Proceedings of IYSC, (2022), vol. 11

Journal homepage: http://journals.sdu.edu.kz/index.php/iysw

GRADE PREDICTING SYSTEM

N. Zhailau

Suleyman Demirel University, Kaskelen, Kazakhstan

Proceedings of IYSC, (2022), vol. 11

Journal homepage: http://journals.sdu.edu.kz/index.php/iysw

2022 International Young Scholars' Conference

Abstract

Nowadays, prediction of academic performance became necessary for educational entities and universities. As you know most higher educational institutions have a portal system that monitors academic performance. This is necessary in order to assist at-risk students and ensure their retention, as well as to provide exceptional learning resources and experiences, as well as to improve the university's rating and reputation. Predictive analytics employed advanced analytics, including machine learning implementation, to improve achievement and to generate high-quality performance. As a result, the primary goal of this project is to demonstrate the feasibility of training and modeling on a small dataset size, as well as the feasibility of developing a prediction model with a credible accuracy rate. Using visualization and clustering algorithms, this study investigates the possibility of identifying the key indicators in the small dataset that will be used to create the prediction model.

*Keywords***:** Grade Prediction, Machine Learning, Predictive system, Algorithm techniques

Introduction

Every higher education institution has its own student academic management system to keep track of all student data, including academic results such as final test marks and grades in various courses and programs. Today universities are prestigious for higher education and considered to be the first step towards adulthood. It is a critical time for everyone because it determines our future. Because of this, student retention in these universities is a serious challenge. As a result, the university's tasks are to support students everywhere and provide them with everything they require. Everything will be new for first-year students at the start of their university careers. It is necessary to develop a solution that would help students stay in higher education institutions. So, early grade prediction is one of the solutions that has a propensity to monitor students' performance and will lead to an improvement in the students' learning process based on anticipated grades.

Data analysis is a powerful tool for understanding what is going on with data in the background. Machine learning, through data visualization, allows us to easily understand and analyze processes, as well as make decisions. Using Machine learning algorithms can improve the student achievements. Machine learning techniques can be used to predict student grades across courses. Such methods will help students improve their performance based on predictive grades

and enable teachers to identify individuals who may need assistance with courses. There are several types of algorithms used for prediction such as Collaborative Filtering, Matrix Factorization and Restricted Boltzmann Machines.

The paper is written out as follows. The second describes the related research work of the grade prediction system. The methodology of prediction models is explained in Section 3. Section 4 describes results of predictions and analysis. Finally, the main conclusions of the paper are highlighted, along with some future directions.

Literature review

Several research in higher education have been undertaken to predict student grades using various machine learning approaches. According to research undertaken by, a strategy for predicting future course grades received from the Computer Science and Engineering department has been devised. this study uses algorithms such as Collaborative Filtering, Matrix Factorization and Restricted Boltzmann Machines, to improve the accuracy of students' final grade prediction. They work on 225 real data of undergraduate students to predict student grades in different courses. In this study, they remark that using Collaborative Filtering does not provide an accurate prediction for a sparse database. Moreover, their study results indicate that the proposed RBM, especially for modeling tabular data, gives efficient learning and superior prediction accuracy than Collaborative Filtering and Matrix Factorization with a minimum Root Mean Squared Error (RMSE) of 0.3.

And also, another study tested 3 models, such as fivehidden-layers neural network, linear regression for supervised learning and linear regression for deep learning. This study used a dataset which was collected in Kaggle. The best mean average error (MAE) was 3.26 in the linear regression.

Methodology

This section discusses the dataset we collected for this study. In this chapter, the article presents data collection and data pre-processing and Machine learning algorithms and tools.

Data collection and preprocessing

As this research study is about the university grading system, we began by collecting the data from the Suleyman Demirel University portal. The data contains all speciality courses of 2016/2018 year students of Information System and Computer Science(Engineering faculty). The dataset was compiled from comma-separated value (CSV) files. 2 dataset files were collected. 1st to predict GPA by ENT score, 2nd to predict similar courses' final grade. Dataset consists of about

1200 students, and we have 5000 records for our analysis. The file record has the following rows: student name, student id, gender, final gpa, ent scores, teacher name, course name, mid 1, mid 2, final grade. And we have some issues like grading policy changes, and ent score point was also changed.

