Iraqi Journal of Science, 2022, Vol. 63, No. 9, pp: 3954-3965
DOI: 10.24996/ijs.2022.63.9.27

ISSN: 0067-2904

Educational Data Mining For Predicting Academic Student Performance Using Active Classification

Rasha H. Ali
Computer Department , College of Education for Women-, University of Baghdad, Baghdad, Iraq

Received: 4/9/2021 Accepted: 27/11/2021 Published: 30/9/2022

Abstract

The increasing amount of educational data has rapidly in the latest few years. The Educational Data Mining (EDM) techniques are utilized to detect the valuable pattern so that improves the educational process and to obtain high performance of all educational elements. The proposed work contains three stages: preprocessing, features selection, and an active classification stage. The dataset was collected using EDM that had a lack in the label data, it contained 2050 records collected by using questionnaires and by using the students' academic records. There are twenty-five features that were combined from the following five factors: (curriculum, teacher, student, the environment of education, and the family). Active learning had been utilized in the classification. Four techniques had been applied for classifying the features: Random Forest (RF) algorithm, Label Propagation (LP), Logistic Regression (LR), and Multilayer Perceptron (MLP). The accuracies of prediction were 95.121%, 92.195%, 92.292%, and 93.951% respectively. Also, the RF algorithm has been utilized for assorting the features depending on their importance.

Keywords: Educational Data Mining, Active classification, Students' Prediction, Feature Importance, Random Forest, Multilayer Perceptron.

التنقيب في البيانات التعليمية للتنبؤ بأداء الطالب الأكاديمي باستخدام التصنيف النشط

رشا حسين علي
قسم الحاسبات، كلية التربيه للبنات، جامعة بغداد، بغداد، العراق

الخلاصة

تزايدت كمية البيانات التعليمية بسرعة في السنوات القليلة الماضية. يتم استخدام تقنيات التنقيب عن البيانات التعليمية (EDM) للكشف عن النمط القيم من أجل تحسين العملية التعليمية والحصول على أداء عالٍ لجميع العناصر التعليمية. يحتوي العمل المقترح على ثلاث مراحل: المعالجة المسبقة ، واختيار الميزات ، ومرحلة التصنيف النشط. تم جمع مجموعة البيانات باستخدام EDM التي تفتقر إلى بيانات التسمية ، وتحتوي على 2050 سجلًا تم جمعها باستخدام الاستبيانات وباستخدام السجلات الأكاديمية للطلاب. هناك خمسة وعشرون سمة تم جمعها من خمسة عوامل التالية: (المنهج ، المعلم ، الطالب ، بيئة التعليم ، والأسرة). تم استخدام التعلم النشط في التصنيف. تم تطبيق أربع تقنيات لتصنيف الميزات: خوارزمية الغابة العشوائية (RF) ، انتشار الملصق (LP) ، الانحدار اللوجستي (LR) ، والإدراك متعدد الطبقات (MLP). وكانت دقة التنبؤ

*Email: rashaha2003@coeduw.uobaghdad.edu.iq

95.121٪ , 92.195٪ , 92.292٪ , 93.951٪ على التوالي. أيضًا ، تم استخدام خوارزمية التردد اللاسلكي لتصنيف الميزات حسب أهميتها.

1. Introduction

The quality of education is compulsory in the development of each country, the amount of data in the education field is getting increased increasingly. The data gathered from students, teachers, and managers of schools are usually used to make a decision. But most of the data remain unused due to complexity and a large amount of data. Therefore, analyzing this huge amount of educational data is of great interest to predict the students' performance[1].

Recently, the institutions of education have been facing a quite competitive environment. It is very important to ensure that the resource is used efficiently and effectively to improve the experience of students' learning and to promote the variables affecting the learning process and the students' performance. Recently, researchers have focused on the educational data to that analyze the factors that are affected by the students' academic performance and to improve the teaching and learning practices[2]. The educational institutions are more interested in finding the factors that affect the students' performance, large numbers of factors may affect the students' performance such as social, economic, cultural, demographic, and academic background[3]. Generally, the data are great available but the information suffers from deficiency which leads to uncertain decisions making. The great challenge is the effective transformation of huge of data into knowledge to enhance the quality of managerial decisions. The concept of Data Mining (DM) refers to a crucial and significant part of the Knowledge Discovery in Databases (KDD) process[4].

The concept of DM is depending on extracting the hidden patterns and exposing the relationships between parameters for huge amounts of data. Applications of data mining in educational environments are the most popular area of the research which is called Educational Data Mining (EDM) and there is an increasing number of studies that published in the latest years. Researches areas in the EDM focus on different aspects of an educational process: students, teachers, teaching aids, organization of teaching, etc. The main aims of EDM applications can be categorized as the following: prediction of students' success, organization of teaching programs, predicting the success of students to the higher level of the educational program. To achieve these goals, different algorithms of DM are used, such as decision trees (DT), artificial neural networks (ANNs), k-nearest neighbors, Support Vector Machine (SVM)[5]. Early forecasting of students' performance might help teachers for identifying the students who can succeed in other courses and the students who need additional education, additional effort, or more time to improve their knowledge[6].

