Evaluation of Data Mining Techniques for Predicting Student's Performance

Mukesh Kumar
Himachal Pradesh University, Summer-Hill, Shimla (H.P) Pin Code: 171005, India
Email: mukesh.kumarphd2014@gmail.com

Prof. A.J. Singh
Himachal Pradesh University, Summer-Hill, Shimla (H.P) Pin Code: 171005, India
Email: aj_singh_69@yahoo.co.uk

Abstract—This paper highlights important issues of higher education system such as predicting student's academic performance. This is trivial to study predominantly from the point of view of the institutional administration, management, different stakeholder, faculty, students as well as parents. For making analysis on the student data we selected algorithms like Decision Tree, Naive Bayes, Random Forest, PART and Bayes Network with three most important techniques such as 10-fold cross-validation, percentage split (74%) and training set. After performing analysis on different metrics (Time to build Classifier, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error, Root Relative Squared Error, Precision, Recall, F-Measure, ROC Area) by different data mining algorithm, we are able to find which algorithm is performing better than other on the student dataset in hand, so that we are able to make a guideline for future improvement in student performance in education. According to analysis of student dataset we found that Random Forest algorithm gave the best result as compared to another algorithm with Recall value approximately equal to one. The analysis of different data mini g algorithm gave an in-depth awareness about how these algorithms predict student the performance of different student and enhance their skill.

Index Terms—Educational Data Mining, Random Forest, Decision Tree, Naive Bayes, Bayes Network

I. INTRODUCTION

Data mining is a process which is used to extract the useful information from the database. This information is further used to make some decision for improvement in near future. Data mining is further categorised into the different field like education, medical, marketing, production, banking, hospital, telecommunication, supermarket and bioinformatics etc. In this entire field lots of data are generated day by day, and if that data is not processed properly then that data is useless. But if that data is processed properly then it will be helpful in making some decision for any business organisation. To properly analyse the database some important and frequently used data mining techniques are applied to get hidden information. An education system is one of the most important parts for the development of any country. So it should be taken very seriously from its start. Most of the developed countries have their own education system and evaluation criteria. Now a day's education is not limited to only the classroom teaching but it goes beyond that like Online Education System, MOOC course, Intelligent tutorial system, Web-based education system, Project based learning, Seminar, workshops etc. But all these systems are not successful if they are not evaluated with accuracy. So for making any education system to success, a well-defined evaluation system is maintained.

Now a day's every educational institution generate lots of data related to the admitted student and if that data is not analysis properly then all afford is going to be wasted and no future use of this data happens. This institutional data is related to the student admission, student family data, student result etc. Every educational institution applies some assessment criteria to evaluate their students. In modern education, we have lots of assessment tools which are used to observe the performance of the student in their study. Data Mining is one of the best computer based intelligent tool used to check the performance of the students. Educational data mining is one of the most important fields to apply data mining. Because, Now- a-days the most important challenges for every educational institution or universities have to maintain an accurate, effective and efficient educational process for the betterment of the student and institution. Data mining technology going to fill the knowledge gaps in higher education system. As already mentioned above that data mining applied in every field of education like the primary education system, secondary education system, elementary education system as well as higher education system also. At present scenario, lots of student data are generated in the educational process like student registration data, student sessional marks, student's sports activities, student's cultural activities, student's attendance

detail etc. By applying some data mining techniques on these data, some most interesting facts about student's behaviour, student's interest in the study, student's interest in sports etc may come out and further according to these information students may be guided for improving their performance. As a result of this improvement further bring lots of advantages to education system such as maximising the student retention rate, success rate, promotion rate, transition rate, improvement ratio, learning outcome and minimising the student's dropout rate, failure rate and reduce the cost of education system process. And in order to achieve the above-mentioned quality improvement, we need to apply a suitable data mining techniques, which can provide deep insight into the student's database and provide a suitable needed knowledge for making the decision on the education system.

