
Production Systems and Information Engineering
Volume 9 (2020), pp. 36–53

STUDENT ACADEMIC PERFORMANCE PREDICTION

Jawad Alshboul
University of Miskolc, Hungary
Department of Information Technology
alshboul.jawad@student.uni-miskolc.hu

Erika Baksa-Varga
University of Miskolc, Hungary
Department of Information Technology
vargae@iit.uni-miskolc.hu

[Received ... and accepted ...]

Abstract. Given the increasing number of students who attend traditional and non-traditional classes that deploy internet-based educational resources and environments, large volumes of data are being generated on a daily basis. As a result, more researchers are now working with Educational Data Mining (EDM) methods to understand learning processes and behaviors of learners. The problem that led to this research is the need to make use of unused data that is collected during education and learning processes by gaining insights in order to support students in regards to their academic performance and in taking actions to prevent or warn students from failure. The main focus of this research is on how EDM can support student learning in regards to student academic performance, engagement, and intervention. The research mainly addresses the appropriate EDM methods used to predict student academic performance. Modeling and evaluation of several classifiers were conducted. As a result, Random Forest classifier has been chosen as the best model to be deployed in an interactive R Shiny application.

Keywords: educational data mining, academic performance prediction, classification

1. Introduction

Due to the growing use of educational resources and technologies, educational data are being generated in huge amounts on a daily basis. Data-driven decision making (DDDM) refers to the systematic processes of collection, analysis, and interpretation of data to help in decision making [1]. Educational data

can be used in DDDM at different educational levels to achieve effective educational data decision making. DDDM for educational data can be related to educational resources, and human resources decisions.

Data-Driven Education enables institutions to leverage educational data to get insights about teaching-learning process and to make data-driven educational decisions based on student needs [2]. DDDM includes exploiting available data, such as the kind offered in virtual learning environments or Learning Management Systems (LMS), to make teaching decisions [3, 4]. Values underlying educational data mining are to analyze student learning data and its contexts in order to better understand and personalize student learning experiences [5, 6].

According to [7], Data Mining (DM) is a computerized information system dedicated to handle large amounts of data, produce information, and discover hidden patterns. The demand on using DM in educational settings led to the establishment of Educational Data Mining (EDM) as a new field of knowledge and line of research [8]. The growing acceptance of the emergent field, EDM, is due to its ability to elicit valuable insights from data for either students or staff [9]. The authors in [10] define EDM as a multidisciplinary field of study that combines skills and knowledge from machine learning, statistics, DM, psychology, information retrieval, cognitive science, and recommender systems techniques to support resolving issues related to education.

The main reason for the late emergence of data mining in education, compared to all other fields, was that the availability of large educational datasets in machine readable formats emerged later in education [11].

2. Educational Data Mining Methods and Applications

2.1. Regression Techniques

There is a number of regression algorithms such as: Single Linear Regression, and Multiple Linear Regression. The regression technique is used to predict values and it has been applied in education domain to: predict students' grades [12], and predict academic GPA of graduated student [13].

2.2. Classification Techniques

There is a number of classification algorithms such as: Decision Trees, Neural Networks, Logistic Regression, and Nave Bayes classifiers. Classification

techniques have been applied in education domain to: analyze the academic performance of undergraduate students [14], assess how effective EDM techniques are for students early prediction failure [15], and develop a model to prevent academic dropout [16].

2.3. Clustering Techniques

There are different clustering algorithms such as: K-Means and Expectation Maximization. Clustering techniques have been applied in education domain to: generate a model for student dropout by exploring student categories and characteristics [17], group university students into careers by analyzing their performance and outcomes of the self-evaluation test beginning from their first year [18], associate students and teachers [19], and group competent students of an educational institution in regards to their skills and abilities [20].

2.4. Association Rules Techniques

The association rule mining techniques are applied to identify associations or dependencies between attributes in the datasets [21]. Association Rules have been applied in education domain to: propose a quantifiable measure that shows degradation in regard to students expected performance [22], analyze students' performance based on real time patterns in students' data [23], investigate on association between self-esteem and performance of students [24], and discover the impact of teaching on improving how student performs [9].

2.5. Social Network Analysis and Visualization Techniques

Social Network Analysis (SNA) and Visualization can reduce the size of the datasets and their complexity in case they are multidimensional datasets. SNA and visualization have been applied in education domain to: introduce a model based on visual analytics, and learning analytics, in addition to a tool, to perform confirmatory and exploratory data analysis through interaction between information gathered [25], process the interaction networks of students in a forum [26], and check the progress of online collaborative learning and provide informed interventions when needed [27].

2.6. Process Mining Techniques

Process Mining techniques are used to deal with log files and events and they have been applied in education domain to: analyze events flow logs in an adaptive learning model [28, 29], and provide feedback on the basis of behavioral data [30].

