Web-based student academic performance predictor based on study skills and habits

Alan Karl S. GATDULA, Lanie B. LAUREANO,
Rich Anjo M. CAPILOYAN, Apple Erika M. MEJOS

Faculty of Computing, Engineering, and Technology, Davao Oriental State University, Guang-guang, Dahican, City of Mati, Davao Oriental, Philippines

lanie.laureano@dorsu.edu.ph

Abstract: *The purpose of the study is to develop a web-based student academic performance predictor using the study skills and habits. The development started with the creation of the predictive model for student academic performance based on study habits and the second stage was the development of the web-based student academic predictor which allows the capturing of current study habits and displays the predicted academic performance. A study habits dataset with a total of 373 instances was uploaded to the Statistical Package for the Social Sciences software where the multiple regression analysis was applied. Results revealed that out of 12 variables, only four variables were significant in determining the academic performance which are the following: health habits (H), goal setting (GS), preparation and follow-up (PF), and comprehension (C). The model reveals a strong positive relationship (R = 0.752) with 56.65% of GPA variance explained. The model was successfully integrated in the web-based student academic predictor.*

Keywords: Predictive Model, Academic Performance, Study Habits, Multiple Regression Analysis.

1. Introduction

Issues regarding student academic performance have always been on the list when talking about the problems of students, parents, faculty, and staff within the institution (Lamas, 2015). Indeed, some authors noted student academic performance as one of the leading indicators for evaluating the quality of education in universities (Lawrence, 2014; Odiri, 2015). It has been the basis of teachers' evaluation and grading and information. Likewise, academic performance has been the basis on students' weaknesses and strengths and, at the same time, students' learning skills in their study (Dullas, 2010).

In the Philippines, many higher learning institutions used final grades to evaluate students' academic performance. Final grades are based on the scores obtained from quizzes, oral recitations, summative tests, lab activities, performance tasks, major exams, and various academic activities. However, study habits lost its importance due to the bad influence of mainstream and social media which led to

https://doi.org/10.58503/icvl-v20y202518

having inadequate study habits (Palani, 2012). Such a situation shows that academic performance is influenced by students' study habits (Odiri, 2015).

Recent studies have explored the development of web-based systems and predictive models aimed at assessing and enhancing student academic performance. For instance, the study of Yagci (2022) focused on predicting final exam grades of undergraduate students by employing machine learning techniques like random forests, nearest neighbor, support vector machines, logistic regression, Naïve Bayes, and k-nearest neighbor algorithms, using midterm exam grades as input data. On the other hand, a study developed a web-based system that leverages machine learning algorithms — such as Support Vector Machine (SVM), Random Forest (RF), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Linear Regression (LR) — to predict students' total course scores at early stages, utilizing academic and demographic factors (Alboaneen et al., 2022).

Despite the significant advancements in web-based academic performance predictors and machine learning models, several research gaps remain. Many existing studies focus on general academic predictors based on demographic (Alboaneen et al., 2022; Rajendran et al., 2022; Sekeroglu et al., 2019) and performance-based factors (Alboaneen et al., 2022; Yagci, 2022; Alhassan et al., 2020; Sekeroglu et al., 2019), but only few studies integrate study habits and skills as primary variables in predictive models (Orji & Vassileva, 2023). While prior research has demonstrated the effectiveness of machine learning algorithms — such as Support Vector Machines (SVM), Random Forest (RF), and Artificial Neural Networks (ANN) — in predicting student performance, there is limited work on deterministic predictive models that leverage structured frameworks like multiple regression specifically tailored to study habits. Additionally, recent models (Waheed et al., 2020; Lau et al., 2019; Zohair & Mahmoud, 2019) tend to rely heavily on quantitative performance indicators, such as test scores and course completion rates, rather than exploring qualitative aspects like study habits, which are crucial for academic success.

To bridge the research gaps, the proponents developed a Web-Based Student Academic Performance Identifier, a web-based application designed to provide a structured and deterministic approach to predicting student academic performance through an embedded determinsitic predictive model. Unlike conventional models that rely heavily on quantitative indicators such as grades and test scores, this system integrates study habits and skills as primary factors in its predictive framework. Through this platform, students gain deeper insights into their academic strengths and weaknesses, enabling them to adopt more effective study strategies. By offering personalized recommendations, the application empowers students to enhance their learning habits and exceed their expected academic performance.