ID	STUD_ID	GENDER_ID	PROG_YEAR	YEAR	PROG_CODE	DERS_KOD	EMP_ID	LETTER_GRADE	GRADE
1	160103002	2	2016	2018	10103	INF 321	10120	C-	63
2	160103095	2	2016	2018	10103	INF 321	10120	D	50
3	160107011	2	2016	2018	10107	CSS 324	10514	F	0
4	160107020	2	2016	2018	10107	CSS 206	10126	B	80

Figure 1. Sample of student dataset

First, we deleted records with missing value, indicating that students dropped the course or did not receive a grade. To obtain appropriate data for various predictions, we performed some pre-processing on the collected data. We created a dataset to predict GPA using ENT scores, final exam grades, make a correlation between similar courses, and analyze elective Data Science courses. Because the algorithm learns from previous years, the predictions become more accurate as more data from previous years becomes available.

Furthermore, we wanted to predict the final grade of the course using absence count, so we pre-processed data with absence and participation count. However, we started having problems such as an insufficient data. According to our findings, the digital attendance system was first used in 2017 and has changed the limit of absences counts two or more times. That is why we do not predict course grades based on absence, but it'll be a great way of predicting final grades in the future.

Machine learning algorithms

I have used linear regression and decision tree techniques to predict GPA of the student for the courses.

Linear regression is a method for quantifying the relationship with one or more predictors and one or more target variable. Fitting a regression model to predict the values of new findings is one of the most common reasons.

Decision Trees (DTs) are a type of non-parametric supervised learning method that can be used for classification and regression. The focus is to create a model that predicts the value of the dependent variable using simple decision rules obtained from data features. A tree is an example of a piecewise constant approximation.

Results

Correlation and prediction gpa

I had the suggestion to predict GPA using ent score at the start of this work to see how ent results impact GPA and if students' school achievements are similar to their university performances. I received data from the university portal. It contains the most important information for students in the 2016–2018 school years. I deleted rows with a GPA of less than 1.47 and an ent score of less than 50. After data preparation, we have about 3098 records. First, we decided to investigate the relationship between ent and GPA. In this work, correlations measured by pearson correlation coefficient using Python.

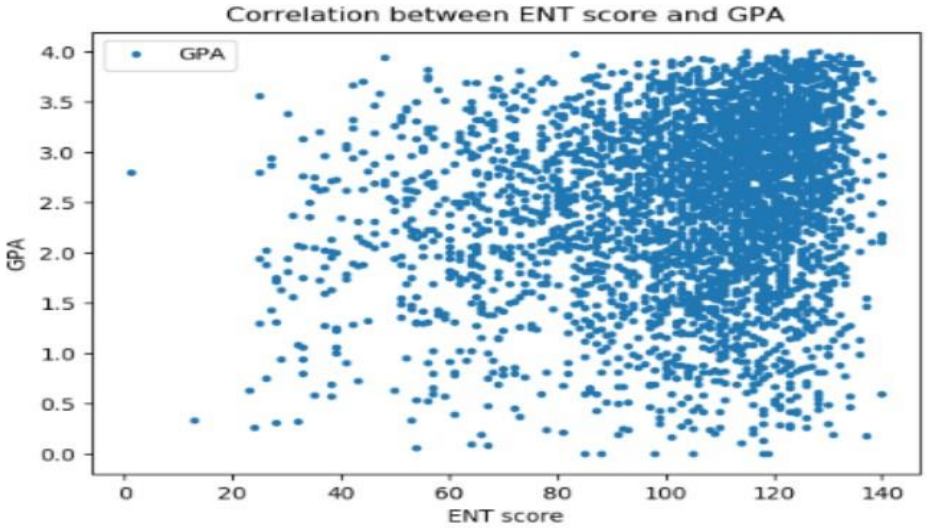

Figure 2. Correlation between ent and gpa

There you can see results, and correlation coefficient was low positive r=0.18 and relationship between them low and insignificant. In the figure 3, you can see the predicted