2.7. Text Mining Techniques

Text Mining techniques are used to deal with unstructured data by capturing key terms and uncover hidden patterns. Text Mining techniques have been applied in education domain to: analyze students' online interaction via online questions and chat messages [31], extract knowledge from students' evaluation comments that help instructors and administrators obtain understanding of student sentiments and views [32], and rate educational institute faculty members based on the feedback submitted by students [33].

2.8. Outlier Detection Techniques

Outlier Detection is used to check whether there is any deviation in any observation away from all other observations using data mining algorithms based on association, classification, clustering, visualization, or statistics-based approach. It has been used in education domain to: discover any anomaly or abnormal observations [34, 35], and predict dropouts by clustering outlier data with unsupervised learning [36].

2.9. Student Academic Performance Prediction

Student academic performance prediction has been an important research topic for years since students and institutions can benefit from discovering patterns and insights hidden within learning data. Institutions can benefit from it by improving the effectiveness of academic facilities available to their students in order to increase the rates of students who are successful in completing their programs or courses of study. Furthermore, findings can be used to deliver solutions, suggestions, or advices to students to enhance how they perform in the future.

A review of literature on the methods used for student academic performance prediction was conducted. The search focused on Scopus, IEEE, Google Scholar, and ACM for years 2009 to 2019. The number of relevant articles used for the synthesis, after excluding the articles that did not describe the data sets attributes or methods used, is 157 articles.

Table 1 shows some statistics related to the modeling techniques that have been used in the studies related to performance prediction. It is shown that the mostly used classification modeling techniques are: Decision Trees, Bayesian-Based, Neural Networks, Support Vector Machines, Ensemble Methods, K-Nearest Neighbor, and Logistic Regression, respectively.

Table 1. Statistics of modeling techniques used

Modeling Technique	Count	Percentage
Decision Trees	79	50.3 %
Bayesian-based	65	41.4 %
Neural Networks	43	27.4 %
Support Vector Machines	34	21.7 %
Linear Regression	34	21.7 %
Ensemble Methods	29	18.5 %
K-Nearest Neighbor	24	15.3 %
Logistic Regression	22	14.0 %
Others (Hybrid, optimization, statistical, ..etc.)	12	7.6 %
Rule Induction	9	5.7 %

3. Research Methodology

3.1. Research Objectives and Research Questions

The research question (RQ) for each research objective (RO) is explained as follows:
RO1. To investigate the appropriate educational data mining methods used in predicting student academic performance.
RQ1. What are the appropriate techniques that are used to predict student academic performance?
RO2. To apply educational data mining methods to support academic intervention.
RQ2. How can educational data mining be used to support academic intervention?

3.2. Data Collection and Preparation

A quantitative dataset [37] has been chosen based on the literature review conducted to get answers for research questions. The significance of this dataset is due to adopting student behavioral features with academic data during the learning process.

The dataset used was collected from a multi-agent Learning Management System (LMS) called Kalboard 360 using Experience API (xAPI) web service [37].

An activity tracker tool called experience API (xAPI) was used to track learners. The dataset shown in Table 2 consists of 480 student records and 16

Table 2. Student academic performance dataset description

Category	Feature	Data Type	Description
Demographical Information	Nationality	Nominal	Student Nationality
	Gender	Nominal	Student Gender (female or male)
	Place of Birth	Nominal	Student Place of Birth (Jordan, Kuwait, Lebanon, Saudi Arabia, Iran, USA)
	Parent Responsible	Nominal	Student Parent (father or mum)
Academic Information	Educational Stages (School Levels)	Nominal	Stage student belongs such as (primary, middle and high school levels)
	Grade Levels	Nominal	Student Grade (G-01 → G-12)
	Section ID	Nominal	Student Classroom (A, B, C)
	Semester	Nominal	School Year Semester (First or Second)
	Topic	Nominal	Course Topic or Subject (Math, English, IT, Arabic, Science, Quran)
	Student Absence Days	Nominal	Student Days of Absence (Above-7, Under-7)
Parents Participation in Learning Process	Parent Answering Survey	Nominal	Parent Answering School Surveys or Not.
	Parent School Satisfaction	Nominal	Parent Satisfaction Degree about School (Good, Bad)
Behavioral Information	Discussion Groups	Numerical	Student Behavioral Interaction with E-Learning System.
	Visited Resources	Numerical	
	Raised Hand on Class	Numerical	
	Viewing Announcements	Numerical	
Target	Student Mark	Ordinal	L: Low-Level values from 0 to 69. M: Middle-Level values from 70 to 89. H: High-Level values from 90-100.

features in addition to the target variable which represents student academic performance. The 16 features are grouped into four categories: features that constitute Demographic Information, features that constitute Academic Information, features that constitute Parents Participation in Learning Process Information, and features that constitute Behavioral Information.