2. Creating the deterministic predictive model

In creating a deterministic predictive model, the researchers applied framework of the study illustrated in Figure 1.

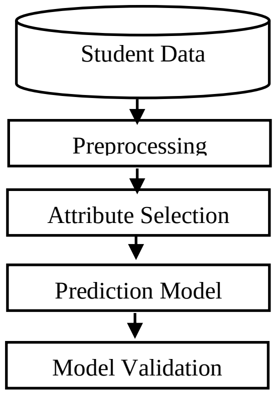

Figure 1. Framework for model creation

2.1 Dataset and preprocessing

The dataset was obtained through the administration of an adopted study skills and habits questionnaire from Queen's University (2013) which was available online. The adopted questionnaire underwent a validation process before it was administered to 373 students at Davao Oriental State University. The number of respondents was computed from the total population of DOrSU students on the 2nd semester of AY 2019-2020 through the Slovin's Formula. The respondents were also asked of their General Weighted Average for the said semester. Prior to the conduct of the survey, the consent of the respondents was obtained and was informed that the information collected follows the Data Privacy Act.

After tabulating the collected responses, data preprocessing was conducted to ensure consistency and accuracy. During this stage, missing values were identified, particularly in the GPA attribute, where several respondents had not provided their GWA. To address this issue, the ReplaceMissingValues filter in Weka was applied, replacing missing GPA values with the mean of the existing GPA distribution. This approach minimized data bias while preserving the overall characteristics of the dataset. In addition to handling missing values, further preprocessing steps were implemented to prepare the data for analysis. Categorical responses from the survey, such as study habits and time management strategies, were transformed into numerical representations using One-Hot Encoding or Label Encoding, ensuring compatibility with statistical and machine learning models.

To maintain data integrity, outlier detection was performed using box plots and Z-score analysis to identify extreme values in the GPA attribute. Any detected outliers were either validated, winsorized, or removed, depending on their impact on the dataset. By applying these preprocessing techniques, the dataset was refined and prepared for statistical analysis and modeling, ensuring its reliability and consistency while minimizing errors and biases.

2.2 Attribute selection and predictive model generation

SPSS is the software used in attribute selection and generation of the predictive model. The study habits dataset was uploaded, and multiple regression analysis was utilized in analyzing the relationship between a single dependent (criterion) variable and several independent (predictor) variables (Hair et al., 2010).

2.3 Validation procedures

After the creation of the predictive model, the mean absolute percentage error (MAPE) was computed to test the accuracy of the model. The proponents then chose a new set of 30 respondents to answer the same survey questionnaire but this time, they only answered the four indicators of study habits that were considered as significant towards the prediction of GPA. The formula in getting the mean absolute percentage error (MAPE) is:

$$MAPE = (1/n) * \Sigma(|actual – forecast| / |actual|) * 100$$

where:
Σ – a fancy symbol that means “sum”
n – sample size
actual – the actual data value
forecast – the forecasted data value

3. Development of the web-based student academic performance identifier

3.1 Software methodology

The software methology utilized for the web-based development is Scrum Methodology as shown in Figure 2. According to Porras (2019), Scrum refers to an agile framework for managing a process such as software development that uses incremental and iterative practices. It consists of development cycles called ‘Sprints’ that help speed up delivery, ensure quality, and mitigate risks.

