Received: October 22, 2021. Revised: December 14, 2021.

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Using Socio-Demographic Information in Predicting Students' Degree Completion based on a Dynamic Model

Mahmood Shakir Hammoodi[1] **Ahmed Al-Azawei[1]***

[1]*University of Babylon, Iraq*
* Corresponding author's Email: ahmedhabeeb@itnet.uobabylon.edu.iq

Abstract: Educational institutions prioritize identifying variables and features that can help understand students' completion in higher education. This research aims at predicting students' completion based on socio-demographic information. Unlike previous work, a dynamic prediction model was proposed here. Real data of 13262 undergraduate students at a public university in Iraq was integrated and cleaned first. This was followed by applying several feature selection techniques to identify relevant socio-demographic features that can affect the prediction of students' completion. This includes using correlation-based feature subset selection (CfsSubsetEval), symmetrical uncertainty with respect to the class (SymmetricalUncertAttributeEval), GainRatioAttributeEval, ReliefFAttributeEval, and CorrelationAttributeEval. Finally, a prediction model was built based on six prediction approaches which are Bagging, DecisionTable, HoeffdingTree, IBK, J48, RandomForest, and RandomTree. Overall, several different features showed a significant effect on students' completion in which societal regression and the type of secondary school had the highest weight. Furthermore, comparing the accuracy of the implemented classification techniques can reveal that the Bagging classifier outperforms other approaches. The accuracy of predicting students' completion based on this method was 87.56%. Drawing on the research outcomes, educational institutions should pay further consideration to the identified features to ensure students' success and degree completion.

Keywords: Socio-demographic features, Dynamic prediction model, Degree completion, Educational data mining, Knowledge-based systems.

1. Introduction

Higher education represents a key pillar in the enhancement process of societies. This encourages governments worldwide to invest heavily in developing this sector. Higher education refers to a field that provides a significant type of data about the key pillars of the educational process such as curricula, teachers, and students [1]. A primary aim of educational systems, therefore, is to provide appropriate knowledge and skills for all students. This can lead to obtaining successful careers within a specific time, but the ability of the global educational system to come upon this aim is a key determinant of social progress and economy [2]. Accordingly, students are the main stakeholders of all educational institutions in which their academic degree completion is the main priority of such organizations [3]. Degree completion means that students will either complete their degree or not. To identify features that may have a significant effect on students' completion, educational data mining (EDM) is widely used. Although data mining refers to extracting useful knowledge from raw data that can affect the decision-maker, EDM is a method or technique to extract significant information that has a potential influence on educational institutes [2].

EDM is necessary for contemporary education as a large amount of data available on students' databases overrides the ability of a human to extract useful knowledge without applying an automated analysis method. Automated prediction is a method that has dominated in many fields such as medicine, politics, biology, finance, and education in which its prevalence has been attributed to the great enhancement in machine learning techniques [4]. Accordingly, most educational institutions use these techniques to improve their educational systems and teaching practice [5]. Four major areas have been suggested for EDM which are: (1) enhancing students' models, (2) promoting domain models, (3) highlighting the pedagogical support equipped by educational software, and (4) carrying out scientific research [2, 6]. In this regard, earlier research categorizes EDM into four main fields. This includes applications that consider the evaluation of learners' performance, adaptive educational hypermedia systems, assessing learning materials in online settings, and detecting students' behavior [7].

Based on such advantages of EDM, educational institutions have to exploit students' datasets to extract useful knowledge that can help multipurpose decision-making. Students' completion represents one of the main concerns that educational institutions aim to achieve in different learning settings and environments [8]. However, there is a necessity to build a dynamic prediction model to consider the significance of each attribute and its role in the prediction process where the majority of previous literature was based on a static prediction model. This research, therefore, aims at using socio-demographic features in predicting students' degree completion based on a dynamic model and using a database of students' information in Iraq. It identifies features that may have a significant effect by calculating the weight of each feature. After that, the performance of several different classification techniques was compared to highlight the most accurate one.