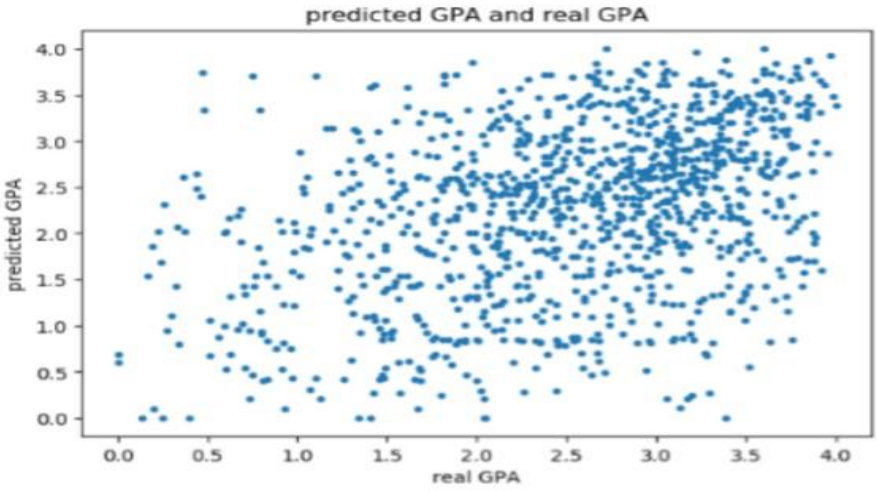

Figure 3. Predicting GPA with ent

Gpa, and to predict we have used the Decision tree algorithm on Python with Scikit learn. Using an ENT score alone to predict GPA returns no results. So we took GPA as an attribute value and introduced some elements such as class and program code that makes a more likely prediction. Because GPA keeps changing each semester, students' chosen courses and faculties also influence the change in GPA.

The possibility is extremely low. It is difficult to predict GPA with +50% accuracy because GPA is such a tiny number, and the correlation results show that the ENT score has little impact on GPA change. However, if we have more data for testing, the probability rises and the prediction becomes more accurate.

	Actual	Predicted
0	2.71	2.72
1	2.12	2.21
2	3.00	2.73
3	2.96	2.59
4	3.31	2.63
5	3.10	2.62

Figure 4. Predicting GPA with ent

Correlation and prediction final grade

There you can see predictions for final grades by similar courses. MAT153 and MAT 158(Mathematics for CS 1 and 2). I made a correlation between these courses using the Python Pandas package. The result was high positive correlation with Pearson correlation coefficient r=0.65. It is a very good result to predict. You can see that the relationship between these courses is very good, even if the teachers have changed. I built it using the DecisionTreeClassifier() class from the Scikit-learn library. We predicted the final grade with the accuracy 40%. The results shown in below:

Finally, I also predicted final grades. I used the Linear Regression class from the Scikit-learn library, to predict final grades using midterm grades as a regression model. Made correlation between midterm grades and final grade. Midterm1 grades' coefficient is 0.55, midterm-2 0.45. But this always changes depending on the chosen course. We got a good result as shown in Figures. But when grades are numeric values, it is hard to predict 100% accurately. So, to get better results I decided to predict grades using letter grades. I built it in the DecisionTreeClassifier() class in the

Scikit-learn library. Added features like year, term, teacher ID and used letter grade as a target variable. So, we predicted the letter grade using a classification method. The accuracy was 35%. These results are better than regression model results.