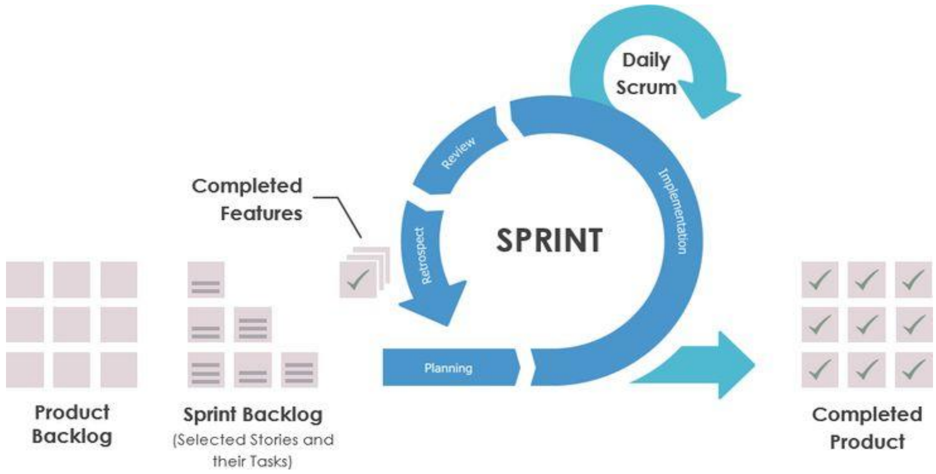

Figure 2. Scrum methodology (Porras, 2019)

3.2 Requirement documentation

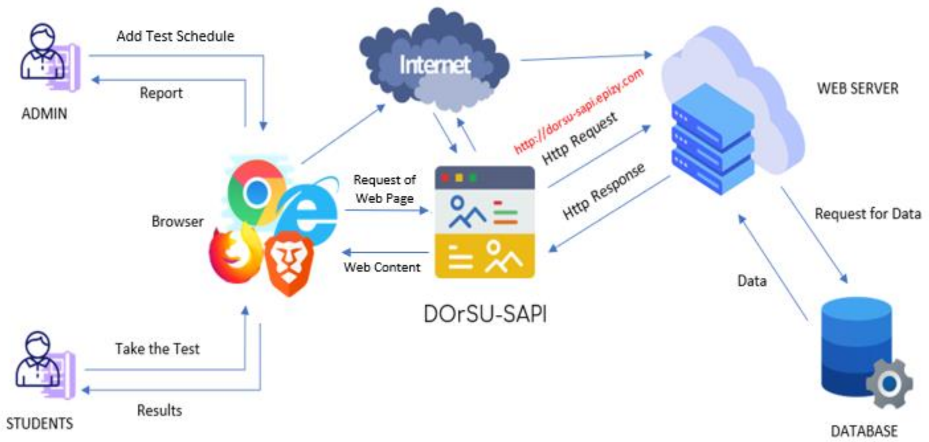

Figure 3. System architecture of web-based student academic performance predictor

Illustrated in Figure 3 is the system architecture of Web-based Student Academic Performance Predictor. There are two types of users who interact with the system. The first one is the administrator. This type of user manages the study skills and habits questionnaire including its schedule if he/she logged in the system. Furthermore, this user has access to students list and report. All the inputs were saved in the web server and the system displays it through user command. The second type of user is the student. This user can take the test if he/she logged in the system. The results were saved in the web server and could be retrieved and displayed by the system through user command.

4. Results and discussion

4.1 Predictive model based on study habits

Table 1. Coefficients of the model

Model	Unstandardized Coefficients		Standardized Coefficients	t	p-value	Decision on H_o
	B	SE	Beta			
(Constant)	2.222	.118		18.815	.000	Significant
Health	**.068**	**.030**	**.158**	**2.301**	**.022**	**Significant**
Time Management	.030	.038	.066	.784	.434	Not Significant
Attitude	-.006	.040	-.012	-.145	.884	Not Significant
Concentration	.028	.040	.065	.704	.482	Not Significant
Academic Stress	-.028	.037	-.071	-.753	.452	Not Significant
Goal Setting	**-.087**	**.041**	**-.199**	**-2.105**	**.036**	**Significant**
Preparation and Follow-up	**.083**	**.037**	**.207**	**2.255**	**.025**	**Significant**
Comprehension	**-.090**	**.044**	**-.209**	**-2.077**	**.039**	**Significant**
Selecting Main Ideas	.037	.043	.080	.862	.389	Not Significant
Use of Resources	-.005	.035	-.010	-.130	.897	Not Significant
Exam Preparation	-.016	.045	-.036	-.365	.715	Not Significant
Exam Writing	-.013	.042	-.029	-.313	.754	Not Significant
Model Summary: R = 0.752; R square = 0.5655; F-value = 72.706; p = 0.000						

Shown in Table 1 are the coefficients of the model. Upon applying the multiple regression analysis, results revealed that there were only four variables that significantly predict the value of the grade point average (GPA) of the students. These are health habits (H), goal setting (GS), preparation and follow-up (PF), and comprehension (C). The model is now:

$$GPA = 2.222 + 0.068H - 0.087GS + 0.083PF – 0.090C$$

This model suggested that the GPA of the students without the variables is 2.222. For every unit increase of their health, there will be an increase of 0.068. For every unit increase of their goal setting, there will be a decrease of 0.087. For every increase of their preparation and follow-up, there will be an increase of 0.083. And for every unit increase of their comprehension, there will be a decrease of 0.090.