This study extends previous literature on the role of socio-demographic variables in exploring higher education degree completion. Moreover, it explains features that can affect degree completion in the Iraqi higher education context in which this has not been investigated in previous work, to the best of the authors' knowledge. Accordingly, the key contributions of this research are threefold. First, it is based on real data and high number of students. Furthermore, the present study proposes a dynamic prediction model in which if the classification accuracy dropped down, some irrelevant features would be replaced with others. Hence, the classification accuracy was used to re-evaluate the attributes and re-build the model accordingly. This can clearly reflect the dynamicity of this research. Finally, the study uses information from different academic years, and includes different socio-demographic information. As such, the research outcomes contribute significantly in understanding the key variables that should be considered by the Iraqi higher education to enhance students' performance and ensure their courses' completion.

The rest of this paper is structured as follows. The next section provides general background on higher education in Iraq, related literature, and the main techniques implemented in this work. The explanation of the research methodology is presented in section three. Section four demonstrates the research findings along with their discussion. Finally, the key outcomes and recommendations drawn based on this investigation are concluded in the last section.

2. Previous literature

2.1 Higher education in iraq

The location of Iraq is in the east of the Arabian peninsula. Its previous civilizations (e.g., Assyrians, Babylonians, and Chaldeans) are considered the first of their kind worldwide [9]. The primary forms of education had begun parallel in both Egypt and Babylon [10]. Recently, the Iraqi education system has been recorded the best in the Arabian region [11]. After the second Gulf war in 2003, Iraqi higher education has been influenced by the violence in its governorates. From 2003 to 2012, about 500 academic staff were killed, whereas approximately 600 attacks against schools and universities were performed [12]. However, the Iraqi ministry of higher education has worked hard to overcome such crises and improve the quality of higher education [11]. The stages of higher education study in Iraq are divided into three major levels:

- Bachelor’s level: this is granted after four, five, or six years of study according to the discipline. Engineering, science, and education degrees, for example, are awarded within four years, whereas veterinary and dentistry medicine within five years, and finally medicine within six years.
- Master’s level: this is granted conferred within a minimum of two years. In the first year, students have to study different advanced modules based on their discipline, whereas research only has to be performed in the second year.
- Doctoral level: this is similar to an MSc degree, except for the study's period in which it is at least three years and should result in a novel work.

Iraq's educational system is based on providing free education at all levels, including higher education [9]. Accordingly, it may be difficult to provide high-quality education to such a huge number of students. This may affect students' degree completion within the specified periods. EDM therefore can represent a comprise solution to provide a clear picture of particular factors that can influence students' learning and achievement. Applying data mining techniques to investigate students' socio-demographic information can help highlight potential variables for degree completion.

2.2 Related literature

Several different studies have investigated factors that may contribute to explaining students' academic performance and variables that may affect students' degree completion in higher education. This direction of research is highly developed in recent years because of the widespread use of educational systems that can record and digitalize students' information effectively. Earlier literature has examined students at risk of completion or low achievement based on several factors such as socioeconomic status, peer pressure, low score at a particular level of study, and class size [13, 14]. Our study adds to such a line of research by investigating the role of socio-demographic factors in predicting higher education degree completion. This section reviews some of the previous literature that used machine learning to predict academic performance.

In [15], the bayesian profile regression was used to identify students who would more likely to drop out or achieve good academic results. The features considered in this research were the students' performance, resilience, and motivation. Overall, the research findings showed that the Academic Resilience Scale, the scoring of the motivational items, the scoring on the difficulties, and satisfaction during the academic life were the most influential attributes on students who were at risk of dropout. On the other hand, students who were more satisfied, resilient, and faced fewer difficulties in their academic study were less invaluable to the risk of dropout.

In [3], several features were used to predict students' completion in which the selected factors were based on family expenditures and students' personal information. Two types of methods were applied which are discriminative (e.g., Support Vector Machine (SVM), C4.5, and Classification and Regression Tree) and generative (e.g., Bayes Network and Naive Bayes) classification models. Overall, the results indicated that SVM outperforms other techniques in which the achieved F1 score was 0.867. Moreover, it was found that expenditures such as natural gas, telephone, electricity, accommodation, water, and miscellaneous expenditures were predictors of students' academic performance. However, the most important feature was family expenditure on education. Moreover, the best prediction accuracy was obtained from a hybrid of family expenditures and students' personal information.