The problem of this study, Iraqi students' performance in secondary education had been decreased in the last years. The educational statistics in Iraqi stated there is a decrease in the performance of the Iraqi students in secondary education. So, this study suggests a system to prophesy the students' performance and detect which factors are more affected on the students' performance. The present study aims at analyzing the educational data had lacked in label data by building a model used to classify the students' performance of the secondary education by using active learning in the classification. Also, the sorting of the attributes depending on the importance that affect the student's performance by using the RF algorithm. The organization of this work is as follows: Section one shows the introduction. The related works are present in section two. Section three shows the methodology of the proposed system. The importance of the features presents in section four. The results and discussion shows in section five. The proposed work concludes in section six.

2. Related Works

The following studies have presented techniques for predicting the students' performance. In ref.[5], techniques of data mining have been used in the performance prediction of first-year bachelor students in a computer science course, the data set contained the 497 records that contained the following information: the students' demographics, previous academic records, and family background information, the methods that were used in the classification stage were Decision Tree, Naïve Bayes, and Rule-Based classification (JRip), the results were 68.8%, 67.0%, and 71.3% respectively. In ref [7], the extended profile has been applied for the prediction of the final degree for students. In this study, the features were combined from the following factors are (economic, institutional factor, demographic, psychological factor, and academic) factors. While in the classification has been applied using a multi-layered neural network (NN) to assort students' degrees into a good or basic degree class. the size of the dataset was 470 while the accuracy was 83.7%. In ref.[2], An academic performance prediction by student heterogeneity had been presented. The dataset contained the enrollment data of the student at Australian University which contained all information about the student, most of this information depended on the observation for constructing socio-demographic features. The suggested system included three stages were preprocessing, features extraction, and a classification stage. Four methods were applied in the classification: Naive Bayesian, Sequential minimal optimization (SMO) for SVM, decision tree (J48), and classification-based rule. The results were 74.2%, 73.87%, 76.19%, and 73.84%. In ref.[1], the techniques of supervised learning were used in this work for predicting the students' performance. Three techniques have been applied in the features classification which was Multilayer perceptron(MLP), NNge, and decision tree (J48). The dataset belonged to the Alentejo region of Portugal. The J48 obtained the highest accuracy with 95.78% compared with the other techniques. In [8], the predicting of the engineering student attrition risk using the probabilistic neural network (PNN) by using deep learning NN to predict students' performances and compare its classification accuracy with MLNN and other classification models such as logistic regression. The model utilized 58 student variables related to demographics and academic background, the size of the dataset was 682 while the accuracy was 90.1%. In [9], the prediction of the student academic performance using the techniques of data mining had been presented. The dataset in this study contained 480 records, the number of features was 16 features. Five techniques of data mining were used while the accuracy were (69.7% for Random Forest, 67.7% for Logistic regression, 64.5% for Perceptron Neural Network, 46.8 % for Decision tree, and 70.8% for SVM).

3. The Methodology of the Work

This work suggests the prediction of the students' academic performance by using EDM and active learning for classifying the features. Also, machine learning techniques such as the Random Forest algorithm have been utilized for classifying the factors that have more effect on the students' academic performance. The suggested work contains four stages were (educational data collection, data preprocessing, feature selection, and a classification stage.). Figure 1 illustrates the general structure of the suggested work.