It must be noted that this formatting of GPA infers that the lesser the GPA, the higher the value is.

Furthermore, the model summary stated that the correlation coefficient, denoted by R, has a value of 0.752 which indicates a strong positive linear relationship between variables (Puth, Neuhäuser & Ruxton, 2014). Meanwhile, the R-square value of the model is 0.5655, which implies that 56.65% of the variance of students' GPA is explained by the variables of the model. This makes the model reliable to be used because according to Moore et al. (2013), an R-square that is above 50% is acceptable.

The t-value of the model was used to calculate the p-value for testing whether the coefficient is significantly different from 0. Thus, the p-value determines whether the decision on null hypothesis is significant or not. The level of statistical significance is often expressed as a p-value between 0 and 1. A p-value less than 0.05 (typically ≤ 0.05) is statistically significant. It indicates strong evidence against the null hypothesis, as there is less than a 5% probability the null is correct. The results showed that only Health, Goal Setting, Preparation and Follow-up, and Comprehension obtained a p-value of less than 0.05. Thus, the decision on H0 for the mentioned variables is significant.

After the creation of the predictive model, the mean absolute percentage error (MAPE) was computed to test the accuracy of the model. The proponents then chose a new set of 30 respondents to answer the same survey questionnaire but this time, they only answered the four indicators of study habits that were considered as significant towards the prediction of GPA. The formula in getting the mean absolute percentage error (MAPE) is:

$$\text{MAPE} = (1/n) * \Sigma(|\text{actual} - \text{forecast}| / |\text{actual}|) * 100$$

where:
Σ – a fancy symbol that means "sum"
n – sample size
actual – the actual data value
forecast – the forecasted data value

With the formula of MAPE applied, the proponents were able to compute for the absolute percent error which can be seen on Table 2.

Table 2. Mean Absolute Percentage Error Computation

ACTUAL	FORECAST	Absolute Percent Error
1.9375	2.1316	10.01806452
1.8100	2.0366	12.51933702
1.7000	1.956	15.05882353
2.3020	2.0184	12.31972198
2.0000	2.124	6.2

1.6250	1.997	22.89230769
1.7500	2.137	22.11428571
2.2500	2.1038	6.497777778
2.5000	2.0916	16.336
2.2000	2.2242	1.1
1.9062	2.1338	11.93998531
2.3860	2.1274	10.83822297
2.4500	2.0504	16.31020408
2.0625	1.982	3.903030303
1.7500	2.0582	17.61142857
2.5500	2.1076	17.34901961
1.5000	2.05	36.66666667
1.6730	2.035	21.63777645
2.2917	2.0896	8.818780818
2.2188	2.1198	4.461871282
2.1354	1.8136	15.06977615
2.0833	1.9952	4.228867662
2.0658	2.0896	1.15209604
1.8333	2.1102	15.10391098
2.3750	2.0304	14.50947368
1.8571	2.0022	7.813257229
2.0938	2.288	9.275002388
2.6875	2.1536	19.86604651
2.7500	2.1542	21.66545455
2.1071	1.9516	7.379811115
	MAPE	**13.02190002**

In finding the mean absolute percentage error, the average values in the Absolute Percent Error must be calculated. In this case, the MAPE of this model is 13.02%. That means the average difference between the forecasted value and the actual value is 13.02%. This makes the predictive model more reliable because according to Lewis (1982), a MAPE between 10% and 20% is considered as a good MAPE score.

4.2 Web-based student academic performance identifier

Figure 4 shows the online survey system used in capturing the current study skills and habits of the students. The survey questions used was the validated adapted questionnaire.

Figure 4. Online Survey System

Once the student was able to answer the survey question, the predicted academic performance is generated along with the interpretation remarks.