A hybrid of demographic and performance features, as well as attributes that related to particular students' modules, were used in [16] to predict students at risk in higher education and those who can achieve a good honor degree. The first group included gender, age at entry, level of widening participation in higher education, disability (Yes/No), nationality, overall score in year one, overall score in year two, overall score in year three, the year they obtained their academic award, the award class, fee status as in Home, Overseas or European, and finally the name of the course of enrolment, whereas the second encompassed name of a module, module code, number of students enrolled on the module, average mark for the module, and individual mark for the module. The research outcomes successfully classified the students into two groups based on their possible performance. It was reported that Home students achieved significantly better than overseas learners. Furthermore, many of the attributes used were predictors of students' degree completion and academic achievement.

Another research study used socio-demographic features such as gender, age, education, ethnicity, disability, and work status as well as study environment attributes (e.g., course program and course block) in predicting academic performance [17]. The empirical analysis showed that ethnicity, course program, and course block were the most important features in the prediction process. Moreover, regression tree (CART) successfully classified 60.5% of the research data. The study concluded that integrating socio-demographic and environmental attributes can achieve better results than depending on enrolment data only.

Al-Azawei and Al-Masoudy [18] used both demographic and behavioral features to determine students' learning performance in a virtual learning environment (VLE). After building the prediction model based on the M5P regression algorithm, the results showed that some of the demographic features and all online behavioral variables were predictors of students' performance. The significant demographic attributes were region, highest education, disability, gender, and age band. This supports the effectiveness of socio-demographic information on learners' academic achievement.

Another research study compared the accuracy of several methods namely, k-nearest neighbor (KNN), support vector machine (SVM), decision tree, logistic regression, multilayer perceptron (MLP), adaboost, and extra tree classifier in the prediction of students'

Received: October 22, 2021. Revised: December 14, 2021.

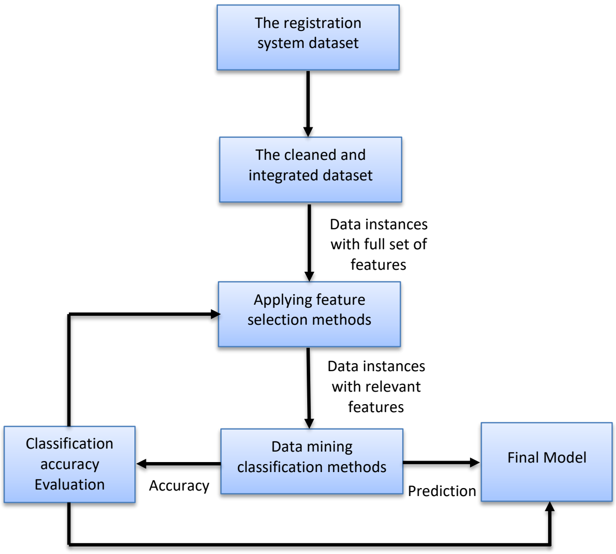

Figure. 1 The proposed model of predicting students' performance

achievement [19]. Although the first classification included predicting four classes which are excellent, good, poor, and fail, the accuracy was highly improved after categorizing the excellent and good into one class and the poor and fail into the other class. This research, however, did not clearly identify the important features that played a significant role in the prediction process.

To sum up, it cannot be generalized that applying a hybrid set of features could always present better findings as this may be attributed to other factors that may influence the strength of demographic, performance, and/or environmental attributes. Such features may include but are not limited to, society, culture, and individual circumstances. Thus, our research focuses on socio-demographic information that belongs to students' profiles on registration to understand if this individual student will complete his/her degree or not.

3. The proposed system

This research sought to achieve three key aims. First, it identifies the most influential socio-demographic features on students' completion in higher education. Second, the study implements several feature selection methods to weigh the effect of each attribute on predicting degree completion. Finally, seven classification models are built and their prediction accuracy is compared to highlight the best predictor of students' completion. To achieve these aims, a dynamic prediction model was proposed and different stages were implemented as explained in this section.