II. LITERATURE SURVEY

Position In education system of any country, Quality education is an important fact and every educational organization working hard to achieve this. After searching and reading of almost thirty odd research paper on educational data mining, we find that student academic prediction, educational dropout prediction, student placement prediction, student success prediction, institution admission prediction etc are the mostly used for research purpose. By taking these topics into consideration, we select the student's academic prediction as one of the most interesting topics for research.

Raheela Asif, Agathe Merceron, Syed Abbas Ali, Najmi Ghani Haider (2017) in his study author mainly focuses on two different aspects of student performance. First they tried to predict the final result of the student in their fourth year of degree program with the pre-university data and result of first and second year of their degree. Secondly they analyzing the student progress throughout their degree program and then combine that progress with the prediction result. With this research they try to divide the students into two different groups like high and low performer group.

Raheela Asif, Agathe Merceron and Mahmood Khan Pathan (2015) in his they found that it is possible to predict the performance of the final year student with the help of pre-university and result of first and second year of their degree program. They were not using any social-economic or any other demographic attributes to predict the final result with a reasonable accuracy. For their research they used Decision tree algorithm with Gini Index (DT-GI), Decision tree with Information Gain (DT-IG), Decision tree with accuracy (DT-Acc), Rule-Induction with Information Gain (RI-IG) 1-Neural Networks (1-NN), Naive Bayes and Neural Networks (NN). They got the overall accuracy of 83.65% with the help of Naive Bayes on data set II.

Raheela Asif, Agathe Merceron and Mahmood Khan Pathan (2015) in here research they grouped the students according to the marks taken every year. These groups are created according to the range of the marks or percentage taken by the student in the examination every year (like 90-100 for group 1, 80-90 for group 2 etc.). In this paper the analysis the progress of the student every year and check whether student upgrade their group or not. They used k-mean clustering or x-mean clustering for finding the group of the students.

Mashael A. Al-Barrak and Mona S. Al-Razgan (2015) in here research collected dataset of student's from the Information Technology department at Kin Saud University, Saudi Arabia for their analysis. They further used the different attribute for the prediction like student ID, student name, student grades in three different quiz's, midterm1, midterm2, project, tutorial, final exam, and total points obtained in Data structure course of computer science department.

Raheela Asif and Mahmood K. Pathan (2014) in his study they used four academic batches of Computer Science & Information Technology (CS&IT) department at NED University, Pakistan. They used HSC marks, marks in MPC, Math's marks in HSC, marks in various subject studied in the regular course of a programming language, CSA, Logic design, OOP, DBMS, ALP, FAM, SAD, Data Structure etc for their analysis.

Mohammed M. Abu Tair and Alaa M. El-Halees (2012) in his study tried to extract some useful information from student's data of Science and Technology College – Khan Younis. They initially selected different attributes like Gender, date of Birth, Place of Birth, Specialty, Enrollment year, Graduation year, City, Location, Address, Telephone number, HSC Marks, SSC school type, HSC obtained the place, HSC year, College CGPA for analysis. But after preprocessing of the data they found that attribute like Gender, Specialty, City, HSC Marks, SSC school type, College CGPA are most significant.

After reviewed different research paper, we found that in most of the cases, the important personal attributes of the student like Age, Gender, Home Location, Communication skill, Sportsperson, Social Friends, Smoking habits, Drinking habits, Interest in study, Internet and Hosteller, Day Scholar are taken into consideration for further research. The family attributes like Father's Qualification, Mother's Qualification, Father's Occupation, Mother's Occupation, Total Family Member, family Structure, Family Responsibilities, Family supports, Financial Condition are also taken as important for the academics prediction. Whereas for academic attributes like 10th %age, 10th Board, 10th Education Medium, 12th %age, 12th Board, 12th Education Medium, JEE Rank, Admission Type, Institution Type, Branch Taken, Attendance during Semester, Internal Sessional Marks, External Marks, Technical Skill, Logical Reasoning skill, Extra Coaching are taken into consideration and for institutional attributes most the researcher is taken Institution Type, Institution Location, Transportation facility, Library Facilities, Lab Facilities, Hostel Facilities, Sanitary facilities, Sports facilities, Canteen Facilities, Internet facilities, Internet facilities, Teaching Aids, Institution

Placements, Student Counselor, Mentorship cell into consideration.

