Hybrid Feature Selection Framework to Support Learning Analytics in Educational Data Mining

Maryam Zaffar, Manzoor Ahmed Hashmani, Savita K. Sugathan

EDUCATIONAL DATA MINING

Quality education is the main responsibility of any university to its stakeholders. It is not only concern to the construction of higher level of knowledge but also upright provision of education so that students achieve their learning objectives without any problem [1]. One way of improving knowledge is to discover hidden information from the available data in various educational institutions. This new research field, concerned with exploring the information from educational settings to help all the educational stakeholders, is called Educational Data Mining (EDM) [2]-[3]. The point to ponder is that mining the gold from gold mines is referred as gold mining. Whereas the mining the data (information) from educational repositories (data) is termed as educational data mining. These educational databases act like gold mines for academic institutions. EDM objectives can be classified into two types [4] that are described as follows:

a. Academic Objectives: These objectives are related to the specific educational institutions related to redesigning of new courses according to real-time requirements. Furthermore, these are also related to the prediction of student performance, student learning, and student satisfaction analysis.

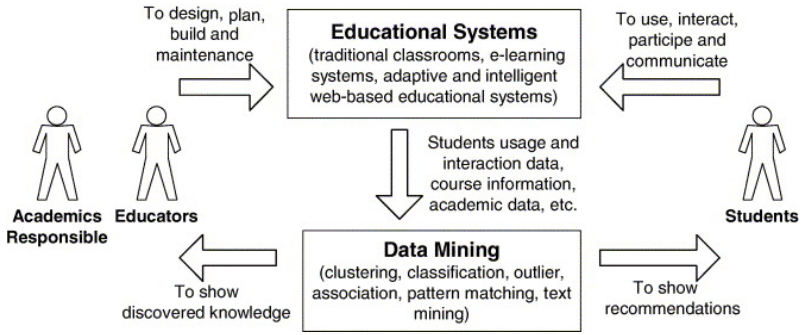

Figure 1 Educational data mining cycle [5]

b. Administrative Objectives: These objectives are related to the involvement of educational analysts, higher authorities of educational organisations and also to find out the challenges of EDM.

Data mining is a process of extracting useful information from the raw data and if the data belongs to educational settings the process is named as Educational Data Mining process. Figure 1 describes the process of educational data mining. The *Educational Systems* is a building block of development of a country with a large set of student data gathered by academic institutions, which can play as a gold mine role if properly processed. Educational institutions store a lot of data that can be helpful in making educational strategies. However, if the data is not processed well, all the efforts on the data will be useless.

EDUCATIONAL DATA MINING (EDM) AS A SOCIETAL (EDUCATIONAL) CHANGE AGENT

EDM plays a positive role in the construction of educated and refined society. Students are the future of any society, so it is very important to provide the students with quality education. The tools and techniques of the EDM not only help to present facts behind the performance of students, but these techniques also indicate the factors that highly affect the performance of students. EDM is an evolving brook where academics, researchers and students can use to check

their performance with the tools provided by EDM. Whereas, the academics can make the predictions and plan proactive decisions to improve the performance of students.

EDM is an interstation research area, which can also extract information regarding reasons of students' good or poor performance, and the effect of teacher's interest on student performance from educational databases. Besides that, the teachers can also plan their strategies of teaching according to the feedback received through EDM techniques. Thus, EDM can play a vital role in the developmental of any nation, which can cause positive societal changes.

TECHNIQUES OF EDUCATIONAL DATA MINING (EDM)

There are different EDM techniques, that may help in obtaining EDM main goals:

i. **Prediction**: Prediction is employed to derive a single variable from an accumulation of variables. The derived variable helps in analysing the data. A prediction model can be constructed through a number of statistical models or data mining models. The performance of classification and clustering models in terms of prediction accuracy is much better than statistical modelling [6]. For example, predictive data mining model for student's performance is essential to study the dominant factors on academic performance and also to know about slow learners. Prediction tries to figure out the patterns that allow prediction of performance or learning outcome based on the data from the available educational data.

ii. **Clustering**: Clustering is an unsupervised process and used to group the objects in different classes. In EDM, clustering is used to group the students according to their cognition [7].

iii. **Relationship Mining**: Relationship mining is utilized to determine the relationship between the variables in a dataset and derive the rules [7].