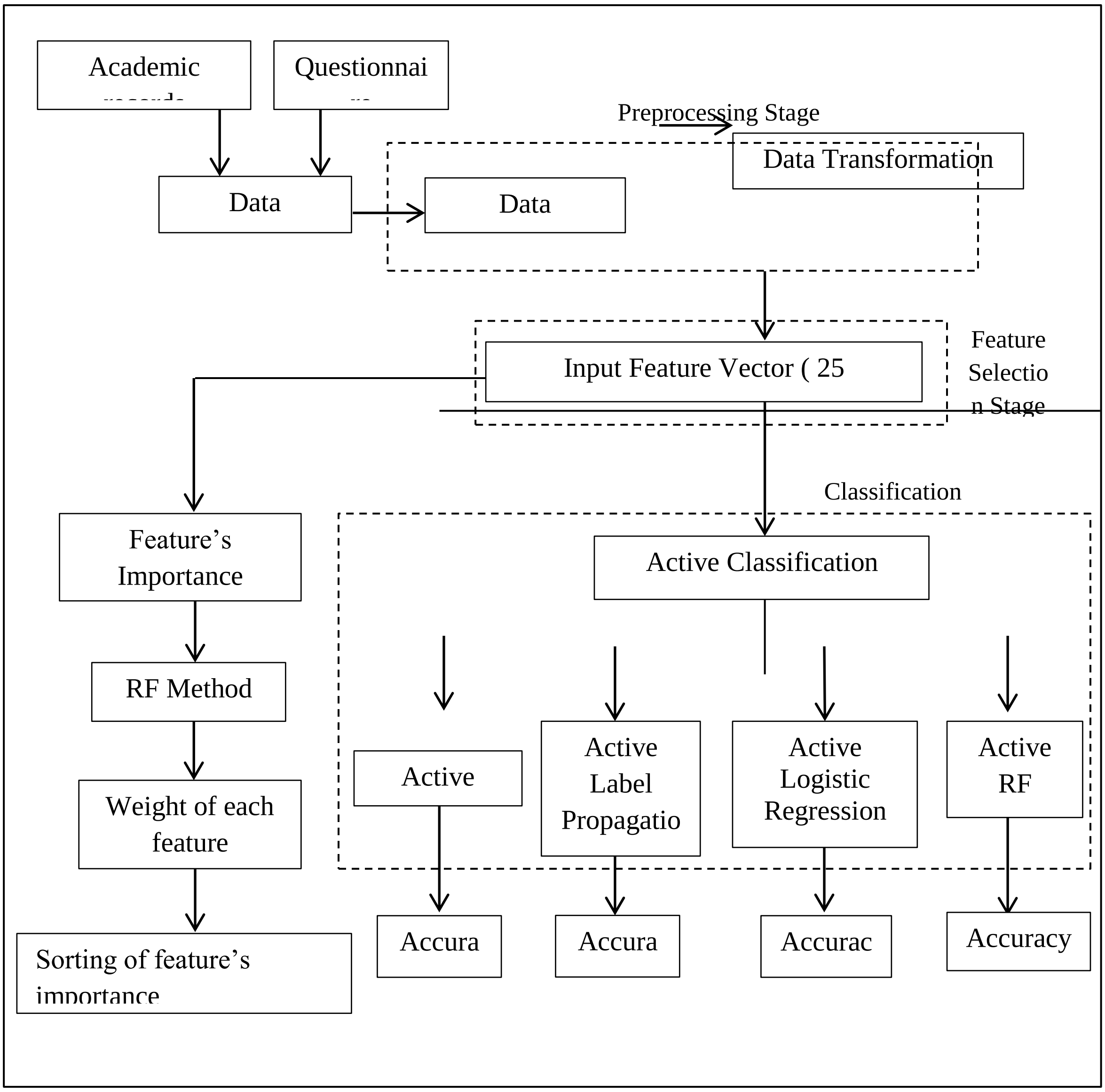

Figure 1- The structure of suggested work

3.1 The Educational Data Collection

The dataset was collected using EDM that contained 2050 records. Each record represents the information for each student. The ages of students between (12-18) years. The data was collected from more than one school from the private and government schools by the questionnaires and the academic records for the students. In general, each school has a record for each student which contains some information about the student and the student's result. The questionnaire contains questions about the following factors (teachers, curriculum, and educational environment, student's characteristics, and student's family). The questionnaires were answered either manually or electronically by the students and teachers depending on the schools because the private schools have electronic pages or websites but most government schools haven't electronic pages or websites. The paragraphs in each factor represent the features of the proposed system. The paragraphs answered by teachers were (student's characteristic, curriculum, and student's family), while paragraphs that answered by the student were (teachers, and educational environment). While the label of data was the final degrees for students. Figure 2 shows the group of factors of each student.