In this meta-analysis, we find that most used data mining techniques for Student's Academic Performance prediction are Decision Tree algorithm, Naive-Bayes algorithm, Random Forest algorithm, Classification and Regression Trees algorithm (CART), J48, Logistic Regression, LADTree and REPTree. In Decision tree algorithm the maximum and minimum accuracy for predicting student's academic performance is 99.9% and 66.8% respectively.

III. STUDENT'S ATTRIBUTE SELECTION

Student's dataset is a collection of different attributes of a student put in a single table. As we mentioned in the literature survey that student's personal attributes, family attributes, academic attributes and institutional attributes are taken into consideration. We are taken student's data of 412 post-graduate for predicting their academic performance in the running semester. During the start of our work we think about the collection of academic and personal attributes of the students. But at the time of personal data collection from the student, we feel that the information providing by the student is not up to the mark and it may affect our prediction result. So at this point of time, we are think about considering only the pre-admission data (like High school grade, Secondary school grade, Admission test result etc) and some academic attributes are also taken into consideration.

Our final student dataset includes attributes like 10th percentage, 12th percentage, graduation marks, Father's and mother's qualification and occupation and financial condition of the family. The predicted class which we consider here are A, B, C and D. Here class A means the student who has marks above 90%, B means students have marks between 75-90%, C means students have marks between 60-74.9% and D means students have marks between 50-59.9%. I just make these predicted classes only for the research purpose and does not relate to any grading system of any organization. For implementation purpose, we are using WEKA tools which are open source software and easily available for use. We are selecting Decision Tree, Naive Bayes, Random Forest, PART and Bayes Network with 10-fold cross-validation, percentage split and training set methods. We are also implementing another classification algorithm also but the above-listed classification algorithm gives the best result.

IV. WORKING WITH DIFFERENT CLASSIFICATIONS DATA MINING ALGORITHMS

In this particular section of this paper, detailed analysis of algorithms with different performance metrics is taken into consideration because I think it gives us a better understanding of the algorithm. The different performance metric such as total time taken by classifier to build a model, Total correctly and incorrectly classified instance, Mean absolute error, Root Mean squared error, Relative Absolute Error, Root Relative Squared Error, True positive rate, True Negative Rate, Precision, Recall, F-Measure and Receiver Operating Characteristics Area (AUC). The comparisons of the different algorithm according to different metrics determine the predictability, goodness and error measurement of the algorithm. The literature study shows that it is difficult to consider in advance that which performance metrics are better for which problem because every problem has their individual attribute and features. So it is recommended that different metrics with combination are used for the better result of the algorithm. For example, True positive rate metrics are taken higher values for the better result of the algorithm and Mean Absolute Errors are taking lower values.

V. PERFORMANCE METRICS FOR EVALUATION OF ALGORITHM

In the literature of data mining, the performance metrics are divided into three different categories like the probabilistic error, Qualitative Error and visual metrics. For finding the result of these metrics we are using some terms like TP (True Positive), TN (True Negative), FP (False Positive), FN (False Negative), ROC (Receiver Operating Characteristics), AUC (Area under Curve) which are important to understand for finding the result of different performance metrics. We discuss the formula for finding the result of different performance metric below:

Probabilistic Error: These types of metrics are totally based on probabilistic perceptive of predictions output (PO) and of errors (PO-AO), where PO is known as predicted outcome by an algorithm, AO is the actual outcome of that algorithm. There are different types of probabilistic error such as Mean Absolute Error (MAE), Log Likelihood (LL) and Root Mean Square Error (RMSA). These types of evaluation metrics are significant for the better understanding of the prediction result. Lowest value of prediction result is best for Mean Absolute Error and Root Mean Square Error and higher value of prediction result is best for the Log Likelihood. Table 1.below gave the formula to find out all types of probabilistic error.