The main areas of research in EDM are based on learning environment such as formal, in-formal and blended learning. Different techniques are used to evaluate educational data from different learning environments. Formal/offline learning is the traditional mode of enlightening knowledge by face to face interaction and the data is extracted from the institutional databases [8]. Informal/e-learning provides online instruction, and the data is stored in log files and database. Furthermore, web mining techniques are used in dealing with data related to informal learning environments [9].

Major techniques used in EDM are as follows, association, classification, clustering and regression and sequential patterns. The classification methods are frequently used in EDM. Meanwhile, Neural Networks and Decision Tree methods are frequently used for predicting student performance [10]. Classification techniques are more appropriate for prediction of student performance than regression methods [11]. The classification efficiency can be improved by selecting the most relevant features out of a large number of features from a dataset [12].

Improvement of student predictive model is the principal goal of EDM. The predictive model helps in prediction of the performance of students in any educational institution [2],[4]-[5],[13]. EDM can also assist the educational institutions in making future plans, predicting the ratio of dropouts[14], identifying academically weak students, and plan regarding enrolment of students in particular courses etc. [1],[15]-[17]. The concepts of a large number of variables for the development of prediction model is not standardized at present. Consequently, feature selection techniques for the development of prediction model is getting prominence [18].

The focus of the educational institutions is to build their distinctive position in academic world, and to build a strategic plan for future actions. Therefore, the EDM techniques support the academic stakeholders in building new academic interventions in order to enhance the performance of students as well as an educational institution. This gives the attention to analyse academic data in order to support new interventions in teaching skills, and mitigations of important factors that may affect student academic performance. A lot of research are being done in this area but still, it requires more attention.

CHALLENGES IN EDUCATIONAL DATA MINING (EDM)

Data pre-processing is the first step of any data mining problem. Data pre-processing is described as any pre-processing performed on raw data to prepare it for further analysis. As we know, educational institutions store huge amount of data with large attributes. According to the study in [10], there is a need for investigation on pre-processing stage to improve the quality of data in education sector. However, to solve a particular educational domain problem the data usually is not in an appropriate form so that the results generated through the desired model may become useless and often face over fitting problem [20]. Here, the role of EDM pre-processing is to transform available raw educational data into a suitable format for a data mining algorithm to solve a particular educational domain problem.

Furthermore, the identification of a representative set data to develop student performance prediction model plays a vital role in making future educational strategies, so that the negligence in that step may cause severe effects on the educational constraints of a country.

The main challenge of an EDM is to pre-process student data to make sure not a single irrelevant student information is ignored and removed, in order to obtain the highly accurate prediction of student performance prediction models. Therefore, the main focus of our study presented in this book chapter is to improve the quality of educational dataset by proposing a framework at data pre-processing stage and to develop a prediction model to predict the students' grades. The resultant prediction model is expected not only to help the students, but also the educational institutions to make future decisions and strategies according to the results.

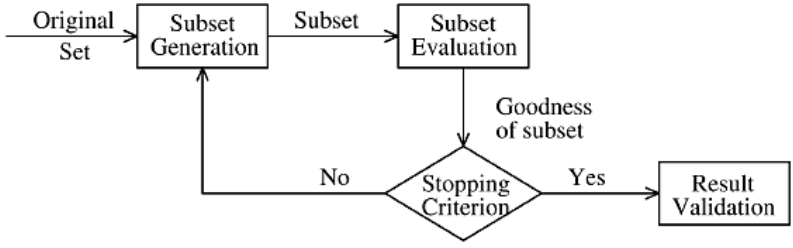

Figure 2 Basic steps of feature Selection Process [21]

FEATURE SELECTION ALGORITHMS AND EDM

The main aim of Feature Selection process is to find the most relevant features in any problem domain. Feature Selection Algorithms increase the prediction accuracy of the prediction model. But the identification of most relevant features is a tough task. There are four main steps of Feature Selection process; Subset Generation, Subset Evaluation, Stopping Criterion and Result Validation [18] as shown in Figure 2.