3.2.1. Data Preprocessing

Checking missing values, renaming some attributes, data type conversion of some attributes to factors, and correcting some misspelled country names by using the standard names were required to prepare data for the next steps.

3.2.2. Feature Selection

Feature Selection is the process used for selecting those dataset features that will contribute most to the prediction task. Correlation matrix and recursive selection are used for this purpose.

1. Correlation Matrix: Figure 1 shows the output for feature selection based on correlation matrix setting with a cutoff 0.75 for highly correlated attributes. It is clear that there are no two features correlated with at least 0.75 to be considered highly correlated.
2. Recursive Selection: Recursive Feature Elimination (RFE) method is used to identify the features that can be eliminated without affecting the accuracy of classification models. Figure 2 and Figure 3 show the result of performing REF on the features of the data set. Based on the result of performing REF on the features of the data set, it is possible to either keep or remove semester feature from the data set. Consequently, it is kept for the further phases in modeling. As a result, the 16 features will be used to build models for predicting student academic performance.

	raisedhands	VisITedResources	AnnouncementsView	Discussion
raisedhands	1.0000000	0.6915717	0.6439178	0.3393860
VisITedResources	0.6915717	1.0000000	0.5945000	0.2432918
AnnouncementsView	0.6439178	0.5945000	1.0000000	0.4172900
Discussion	0.3393860	0.2432918	0.4172900	1.0000000

Figure 1. Correlation matrix result

```
Recursive feature selection

Outer resampling method: Cross-Validated (10 fold)

Resampling performance over subset size:

 Variables Accuracy  Kappa AccuracySD KappaSD Selected
         1   0.5165 0.2468    0.03786 0.06744
         2   0.6545 0.4751    0.08670 0.13188
         3   0.6980 0.5389    0.07575 0.11413
         4   0.7143 0.5585    0.05299 0.08181
         5   0.7668 0.6410    0.04381 0.06705
         6   0.7669 0.6415    0.05429 0.08268
         7   0.7751 0.6531    0.05673 0.08574
         8   0.7812 0.6629    0.03714 0.05495
         9   0.7958 0.6842    0.05619 0.08642
        10   0.8022 0.6946    0.05460 0.08342
        11   0.7959 0.6851    0.06481 0.09831
        12   0.8021 0.6937    0.06229 0.09508
        13   0.8208 0.7228    0.05031 0.07645
        14   0.8104 0.7076    0.05514 0.08357
        15   0.8147 0.7138    0.04487 0.06779
        16   0.8250 0.7302    0.05310 0.08047        *

The top 5 variables (out of 16):
   StudentAbsenceDays, VisitedResources, RaisedHands, AnnouncementsView, ParentAnsweringSurvey

> # list the chosen features
> predictors(results)
 [1] "StudentAbsenceDays"       "VisitedResources"         "RaisedHands"
 [4] "AnnouncementsView"        "ParentAnsweringSurvey"    "Relation"
 [7] "Topic"                    "Discussion"               "Nationality"
[10] "ParentschoolSatisfaction" "PlaceofBirth"             "Gender"
[13] "GradeID"                  "StageID"                  "SectionID"
[16] "Semester"
```

Figure 2. Recursive feature elimination performed on the data set

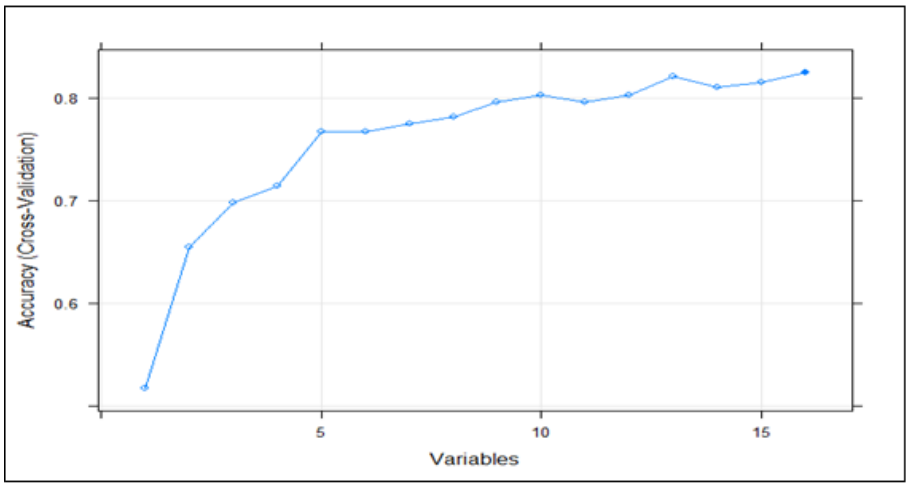

Figure 3. Accuracy based on importance of features in modeling classifiers

3.3. Modeling

The classification techniques used are: Decision Trees, Naive Bayes, K-Nearest Neighbors, Support Vector Machines, Logistic Regression, Ensemble Methods (Random Forest), and Neural Networks.