Table 1. Types of Probabilistic Error of performance metrics

Types of Probabilistic Error	Formula used for the calculation
Mean Absolute Error	$(1/)\Sigma|AO-PO|$ where AO = Actual Output, PO = Predicted Output
Root Mean Square Error	$((1/)\Sigma(AO-PO)^2)$ where sqrt = Square root
Log Likelihood (LL)	$\Sigma AO(\log PO)+(1-AO)(\log(1-PO))$

Qualitative Error: These types of metrics are totally based upon qualitative perceptive of errors i.e. whether the prediction is acceptable or not acceptable. In the case of student predictive modeling, this approach is suitable for predicting the student state in future. In this metrics only two possible classes are possible for classification like true/false or positive/negative by using a confusion matrix. The diagonal elements of the confusion matrix are always having correct prediction for classification and the other element of the matrix are known as model errors. Some most commonly used metrics for the qualitative performance are accuracy, sensitivity or recall value, specificity, precision and F-Measure. These types of statistical metrics are usually used for inspecting that how excellent and reliable was the classifier. In the confusion matrix, True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) rate determine the predictive efficiency of the data mining algorithm

Table 2. Types of Probabilistic Error of performance metrics

Types of Qualitative Error	Formula used for the result calculation
Overall Accuracy	$=((TP+TN)/(TP+TN+FP+FN))*100\%$
Sensitivity (Recall)/ Positive Class accuracy	$=(TP/(TP+FN))*100\%$
Specificity/ Negative class accuracy	$=(TN/(TN+FP))*100\%$
Precision	$=(TP/(TP+FP))*100\%$
F-Measure	$=(2\ Precision * Recall)/(Precision + Recall)$

VI. Performance Evaluation of Algorithms Taken Into Consideration

In this particular section, we have examines the performance of different selected data mining algorithm on the dataset in hand. For example, in the below-mentioned table, the performance metrics like correctly classified instance measure the accuracy of the algorithm. Other such metrics such as Sensitivity (Recall), Specificity, Precision and Accuracy all depends upon True Positive, True Negative, False Positive and False Negative. Below listed table evaluated these performance metrics on different classification algorithms with three different test mode like 10-fold cross validation test, training data set from the means dataset in hand and percentage split from the dataset. For all these evaluations we have used WEKA Tool with mostly inbuilt classification algorithm for use.

Evaluating with Training set mode in WEKA means that you are using the pre-processed data uploaded through Explorer interface for testing purpose. Secondly, supplied test set used for the testing of the classifier build by the training data set. So we can say that supplied data set is only used for the testing purpose only. For this propose, we select an option and click the set button. After that we can get a small window through which you can enter you test dataset and then this show you the name of the relation, total number of attribute total instance and the relation between different attributes.

Table 3. Evaluating with Training set mode on the student's dataset.

Performance Metrics	J48	Naive Bayes	Random Forest	PART	Bayes Network
Time to build Classifier	0.05 Sec	0.00 Sec	0.1 Sec	0.05 Sec	0.01 Sec
Correctly Classified	253(61.40%)	364(88.34%)	412(100%)	283(68.68%)	399(96.84%)
Incorrectly Classified	159(38.59%)	48(11.65%)	0(0.00%)	129(31.31%)	13(03.15%)
Mean Absolute Error	0.25	0.25	0.12	0.21	0.21
Root Mean Squared Error	0.35	0.31	0.15	0.32	0.26
Relative Absolute Error	75.71%	76.62%	37.37%	63.49%	63.87%
Root Relative Squared Error	87.05%	77.53%	37.75%	79.72%	64.48%
TP Rate	0.614	0.883	1.00	0.687	0.968
FP Rate	0.228	0.081	0.00	0.174	0.021
Precision	0.621	0.903	1.00	0.688	0.970
Recall	0.614	0.883	1.00	0.687	0.968
F-Measure	0.598	0.863	1.00	0.678	0.962
ROC Area (AUC)	0.788	0.994	1.00	0.862	0.999