1. **Subset Generation:** Subset generation step is divided into two steps – Successor Generation and Search Organization. In Search Organization, a Search Strategy is used to generate the subset of variables. In general, these Search Strategies are called Random, Heuristic and Complete Search Strategies. Some operators measured for Successor Generations are Backward, Forward, Random, Weighing and Weighing operators [21].

2. **Subset Evaluation:** Different criteria are used for Subset Evaluation. There may be different subsets of features with different criteria. There are two main types of evaluation criteria. One type is Independent Criteria that is being used by filter feature selection model, as they do not include any learning algorithm. Some of them are Euclidean Distance, Entropy, Dependency Measure (Chi Square Statistics), and Consistency Measures. The second type, Dependent Criteria, require data mining algorithms. Wrapper model is used for such type of criteria [22].

3. **Stopping Criterion:** Some of the search used in Stopping Criteria is to stop the search for new subsets that are identified by Bagherzadeh et al. [18] and Kumar and Minz [22] as follows:
   Completion of search,
   a. When a threshold is reached (threshold can be maximum number of iterations or minimum number of variables)
   b. Next iteration fails to produce better results
   c. Acceptable subset is selected according to evaluation criterion

4. **Result Validation**: It's not part of feature selection process but feature selection process must be validated with executing and comparing some of the tests with earlier established results.

In educational context data pre-processing phase is most important step in whole EDM process [20]. Feature Selection Algorithms are the main techniques to be applied in data pre-processing stage. Feature selection is previously used in EDM, as to improve the classification accuracy of the student performance prediction models. However, the focus of the researchers was on Filter Feature Selection Algorithms.

In this chapter we have described some of the previous study focusing on Feature Selection Algorithms in EDM. There may be different features affecting the success and failures of students, as intra class attribute must be different from the attributes of the other class[23].

The Osmanbegović, Suljić and Agić [24] takes 1210 secondary school students dataset in order to predict their performance. Filter feature selection techniques, Information Gain and Gain Ratio, subsets that are applied on data to reduce the dimensionality of data. Highly ranked attributes are selected from both of these feature selection techniques. The model generated through J48 and Random Forest shows accuracy higher than 71%. The main focus of the study is limited to secondary school decision system. Furthermore, the size of dataset is not so high. But according to the study decision tree has over fitting problem. Meanwhile, the study in Chirunmilla et al. [25] takes a dataset of 120 university students from India to predict the performance and placement chances of students. The study uses Filter Feature Selection technique on pre-processing techniques (Information Gain) and C4.5 classification algorithm to build prediction model. The result reveals 77.78% precision of the model.

The study in Ramaswami and Bhaskaran [6] compares six Filter Feature Selection Algorithms to predict the performance of the 1969 higher secondary school students in India. The study evaluated the Feature Selection Algorithms on the basis of ROC curve value. Furthermore, it is observed from the result that F1-measure and ROC values for information gain feature selection method showed better results. While, the investigation in Doshi [26] used dataset of 380 first

year students from a Mumbai College India in order to predict the student admission in engineering stream. The study used four filter feature selection methods. The students build two prediction models, the first one was based on attributes from averaging Chi-Square, Info Gain and Gain Ratio filter feature selection methods, whereas the second model used FCBF (Fast Correlation-Based Filter). The results of the investigation showed that FCBF along with Instance based K-nearest neighbour (IBK) classifiers produce better results than other models. The study identified family income and father qualification as most important factors for student academic success.

Filter Feature Selection Techniques in EDM

Filter Feature Selection is a very important type of Feature Selection Algorithms to apply in pre-processing stage of data mining. Calculation time of Filter Feature Selection Algorithm is less. These do not use any classification algorithm for selecting the most relevant features from a dataset [12]. The Filter Feature Selection Algorithms are preferred because these can perform better with any classification algorithm since they required less computational complexity. Table 1 shows the related literature on filter feature selection algorithms used to predict the performance of students in EDM.

Wrapper Feature Selection Method in EDM

Wrapper feature selection is not fast when applied to huge datasets, but it gives assured results. On average, the accuracy of the wrapper model is much better than the filter model. It evaluates the combination of feature subsets and can identify the relations between all the features simultaneously. Wrapper models are better than filter as in wrappers the attribute selection is optimized for a model being used. Table 2 presents different studies in EDM to predict the performance of students.