Fig. 1 shows the proposed framework for predicting students' completion. The framework consists of five key steps. This encompasses selecting the original database, cleaning it and integrating the research dataset, applying feature selection methods, using classification data mining algorithms to build a model, and evaluating the proposed model performance.

A full set of features were required to be ranked based on the well-known data mining feature selection techniques. A group of features was identified as the best candidate features based on such statistical methods (i.e., ranking). Subsequently, this group was sent to data mining classifiers to build the dynamic model. Hence, students' performance was connected directly to this group of features. However, this traditional process was applied in most of the previous approaches. This, in turn, leads to evaluating students' performance using specific features with a static model which is not realistic in the real application. In such a situation, students are affected by a variety of unexpected issues, daily. This needs to be taken into consideration when building a prediction model using data mining techniques. As such, the classification accuracy evaluation is applied in this research study to build the dynamic prediction model. If the classification accuracy dropped down, this means that some features became irrelevant and should be replaced with relevant features. In this research, therefore, the classification accuracy evaluation was proposed to re-evaluate the features and re-build the model accordingly. This process reflects the dynamicity of this present research.

3.1 The original dataset

The dataset used in this research was retrieved from the undergraduate registration system at a public university in Iraq. It provides main information about students such as name, address, age, gender, family information, health, etc. The original dataset consists of 62 features. Most of these are considered as sub information which could be varied from a student to another such as hobbies. Some students, for example, preferred football, whereas others preferred basketball and vice versa. Hence, such kinds of features would be not relevant in the case of students' performance evaluation and prediction. The greater the impact/relevant features on the largest number of students, the best evaluation and prediction could be achieved. However, evaluating students' performance using a specific set of features without taking into consideration the ignored features that would become relevant over time is unrealistic. This is because such features may have a significant effect on students' performance and

Received: October 22, 2021. Revised: December 14, 2021.

Table 1. The main attributes of the research dataset

Feature	Description
StudyType	Type of student registration (Morning Study or Evening Study)
Admission	Type of student admission (Morning Admission or Evening Admission)
Income	Monthly Students' Family Income
Marital status	Marital Status (Married, Single, Divorced, Widow)
Gender	Gender (Male or Female)
SchoolType	Type of Secondary School (Government, Private, Organizations, Others)
Health	Health (Good Health, Blind, Others)
SocialReg	Societal regression (Employment, Workers, Agricultural, Economic, Industrial, Others)
Label	Class labels (Pass or Fail)

it could make the generated model unable to interact.

3.2 Preprocessing of the original database

It was found that a user data entry may input zero value in the field of monthly income, or a student preferred not to mention his/her family income. Some features also consisted of string values (i.e., categorical) in combination with numerical values such as Emails and postcodes. Such features were also ignored. For example, a user data entry may input invalid data with or without spaces such as place of birth, the nearest place to a student's home address, student's hobbies, student's favourite sport, etc. In some important fields in the database such as students' age, it was also found that this field includes null values or invalid numbers. Another group of features are already irrelevant and ignored directly without any pre-processing technique such as phone number and mobile number. Handling the aforementioned issues might require advanced pre-processing techniques using expensive processing procedures before evaluating students' performance. This is because the prediction of students' performance relies on the relevant features which have a significant effect on a student during his/her daily study.

3.3 The research dataset

Only possible relevant features with numeric values and had not too much null entry were selected in advance, as shown in Table 1. However, in terms of data instances (rows), each data instance with at least one of the issues mentioned above was removed. Hence, the resulting dataset used here consists of 13262 data instances with two class labels and eight features. It is worth mentioning here that the features described in Table 1 are the best sample of features that may have a general effect on most students. However, the proposed framework evaluates the classification accuracy every time. Once the accuracy dropped down, the framework re-evaluates the features and re-builds the model accordingly.

3.4 Applying feature selection methods

Relevant features were identified through this component and then they were sent to data mining algorithms to build the classification model. The received instances (i.e., set of features) from the database were evaluated using some traditional feature selection methods such as correlation-based feature subset selection (CfsSubsetEval), symmetrical uncertainty with respect to the class (SymmetricalUncertAttributeEval), GainRatioAttributeEval, ReliefFAttributeEval, and CorrelationAttributeEval. The methods are available in an open-source framework called Weka which consists of a group of machine learning algorithms for data mining tasks.

