AN EMPIRICAL STUDY OF MACHINE LEARNING ALGORITHMS TO PREDICT STUDENTS' GRADES

M.U. Ahmed and A. Mahmood[*]

Department of Computer Science, Allama Iqbal Open University
[*]Department of Computer Science, Abasyn University, Islamabad Campus
Corresponding author's email: moiz.ahmed@aiou.edu.pk

ABSTRACT

Machine learning algorithms provide an opportunity to analyze the existing educational data and predict futuristic needs. In present study a model was proposed to classify students' grades by employing machine learning algorithms. The important parameters from profiles and preferences were considered for classification purpose. Five classification algorithms *i.e.* Decision Table, OneR, J48, Random Forest and Random Tree were used for model construction and prediction of grades. The pattern analysis was done by WEKA open source data mining tool. The J48 was found to be the best algorithm in predicting grades with highest accuracy of 78 %. The accuracies obtained from Random Forest, Random Tree, Decision Table and OneR were 73, 72, 58.28 and 39.9 % respectively, which were lower than the accuracy obtained by J48. The results showed the effectiveness of machine learning algorithms to predict the performance of students.

Keywords: Classification algorithm, Grade Prediction, J48 algorithm, Machine learning algorithms and Student grades.

(Received   23-08-2017   Accepted 15-01-2018)

INTRODUCTION

The amount of data in the information industry is of no benefit for the management unless it is converted into meaningful information. The students' data is growing and providing us with an opportunity to dig out interesting patterns and useful information, generally called Education Data Mining (Katare and Dubey, 2017). Patterns can be used to predict students' grades to enhance the quality of academic programs (Majeed and Junejo, 2016). EDM is based on the data mining and machine learning (Jiawei and Kamber, 2011), which has recently got a lot of attention. Machine learning algorithms facilitate computers to learn without being explicitly programmed (Munoz, 2014). These algorithms manipulate raw data, iteratively learn from experience and discover hidden knowledge (Michalski *et al.,* 2013). Machine learning algorithms are often categorized as being supervised and unsupervised (Lantz, 2013, Bird *et al.,* 2009).

There are a variety of research studies available on using data mining and machine learning techniques to analyze the educational data. The commonly used features are the comparison of algorithms' accuracy, attributes of students affecting the success and failures, impact of academic history, domain of the study and performance of algorithms (Ilic *et al.,* 2016, Zimmermann *et al.,* 2011). These studies include analysis of socio demographic variables and educational environment parameters using Classification Trees (Kovaic, 2010), daily attendance records, assignment marks and seminar evaluations using Iterative Dichotomiser 3 (ID3), C4.5 and Regression Tree (CART) algorithms (Yadav *et al.,* 2011), performance of 1-Nearest Neighbor (1-NN) and Naïve Bayes for predicting grades (Koutina and Kermanidis, 2011), performance of Decision Tree algorithm (Abutair and El-Halees, 2012), factors affecting the success and failures of students (Lakshmi *et al.,* 2013), enhancing the quality of educational programs by discovering hidden knowledge in student records (Chalaris *et al.,* 2014), analysis of semester activities (Agrawal and Mavani, 2015) and performance of ID3, CART, Chi-squared Automatic Interaction Detector (CHAID), C4.5 and Naïve Bayes algorithms (Abusaa, 2016). Classification models work with different set of features and different combination of algorithms.

Keeping in view the futuristic use of machine learning algorithms and to enhance the quality of educational programs, the performance of various machine learning algorithms to investigate the parameters/attributes having direct impact on the students' performance were evaluated.