Hybrid Feature Selection Method

A recent study by Shahiri & Husain [27] used a hybrid feature selection algorithm to predict the performance of students. The study used student dataset from University California Irvine (UCI) repository [28]. The study used a hybrid model of Fast Correlation Based Feature

Selection (FCBF) and Wrapper Sequential Forward Selection (WSFS) on pre-processing stage in order to improve the quality of data. Table 3 presents the previous study on predicting the performance of students applying wrapper feature selection algorithms. There are advantages as well as disadvantages of both filter and wrapper feature selection methods. A hybrid model tries to combine the advantages of both techniques.

Table 1 Study of filter Feature Selection algorithm in EDM

Feature Selection Algorithm Type	Algorithm	Results	Drawbacks
Filter Feature Selection Algorithm	Correlation Based (CBF) [6], [31]-[32]	58.4% 70% 84.2%	CFS generates discrepant results that make difficulty in discovering right features [29]. CBF gives accurate results on small datasets; it cannot handle the numeric values [30].
	IGATE Gain Ratio [6], [31]-[32]	58.8% 76% 74.4%	
	Chi squared (CHI) [6], [31]	59.2% 73%	Ignores interaction with classifiers.
	Symmetrical Uncertainty [6]	59.2%	
	Relief Attribute [6]	58.2%	It cannot eliminate the redundant feature [30].
	Information Gain [6], [25]	59.2% 77.8%	Information gain ignores feature dependencies, ignores interaction with classifiers [12].
	Fast Correlation-Based Filter [26]	87.5%	FCBS is not much scalable. Ignores interaction with classifiers [26]. Furthermore it cannot handle feature redundancy [30].

Table 2 Study of wrapper Feature Selection algorithm in EDM

Feature Selection Algorithm	Algorithm	Result	Drawbacks
Wrapper	Voting Frequency Intervals [33]	81.33	Wrapper feature selection methods have high computational cost for large dataset. Whereas the execution is much slower than filter methods. Furthermore, they are dependent on classification algorithms.
	Naïve Bayes [31]	75	
	1-Nearest Neighbour(1-NN) [31]	76	

Table 3 Study of hybrid Feature Selection algorithm in EDM

Algorithm	Result	Drawbacks
FCBF+WSFS [27]	88%	WSFS is not being able to re-evaluate feature usefulness together with other features, once a feature is removed. It has a risk of over fitting, and its solution is not considered to be optimal [34]. Whereas FCBF is not much scalable. The prediction Accuracy needs to be improved. The classification accuracy of the model is not mentioned.

Contribution in EDM

This research study is primarily conducted on Feature Selection Algorithms, i.e., to choose the best feature sets from student data to help the educational institutions in taking proactive decisions. According to the proposed model, a student dataset is taken from the academic repository and then a suitable hybrid feature selection technique is

applied to enhance the quality of the dataset and to account for the scalability issue.

Framework of Proposed Methodology

The main workflow of this study is divided into three main steps. The study conducted as explained below.

Table 4 Description of datasets used in the study

Dataset	Number of Instances (Students)	No of Features
Dataset 1 [35]	500 students record	16
Dataset 2 [36]	344 students record	12
Dataset 3 [37]	300 students record	21

Description of Datasets

Three benchmark student datasets are collected to conduct the study. The information about the datasets is explained in Table 4. Table 4 describes the source of datasets, the number of instances and number of features in the datasets. Three different benchmark students datasets with variations in number of instances and number of features are taken from the sources mentioned in Table 4.

Comparing Different Filter Feature Selection on Student Datasets

Different filter feature selection algorithms are applied in this study to evaluate best performance of the feature selection algorithms. Due to the limitations of space we present the most appropriate ones to highlight the deducted points. Two classification algorithms MLP (Multi-layer Perceptron) and Support Vector Machine (linear) are selected to calculate the prediction accuracy of the filter feature selection algorithms. MLP is a classification algorithm with good performance in terms of prediction accuracy on different student performance prediction models[31], [38- 39]. Whereas, Support Vector Machine (SVM) is a classification algorithm that has been used in many prediction models [38]-[39] in the area of EDM.