3.5 Building the research classifier

The theoretical base of this research is grounded on educational data mining techniques. To predict students' degree completion accurately, seven classification algorithms were applied with 10-fold cross-validation. These are: Bagging, DecisionTable, HoeffdingTree, IBK, J48, RandomForest, and RandomTree. The performance of these classifiers was compared in order to choose the best one that can achieve higher accuracy and the key features that can predict degree completion in the proposed model.

3.6 Evaluating the research outcomes

In this stage, the performance of the proposed framework was evaluated in terms of classification accuracy. The conducted analysis aims to present the ability of the proposed framework with the best candidate features to predict study completion. The best features were identified by applying a set of well-known feature selection methods. Besides, the relevant features were used with the classification algorithms to measure the accuracy of the proposed model. However, once the accuracy drops down, the original full set of features was re-analysed again by applying the aforementioned feature selection methods. The selected features were evaluated by measuring the precision of the classification algorithm as shown in Eq. (1).

$$\mathrm{Precision} = \mathrm{CC}/\,(\mathrm{CC} + \mathrm{IC}) \qquad (1)$$

Received: October 22, 2021. Revised: December 14, 2021.

Table 2: Feature weight using feature selection methods

Feature selection method	No.	Features	Feature weight
CfsSubsetEval	1	Gender	0.09141
	2	SchoolType	0.04341
	3	SocialReg	0.01240
	4	StudyType	0
	5	Admission	0
	6	Income	0
	7	MaritalStatus	0
	8	Health	0
SymmetricalUncertAttributeEval	1	SchoolType	0.03246
	2	SocialReg	0.03232
	3	Income	0.01092
	4	Gender	0.00785
	5	Admission	0
	6	MaritalStatus	0
	7	Health	0
	8	StudyType	0
GainRatioAttributeEval	1	SchoolType	0.12662
	2	SocialReg	0.02864
	3	Income	0.00665
	4	Gender	0.00622
	5	Admission	0.00407
	6	MaritalStatus	0
	7	Health	0
	8	StudyType	0
ReliefFAttributeEval	1	Admission	0.00626
	2	SocialReg	0.00259
	3	MaritalStatus	0.00207
	4	StudyType	0.00076
	5	Income	0.00057
	6	Gender	0.00027
	7	SchoolType	0.00016
	8	Health	0
Correlation AttributeEval	1	SocialReg	0.17320
	2	SchoolType	0.14282
	3	Gender	0.09273
	4	Admission	0.05935
	5	MaritalStatus	0.01555
	6	Income	0.01286
	7	StudyType	0.01254
	8	Health	0.00852

where CC is the correct classification and IC is the incorrect classification.

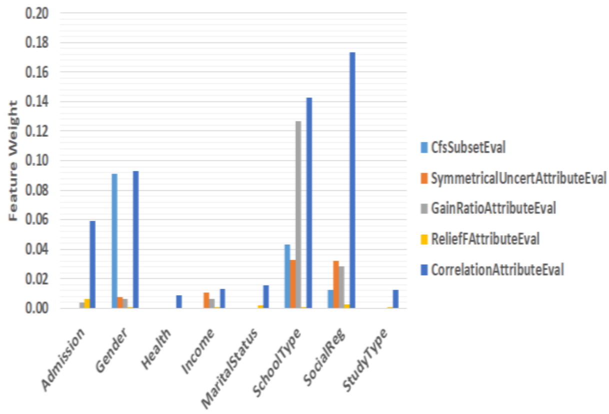

Figure. 2 A Comparison of feature weight using feature selection methods

4. Results and discussion

This research aims at predicting students' degree completion based on socio-demographic information recorded on the students' registration database. To meet this key aim, the research weights and evaluates all features first. This was followed by building the classification model. It was based on a dynamic prediction technique as all features were weighted and re-ranked over time.