3.3.1. K-Nearest Neighbor (KNN)

It is a non-parametric, instance-based supervised learning algorithm, and easy to implement and understand but computationally expensive. KNN algorithm works as follows:

1. Choose K as the number of neighbors.
2. Select the K nearest neighbors of the unknown data point based on their Euclidean distances (Or other distance measure if the points are categories) from that unknown data point.
3. From the selected K neighbors, compute the number of data points in each category.
4. Allocate the new data to the category that has the largest count of neighbors among other categories.

3.3.2. Decision Tree (DT)

It is a supervised learning algorithm and It works for both continuous and categorical data. It splits the nodes based on all features then selects the

appropriate split using some criteria such as Gini index, Information Gain, and Variance. DT algorithm works as follows:

1. Place all training examples at the root node.
2. Categorize the data set attributes.
3. Split examples based on specific selected attributes.
4. Select test attributes based on a specific measure like statistics or heuristic.
5. Stop when: all examples being members of the same class, no attributes left for partitioning, or no examples left for classification.

3.3.3. Support Vector Machines (SVM)

They are supervised learning algorithms that can be used for both classification and regression problems. SVM algorithm works in the following steps:

1. Plot each data point in an m-dimensional space where m represents the number of attributes in the dataset.
2. Perform classification by trying to find the suitable hyperplane that separates the classes.

3.3.4. Logistic Regression (LR)

It is a supervised learning algorithm that can be used for binary classification but can deal with multi-class classification problems as well by using one-vs-all principle. A logistic function called sigmoid function is used to map the outputs to probabilities. One-vs-all classification is performed by training M distinct binary classifiers in which each trained binary classifier can recognize a particular class. Consequently, those M classifiers are combined together to be used for multi-class classification.

3.3.5. Random Forest (RF)

It is an ensemble decision tree that creates and combines many decision trees. Creating and combining many decision trees allow weak decision trees on their own to be used in order to create a stronger decision tree with better accuracy. It is called random because the attributes are chosen randomly during model building and training. Furthermore, it is called forest because it takes outputs of many decision trees to create a better decision tree. RF algorithm works in the following steps:

1. Choose n samples randomly from the training set.
2. Grow a decision tree from the chosen sample by selecting a number of features randomly.

3. Split the node based on a chosen feature which has the highest information gain.
4. Repeat the previous steps K times (K represents the number of trees to be created).
5. Aggregate the trees and then choose the majority class based on voting.

3.3.6. Nave Bayes (NB)

It is a Bayes theorem-based supervised learning algorithm. It is called nave because it assumes that an attribute is independent in terms of probability to happen from all other features. NB formula is shown as follows:

$$P(C|f) = \frac{P(f|C)P(C)}{P(f)} = P(f_1|C) \times P(f_2|C) \times ... \times P(f_n|C) \times P(C) \quad (3.1)$$

Where C is the class and f_i is any feature.

3.3.7. Neural Networks (NN)

They are machine learning algorithms that are built based on the human brain. They can be used for multi-class classification problems by considering one-vs-all principle. An activation function is used to take a number of inputs to produce an output. Assume there are M classes, one-vs-all classification is performed by training M distinct binary classifiers in which each trained binary classifier can recognize a specific class. Consequently, those M classifiers work together to be used for multi-class classification.

3.4. Evaluation

A decision on the adoption of the EDM outcomes should be delivered based on the evaluation phase. The Confusion matrix will be used for calculating the correctness and accuracy of the model.

Since target variable classes are nearly balanced, accuracy will be used as a performance metric to compare the models. The Confusion matrix is an easy and intuitive metric used for finding the correctness and accuracy of the model. It is used for classification problems where the output would be of two or more types of classes. Some Terminology and derivations from a confusion matrix are shown as follow:

1. True Positive (TP): Cases that are TRUE and predicted correctly as TRUE.

2. True Negative (TN): Cases that are FALSE and predicted correctly as FALSE.
3. False Positive (FP): Cases that are FALSE but predicted incorrectly as TRUE (known as "Type I error").
4. False Negative (FN): Cases that are TRUE but predicted incorrectly as FALSE (known as "Type II error")

Accuracy is a metric for evaluating classification models and it refers to the percentage of predictions that happen to be right. Based on the contents of the confusion matrix it is possible to extract the accuracy of the model as shown in following formula:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \tag{3.2}$$

The research tools need to verify the reliability and validity. In quantitative research, enhancing and verifying of experiments are achieved through measurement of the validity and reliability [38]. To make sure that the study is reliable and valid, the experiments conducted using 10-fold cross validation repeated for 10 times.