Table 5 Prediction accuracies of filter Feature Selection algorithms on three datasets

Filter Feature Selection Algorithm	Classification Algorithm	Dataset 1	Dataset 2	Dataset 3
FCBF	SVM	68.75%	85.51%	42.50%
	MLP	26.56%	74.64%	24.17%
Chi-Square	SVM	75.60%	85.51%	46.67%
	MLP	60.00%	40.60%	38.33%

Table 5 presents the prediction accuracies of filter feature selection algorithms FCBF and Chi-Square against two classification algorithms, SVM and MLP, on three datasets. FCBF is a feature selection algorithm that tries to find out the best subset of features based on the goodness of features [41]. FCBF has been used in different research studies [42]-[43] in the area of EDM to predict the performance of students.

Whereas, Chi-Square algorithm is a uni-variant Feature Selection Algorithm. Several of studies [6], [44]-[46] in EDM use this algorithm as Feature Selection Algorithm for predicting the performance of students.

Figure 3 is a clear graphical representation of results of Table 5. Figure 3 clearly shows that SVM classifier performs better than MLP on all three datasets. Whereas the performance of Chi-Square feature selection algorithm is better as compared to FCBF. Meanwhile, Table 6 presents a review of the results of Chi-Square Feature Selection Algorithm on Dataset 1. In this table, we present all details of extracted number of features and classification accuracy of Filter Feature Selection on Dataset 1. The result in Table 6 shows that the performance of Chi-Square is better than FCBF and that SVM classification algorithm performs better than MLP classification algorithm.

Table 6 Results of Chi-Square filter Feature Selection algorithm on Dataset 1

Filter Feature Selection Algorithm	Dataset	Total Number Of features	Features Extracted by Filter	List of Features	Classification Accuracy
Chi-Square	Dataset 1	16	11	0, 2, 6, 8, 9, 10, 11, 12, 13, 14, 15	75.60%



Figure 3 Comparison of performance of FCBF and Chi-Square on students' datasets

Comparing the Results of Wrapper Feature Selection Algorithm

Wrapper Sequential Forward Search (SFS) algorithm is also used in this study to conduct the tests. As it is previously [47] used in EDM and shows better results in terms of prediction accuracy on student data. Therefore, we selected SFS as wrapper algorithm in our study. SFS starts with an empty set of features. With every iteration features are selected and combined with the features on new dataset.

Table 7 Results of wrapper Feature Selection algorithm on Dataset 1

Wrapper Feature Selection Algorithm	Dataset	Total Number Of features	Features Extracted by Wrapper	List of Features	Classification Accuracy
SFS	Dataset 1	16	13	1, 2, 3, 4, 5, 6, 8, 9, 10, 11, 13, 14, 15	73.44%

The results in Table 7 show that there is no increase in prediction accuracy when using SFS as wrapper algorithm. The number and types of features are different as selected by the filter feature selection algorithm (*refer* Table 6). However, we cannot neglect the advantages of wrapper feature selection algorithm.

Proposed Methodology: Hybrid Feature Selection Framework

Feature Selection Algorithms have two main types as discussed in above sections, i.e., Filter Feature Selection algorithms and Wrapper Feature Selection algorithms. We cannot ignore the advantages and disadvantages of Filter Feature Selection Algorithms and Wrapper Feature Selection Algorithms. This research is an effort to enhance the feature selection technique by combining Filter and Wrapper Feature Selection Algorithms. In this chapter, we propose a Hybrid Feature Selection framework for EDM. Figure 4 represents the main flow of our proposed methodology. A student dataset having *N* number of features is taken in first step. Then the Filter Feature Selection Algorithms, Chi-Square, is applied on the selected dataset to retrieve the most suitable feature set that may affect the performance of students. These selected features may be Selected Features {Fn, Fm,…}. Wrapper Feature Selection Algorithm is then applied on these selected set of features to overcome the shortcoming of the Filter Feature Selection Algorithms. As a result, final set of selected features is obtained. SVM classification algorithm is then applied to find the prediction accuracies. Retrieved results give a better insight on the features affecting the performance of students.