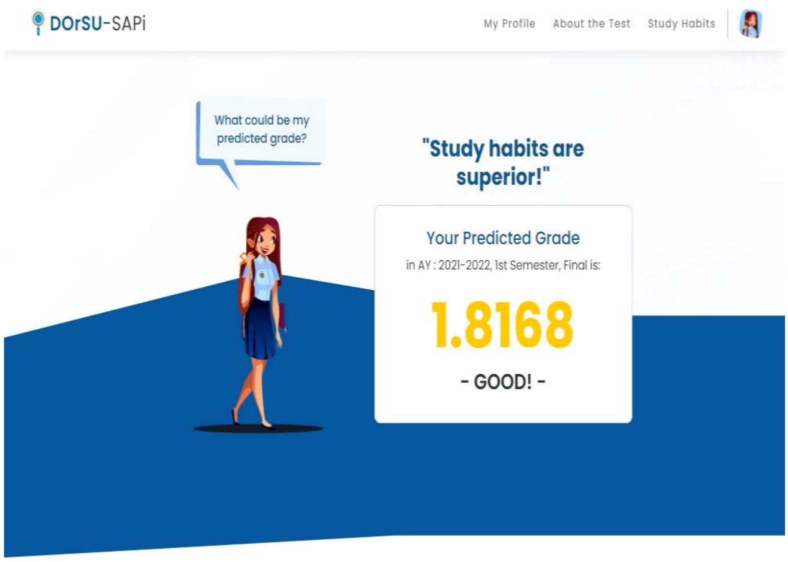

Figure 5. Predicted GPA

Shown in Figure 5 is the predicted GPA that can be found in the user interface of the Results Page. From the figure above, the web-based application generated the predicted GPA of the user and projected it immediately along with its interpretation right after the user completed the test.

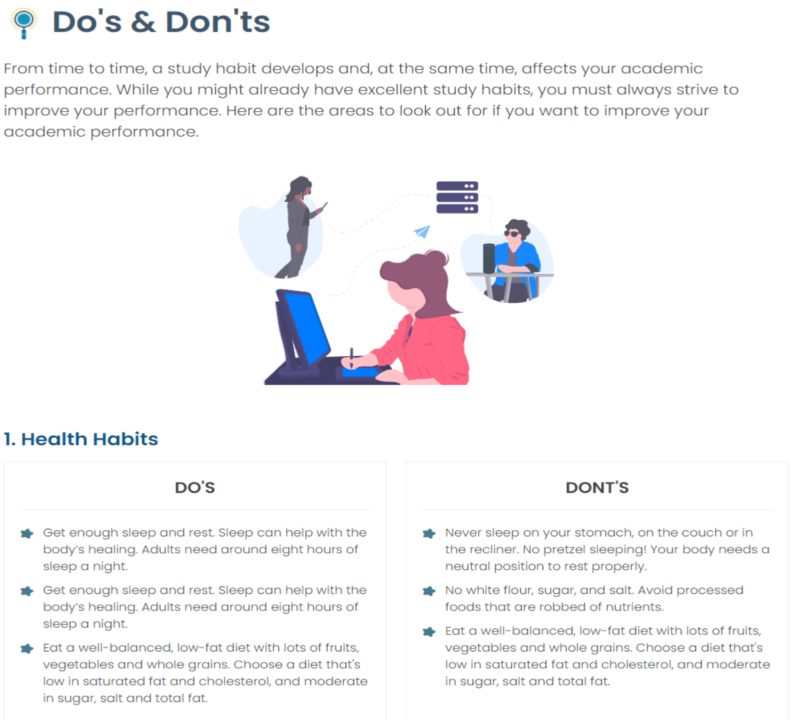

Figure 6. Do’s and Don’ts

Shown in Figure 6 are the Do’s and Don’ts in improving the study habits of the user. After the user completed the survey, the web-based application provides a tabular list of Do’s and Don’ts just below his/her predicted GPA.