MATERIALS AND METHODS

The proposed model of grade prediction is shown in Figure 1. The model had three important components: data collection, data preprocessing and pattern analysis.
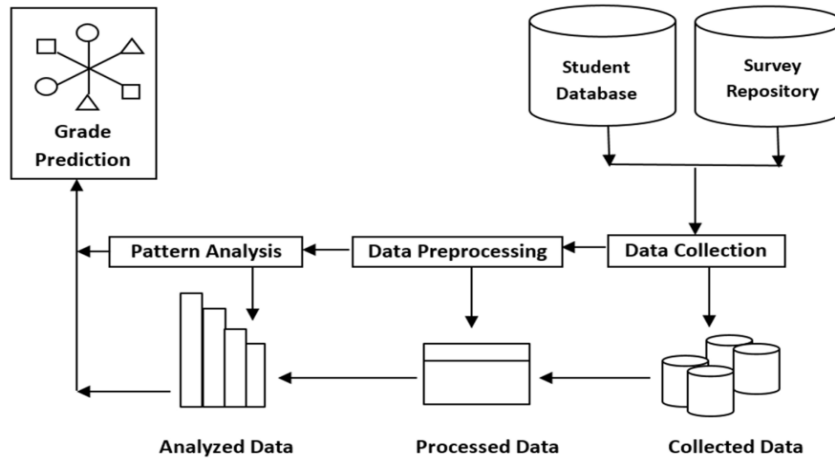
Figure-1: Proposed grade prediction model showing processing steps and data sources

In the data collection step, data was extracted from two data sources including student data from the main university database and survey data stored in an oracle database managed at Allama Iqbal Open University, (AIOU), Pakistan. The open source tool PHP was used as a front end and MySql database at the backend. The data preprocessing step was automatically performed by the Waikato Environment for Knowledge Analysis (WEKA) tool (Rangra and Bansal, 2014). In the data preprocessing step, data was cleaned by removing all the unwanted attributes and data rows with null values (Rangra and Bansal, 2014). This involved extracting only those attributes that were used by the machine learning algorithms. The incomplete surveys and missing values in student database were also removed. Nine attributes were selected for pattern analysis (Table 1) as described in literature (Moiz, 2015).

Table-1: Students' Attributes selected for training model.

No.	Name	Description	Values
1.	Gender	Gender category	Male, Female
2.	Location	Location of study	Urban, Semi urban, Rural
3.	Age	Age of student	Less than or equal to 20 years, Between 21 to 30 years, Between 31 to 40 years, More than 40 years
4.	Profession	Job status	Private employee, Government employee, Household, Unemployed
5.	Access to Computers	Accessibility to computers	Very Often, Often, Normal, Some Times, Never used
6.	Access to Internet	Accessibility to Internet	Very Often, Often, Normal, Some Times, Never used
7.	E-learning Preference	Inclination of students about e-learning	Strongly Agree, Agree, Normal, Do Not Agree, Strongly Disagree
8.	Content Preference	Favorite format of learning contents	Tutorials, Animations, Homework
9.	Grades	Examination Grades	A+ : 80% and above, A : 70% - 79%, B : 60% - 69%, C : 50% - 59%, D : 40% - 49%, F : less than 40 %

The third step was the analysis of data using graphical interface of WEKA (Rangra and Bansal, 2014) that provided built-in machine learning algorithms for model construction and prediction of grades (Singhal and Jena, 2013). The five machine learning algorithms used in the proposed work were Decision Table (Chen, 2017), OneR (Alam and Pachauri, 2017), J48 (Katare and Dubey, 2017), Random Forest (Beaulac, and Rosenthal, 2017) and Random Tree (Sutera, 2013).

RESULTS AND DISCUSSION

The student data of four semesters of a computer science diploma program was collected to study the accuracy of various machine learning algorithms. The dataset contained 163 instances and 09 attributes (Table 1). The demographic results showed that out of 163 respondents, 87 (53 %) were males and 76 (47%) females. The majority 49 (30 %) were government employees and 37 (23%) doing private jobs. The 57 (35 %) of respondents belonged to urban areas while 50 (31 %) were from rural area and the remaining 56 (34%) from semi-urban area. The age group illustrated that 87 (53 %) of the students belonged to 21 – 30 years of age followed by 74 (45 %) in 31 – 40 years and the rest were in other age groups.