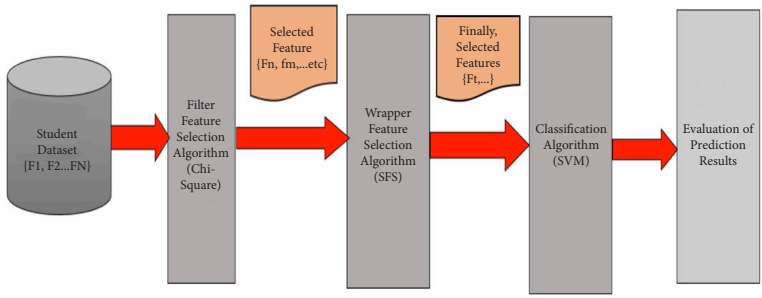

Figure 4 Framework of proposed methodology

As an experimental study the proposed Hybrid Feature Selection is applied on Dataset 1. Table 8 summarizes the results of proposed framework on Dataset 1. Comparing the classification accuracy results of Table 8 with the results in Table 6 and Table 7, it is observed that the prediction accuracy of our proposed approach is better.

In future, we will also publish the results of our proposed framework on different datasets. Table 8 shows that 10 features are selected by our proposed framework.

Table 8 Results of proposed framework on Dataset 1

Dataset	Total Features	Feature Selected through Filter Feature Selection	Features Selected through Wrapper Feature Selection	Classification Accuracy
Dataset 1	16	11 (0 2 6 8 9 10 11 12 13 14 15)	10 (0 2 6 8 9 10 11 12 13 15)	76.04%

SUMMARY

Educational Data Mining (EDM) is an area of research not only helpful for all educational stakeholders, but also this can play an important role in the development of a nation. Good quality education is a prominent factor in the development of nation because it produces skilled, capable, and problem-solving brilliant graduate. These graduates can survive better in the society and contribute towards the development of a nation. EDM helps educational institutions in giving students a quality education, by identifying the factors affecting the performance of students. EDM also helps educational stakeholders in taking proactive decisions by applying different techniques on the educational data. The role of educational data is very important in EDM. It is very important to refine the quality of educational data, to take right decisions. As a wrong selection of feature from educational data, may lead to wrong results and these results not only waste the time and money of educational stakeholders but also may contribute negatively towards the development of a nation. But the appropriate selection of features not only helps the educational decision makers in planning the strategies to increase the quality of education but also development of any nation. Predicting the academic performance of students is one of the most important applications of EDM. In this research we focused on increasing the quality of educational data to increase the prediction accuracy of student performance prediction model. There are two main types of feature selection algorithms, i.e., Filter and Wrapper Feature Selection algorithms. Educational data miners cannot ignore the pros and cons of both filter and wrapper algorithms to retrieve a better dataset for learning analytics. Our proposed framework uses a combination of filter and wrapper algorithms and develops a hybrid feature selection algorithm. Our framework shows better results as compared to existing approaches. Our research can be helpful for researchers and educational analysts to better investigate the factors affecting the performance of students and to be helpful in developing strategies, developing curriculum and training the teachers accordingly.

REFERENCES

[1] S.K. Yadav & S. Pal, "Data mining application in enrollment management: A case study*", International Journal of Computer Applications*, 41, 5, 2012.

[2] M.M.A. Tair & A.M. El-Halees, "Mining educational data to improve students' performance: a case study*", International Journal of Information*, 2, 2, pp. 140-146, 2012.

[3] M.Chalaris, S. Gritzalis, M. Maragoudakis, C. Sgouropouloum & A. Tsolakidis, "Improving quality of educational processes providing new knowledge using data mining techniques", *Procedia-Social and Behavioral Sciences*, 147, pp. 390-397. 2014.

[4] R. Jindal & M.D. Borah, "A Survey on Educational Data Mining and Research Trends", *International Journal of Database Management Systems*, 5, 3, pp. 53, 2013..

[5] Romero, C. & S. Ventura, "Educational data mining: A survey from 1995 to 2005", *Expert systems with applications*, 33, 1, pp. 135-146, 2007.

[6] M. Ramaswami & R. Bhaskaran, "A study on feature selection techniques in educational data mining", *arXiv,* abs/0912.3924, 2009.