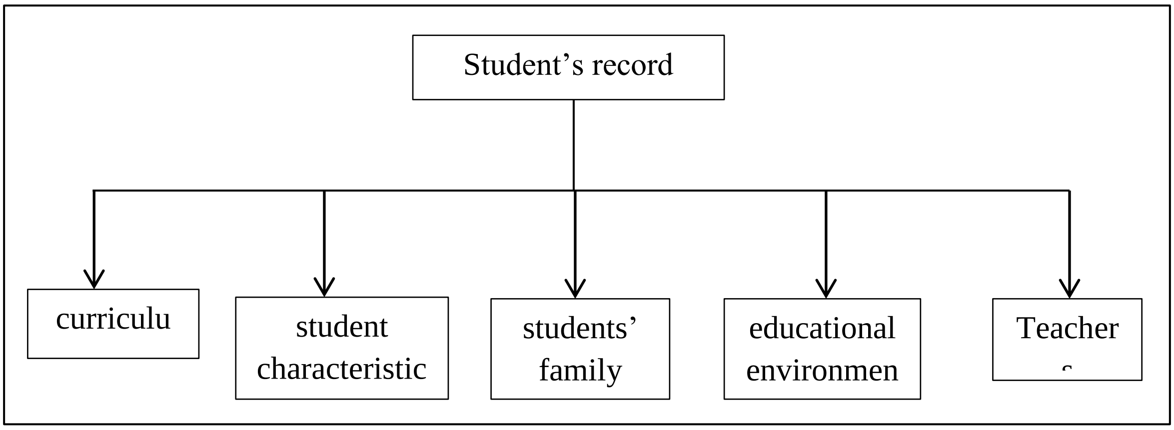

Figure 2- The student's record

3.2 The Preprocessing stage

The preprocessing step is an important part of the prediction process to repair the data so can be used in the features selection in hence in the classification stage. Two processes were performed in this stage. Firstly, the discretization process was performed for attributes. In the questionnaire, the values of each attribute were categorized into five quartiles (agreement with a very big degree, agreement with a big degree, agreement with an average degree, no agreement, no agreement with a big degree). The value of scaling digital is presented in Table 1. The value of label data was either pass or fail, the value of a result class was one and zero respectively. Secondly, in the process of transformation for data, the gathered data were transformed from Excel format into CSV format so that can be applied in the classification stage.

Table 1- Digital value of attributes

The attribute value	***The digital value of an attribute***
agreement with a very big degree	5
agreement with a big degree	4
agreement with an average degree	3
no agreement	2
no agreement with a big degree	1

3.3 Features Selection Stage

The features selection technique means selecting the attributes directly which are used in the classification stage. The dataset contained twenty-five attributes, these attributes were collected from five groups (attributes for the student, the teachers, the student's family, the curriculum, and the educational environment). Some of these attributes had been filled by the teachers and students by using questionnaires as mentioned in the previous section, while the values of some other attributes have been filled by using the student's academic records. Each feature has a score ranging (1-5). Table 2 present the attribute, its description, and variable name of features.

Table 2- The Attribute and it's description

No.	***Attribute***	***Description***	***Variable Name***
1	Difficulty of curriculum	The curriculum is very difficult	F31
2	Equivalent of teacher	Lacking the knowledge level for the teacher	F15
3	Ambiguity	The scientific material is not clear	F21
4	No Consideration	The teacher isn't considering the Individual differences	F25
5	Practically	Lacking the practical side for the scientific material	F41

6	Teaching methods	Lacking the teacher for the suitable teaching methods	F13
7	No agreement	No agreement between the curriculum and the general level for the student.	F51
8	Motivation	Lacking motivation of the student	F45
9	Educational level	Lacking educational and a knowledge level for the student’s family	F53
10	Interactive	Lacking the interaction between the teacher and the student	F35
11	Multi concepts	The many concepts in the one scientific material and many materials in the one stage	F12
12	Lack of experience	Lacking experience for the teacher	F36
13	Coordination	Lacking coordination between the management of the school and student’s family	F42
14	Gamming	The student spends the most time playing games and watching T.V.	F52
15	Stability	Lacking the social stability of student's family	F33
16	Health	Lacking the healthy state of the student	F14
17	Absence	The lack of a class attendance	F43
18	Contention	The contentions in the student’s family	F54
19	Stress	The psychological- social stress for the student	F32
20	The classroom suitability	The classroom is not suitable for increasing the number of students	F34
21	Concentricity	Lacking the focus of the student in the classroom	F22
22	Educational devices	Lacking the educational devices in the school.	F23
23	Economical level	The economic level for the student’s family	F44
24	Laboratories	Lacking the laboratories in the school.	F24
25	Infrastructure of school	The infrastructure of the school is not suitable	F11

3.4 The Classification Stage

This stage aims to classify the features for predicting the academic performance of the students. Generally, machine learning has three types: supervised, unsupervised, and semi-supervised learning (SSL). Active learning is a technique that depends on SSL that is used when the dataset has a lack of labeled data and availability of unlabeled data. The key point in active learning is building a high accuracy classifier (active learner) without making too many queries using a small labeled training set[10]. The connotation of active learning is to maximize the achievement of the classifier by selecting the better example to label. The pool-based sampling scenario was used in this work. This scenario assumes there is a small labeled dataset L and a large unlabeled dataset U (unlabeled data pool). Queries are selected from U according to specific evaluation criteria. The following algorithm is the pseudo-code of active learning[11].

Algorithm 1- The pseudo-code of active learning

Input: labeled dataset *L*, unlabeled dataset *U*.

1. Initially, apply base learner *B* to the training set *L* to obtain classifier *C*.
2. Apply *C* to unlabeled dataset *U*.
3. From *U*, select the most informative *m* instances to learn from (*I)*.
4. Ask the teacher/expert for labeling the *m* instances (*I)*.
5. Move *I*, with supplied classifications, from *U* to *L (U=U-I)*.
6. Re-train using *B* on labeled set (*L=L+I)* to obtain a new classifier, *C*.
7. Repeat steps 2 to 6, until *U* is empty or until some stopping criterion is met
8. Output a classifier that is trained on *L*.