The visualization of distribution of data for different attributes is shown in Figure 2. It was observed that the training and test data used in the experimentation were well distributed for different attributes indicating validity and adequacy of data for experimentation.
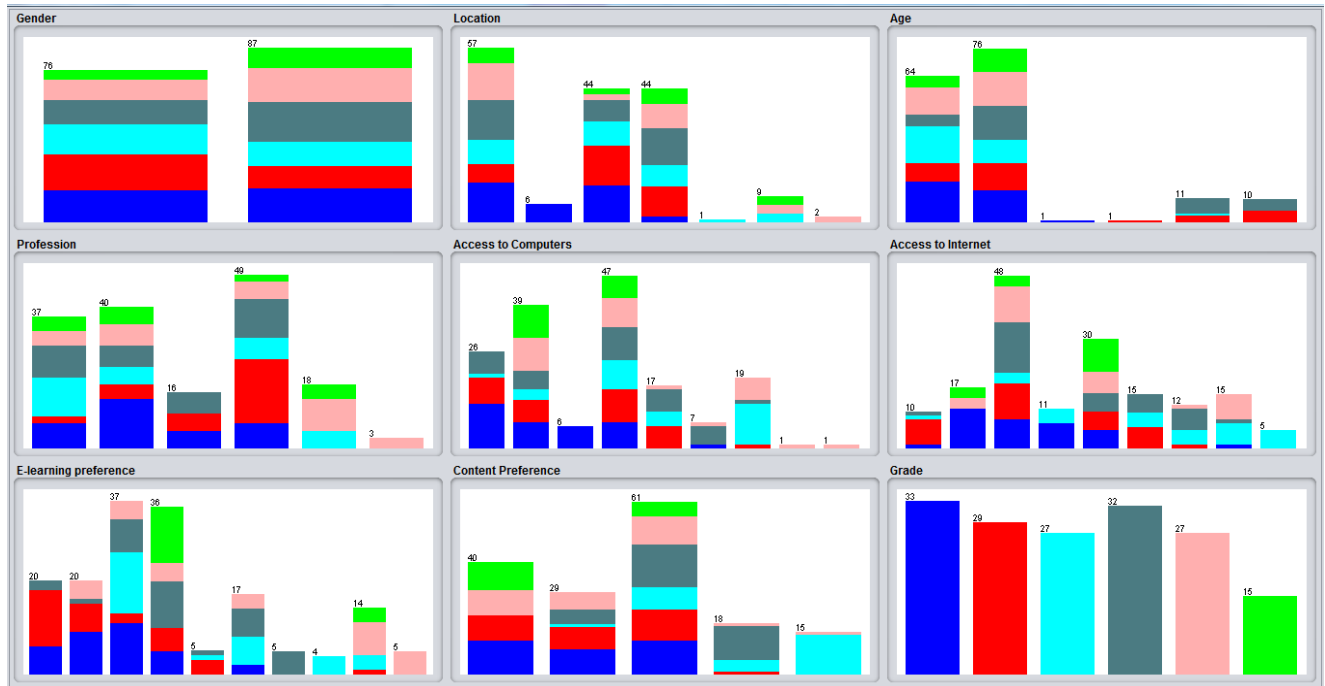

Figure-2: Model Visualization showing distribution of attribute values

The tenfold cross validation, a commonly used method in machine learning experiments, was followed to split the data into disjoint sets of equal size. The machine learning algorithms studied in the experimentation used nine disjoint data sets for training while one data set was used for testing. Five different machine learning algorithms i.e. Decision Table, OneR, J48, Random Forest and Random Tree were compared. The accuracy of each algorithm when all the nine attributes are used is shown in Table 2.

Table-2: Accuracy Comparison of ML algorithms with all nine attributes.