From table 3 we find that the **Sensitivity** (Recall) value of the all the algorithm is good but Random Forest algorithm gave the best result as compared to another algorithm with the value approximately equal to one. Bayes Network algorithm performs with second highest value 0.968, but J48 and PART algorithm were not performed best according to our dataset with recall value close to 6. The correctly classified instance values are also greater than 60% in all cases with 100% highest value for Random Forest algorithm.

Evaluating with 10-fold cross-validation mode means that the classification results will be evaluated by cross-validation. In this mode you can also change the number of folds.

Table 4. Evaluating with 10-fold cross-validation mode on the student's dataset.

Performance Metrics	J48	Naive Bayes	Random Forest	PART	Bayes Network
Time to build Classifier	0.01 Sec	0.00 Sec	0.10 Sec	0.05 Sec	0.01 Sec
Correctly Classified	131(31.79%)	139(33.73%)	173(41.99%)	120(29.12%)	143(34.70%)
Incorrectly Classified	281(68.20%)	273(66.26%)	239(58.00%)	292(70.87%)	269(65.29%)
Mean Absolute Error	0.35	0.34	0.34	0.35	0.34
Root Mean Squared Error	0.48	0.42	0.41	0.51	0.42
Relative Absolute Error	103.7%	103.2%	100.8%	106.0%	103.4%
Root Relative Squared Error	118.4%	103.2%	100.9%	124.4%	102.3%
TP Rate	0.318	0.337	0.420	0.291	0.347
FP Rate	0.367	0.401	0.423	0.362	0.407
Precision	0.295	0.252	0.178	0.279	0.248
Recall	0.318	0.337	0.420	0.291	0.347
F-Measure	0.305	0.281	0.250	0.284	0.286
ROC Area (AUC)	0.459	0.429	0.386	0.442	0.440

In the last evaluating dataset with Percentage Split mode means that classification results will be evaluated on a test set that is a part of the original data. The default percentage split of the data is 66% in the WEKA data mining tool. Which further means that 66% of the data from the dataset are used for training purpose and 34% of the data are used for testing? We may further change this value according to the situation in hand. In our problem we change this percentage split upto 74% for our training purpose and remaining 26% for our testing purpose.

Table 5. Evaluating with Percentage Split mode on the student's dataset.

Performance Metrics	J48	Naive Bayes	Random Forest	PART	Bayes Network
Time to build Classifier	0.01 Sec	0.00 Sec	0.07 Sec	0.04 Sec	0.00 Sec
Correctly Classified	34(39.08%)	35(40.22%)	36(41.37%)	27(31.03%)	35(40.22%)
Incorrectly Classified	53(60.91%)	52(59.77%)	51(58.62%)	60(68.96%)	52(59.77%)
Mean Absolute Error	0.33	0.34	0.33	0.33	0.34
Root Mean Squared Error	0.42	0.41	0.40	0.49	0.41
Relative Absolute Error	99.69%	103.2%	100.1%	101.2%	102.7%
Root Relative Squared Error	105.4%	103.3%	100.4%	122.7%	101.6%
TP Rate	0.391	0.402	0.414	0.310	0.402
FP Rate	0.376	0.388	0.414	0.307	0.396
Precision	0.342	0.416	0.171	0.338	0.302
Recall	0.391	0.402	0.414	0.310	0.402
F-Measure	0.352	0.361	0.242	0.322	0.336
ROC Area (AUC)	0.530	0.461	0.432	0.516	0.492

From Table 4 to Table 5, we find that the **Sensitivity** (Recall) value of the all the algorithm are not very good but Random Forest algorithm gave the best result and second best result by Bayes Network compared to another algorithm. But compared with the **Sensitivity** (Recall) value of Table 3, this value is very low. The correctly classified instance values are also less than 40% in all cases which are very less as compared to 100% for Random Forest algorithm in Table 3. From the above table, we can conclude that the Mean Absolute Error and Root Mean Square Error values of Random Forest algorithm are lowest in all the three test set taken into consideration. Thus we can say that Random Forest algorithm gave us the best predictive result in all algorithms like the decision tree, Naive Bayes, CART and Bayes Network.