In this work, the dataset which contained 2050 instance were split into 620 records had label data and 1430 records had unlabeled data. The dataset had been divided into two groups: 50% for the training that had been utilized to build the model, and 50% for the testing that had been used to validate the model. The classification process was achieved by using active learning by using four methods were: Random Forest algorithm (RF), Label Propagation (LP), Logistic Regression (LR), and Multi-layer perceptron neural network (MLP).

3.3.1 Label Propagation (LP)

LP is a semi-supervised learning technique based on a graph. It uses the labels from the labeled data to transduce the labels to unlabeled data. in LP there are two sets: labeled set denoted as L, and unlabeled set denoted as U. Where $(x_1, y_1)...(x_l, y_l) \in L$, $(x_{l+1}, y_{l+1})...(x_n, y_n) \in U$, $\{x_1...x_n\} \in \Re^D$, $\{y_1...y_n\} \in \{1..C\}$. C is the number of classes, and L consists of all classes. The algorithm of LP uses graph algorithm constructs as $G = (V, E)$. Where L represents the V that represents the number of vertices, and U represents the E in the G that represents the number of the edges. E represents the similarity between two nodes i and j with the weight w_{ij}. The weight is calculated such that nodes will similar nodes (similar distances) will have larger weights. The probabilistic transition matrix T is used in this algorithm:

$$T_{ij} = P(i \to j) = \frac{w_{ij}}{\sum_{k=1}^{n} w_{kj}} \tag{1}$$

The algorithm updates the label $Y \leftarrow TY$ iteratively, clamp the label until Y converges[12].

3.3.2 Logistic Regression (LR)

It's one of the statistical analysis methods. It's used for estimating the probabilities to understand the relationship between the dependent variable and one or further independent variables using Eq.(2), while the probability model can be computed using Eq.(3).

The type of active learning is pool-based active learning for logistic regression using an algorithm (2)[13].

$$\sigma(\theta) = \frac{1}{1 + \exp[-\theta]} \tag{2}$$

$$P(Y_n = 1 | X_n) = \sigma(w . X_n) \tag{3}$$

Algorithm 2- The active logistic regression

```
Require: partial training set, pool of unlabeled examples.
repeat several times
Select T random examples from pool
Rank these T examples according to active learning rule
Present the top-ranked example for labeling
Augment the training set with the new observation
until Training set reaches desirable size
```

3.3.3 Random Forest (RF)

RF is a classification method, it encompasses random decision forests and bagging. RF is a set of decision trees algorithm that depends on the voting technique in the classification decision[14]. The RF classifier selects an input vector and classifies it by every single tree in the forest. The final decision about the class of input vector is then decided by considering the plurality of votes.

Let *T* be the training set having L labeled samples as $(x_1, y_1), (x_2, y_2), ..., (x_L, y_L)$. The vector $X = (x^1, x^2, ..., x^p)$ is a vector consisting of explanatory variables and $X \in R^p$. The label class of samples is represented by $Y \in (+1, -1)$ then the goal is to design the classifier which maps the input data to the output that $h : R^p \to Y$. RF is a set of m classifiers as shown in the following equation:

$$R_F = h(X, \Omega_k), \qquad k = 1, ..., m. \tag{4}$$

where *x* is *n* dimensional random vector, sampled from vector *X*. Here, each classifier h is trained using the random vector Ω_k, and each Ω_k is selected from the bootstrapped training set *T*. The individual trees constructed are not pruned, one of the reasons for this is averaging or the majority of votes eliminate the prediction error of individual tree[15].

3.3.4 Multilayer Perceptron (MLP)

The MLP is the feed-forward ANN that consists of three layers or more. Simply, It contains three layers: the visible input layer, the hidden layer, and the output layer. Each neuron is transforming information and carries it to the neurons in the next layer and contains an activation function[16]. In this work, the following parameters have been used for building the MLP classifier using active learning are: (number of the hidden layers are three, the number of neurons are100 neurons, an alpha value equal to 0.00001).

4. Features Importance

Each feature has relative importance on the prediction process. In the training phase, each feature has a value that helps in choosing the extremely important features and dropping the least important ones for model building. Finding out the feature's importance has more than one benefit such as improving the accuracy, reducing the computation time for building the model, finding important features makes your model more interpretable and easy to comprehend in general[17]. In this work, the RF algorithm is used to locate the features' importance by applying the Gini function which is computed by describing the total power of the variable. In this work, the number of the decision tree was 100 trees. Table 3 presents each feature's value according to importance. Figure 3 shows the sorting of features depending on importance.