[7] M. Anoopkumar & A.M.Z. Rahman, "A Comprehensive Survey on Educational Data Mining and Use of Data Mining Techniques for Improving Teaching and Predicting Student Performance", In Advances in Innovatice Engineering and Technologies, 2015. Doi: 10.18797/CAASR/ICIET/2015/11/27/06

[8] K.Sukhija, M. Jindal & N. Aggarwal, "The recent state of educational data mining: A survey and future visions", In *MOOCs, Innovation and Technology in Education (MITE), 2015 IEEE 3rd International Conference*, IEEE, pp. 354-359, 2015.

[9] C. Romero & S. Ventura, "Educational data mining: a review of the state of the art", *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews*), 40, 6, pp. 601-618, 2010.

[10] A.M. Shahiri & W. Husain, "A review on predicting student's performance using data mining techniques", *Procedia Computer Science*, 72, pp. 414-422, 2015.

[11] R.L. Wilson & B.C. Hardgrave, "Predicting graduate student success in an MBA program: Regression versus classification", *Educational and Psychological Measurement*, 55, 2, pp. 186-195, 1995.

[12] M. Doshi & S.K. Chaturvedi, "Survey of Feature Selection Algorithms in Higher Education", *International Journal of Computer Applications in Engineering Sciences*, 4, 1, p. 5-8, 2014.

[13] V. Ramesh, P. Parkavi & P. Yasodha, "Performance analysis of data mining techniques for placement chance prediction", *International Journal of Scientific & Engineering Research*, 2, 8, pp. 1, 2011.

[14] T. Hasbun, A. Araya & J. Villalon, "Extracurricular activities as dropout prediction factors in higher education using decision trees", *Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conferenc*e, IEEE, 2016.

[15] O. Akinola, B. Akinkunmi & T. Alo, "A Data Mining Model for Predicting Computer Programming Proficiency of Computer Science Undergraduate Students", African Journal Computing & ICT, 5, 2, pp. 43-54, 2012.

[16] Z.J Kovacic, "Early prediction of student success: Mining students' enrolment data", Proceeding of Informing Science & IT Education Conference (InSITE), 2010.

[17] L. Ramanathan, S. Dh & S. Kumar, "Predicting Students' Performance using modified ID3 algorithm", *International Journal of Engineering aand Technology*, 5, 3, 2013.

[18] F. Bagherzadeh-Khiabani, A, Ramednzankhani, F. Azizi, F. Hadeegh, EW. Steyerberg & D. Khalili, "A tutorial on variable selection for clinical prediction models: feature selection methods in data mining could improve the results", *Journal of Clinical Epidemiology*, 71, p. 76-85, 2016.

[19] D. Kabakchieva, "Predicting student performance by using data mining methods for classification", *Cybernetics and Information Technologies,* 13, 1, pp. 61-72, 2013.

[20] C. Romero, J.R. Romero & S. Ventura, "A Survey on Pre-Processing Educational Data, *Studies in Computional Intelligence*, Springer: Switzerland, 2014.

[21] L. Ladha & T. Deepa, "Feature selection methods and algorithms. International journal on computer science and engineering", 1, 3, pp. 1787-797, 2011.

[22] V. Kumar & S. Minz, "Feature Selection", *SmartCR*, 4, 3, pp. 211-229, 2014.

[23] M. Aslam, L.B.A. Rahim, J. Watada, and M. Hashmani,, "Clustering-Based Cloud Migration Strategies,". *Journal of Advanced Computational Intelligence, and Intelligence Informatics*, 22, 3, pp. 295-305, 2018.

[24] Osmanbegović, E., M. Suljić & H. Agić, "Determining dominant factor for students performance prediction by using data mining classification algorithms", *Tranzicija*, 16, 34, pp. 147-158, 2015.

[25] V. Chirumamilla, "Novel approach to predict student placement chance with decision tree induction*", International Journal of Systems and Technologies,* 7, 1, pp. 78-88, 2014.

[26] M. Doshi, "Correlation Based Feature Selection (Cfs) Technique To Predict Student Perfromance*", International Journal of Computer Networks & Communications,* 6, 3, pp. 197, 2014.