Figure 4 shows the final result as a comparison based on the accuracy of each classifier. Random Forest (RF) implemented using the Caret package shows the best performance with 85% accuracy.

	Classifier	Accuracy
1	K-Nearest Neighbor (KNN) - Caret	0.7118644
2	Decision Tree (DT-C5) - Caret	0.7881356
3	Support Vector Machines (SVM) - Caret	0.7542373
4	Support Vector Machines (SVM) - e1071	0.8305085
5	Logistic Regression (LR) - Caret	0.7796610
6	Random Forest (RF) - Caret	0.8474576
7	Random Forest (RF) - RandomForest	0.7881356
8	Naïve Bayes (NB) - Caret	0.6186441
9	Neural Networks (NN) - Caret	0.8135593

Figure 4. Accuracy measure of each classifier

4. Discussion

4.1. Deployment

Deployment phase includes the outputs of the experiments as explained in Modeling and it shows a deployment of the best model, found during modeling

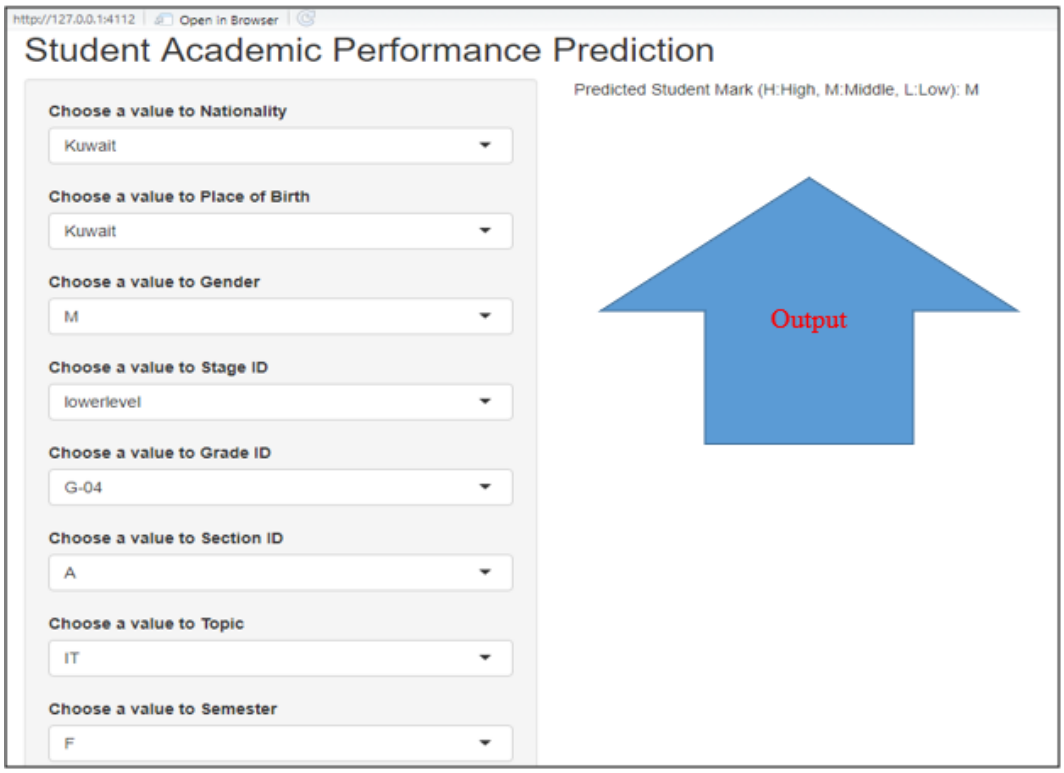

Figure 5. Interactive web application using r shiny part 1

and evaluation phase, in interactive web application using R Shiny package. Figure 5 and Figure 6 show the user interface of the Shiny web application developed to deploy the predictive model.

4.2. Findings

The final predictive model built using random forest revealed some interesting findings which are listed as follows:

1. Based on the literature review conducted, seven classifiers are chosen to model the dataset: Decision Trees, Nave Bayes, K-Nearest Neighbors, Support Vector Machines, Logistic Regression, Random Forest, and Neural Networks.
2. The best model in regards to accuracy is the one built with Random Forest using the Caret package.
3. The best model is deployed into an interactive R Shiny application.