Algorithm	Accuracy (%)	Correctly Classified Values	Incorrectly Classified Values
Decision Table	58.2	95	68
OneR	39.9	65	98
J48	78	127	36
Random Forest	73	119	44
Random Tree	72	117	46

The accuracy of OneR and Decision Table were found to be lowest (39.9 % and 58.2% respectively). The

Decision Table classified the data set into discrete spaces, however it was biased to select a featured subset of items while predicting the success of students. Therefore, its performance was not consistent and varied in different domains and sets of data items. Similar results were also reported in other studies (Lodhi *et al.*, 2011). OneR made predictions based on a single rule. It was useful when the number of attributes was more and Decision Tree of more than one level was to be generated (Anuradha and Velmurugan, 2015). The experimental results showed that the performance of Random Forest and Random Tree algorithms were comparable. This could be explained with the fact that Random Forest built a Decision Tree comprising of possible decisions and their corresponding actions and prediction and classification rules were formed from root to leaf nodes. In case of Random Tree, a Decision Tree was drawn randomly in which each tree got equal chance to occur in the sampling (Wang *et al.*, 2015). In some cases, the performance of Random Forest was better than Random Tree. However better results required more trees for better performance (Mesarić and Šebalj, 2016). J48 showed best classification accuracy for predicting student's grades. There were variations in the performance of J48 algorithm as reported in many previous studies. In some studies, the performance of J48 was worse than Naive Bayes (Kaur and Singh, 2016) and RepTree (Mesarić, and Šebalj, 2016). While in other studies, J48 showed excellent performance in predicting grades as compared to Simple Cart, RepTree and NB Tree (Pandey and Sharma, 2013).

The effect of attributes used on the accuracy of the algorithms was also studied. The experiments were performed by eliminating the attributes 5 to 8 which have not been used in the previous study (Table 3). It is evident from the results that the accuracy of each of the algorithm decreased as compared to accuracy when all the attributes were used.

Table 3: Accuracy Comparison of ML algorithms without attributes 5 to 8.

Algorithm	Accuracy (%)	Correctly Classified Values	Incorrectly Classified Values
Decision Table	51.5	84	79
OneR	32.5	53	110
J48	72.4	118	45
Random Forest	65	106	57
Random Tree	66.3	108	55

Due to its best performance, J48 algorithm was studied for other accuracy measures such as True Positive (TP) and False Positive (FP) and confusion matrix. TP indicated the positive values that were correctly predicted by the classifier, while FP gave negative values that were incorrectly predicted as positive by the classifier. The

results in Table 4 revealed that the TP Rate was high for four of the grades A+ (0.879), A (0.759), B (0.813) and C (0.815). The TP rate was low for grades D (0.667) and F (0.667). The confusion matrix shown in Table 5 reconfirmed the performance of the classifier showing that the proposed classifier was able to predict correct grades with a high accuracy for grades A+, A and C. For instance, 29 grades (out of 33) were correctly predicted for A+ while only 04 were predicted incorrectly. For grade A, the correctly predicted grades were 22 while only 7 grades were incorrectly predicted. Similar results could be seen for C grade. However the performance for grade D and F was a bit lower but still had a good accuracy rate.

Table-4: True Positive and False Negative Rates for J48 algorithm.

TP Rate	FP Rate	Class
0.879	0.031	A+
0.759	0.045	A
0.813	0.038	B
0.815	0.044	C
0.667	0.081	D
0.667	0.027	F

Table-5: Confusion Matrix for J48 algorithm.

Actual Grades	Classified/Predicted as					
	A+	A	B	C	D	F
A+	29	0	2	1	1	0
A	3	22	1	1	2	0
B	0	3	26	1	2	0
C	0	3	1	22	1	0
D	1	0	1	3	18	4
F	0	0	0	0	5	10

Three important parameters were considered during the comparison i.e. selection of machine learning algorithms, attributes of students and accuracy of the predictive model.

Usha (2011) combined the previous examination results with the demographic attributes and got an accuracy of 97 % using the Support Vector Machine algorithm. However, the already known previous examination results contributed in such high accuracy of the model. The study conducted by Ramesh *et al.* (2013) used five machine learning algorithms on twenty nine attributes. The authors used Naive Bayes, Multilayer Perception, Sequential Minimal Optimization (SMO), J48 and REPTree algorithms which were different with the proposed model. The best accuracy found by Multilayer Perception was 72.38 %. The selection of attributes was too wide and the accuracy was also lower than the proposed research.