VII. Conclusion

From the above discussion, we can conclude that Random Forest algorithm for predictive modelling gave me the best result as compared to the Decision Tree, Naive Bayes, Bayes Network and CART. The correctly

classified instance with Random Forest algorithm is 61.40% with a very good recall value equal to 1. With this paper, we try to find out the best result with only selected attribute like academic and family attribute. This paper helps other research to concentrate on academic and family attributes as compared to personal attributes of the student. Because during collection of personal attribute student may not fill the correct information and it may affect our result too. The result of this work help model designer to think about these attribute seriously and do work on it. In our future work, we try to implement our work on the different dataset with different classification algorithm. By doing this we can able to find the best algorithm which is going to perform almost perfect on every dataset in hand.

ACKNOWLEDGEMENTS

REFERENCES

[1] Farhana Sarker, Thanassis Tiropanis and Hugh C Davis, Students" Performance Prediction by Using Institutional Internal and External Open Data Sources, http://eprints.soton.ac.uk/353532/1/Students' mark prediction model.pdf, 2013

[2] D. M. D. Angeline, Association rule generation for student performance analysis using an apriori algorithm, The SIJ Transactions on Computer Science Engineering & its Applications (CSEA) 1 (1) (2013) p12–16.

[3] Abeer Badr El Din Ahmed and Ibrahim Sayed Elaraby, Data Mining: A prediction for Student's Performance Using Classification Method, World Journal of Computer Application and Technology 2(2): 43-47, 2014

[4] Fadhilah Ahmad, Nur Hafieza Ismail and Azwa Abdul Aziz, The Prediction of Students" Academic Performance Using Classification Data Mining Techniques, Applied Mathematical Sciences, Vol. 9, 2015, no. 129, 6415 - 6426HIKARI Ltd, www.m-hikari.com http://dx.doi.org/10.12988/ams.2015.53289

[5] Mashael A. Al-Barrak And Mona S. Al-Razgan, predicting students" performance through classification: a case study, Journal of Theoretical and Applied Information Technology 20th May 2015. Vol.75. No.2

[6] Edin Osmanbegović and Mirza Suljic, DATA MINING APPROACH FOR PREDICTING STUDENT PERFORMANCE, Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

[7] Raheela Asif, Agathe Merceron, Mahmood K. Pathan, Predicting Student Academic Performance at Degree Level: A Case Study, I.J. Intelligent Systems and Applications, 2015, 01, 49-61 Published Online December 2014 in MECS (http://www.mecs-press.org/) DOI: 10.5815/ijisa.2015.01.05

[8] Mohammed M. Abu Tair, Alaa M. El-Halees, Mining Educational Data to Improve Students" Performance: A Case Study, International Journal of Information and Communication Technology Research, ISSN 2223-4985, Volume 2 No. 2, February 2012.

[9] Azwa Abdul Aziz, Nor Hafieza Ismailand Fadhilah Ahmad, First Semester Computer Science Students" Academic Performances Analysis by Using Data Mining Classification Algorithms, Proceeding of the International Conference on Artificial Intelligence and Computer Science(AICS 2014), 15 - 16 September 2014, Bandung, INDONESIA. (e-ISBN978-967-11768-8-7).