4.3. Fulfilment of Research Objectives

RO1. Investigating the appropriate educational data mining methods used in predicting student academic performance has been achieved by describing the techniques of educational data mining first, then exploring educational data mining techniques used in predicting student academic performance in depth as discussed in Section 2. Educational Data Mining techniques explored are regression, classification, clustering, association rules, social network analysis and visualization, process mining, text mining, and outlier detection. A synthesis of literature related to the appropriate educational data mining techniques relevant to student academic performance prediction shows that there are two relevant educational data mining techniques which are regression (Linear Regression) and classification (Decision Trees, Bayesian-Based, Neural Networks, Support Vector Machines, Ensemble Methods, K-Nearest Neighbor, and Logistic Regression).

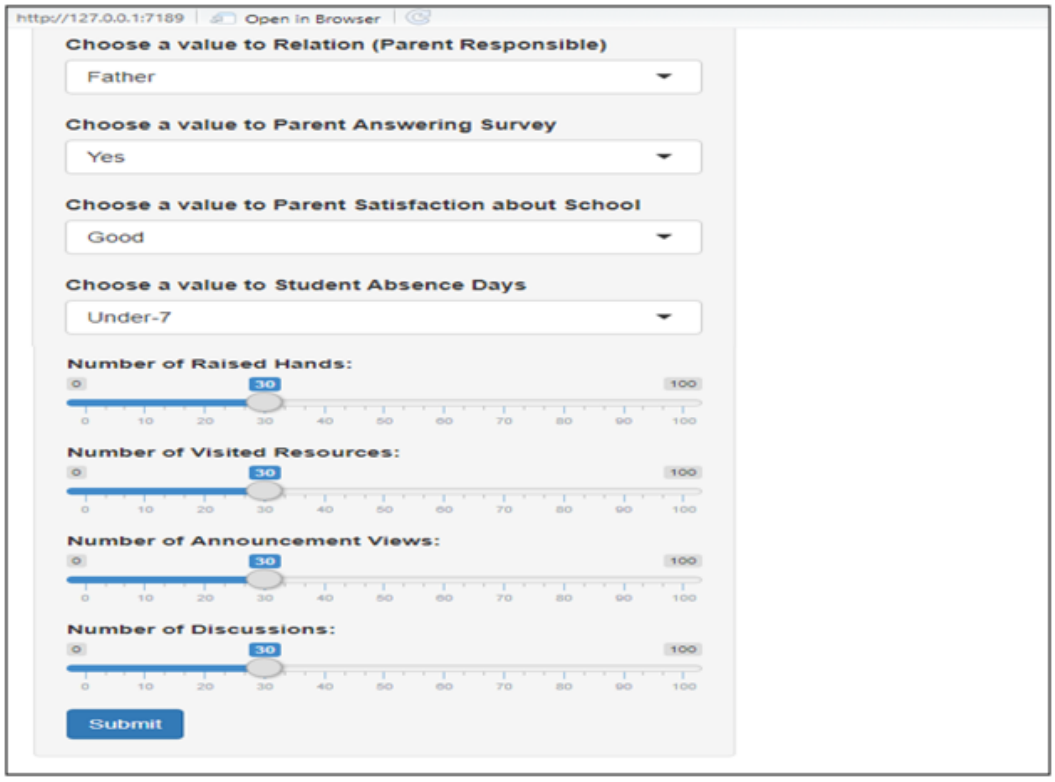

Figure 6. Interactive web application using r shiny part 2

RO2. To apply educational data mining methods to support academic intervention by applying the relevant educational data mining classification techniques explained in the synthesis of the literature review performed in Section 2 on an appropriate dataset and building a model using R Shiny to gain insights from learning data in order to support students in regards to their academic performance and in taking actions to prevent or warn students from failure which leads to grade improvement that in turn will drive the overall degree success as discussed in Sections 3.2, 3.3, 3.4, and 4.1.

5. Conclusion

5.1. Summary

Mining educational data is far from conclusive, yet it has been evolving and growing continuously. Using the appropriate tools and the research lines in this area are not only going to help students and instructors but also the other stakeholders/users and that impact has been extended to parents, society, and the public in general. The main focus of this research is on how EDM can support student learning in regards to student academic performance, engagement, and intervention through predicting the academic performance of students.

Student academic performance prediction has been a research topic of a significant value for many years because students and institutions can gain findings and insights uncovered from learning data. Institutions can gain from it by trying to enhance the quality of academic services and resources made available to their students in order to increase the rates of students who can progress with their study and be successful in completing their programs or courses of study. In addition, findings or insights can be deployed to deliver solutions, suggestions, or advices to students in order to improve their performance in the future.

5.2. Future Work

Since random forest has been the best model implemented and it gave an accuracy of 85%, it would be advisable to try adopting advanced methods for classification models like genetic algorithms and see if they can further improve the accuracy/performance of the model. Furthermore, with new data added, models can be tested again and see if there is any improvement in accuracy. Once a model has shown better performance, it is easy to adopt and use it for deployment on Shiny and Azure Machine Learning Studio.