Aguiar *et al.* (2015) used Random Forest and Logistic Regression algorithms and obtained the accuracies of 75% and 74% respectively. Even though demographics, attendance, behavior and performance attributes were used in the study but accuracies of the algorithms were well below than that of obtained in present study. Agrawal (2015) used neural networks and the accuracies varied from 50% to 70% depending upon the training dataset size. In addition to demographics features, Abusaa (2016) used twenty one (21) attributes related to social, personal and academic data to predict students' grade using ID3, CART, CHAID, and Naïve Bayes algorithms. The accuracies were 33.33 %, 40 %, 34.07 % and 36.40 % respectively, which were lower than the accuracy of proposed research.

Mesarić and Šebalj (2016) analyzed enrolment and previous examination results for performance prediction using ID3, J48, RepTree, RandomTree and Random Forest algorithms. The RepTree algorithm gave best accuracy of 79% for one set of data, but its accuracy was not good for the second data set indicating inconsistent performance of the algorithm. However, the average accuracy of J48 was 60.7% which was lower than the accuracy of proposed research.

Conclusion: It was concluded that J48 algorithm was the best for predicting the students' grades as compared to Decision Table, OneR, Random Forest and Random Tree with the nine attributes related to students' profile and learning preferences. Such predictive models may be used for the quality assurance of academic programs.

REFERENCES

Abusaa, A. (2016). Educational data mining and students' performance prediction. Int. J. Adv. Com. Sci. App. 7(5): 212-220.

Abutair, M.M. and A.M. El-Halees (2012). Mining educational data to improve students' performance: a case study. Int. J. Comm. Tech. Res. 2(2): 140-146.

Agrawal, H. and H. Mavani (2015). Student performance prediction using machine learning. Int. J. Eng. Res. Tech. 4(3): 111-113.

Aguiar, E., H. Lakkaraju, N. Bhanpuri, D. Miller, B. Yuhas and K. L. Addison (2015). Who, when, and why: a machine learning approach to prioritizing students at risk of not graduating high school on time. Proc. 5th Int. Conf. LAK. ACM, 93-102.

Alam, F. and S. Pachauri (2017). Comparative study of J48, naive bayes and one-r classification technique for credit card fraud detection using weka. Adv. Comp. Sci. Tech. 10(6): 1731-1743.

Anuradha, C. and T. Velmurugan (2015). A comparative analysis on the evaluation of classification algorithms in the prediction of students performance. Ind. J. Sci. Tech. 8(15): 1-12.

Beaulac, C. and J.S. Rosenthal (2017). Predicting university students' academic success and choice of major using random forests. probability.CA. 1-18.

Bird, S., E. Klein and E. Loper (2009). Natural language processing with python: analyzing text with the natural language toolkit. O'Reilly Media, Inc.221-225.

Chen, H., R.H. Chiang and V.C. Storey (2012). Business intelligence and analytics: from big data to big impact. MIS quarterly. 36(4): 1165-1188.

Chen, X. (2017). Building hierarchies of probabilistic decision tables. In Proc. South East Conf. ACM, 142-143.

Chalaris, M., S. Gritzalis, M. Maragoudakis, C. Sgouropoulou and A. Tsolakidis (2014). Improving quality of educational processes providing new knowledge using data mining techniques. Procedia-Social and Beh. Sci. 147: 390-397.

Gray, G., C. McGuinness and P. Owende (2014). An application of classification models to predict learner progression in tertiary education. Adv. Comp. Conf. (IACC): 549–554.

Ilic, M., P.Spalevic, M. Veinovic and W.S. Alatresh (2016). Students' success prediction using weka tool. INFOTEH-JAHORINA. 15: 684-688.

