate	Substantial	Almost Perfect
Kappa	0.0	0 .20	0.4	0.6	0.8	1.0

5. RESULTS AND ANALYSIS

Experimentations on the training data with cross validation inferred that tree based algorithms predicted more accurately compared to Lazy classifiers and classifiers under functions. Experimentations on the training data proved that tree based algorithms out performed with an average accuracy of 97%. Lazy classifiers (Kstar, IBK and IB1) and classifiers under functions (Logistic, Multilayer Perceptron(MPL) and Simple Logstic(SL)) showed consistency but with a lesser prediction rate of 96% compared to tree based classifiers. Experimentations on the test data also proved that tree based algorithms (Random Forests, J48 and Random tree) out performed with an average accuracy of 95%, whereas Lazy classifiers(Kstar, IBK and IB1) and classifiers under functions (Logistic, Multilayer Perceptron(MPL) and Simple Logstic(SL)) showed consistency but with a lesser prediction rate compared to tree based classifiers. The percentage of correctly and incorrectly classified instances for training dataset and test dataset are visualized in Figure 5.4 and Figure 5.5 respectively.

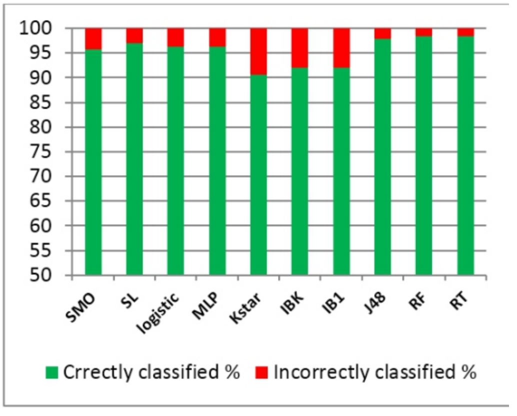


Figure 4. Training Data set

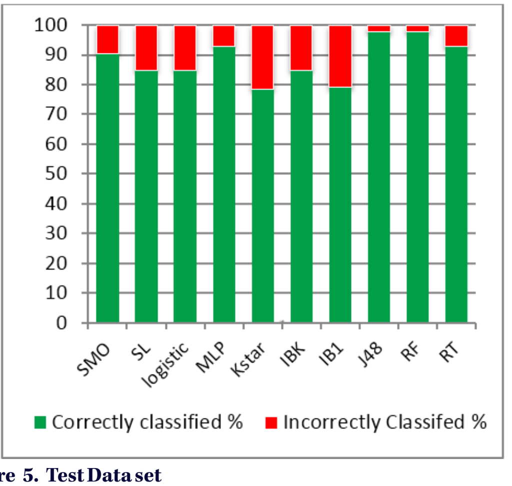


Figure 5. Test Data set

From the experimental results tabulated in table 3 and table 4, the accuracy measure parameters Precision, Recall, F-Measure and ROC Area for the top classifiers J48, Random forest, random tree and MLP are visualized in the form of graphs in Figure 6 and Figure 7 respectively.

Figure 6. Training Data set

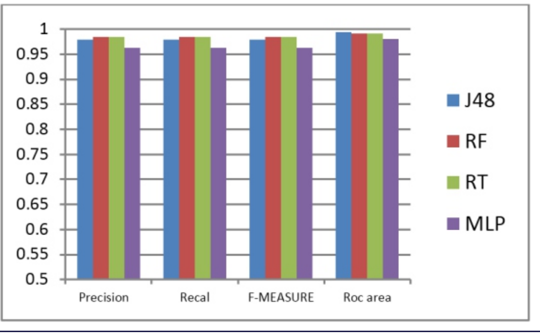
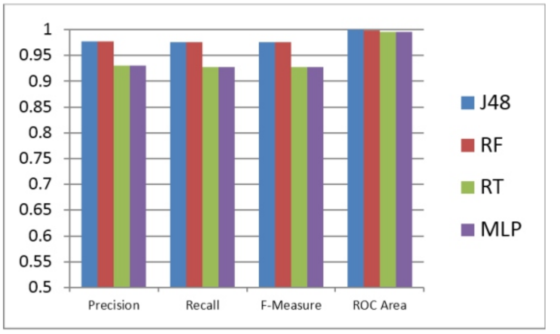


Figure 7. Test Data set



Precision is the fraction of retrieved instances that are relevant, while **recall** is the fraction of relevant instances that are retrieved. This is also represented in the form of confusion matrix where all the diagonal elements represent correctly classified instances and non-diagonal elements represent incorrectly classified instances.

Table 6. Confusion Matrix

Beginner	Intermediate	Expert	Advanced	Classified as
27	2	0	0	Beginner
5	**41**	1	0	Intermediate
0	0	**29**	1	Expert
0	0	0	**19**	Advanced
32	43	30	**20**	

Confusion matrix of test data shown in table 6 contains 125 instances. The MLP algorithm classified 116 instances correctly and 9 instances incorrectly. From the 9 incorrectly classified instances, 2 instances which actually belong to beginner class but incorrectly classified as intermediate class, 5 instances which actually belong to intermediate class but incorrectly classified as beginner class, 1 instance which actually belong to intermediate class but incorrectly classified as expert class and 1 instance which actually belong to expert class but incorrectly classified as Advanced class.