REFERENCES

[1] Mandinach, E. B.: A perfect time for data use: Using data-driven decision making to inform practice. *Educational Psychologist*, **47**(2), (2012), 71–85.

[2] Slade, S. and Prinsloo, P.: Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist*, **57**(10), (2013), 1510–1529, URL <Go to ISI>://WOS:000324103200009.

[3] Picciano, A.: The evolution of big data and learning analytics in american higher education. *Journal of Asynchronous Learning Network*, **16**.

[4] Zeide, E.: The structural consequences of big data-driven education. *Big Data*, **5**(2), (2017), 164–172, URL <Go to ISI>://WOS:000403939100008.

[5] Patwa, N., Seetharaman, A., Sreekumar, K., and Phani, S.: Learning analytics: Enhancing the quality of higher education. *Research Journal of Economics*, **2**(2).

[6] Roberts, L. D., Howell, J. A., Seaman, K., and Gibson, D. C.: Student attitudes toward learning analytics in higher education: "the fitbit version of the learning world". *Frontiers in psychology*, **7**, (2016), 1959–1959.

[7] Pena-Ayala, A.: Educational data mining: A survey and a data mining-based analysis of recent works. *Expert Systems with Applications*, **41**(4), (2014), 1432–1462, URL <Go to ISI>://WOS:000330158700045.

[8] Romero, C. and Ventura, S.: 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), (2010), 601–618, URL <Go to ISI>://WOS:000283447800001.

[9] Khan, A. and Ghosh, S. K.: Data mining based analysis to explore the effect of teaching on student performance. *Education and Information Technologies*, **23**(4), (2018), 1677–1697.

[10] Dutt, A., Ismail, M. A., and Herawan, T.: A systematic review on educational data mining. *IEEE Access*, **5**, (2017), 15991–16005.

[11] Baker, R. S.: Educational data mining: An advance for intelligent systems in education. *IEEE Intelligent Systems*, **29**(3), (2014), 78–82, URL <Go to ISI>://WOS:000341575700014.

[12] Lopez, S. L. S., Redondo, R. P. D., and Vilas, A. F.: Predicting students' grade based on social and content interactions. *International Journal of Engineering Education*, **34**(3), (2018), 940–952, URL <Go to ISI>://WOS:000443168300010.

[13] Nasiri, M., Minaei, B., and Vafaei, F.: Predicting gpa and academic dismissal in lms using educational data mining: A case mining. In *3rd International Conference on E-Learning and E-Teaching (ICELET)*, IEEE International Conference on E-Learning and E-Teaching, IEEE Computer Soc, LOS ALAMITOS, ISBN 978-1-4673-0958-5; 978-1-4673-0956-1, 2012, pp. 53–58, URL <Go to ISI>://WOS:000310432500008.

[14] Asif, R., Merceron, A., Ali, S. A., and Haider, N. G.: Analyzing undergraduate students' performance using educational data mining. *Computers and Education*, **113**, (2017), 177–194, URL <Go to ISI>://WOS:000406728400013.

[15] Costa, E. B., Fonseca, B., Santana, M. A., de Araujo, F. F., and Rego, J.: Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses. *Computers in Human Behavior*, **73**, (2017), 247–256, URL <Go to ISI>://WOS:000403625400025.

[16] Burgos, C., Campanario, M. L., de la Pena, D., Lara, J. A., Lizcano, D., and Martinez, M. A.: Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers and Electrical Engineering*, **66**, (2018), 541–556, URL <Go to ISI>://WOS:000429760300041.

[17] Iam-On, N. and Boongoen, T.: Generating descriptive model for student dropout: a review of clustering approach. *Human-Centric Computing and Information Sciences*, **7**, (2017), 24, URL <Go to ISI>://WOS:000396492500001.

[18] Campagni, R., Merlini, D., and Verri, M. C.: University student progressions and first year behaviour. *Proceedings of the 9th International Conference on Computer Supported Education (Csedu), Vol 2*, pp. 46–56, URL <Go to ISI>://WOS:000444908800004.

[19] Najera, A. B. U., de la Calleja, J., and Medina, M. A.: Associating students and teachers for tutoring in higher education using clustering and data mining. *Computer Applications in Engineering Education*, **25**(5), (2017), 823–832, URL <Go to ISI>://WOS:000410722900013.

[20] Ashok, M. V. and Apoorva, A.: Clustering proficient students using data mining approach. *Advances in Computing and Data Sciences, Icacds 2016*, **721**, (2017), 70–80, URL <Go to ISI>://WOS:000434872100008.