4.1 Results of feature selection methods

Table 2 shows features' weight using feature selection methods. Features are listed from high to low weight according to the feature selection method. Fig. 2 shows a comparison of features' weight. It can be seen that the feature SocialReg had the maximum weight, and SchoolType comes next. The Health feature had the minimum weight. However, each group of features was evaluated in terms of classification accuracy, .i.e, Precision, as shown in Table 3 and Fig. 3. Features that reported the lowest weights were ignored.

Table 3 presents the selected features based on several different feature selection methods. Many features were significant in the prediction process regardless of a particular method. In each method, there is a selected group of features. The relevance of the selected features is identified by measuring the classification accuracy. Although higher precision was reported with some classification algorithms as shown in Fig. 3, this may change over time in real-life data and features may become relevant in some circumstances. Therefore, each group of features is used to evaluate the performance of the classification algorithms to predicate students' academic completion.

Received: October 22, 2021. Revised: December 14, 2021.

Table 3. Evaluation of feature selection methods

Method of FS	**No. of selected features**	**Classification algorithm**	***Precision***
CfsSubsetEval	3 Features: • Gender, • SchoolType, • SocialReg	Bagging	86.48%
		DecisionTable	86.51%
		HoeffdingTree	86.50%
		IBK	86.51%
		J48	86.51%
		RandomForest	86.49%
		RandomTree	86.51%
SymmetricalUncertAttributeEval	4 Features: • Gender, • Income, • SchoolType, • SocialReg	Bagging	87.28%
		DecisionTable	86.86%
		HoeffdingTree	86.29%
		IBK	87.01%
		J48	87.32%
		RandomForest	86.95%
		RandomTree	86.99%
GainRatioAttributeEval	5 Features: • Admission, • Income, • Gender, • SchoolType, • SocialReg	Bagging	87.31%
		DecisionTable	86.86%
		HoeffdingTree	86.36%
		IBK	86.86%
		J48	87.28%
		RandomForest	86.95%
		RandomTree	86.89%
ReliefFAttributeEval	8 Features: • StudyType, • Admission, • Income, • MaritalStatus, • Gender, • SchoolType, • SocialReg	Bagging	87.33%
		DecisionTable	86.81%
		HoeffdingTree	86.29%
		IBK	86.91%
		J48	87.30%
		RandomForest	86.98%
		RandomTree	86.96%
CorrelationAttributeEval	9 Features: All Features	Bagging	87.33%
		DecisionTable	86.78%
		HoeffdingTree	86.38%
		IBK	86.91%
		J48	87.30%
		RandomForest	86.98%
		RandomTree	86.93%

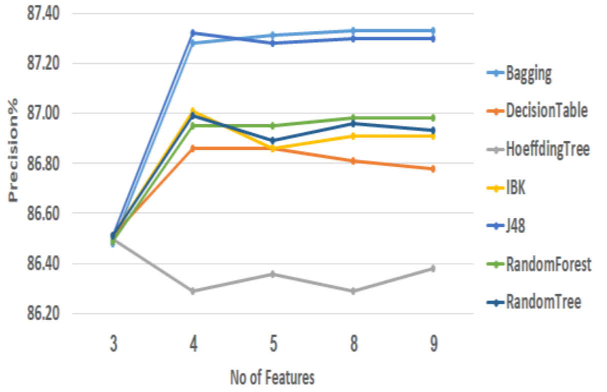

Figure. 3 Evaluation of selected features using feature selection methods and classification algorithms

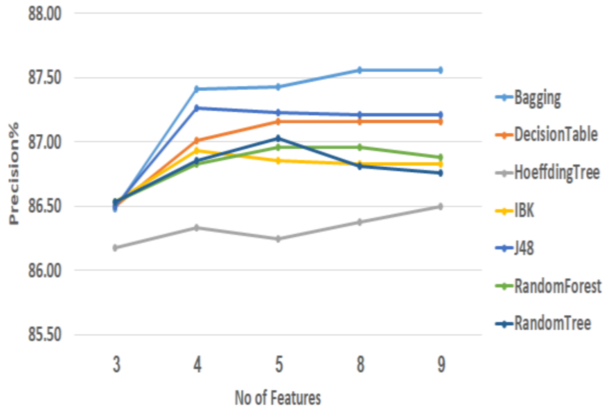

Figure. 4 Evaluation of classification algorithm model (Train and Test) in terms of selected features