Jiawei, H. and M. Kamber (2011). Data mining concepts and techniques. Elsvier. 1-4.

Katare, A. and S. Dubey (2017). A comparative study of classification algorithms in EDM using 2 level classification for predicting student's performance. Int. J. Com. App. 165(9): 35-40

Kaur, G. and W. Singh (2016). Prediction of student performance using weka tool. Int. J. Eng. Sci. 17: 8-16.

Koutina, M. and K.L. Kermanidis (2011). Predicting postgraduate students' performance using machine learning techniques. In Art. Int. App. Inn. Springer Berlin Heidelberg: 159-168.

Kovacic, Z. (2010). Early prediction of student success: mining students' enrolment Data. Proc. Int. Conf. InSITE. 647-665.

Lakshmi, T.M., A. Martin and Venkatesan (2013). An analysis of students performance using genetic Algorithm. J. Com. Sci. App. 1(4): 75-79.

Lantz, B. (2013). Machine learning with R. Packt Pub. Ltd: 5-27

Lodhi, B., S. Jalori and V. Namdeo (2011). Result analysis by decision tree, decision table and RBF network. J. Com. Math. Sci. 2(3): 399-580

Majeed, E.A. and K.N. Junejo (2016). Grade prediction using supervised machine learning techniques. Proc. 4th GSE Summit on Edu. 222-234.

Mayilvaganan, M. and D. Kalpanadevi (2014). Comparison of classification techniques for predicting the performance of students academic environment. Proc. Int. Conf. Comm. Net. Tech. (ICCNT), IEEE. 113-118.

Mesarić, J. and D. Šebalj (2016). Decision trees for predicting the academic success of students. Cro. Oper. Res. Rev. 7(2): 367-388.

Michalski, R.S., J.G. Carbonell and T. M. Mitchell, (Eds.) (2013). Machine learning: an artificial intelligence approach. Springer SBM. 3-6.

Moiz, A. (2015). A learner model for localized and adaptable e-learning. Diss. Def. AIOU. Islamabad, Pakistan.

Munoz, A. (2014). Machine learning and optimization. Courant Institute of Mathematical Sciences. New York, NY.

Musso, M.F., E. Kyndt, E.C. Cascallar and F. Dochy (2013). Predicting general academic performance and identifying the differential contribution of participating variables using artificial neural networks. Fron. Lear. Res. 1(1): 42-71.

Pandey, M. and V.K. Sharma (2013). A decision tree algorithm pertaining to the student performance analysis and prediction. Int. J. of Com. App. 61(13): 1-5.

Rangra, K. and K.L. Bansal (2014). Comparative study of data mining tools. Int. J. Adv. Res. in Com. Sci. Soft. Eng. 4(6): 216-223.

Ramesh, V., P. Parkavi and K. Ramar (2013). Predicting student performance: a statistical and data mining approach. Int. J. Com. App. 63(8): 35-39.

Singhal, S. and M. Jena (2013). A study on weka tool for data preprocessing, classification and clustering. Int. J. Inn. Tech. Exp. Eng. 2(6): 250-253.

Sutera, A. (2013). Characterization of variable importance measures derived from decision trees. Doctoral dissertation, Université de Liège, Liège, Belgique.8-12.

Usha, P. (2011). Predicting student performance using genetic and svm classifier. Int. J. 3(2): 97-102.

Wang, X., F. Guo, K.A. Heller and D. B. Dunson (2015). Parallelizing mcmc with random partition trees. In Adv. Neu. Inf. Pro. Sys. 451-459.

Yadav, S., B. Bharadwaj and S. Pal (2011). Data mining applications: a comparative study for predicting students' performance. Int. J. Inn. Tech. and Cr. Eng. 1(12): 13-19.

Zimmermann, J., K.H. Brodersen, J. P. Pellet, E. August and J. M. Buhmann (2010). Predicting graduate-level performance from undergraduate achievements. Proc. 4th Int. Conf. on EDM. 357-358.