5. Conclusion

In the light of the research results, the following conclusions are taken:

The predictive model that the proponents used in predicting students’ GPA is reliable because the value of R-square, which is 56.65%, is moderately acceptable (Moore et al., 2013), as well as the mean absolute percentage error (MAPE) of 13.02% which is interpreted as a good MAPE score (Lewis, 1982). During the user testing phase, the web-based application was able to capture the current study habit status of the students through online survey, display the expected academic performance of students through their current study habit status, store the survey records and feedback for self-monitoring purposes, and provide suggestions on how students can improve their study habits. Therefore, the Web-Based Student Academic Performance Predictor performed and functioned correctly according to design specifications.

REFERENCES

Alboaneen, D., Almelihi, M., Alsubaie, R., Alghamdi, R., Alshehri, L., & Alharthi, R. (2022) Development of a Web-Based Prediction System for Students' Academic Performance. *Data.* 7(2), 21. doi: 10.3390/data7020021.

Alhassan, A., Zafa, B. & Mueen, A (2020) Predict Students' Academic Performance based on their Assessment Grades and Online Activity Data. *International Journal of Advanced Computer Science and Applications(IJACSA).* 11(4), 185-194. doi: 10.14569/IJACSA.2020.0110425.

Dullas, A. (2010) Academic Performance and Self-Efficacy of Filipino Science High School Students on Mathematics and English Subjects. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2152791.

https://sass.queensu.ca/wp-content/uploads/sites/2/2013/05/Study-Skills-and-Habits-Questionaire.pdf. [Accessed 20th April 2023].

Lamas, H. (2015) School Performance. *Propósitos y Representaciones.* 3(1), 313-386. doi: 10.20511/pyr2015.v3n1.74.

Lau, E., Sun, L. & Yang, Q. (2019) Modelling, prediction, and classification of student academic performance using artificial neural networks. *SN Applied Sciences.* 1, 1–10.

Lawrence, A. (2014) Relationship between study habits and academic achievement of higher secondary school students. *Online Submission*. 4(6), 143–145.

Lewis, C. D. (1982) *Industrial and business forecasting methods: A practical guide to exponential smoothing and curve fitting*. London; Boston: Butterworth Scientific.

Moore, D. S., Notz, W. I. & Flinger, M. A. (2013) *The basic practice of statistics (6th ed.)*. New York, NY: W. H. Freeman and Company.

Odiri, O.E. (2015) Relationship of study habits with mathematics achievement. *J. Educ. Pract.* 6(10), 168–170.

Orji, F. A. & Vassileva, J. (2023) Modeling the Impact of Motivation Factors on Students' Study Strategies and Performance Using Machine Learning. *Journal of Educational Technology Systems*. 52(2), 274-296. doi: 10.1177/00472395231191139.

Palani, K. K. (2012) Promising Reading Habits and Creating Literate Social. *International Reference Research Journal*. 3(2), 91.

Porras, A. (2019, November 11) Scrum Methodology for Digital Product Development. *4Geeks*. https://blog.4geeks.io/scrum-for-digital-product-development/ [Accessed 23rd April 2023].

Puth, M.-T., Neuhäuser, M. & Ruxton, G. D. (2014) Effective use of Pearson's product–moment correlation coefficient. *Animal Behaviour*. 93, 183–189. doi: 10.1016/j.anbehav.2014.05.003.

Queen’s University. (2013) Study Skills and Habits Questionnaire.

Rajendran, S., Chamundeswari, S. & Sinha, A. A. (2022) Predicting the academic performance of middle- and high-school students using machine learning algorithms. *Social Sciences & Humanities Open*. 6(1), 1-15. doi: 10.1016/j.ssaho.2022.100357.

Sekeroglu, B., Dimililer, K. & Tuncal, K. (2019) Student performance prediction and classification using machine learning algorithms. *Proceedings of the 2019 8th International Conference on Educational and Information Technology*. pp. 7–11.

Waheed, H., Hassan, S.-U., Aljohani, N. R., Hardman, J., Alelyani, S. & Nawaz, R. (2020) Predicting academic performance of students from VLE big data using deep learning models. *Computers in Human Behavior*. 104, 106189.

Yağcı, M. (2002) Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learn. Environments*. 9, 11. doi: 10.1186/s40561-022-00192-z.

Zohair, A. & Mahmoud, L. (2019) Prediction of student’s performance by modelling small dataset size. *International Journal of Educational Technology in Higher Education*. 16(1), 1–18.