4.2 Results of the classification algorithms

Data mining classification algorithms were used to build the prediction model of students' degree completion. Five experiments were conducted. For each experiment, the dataset described in Section 3.3 was used with the selected features. Overall, 70% of the dataset was used for training, whereas, 30% was used for testing. Fig. 4 shows a comparison of the classification algorithms performance in terms of precision using the selected features (3, 4, 5, 8, and 9 features). It can be seen that the performance of the classification algorithms was enhanced gradually in some cases, such as bagging. However, building a classification model using a specific group of features rather than the other may have a significant effect on evaluating the students' achievement in the future, as data may change over time. Hence, the model needs to be re-evaluated and this, in turn, represents the dynamicity of the proposed model.

The results showed that the selected features are valid to predict students' degree completion. It was found that societal regression and type of secondary school had the most effect on the prediction models.

Societal regression consists of different cases such as employment, workers, agricultural, economic, industrial, and others. This could indicate that a stable monthly income of a family may motivate students further to progress in their studies. This result is in agreement with another research study which showed that family income was a significant predictor of degree completion [20]. However, the study of Perez and Perez [20] indicated that the high school GPA was a determinant attribute of degree completion, whereas we found that the secondary school type is a predictor of this class. Other cases such as workers, agricultural, economic, and industrial might have a negative effect on students' performance as their overall monthly income is unstable. Hence, supporting students with stable monthly income during their studies may help enhance their overall academic achievement. The results also suggested that the type of secondary school would be connected directly with societal regression. The rationale of this is that students who studied at a private school might have full cost support to be the main reason for the success of their study.

The research findings are consistent with earlier literature. According to kumar and salal [5], the most features that can affect students' academic performance can be categorized into personal, family, and school groups. Our research supports this classification as it showed some of the students' personal features, family income, and secondary school type were predictfors of degree completion. In [21], the key role of demographic features was also confirmed based on a systematic review of previous research. In another research study conducted by Abo Saa [22], it was found that socio-demographic features were also determinants of students' academic achievement. The effect of such attributes can indicate that "the policy-makers can utilize such factors to create focus groups to take care of the students and provide them with special attention during their study in the institution" [21].

5. Conclusion

This research was conducted based on data of undergraduate students in Iraq. It showed that socio-demographic information was significant in predicting degree completion. However, the weight of such features varied based on a particular situation. Overall, the Bagging classifier outperformed other classification models with an accuracy of 87.56%. Thus, educational institutions should pay further attention to the features identified here to ensure students' successful completion of their degrees.

Although this research addressed a significant limitation in previous literature by proposing a dynamic prediction model, many directions still invite further research. This may include analysing the model accuracy dynamically over time, selecting the best relevant features through applying different techniques of feature selection and voting approach and selecting other socio-demographic features that may affect degree completion.

Conflicts of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization, Mahmood and Ahmed; methodology, Mahmood and Ahmed; software, Mahmood; validation, Mahmood and Ahmed; formal analysis, Mahmood; investigation, Mahmood and Ahmed; resources, Mahmood and Ahmed; data curation, Mahmood and Ahmed; writing—original draft preparation, Mahmood and Ahmed; writing—review and editing, Ahmed; visualization, Mahmood; supervision, Mahmood and Ahmed; project administration, Mahmood and Ahmed.

References

[1] P. Leitner, M. Khalil, and M. Ebner, "Learning Analytics in Higher Education — A Literature Review", in *Learning Analytics: Fundaments, Applications, and Trends, Studies in Systems, Decision and Control*, Springer International Publishing, pp. 1–23, 2017.

[2] A. Algarni, "Web Data Mining in Education", *Int. J. Adv. Comput. Sci. Appl.*, Vol. 7, No. 6, pp. 58–77, 2016.