Table 3- The importance values of each feature

The Feature	***Importance feature***	***The Feature***	***Importance feature***	***The Feature***	***Importance feature***
F21	0.097131	F35	0.036670	F34	0.020184
F41	0.085956	F52	0.034580	F33	0.019835
F51	0.084892	F45	0.031608	F24	0.019710
F31	0.084796	F14	0.028909	F42	0.019193
F12	0.066054	F22	0.027678	F54	0.017083
F15	0.059662	F32	0.025569	F11	0.016042
F13	0.056204	F43	0.024994	F44	0.022088
F25	0.037903	F53	0.024423	F23	0.021983
F36	0.036852				

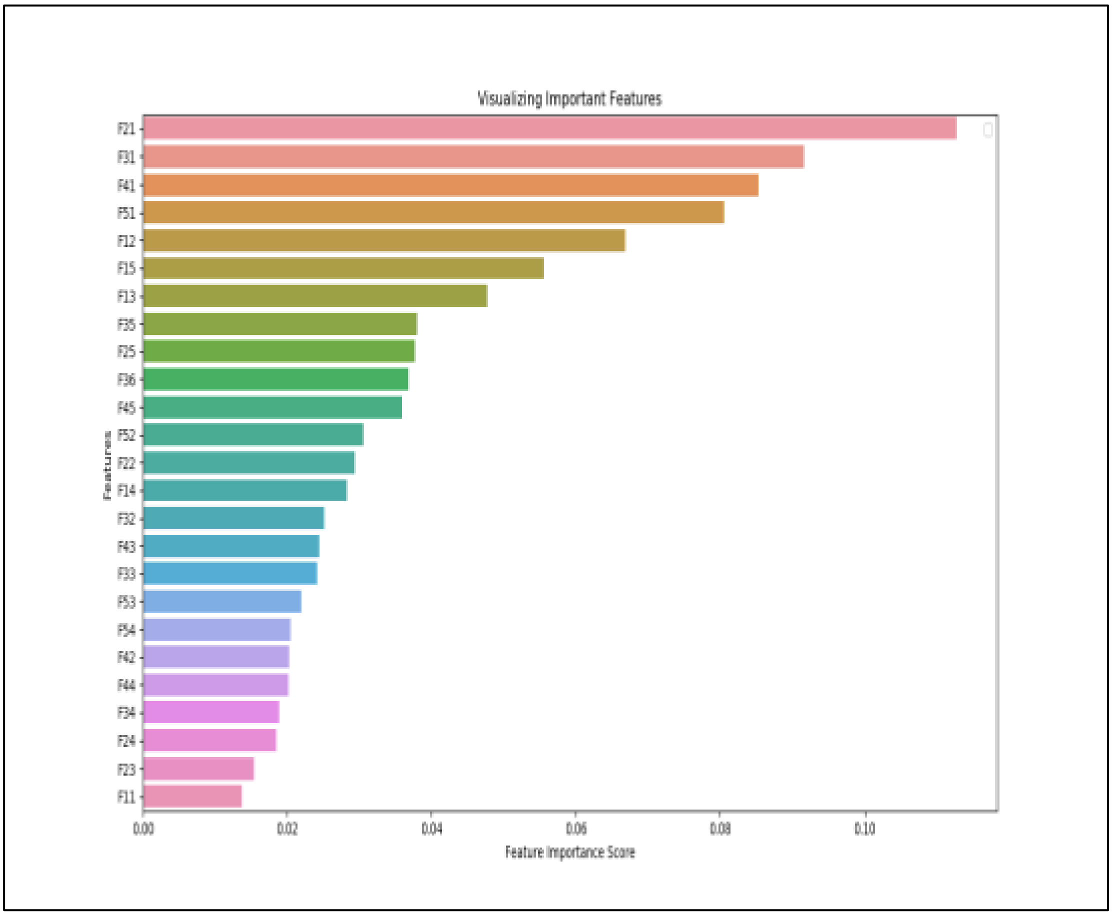

Figure 3- The sorting of the feature's importance

5. Results and Discussion

The proposed work was carried out using Python (V.3.5) as a scripting language, and Py-Charm as a framework. Three stages have been applied to foresee the students' academic performance. The data has been gathered by using academic records of the students in the Iraqi schools and questionnaires. The number of questionnaires was 2050, the number of features was 25 features. The active learning was utilized in the classification, the dataset was divided into the following: labeled training set (512, 25) instance, unlabeled training set (513) instance, and the testing set (1025) instance. Four methods for classification had been applied. Table 4 presents the accuracy for each method, while Figure 4 shows the comparison between the accuracy of methods. Table 5 shows the classification report for each method.