[10] Kolo David Kolo, Solomon A. Adepoju, John Kolo Alhassan, A Decision Tree Approach for Predicting Students Academic Performance, I.J. Education and Management Engineering, 2015, 5, 12-19 Published Online October 2015 in MECS (http://www.mecs-press.net) DOI: 10.5815/ijeme.2015.05.02

[11] Dr Pranav Patil, a study of student's academic performance using data mining techniques, international journal of research in computer applications and robotics, ISSN 2320-7345, vol.3 issue 9, pg.: 59-63 September 2015

[12] Jyoti Bansode, Mining Educational Data to Predict Student's Academic Performance, International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169, Volume: 4 Issue: 1, 2016

[13] R. Sumitha and E.S. Vinoth Kumar, Prediction of Students Outcome Using Data Mining Techniques, International Journal of Scientific Engineering and Applied Science (IJSEAS) – Volume-2, Issue-6, June 2016 ISSN: 2395-3470

[14] Karishma B. Bhegade and Swati V. Shinde, Student Performance Prediction System with Educational Data Mining, International Journal of Computer Applications (0975 – 8887) Volume 146 – No.5, July 2016

[15] Mrinal Pandey and S. Taruna, Towards the integration of multiple classifiers pertaining to the Student's performance prediction, http://dx.doi.org/10.1016/j.pisc.2016.04.076 2213-0209/© 2016 Published by Elsevier GmbH. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/ licenses/by-nc-nd/4.0/).

[16] Maria Goga, Shade Kuyoro, Nicolae Goga, A recommended for improving the student academic performance, Social and Behavioral Sciences 180 (2015) 1481 – 1488

[17] Anca Udristoiu, Stefan Udristoiu, and Elvira Popescu, Predicting Students" Results Using Rough Sets Theory, E. Corchado et al. (Eds.): IDEAL 2014, LNCS 8669, pp. 336–343, 2014. © Springer International Publishing Switzerland 2014.

[18] Mohammed I. Al-Twijri and Amin Y. Noaman, A New Data Mining Model Adopted for Higher Institutions, Procedia Computer Science 65 (2015) 836 – 844, doi: 10.1016/j.procs.2015.09.037

[19] Maria Koutina and Katia Lida Kermanidis, Predicting Postgraduate Students" Performance Using Machine Learning Techniques, L. Iliadis et al. (Eds.): EANN/AIAI 2011, Part II, IFIP AICT 364, pp. 159–168, 2011. © IFIP International Federation for Information Processing 2011

[20] Asif, R., Merceron, A., & Pathan, M. (2014). Investigating performances' progress of students. In Workshop Learning Analytics, 12th e_Learning Conference of the German Computer Society (DeLFI 2014) (pp. 116e123). Freiburg, Germany, September 15.

[21] Asif, R., Merceron, A., & Pathan, M. (2015a). Investigating performance of students: A longitudinal study. In 5th international conference on learning analytics and knowledge (pp. 108e112). Poughkeepsie, NY, USA, March 16-20 http://dx.doi.org/10.1145/2723576.2723579.

[22] Asif, R., Merceron, Syed Abbas Ali, Najmi Ghani Haider. Analyzing undergraduate students' performance using

educational data mining. Computer & Education 113(2017) 177-194, http://dx.doi.org/10.1016/j.compedu.2017.05.007

[23] Mukesh Kumar, Prof. A.J. Singh, Dr. Disha Handa. Literature Survey on Educational Dropout Prediction. I.J. Education and Management Engineering, 2017, 2, 8-19 Published Online March 2017 in MECS (http://www.mecs-press.net) DOI: 10.5815/ijeme.2017.02.02

[24] Mukesh Kumar, Prof. A.J. Singh, Dr. Disha Handa. Literature Survey on Student's performance prediction in Education using Data Mining Techniques. I.J. Education and Management Engineering. (Accepted) in MECS (http://www.mecs-press.net).

Authors' Profiles

Mukesh Kumar (10/04/1982) has pursuing PhD in Computer Science from Himachal Pradesh University, Summer-Hill Shimla-5. India. My research interest includes Data Mining, Educational Data Mining, Big Data and Image Cryptography.