[3] A. Daud, N. R. Aljohani, R. A. Abbasi, M. D. Lytras, F. Abbas, and J. S. Alowibdi, "Predicting Student Performance using Advanced Learning Analytics", in *International World Wide Web Conference Committee*, pp. 415–421, 2017.

[4] M. M. Ashenafi, G. Riccardi, and M. Ronchetti, "Predicting students' final exam scores from their course activities", In: *Proc. of Frontiers in Education Conference, FIE*, vol. 2015, no. October, 2015.

[5] M. Kumar and Y. K. Salal, "Systematic Review of Predicting Student's Performance in Academics", *Int. J. Eng. Adv. Technol.*, Vol. 8, No. 3, pp. 54-61, 2019.

[6] R. S. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions", *JEDM-Journal Educ. Data Min.*, Vol. 1, No. 1, pp. 3–17, 2009.

[7] F. Castro, A. Vellido, À. Nebot, and F. Mugica, “Applying data mining techniques to e-learning problems”, *Stud. Comput. Intell.*, Vol. 62, No. July, pp. 183–221, 2007.

[8] Q. Zhang, F. C. Bonafini, B. B. Lockee, K. W. Jablokow, and X. Hu, “Exploring Demographics and Students’ Motivation as Predictors of Completion of a Massive Open Online Course”, *Int. Rev. Res. Open Distance Learn.*, Vol. 20, No. 2, pp. 140–161, 2019.

[9] A. A. Azawei, P. Parslow, and K. Lundqvist, “Barriers and opportunities of e-learning implementation in Iraq: A case of public universities”, *Int. Rev. Res. Open Distance Learn.*, Vol. 17, No. 5, pp. 126–146, 2016.

[10] T. Saheb, “ICT , Education and Digital Divide in Developing Countries”, *Glob. Media J.*, Vol. 4, No. 7, pp. 1–8, 2005.

[11] N. Kaghed and A. Dezaye, “Quality Assurance Strategies of Higher Education in Iraq and Kurdistan: A Case Study”, *Qual. High. Educ.*, Vol. 15, No. 1, pp. 71–77, 2009.

[12] B. O. Malley, “Education under attack 2014”, 2014.

[13] A. Tamhane, S. Ikbal, B. Sengupta, M. Duggirala, and J. Appleton, “Predicting student risks through longitudinal analysis”, In: *Proc. of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1544–1552, 2014.

[14] K. R. Stevenson, “Educational Trends Shaping School Planning and Design: 2007 National Clearinghouse for Educational Facilities”, 2007.

[15] A. Sarra, L. Fontanella, and S. D. Zio, “Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework”, *Soc. Indic. Res.*, Vol. 146, No. 1–2, pp. 41–60, 2019.

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

[17] Z. J. Kovacic, “Early Prediction of Student Success: Mining Students Enrolment Data”, In: *Proc. of the 2010 InSITE Conference*, pp. 647–665, 2010.

[18] A. A. Azawei and M. A. A. A. Masoudy, “Predicting Learners’ Performance in Virtual Learning Environment (VLE) based on Demographic, Behavioral and Engagement Antecedents”, *Int. J. Emerg. Technol. Learn.*, Vol. 15, No. 9, p. 60, 2020.

[19] M. S. Zulfiker, N. Kabir, A. A. Biswas, P. Chakraborty, and M. M. Rahman, “Predicting students’ performance of the private universities of Bangladesh using machine learning approaches”, *Int. J. Adv. Comput. Sci. Appl.*, Vol. 11, No. 3, pp. 672–679, 2020.

[20] J. G. Perez and E. S. Perez, “Predicting Student Program Completion Using Naïve Bayes Classification Algorithm”, *I. J. Mod. Educ. Comput. Sci.*, Vol. 3, No. June, pp. 57–67, 2021.

[21] A. A. Saa, M. A. Emran, and K. Shaalan, *Factors Affecting Students’ Performance in Higher Education: A Systematic Review of Predictive Data Mining Techniques*, Vol. 24, No. 4. Springer Netherlands, 2019.

[22] A. T. M. A. Saa, “Mining Student Information System Records to Predict Students’ Academic Performance”, *The British University in Dubai*, 2020.