[21] Alharbi, Z., Cornford, J., Dolder, L., and Iglesia, B. D. L.: Using data mining techniques to predict students at risk of poor performance. In *2016 SAI Computing Conference (SAI)*, pp. 523–531.

[22] Khan, A. and Ghosh, S. K.: Analysing the impact of poor teaching on student performance. *Proceedings of 2016 Ieee International Conference on Teaching, Assessment, and Learning for Engineering (Tale)*, pp. 169–175, URL <Go to ISI>://WOS:000400475400029.

[23] Parkavi, A. and Lakshmi, K.: Pattern analysis of blooms knowledge level students performance using association rule mining. *2017 Ieee International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (Icstm)*, pp. 90–93, URL <Go to ISI>://WOS:000426986800016.

[24] Jayanthi, M. A., Kumar, R. L., and Swathi, S.: Investigation on association of self-esteem and students' performance in academics. *International Journal of Grid and Utility Computing*, **9**(3), (2018), 211–219, URL <Go to ISI>://WOS:000441309900001.

[25] Gomez-Aguilar, D. A., Garcia-Penalvo, F. J., and Theron, R.: Visual analytics in e-learning. *Profesional De La Informacion*, **23**(3), (2014), 236–245, URL <Go to ISI>://WOS:000339037000003.

[26] Lotsari, E., Verykios, V. S., Panagiotakopoulos, C., and Kalles, D.: A learning analytics methodology for student profiling. *Artificial Intelligence: Methods and Applications*, **8445**, (2014), 300–312, URL <Go to ISI>://WOS:000352632400024.

[27] Saqr, M., Fors, U., Tedre, M., and Nouri, J.: How social network analysis can be used to monitor online collaborative learning and guide an informed intervention. *Plos One*, **13**(3), (2018), 22, URL <Go to ISI>://WOS:000428093900105.

[28] Okoye, K., Tawil, A. R. H., Naeem, U., Bashroush, R., and Lamine, E.: A semantic rule-based approach supported by process mining for personalised adaptive learning. *5th International Conference on Emerging Ubiquitous Systems and Pervasive Networks / the 4th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare / Affiliated Workshops*, **37**, (2014), 203–+, URL <Go to ISI>://WOS:000349985800025.

[29] Smirnova, E. V., Samarev, R. S., and Willmot, P.: New technology for programming teaching: Process mining usage. In *7th International Conference on Education and New Learning Technologies (EDULEARN)*, EDULEARN Proceedings, IATED, VALENICA, ISBN 978-84-606-8243-1, 2015, pp. 7330–7335, URL <Go to ISI>://WOS:000376685707055.

[30] Sedrakyan, G., De Weerdt, J., and Snoeck, M.: Process-mining enabled feedback: "tell me what i did wrong" vs. "tell me how to do it right". *Computers in Human Behavior*, **57**, (2016), 352–376, URL <Go to ISI>://WOS:000370457800041.

[31] He, W.: Examining students' online interaction in a live video streaming environment using data mining and text mining. *Computers in Human Behavior*, **29**(1), (2013), 90–102, URL <Go to ISI>://WOS:000312684400013.

[32] Koufakou, A., Gosselin, J., and Guo, D. H.: Using data mining to extract knowledge from student evaluation comments in undergraduate courses. In *International Joint Conference on Neural Networks (IJCNN)*, IEEE International Joint Conference on Neural Networks (IJCNN), IEEE, NEW YORK, ISBN 978-1-5090-0619-9, 2016, pp. 3138–3142, URL <Go to ISI>://WOS:000399925503046.

[33] Krishnaveni, K. S., Pai, R. R., and Iyer, V.: Faculty rating system based on student feedbacks using sentimental analysis. *2017 International Conference on Advances in Computing, Communications and Informatics (Icacci)*, pp. 1648–1653, URL <Go to ISI>://WOS:000427645500273.

[34] Rajeswari, A. M., Sridevi, M. R., and Deisy, C.: Outliers detection on educational data using fuzzy association rule mining. In *Int. Conf. on Adv. in Comp., Comm., and Inf. Sci. (ACCIS-14)*.

[35] Weng, C.-H.: Mining fuzzy specific rare itemsets for education data. *Knowl.-Based Syst.*, **24**, (2011), 697–708.

[36] Oeda, S. and Hashimoto, G.: Log-data clustering analysis for dropout prediction in beginner programming classes. In *Procedia Computer Science*, vol. 112, pp. 614–621.

[37] Amrieh, E., Hamtini, T., and Aljarah, I.: Mining educational data to predict student's academic performance using ensemble methods. *International Journal of Database Theory and Application*, **9**, (2016), 119–136.

[38] Heale, R. and Twycross, A.: Validity and reliability in quantitative studies. *Evidence Based Nursing*, **18**(3), (2015), 66.