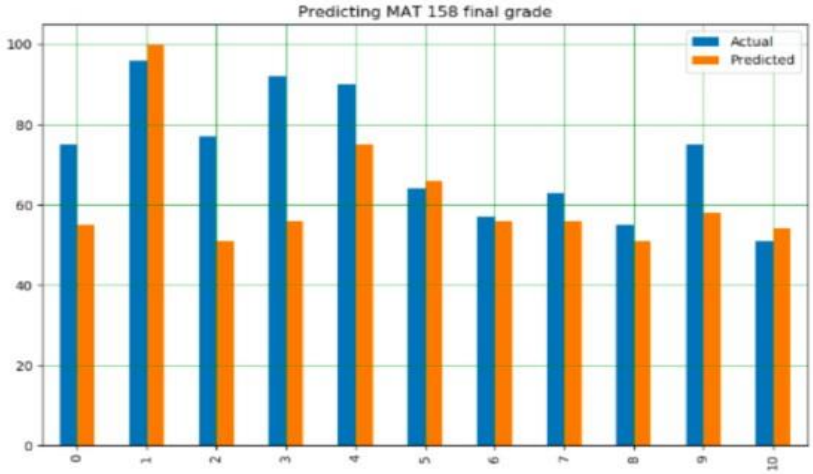

Figure 6. Predicting MAT 158 final grades

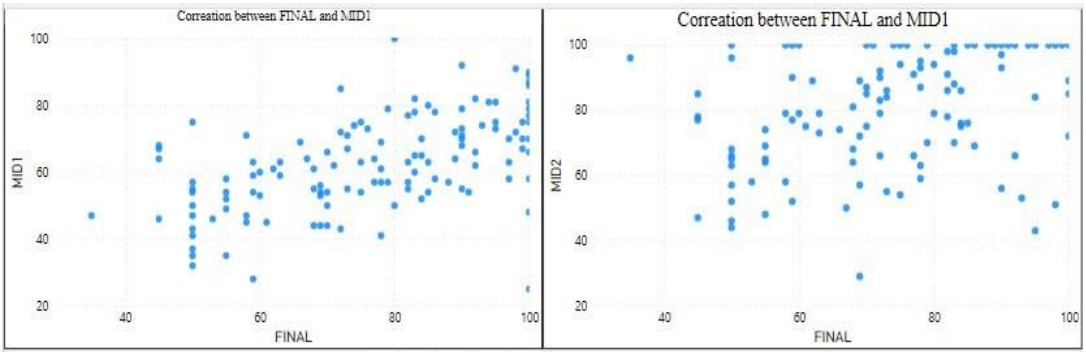

Figure 7. Correlation between final grade and MID1, MID2

Figure 8. Predicting final grade

Conclusion

Early warning system to motivate students and provide advance warning if they need to increase their skills in the courses. It also assists the course instructor in identifying weak students and providing necessary services to increase their performance. Thus, the rate of student retention can be increased. We discovered which characteristics have a significant impact on students' final grades. These predictions allow students to understand their predicted grades prior to the final exam and plan better. Additionally,in the future I want to do grade predictions by absence and teacher effectiveness. and also I'd like to do more predictions by using Ml algorithms in order to achieve really good results.

References

1. Mohammed M Abu Tair and Alaa M El-Halees. "Mining educational data to improve students' performance: a case study". In: Mining educational data to improve students' performance: a case study 2.2 (2012).
2. Mashael A Al-Barrak and Muna Al-Razgan. "Predicting students final GPA using decision trees: a case study". In: International Journal of Information and Education Technology 6.7 (2016), p. 528.
3. Prashant Gupta. "Decision trees in machine learning". In: Towards Data Sci- ence (2017).
4. Zafar Iqbal et al. "Machine learning based student grade prediction: A case study". In: arXiv preprint arXiv:1708.08744 (2017).
5. Emaan Abdul Majeed and Khurun Nazir Junejo. "Grade prediction using supervised machine learning techniques". In: e-Proceedings of the 4th Global Summit on Education (2016).
6. Qasem A Al-Radaideh, Emad M Al-Shawakfa, and Mustafa I Al-Najjar. "Mining student data using decision trees". In: International Arab Confer- ence on Information Technology (ACIT'2006), Yarmouk University, Jordan. 2006.
7. Linlin Zhang and Kin Fun Li. "Education analytics: Challenges and ap- proaches". In: 2018 32nd International Conference on Advanced Information Networking and Applications Workshops (WAINA). IEEE. 2018, pp. 193– 198.