[27] A.M. Shahiri & W. Husain, "Handling High Dimensional Educational Data using Feature Selection Techniques*", Journal of Telecommunication, Electronic and Computer Engineering (JTEC),* 9, 2-12, pp. 89-93, 2017.

[28] P. Cortez & A.M.G. Silva, "Using data mining to predict secondary school student performance", 2008, http://www3.dsi.uminho.pt/pcortez/ student.pdf

[29] K.D. Rajab, "New Hybrid Features Selection Method: A Case Study on Websites Phishing", *Security and Communication Networks*, 2017. doi.org/10.1155/2017/9838169

[30] K. Sutha & J.J. Tamilselvi, "A review of feature selection algorithms for data mining techniques", *International Journal on Computer Science and Engineering*, 7, 6, pp. 63, 2015.

[31] A. Acharya & D. Sinha, "Application of Feature Selection Methods in Educational Data Mining*", International Journal of Computer Applications*, 103, 2, pp 34-38, 2014.

[32] C. Anuradha & T. Velmurugan, "Feature Selection Techniques to Analyse Student Acadamic Performance using Naïve Bayes Classifier", Presented in *The 3rd International Conference on Small & Medium Business*, 2016.

[33] H.M. Harb & M.A. Moustafa, *Selecting optimal subset of features for student performance model.* International Journal Computer Sciences Issues, 9, 5, 1, 2012.

[34] R. Aziz, C. Verma & N. Srivastava, *Dimension reduction methods for microarray data: a review.* AIMS Bioengineering, 4, pp. 179-197, 2017. Doi:10.3934/bioeng.2017.2.179

[35] E.A. Amrieh, T. Hamtini & I. Aljarah, "Mining Educational Data to Predict Student's academic Performance using Ensemble Methods*", I*nternational *Journal of Database Theory and Application*, 9, 8, pp. 119-136, 2016.

[36] G. Kostopoulos, S. Kotsiantis & P. Pintelas, "Estimating student dropout in distance higher education using semi-supervised techniques", *Proceedings of the 19th Panhellenic Conference on Informatics*, 2015.

[37] S. Hussain, N.A. Dahan, F.M Ba-Alwi & N. Ribata, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA*", Indonesian Journal of Electrical Engineering and Computer Science*, 9, 2, 2018.

[38] A. Mueen, B. Zafar & U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques*", International Journal of Modern Education and Computer Science,* 8, 11, p. 36-42, 2016.

[39] S.K.Wanjau, *Data Mining Model for Predicting Student Enrolment in STEM Courses in Higher Education Institutions.* 2016.

[40] W.-X. Liu & C.-H. Cheng, "A hybrid method based on MLFS approach to analyze students' academic achievement", In *12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD) Proceeding,* IEEE, 2016.

[41] L. Yu & H. Liu. *Feature selection for high-dimensional data: A fast correlation-based filter solution*. in *Proceedings of the 20th International Conference on Machine Learning,* 2003.

[42] A. Gopalakrishnan, "A multifaceted data mining approach to understanding what factors lead college students to persist and graduate", *Computing Conference 2017*, IEEE, 2017.

[43] M. Mavrikis, "Modelling student interactions in intelligent learning environments: constructing bayesian networks from data*", I*nternational *Journal on Artificial Intelligence Tools*, 19, 6, pp. 733-753, 2010.

[44] A.Kavitha, R. Kavitha J.V. Gripsy, "Empirical Evaluation of Feature Selection Technique in Educational Data Mining*", ARPN Journal of Science and Technology*, 2, 11, 2012.

[45] N. Rachburee & W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," *Information Technology and Electrical Engineering (ICITEE) 2015 7th International Conference,* IEEE, 2015.

[46] M. Zaffar, M.A. Hashmani & K. Savita, "Performance analysis of feature selection algorithm for educational data mining", *Big Data and Analytics (ICBDA) 2017 IEEE Conference* 2017, IEEE, 2017.

[47] A.M. Shahiri, W. Husain & N.A.A. Rashid, "A proposed framework on hybrid feature selection techniques for handling high dimensional educational data", *AIP Conference Proceedings* 2017, AIP Publishing, 2017.