Table 4- The accuracy for each classifier

The method	***The Accuracy***
Active RF	95.121%
Active Logistic Regression	92.292%
Active MLP	93.951%
Active Label Propagation	92.195%

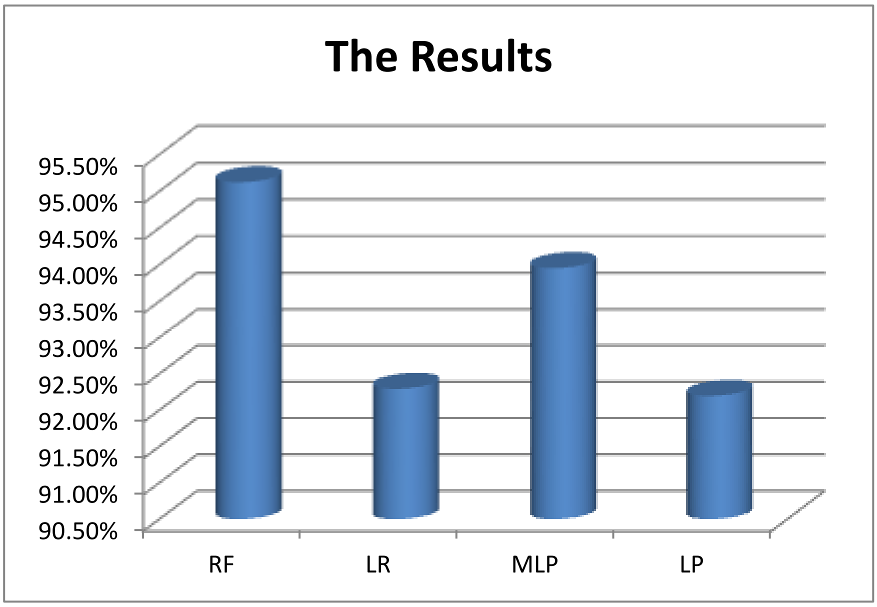

Table 5- The classification report of each classifier

Table 5- The classification report of each classifier

The method	***The Class***	***Precision***	***Recall***	***F1-score***
Active Logistic Regression	Fail (0)	0.93	0.91	0.92
	Pass (1)	0.91	0.93	0.92
Active RF	Fail (0)	0.93	0.97	0.95
	Pass (1)	0.97	0.93	0.95
Active MLP	Fail (0)	0.95	0.93	0.94
	Pass (1)	0.93	0.95	0.94
Active Label Propagation	Fail (0)	0.91	0.93	0.94
	Pass (1)	0.93	0.91	0.94

As shown in the results, the classification using active learning has a positive role in improving the accuracy especially if 0the dataset type that lacked in the label data. By comparing this study with the previous studies, the outcomes of this work had good comparing with the previous studies in the following parameters: The dataset's size in this work is more than the dataset's size in the previous studies, the type of the dataset lacked label data while datasets in the previous studies depend on the label data in hence depends on the supervised learning, the accuracy of the suggested methods in this work more than accuracy in the previous studies. As shown in Table 4, an active RF got more accuracy compared with other methods because depends on many decision trees for classifying the features. Table 6 presents the comparison between the related works and the current work. Also, the sorting importance of features had been applied using RF which is considered the type of machine learning technique. The results have shown that the two following factors: the ambiguity of scientific material, and difficulty of curriculum gotten high importance effect on the achievement of students, while the lacking the educational devices in the schools and not suitable of the infrastructure has less affection as shown in the Figure 3.

Table 6- Comparison related works with proposed work

The study	Data set	The technique	The Results
The prediction of students' academic performance using classification data mining techniques[5]	The dataset contained 497 records.	Decision Tree (DT), Naïve Bayes (NB), and JRip	the results were 68.8% for DT, 67.0% for NB, and 71.3% for JRip
Predicting students' final degree classification using an extended profile[7]	The dataset was 470 records. features were combined from the factors are (institutional, academic, demographic, psychological, and economic)	Multi-layered neural network (NN).	The accuracy was 83.7%
Predicting academic performance by considering student heterogeneity[2]	The dataset contained the enrollment data of the student at Australian University which contained all information about the student.	NB, Sequential minimal optimization (SMO) for Support Vector Machine (SVM), decision tree (J48).	The results were 74.2% for NB, 73.87% for SVM, 76.19%, and 73.84% for J48.
Student academic performance prediction using supervised learning techniques[1]	The dataset belonged to the Alentejo region of Portugal. It includes 1044 instances with 33 attributes including student grades, demographic, social, and school-related features	MLP, NNge, and decision tree (J48)	The accuracy was 92.39% for MLP, 92.81% for NNge, and 95.78% for J48.
Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network [8]	The size of the dataset was 682 of engineering student attrition., The model utilized 58 variables related to demographics and academic background of a student	The probabilistic neural network (PNN) used deep learning	The accuracy was 90.1%
Predicting student academic performance using data mining techniques[9]	The dataset in this study contains 480 records, the number of features was 16 features.	RF, LR, Perceptron, DT, and SVM.	The accuracy were 69.7% for RF, 67.7% for LR, 64.5% for Perceptron, 46.8 % for DT, and 70.8% for SVM
The proposed work	The dataset contained a 2050 record, the number of attributes was 25 attributes related to demographics, grades, demographic, social, and school.	Active learning for the following techniques (RF, LP, LR, MLP)	The accuracies were (95.121% for RF, 92.195% for LP , 92.292% for LR, and 93.951% for MLP).