From the confusion matrix: for the Beginner class

Precision = # correctly classified instances / # instances classified under that class
= 27/32 = 0.844
Recall = # Correctly classified instances / Total number of instances of that class
= 27/29 = 0.931
F-Measure = (precision X recall X 2) / (precision +recall)
= (0.844 X 0.931 X 2.0) / (0.844 + 0.931) = 0.885
Likewise ROC area =0.982

Table 7 ROC area

Agreement	perfect	excellent	good	mediocre	poor	random	Wrong
ROC area	1.0	0.9	0.8	0.7	0.6	0.5	<0.5

Hence from the results shown in Table 3 & Table 4 and ROC area prediction Table 6, it can be inferred that all the top classifiers gave very close to perfect prediction.

6. CONCLUSION

This study deals with classification of learners. An LMS "Predicting Knowledge level of learner in the e-learning environment" was designed for data collection. This data was analyzed by the domain experts. Based on the analysis, learners were classified as Beginner, Intermediate, Expert, or Advanced. Experimentation was carried out using Weka and detailed performance analysis of various classification techniques was tabulated. Experimentations on the training data proved that tree based algorithms outperformed with an average accuracy of 97%. Lazy classifiers (Kstar, IBK and IB1) and classifiers under functions (Logistic, Multilayer Perceptron(MPL) and Simple Logistic(SL)) showed consistency but with a lesser prediction rate of 96% compared to tree based classifiers. Experimentations on the test data also proved that tree based algorithms (Random Forests, J48 and Random tree) out performed with an average accuracy of 95%, whereas Lazy classifiers(Kstar, IBK and IB1) and classifiers under functions (Logistic, Multilayer Perceptron(MPL) and Simple Logstic(SL)) showed consistency but with a lesser prediction rate compared to tree based classifiers. The designed Predicting learner knowledge level in the e-learning environment model helps e-learning content developers to understand the learner types. This provides a base for the content developers, to develop a more personalized content.

REFERENCES

[1] Huimin, Qi, Cui Ming, and Xiao Mingming. "A Personalized Resource Recommendation System Using Data Mining", 2010 International Conference on EBusinessand E-Government, 2010.

[2] Ahmad, Nor Bahiah Hj, and Siti Mariyam Shamsuddin. "A comparative analysis of mining techniques for automatic detection of student's learning style", 2010 10th International Conference on Intelligent Systems Design and Applications, 2010.

[3] Yathongchai, Chusak, Thara Angskun, Wilairat Yathongchai, and Jitimon Angskun. "Learner classification based on learning behavior and performance", 2013IEEE Conference on Open Sy tems (ICOS), 2013.

[4] Feng-Jung Liu. "Design of Self-Directed E-Learning Material Recommendation System with On-Line Evaluation", 2008 International Conference on Convergence and Hybrid Information Technology, 08/2008.

[5] Isa, Sani M.. "Kernel Dimensionality Reduction onSleep Stage Classification using ECG Signal",International Journal of Computer Science Issues (IJCSI)/16940784, 20110701

[6] Bouali, Hanen, and Jalel Akaichi. "Comparative Study of Different Classification Techniques: Heart Disease Use Case", 2014 13th International Conference on Machine Learning and Applications, 2014.

[7] Buckchash, Himanshu, and Gyanendra K. Verma."Texture vs. multiresolution analysis of facialexpressions: application to emotion recognition", International Journal of Applied Pattern Recognition, 2015.

[8] Liu, Qingchao Lu, Jian Chen, Shuyan Zhao."Multiple Naive Bayes classifiers ensemble for traffic incident detection.(Research Article)(Report)", Mathematical Problems in Engineering, Annual 2014 Issue.

[9] Sundar, P. V. Praveen and Kumar, A. V. Senthil. "Evaluation of Regional Benchmark Impact in EDM", International Journal of Computer Science Issues(IJCSI), 2013.

[10] SpringerBriefs in Energy, 2013. ISSN: 2191-5520.

[11] Zanella, Nicola. "Dividend yields are equity riskpremiums: practical implications for financialplanners.", Journal of Wealth Management, Spring 2015 Issue.

[12] Krishna, Gopala, Bharath Kumar, NagarajuOrsu, and Suresh B.. "Performance Analysis andEvaluation of Different Data Mining Algorithms used forCancer Classification", INTERNTIONAL JOURNAL OF ADVANCED RESEARCH IN ARTIFICIAL INTELLIGENCE,2013.

[13] Tesfamariam, S., and Z. Liu. "Seismic risk analysis using Bayesian belief networks", Handbook of seismic risk analysis and management of civilinfrastructure systems, 2013.