6. Conclusion

The EDM technique was applied to predict the academic students' performance by collecting the data from the educational environment. There are many changes in the educational process elements in the last years such as curriculum, teachers, educational environment, students. So, this work presented methods for detecting the affection of educational elements on the prediction of the students' academic performance. The dataset lacked in the labeled dataset so, active learning has been used in the classification of features. Four methods have been applied for classifying were (RF, LP, LR, and MLP). As shown in the results, the RF algorithm presented good accuracy compared with other methods. Also, the important features had been computed that affect the students' academic performance using the Random Forest algorithm.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

[1] M. Imran, S. Latif, D. Mehmood, and M. S. Shah, "Student academic performance prediction using supervised learning techniques," *Int. J. Emerg. Technol. Learn.*, vol. 14, no. 14, pp. 92–104, 2019.

[2] S. Helal *et al.*, "Predicting academic performance by considering student heterogeneity," *Knowledge-Based Syst.*, vol. 161, pp. 134–146, 2018.

[3] Y. Zhang, S. Oussena, T. Clark, and H. Kim, "Use data mining to improve student retention in higher education: A case study," *ICEIS 2010 - Proc. 12th Int. Conf. Enterp. Inf. Syst.*, vol. 1 DISI, pp. 190–197, 2010.

[4] R. Asif, A. Merceron, and M. K. Pathan, "Predicting Student Academic Performance at Degree Level: A Case Study," *Int. J. Intell. Syst. Appl.*, vol. 7, no. 1, pp. 49–61, 2014.

[5] F. Ahmad, N. H. Ismail, and A. A. Aziz, "The prediction of students' academic performance using classification data mining techniques," *Appl. Math. Sci.*, vol. 9, no. 129, pp. 6415–6426, 2015.

[6] V. Oladokun, A. Adebanjo, and O. Charles-Owaba, "Predicting students academic performance using artificial neural network: a case study of an engineering course," vol. 1, pp. 68–72, 2008.

[7] S. Al-Sudani and R. Palaniappan, "Predicting students' final degree classification using an extended profile," *Educ. Inf. Technol.*, vol. 24, no. 4, pp. 2357–2369, 2019.

[8] C. Mason, J. Twomey, D. Wright, and L. Whitman, "Predicting Engineering Student Attrition Risk Using a Probabilistic Neural Network and Comparing Results with a Backpropagation Neural Network and Logistic Regression," *Res. High. Educ.*, vol. 59, no. 3, pp. 382–400, 2018.

[9] L. Masangu, A. Jadhav, and R. Ajoodha, "Predicting student academic performance using data mining techniques," *Adv. Sci. Technol. Eng. Syst.*, vol. 6, no. 1, pp. 153–163, 2021.

[10] M. E. Ramirez-Loaiza, M. Sharma, G. Kumar, and M. Bilgic, "Active learning: an empirical study of common baselines," *Data Min. Knowl. Discov.*, vol. 31, no. 2, pp. 287–313, 2017.

[11] G. Kostopoulos, A. D. Lipitakis, S. Kotsiantis, and G. Gravvanis, "Predicting student performance in distance higher education using active learning," *Commun. Comput. Inf. Sci.*, vol. 744, no. Dm, pp. 75–86, 2017.

[12] A. G. D'Sa, I. Illina, D. Fohr, D. Klakow, and D. Ruiter, "Label Propagation-Based Semi-Supervised Learning for Hate Speech Classification," pp. 54–59, 2020.

[13] A. I. Schein and L. H. Ungar, "Active learning for logistic regression: An evaluation," *Mach. Learn.*, vol. 68, no. 3, pp. 235–265, 2007.

[14] I. Karlsson and I. Karlsson, *Order in the random forest*, no. June. 2017.

[15] J. Maiora, B. Ayerdi, and M. Graña, "Random forest active learning for AAA thrombus segmentation in computed tomography angiography images," *Neurocomputing*, vol. 126, pp. 71–77, 2014.

[16] S. Gumbarević, B. Milovanović, M. Gaši, and M. Bagarić, "Application of Multilayer Perceptron Method on Heat Flow Method Results for Reducing the in-situ Measurement Time," p. 8272, 2020.

[17] S. Wang, J. Tang, and H. Liu, "Feature Selection," no. January, 2016.