Detecting Factors and Exploring Different Techniques that Correlate with Academic Performance

Muhammad Faiz Arsalan   
Department of Computer Science  
Bina Nusantara UniversityJakarta, Indonesia  
muhammad.arsalan@binus.ac.id

Matthew Brian K. Munandar  
Department of Computer Science  
Bina Nusantara UniversityJakarta, Indonesia  
matthew.munandar@binus.ac.id

Novita HanafiahDepartment of Computer Science  
Bina Nusantara UniversityJakarta, Indonesia  
novita.hanafiah@binus.ac.idBagas Aditya Putra  
Department of Computer Science  
Bina Nusantara UniversitySemarang, Indonesia  
bagas.putra001@binus.ac.id

***Keywords— academic performance, machine learning, artificial intelligence, student, data mining, school***

# Introduction (*Heading 1*)

Ever since the COVID-19 pandemic, the world has never been the same. From the government and even to the educational system, students around the world are forced to adapt and restructure the way they live their everyday lives because sadly, one huge effect this situation currently has on students is academic performance.

This impacts students and teachers and especially on how they learn and how they teach will in fact, eventually affect students' marks and will represent instability until an efficient way of teaching and learning is presented since currently, there is no efficient way to gain the ability to adapt to these current factors. This problem induces the idea of Detecting Factors that Correlate with Academic Performance since it is known for a fact that these conditions have impacted everyone's academic performance, whether it be for the good or even the opposite.

The study proposes the use of various methods students are able to enact with datasets taken and observed throughout the students of Bina Nusantara University. The various studies will be held and the results will be presented whilst showing how the different methods can help benefit a student in their everyday academic performance.

Data will be collected throughout a survey and multiple online questionnaires to gain a baseline on how most students spend their time studying. The data mentioned will then be classified to predict and categorize the various student performances. We also seek to elucidate the different factors that affect student performance in relation to other variables in the data set of students by applying machine learning to further accurate the predictions.

# Literature Review

Scientists have regarded growing effectiveness in studying since the advancement of computer science's position in education. Typically, the aim was to predict student behavior and success in the future. There are several approaches that have been considered for accurate prediction. And although the datasets can vary, the most common are score and school activity; however, some papers contain additional factors such as parental status, home addresses, and family size [8].

A student's academic success is influenced by a variety of factors. Academic and non-economic variables can be divided into two categories. Socioeconomic status is one of the non-academic groups. Several socioeconomic factors, such as a student's gender and mother, play a significant role in their academic success [23]. Furthermore, the student's final exam is influenced by alcohol intake, family size, and home internet access [26].

Education plays a very important role in the development of society and predicting performance can be seen as a very helpful tool in aid, this is especially clear in the paper written by Liao, S. N., Zingaro, D., Alvarado, C., Griswold, W. G., & Porter, L. (2019, February). Writing about “Exploring the value of different data sources for predicting student performance in multiple cs courses.” [16] where it explored how the different datasets correlate to each other, specifically to those choosing Computer Science as their main major.

Various researchers realized the importance of data, especially in an academic setting. The majority of the researchers’ main focus is the importance of the students’ well-being in an academic situation. This has successfully resulted in the prevention of poorly-performed students. To overcome such problems, most researchers opted to make a prediction model due to the nature and uniqueness of the dataset that is given.

The papers that were discussed were all about predicting student success using various methods. This is evident throughout, especially in the areas of machine learning (e.g. [10], [11], [19], [20], [21], [22], [23], [24], [25]) and various analytical methods (e.g. [12], [13], [14], [15], [26], [27]). This has been proved accurate with all of the methods mentioned above, as each method outperforms conventional methods in terms of accurately predicting student performance.

The use of all these algorithms have been proven useful due to the many data sets available to collect due to the amount of variety each one provides. A few of the many techniques also include data mining and active learning(e.g. [17], [18], [26], [27]) where the more datasets obtained the more accurate the results will be [16].

Algorithms used also vary, starting from a simple decision tree until a hybrid system combining algorithms to minimize shortcoming compared to a single method. Support Vector Machine and Artificial Neural Network are the one commonly used based not just on the popularity observed from the relative work search, but their accurate prediction in assisting decision making. SVM’s complexity is O(n2) since the calculation of the scalar product among all vectors is required. On the other hand, ANN’s complexity is O(n), since it calculates the gradient for an error in an example, which is independent from the number of samples used [3]. Hybrid algorithms are also proposed in some of the research because integrating different predictive modeling techniques leads to high accuracy predictions compared to using a single approach [6]. Not only that, some of the proposed hybrid models could generate a very small RMSE. The hybrid RF algorithm produced the smallest value of RMSE which shows itself as the best predictive model in this prediction problem [2]. The most widely used technique for predicting students’ behavior was supervised learning, as it provides accurate and reliable results. In particular, the SVM algorithm was the most used by the authors and provided the most accurate predictions. In addition to SVM, DT, NB and RF have also been well-studied algorithmic proposals that generated good results [5].

Hence, the use of various artificial intelligence methods could help schools in predicting the performance of their students. With a reasonable prediction accuracy, there is no doubt in using artificial intelligence to track effectiveness in studying or just regular activities. As shown in Nieto Y.’s paper, graduation-rate predictions rely on students’ academic performance rather than socioeconomic factors [3]. And for the recent online classroom, Rivas A.’s work reflects the importance of interacting with a virtual classroom. However, it would be valuable if we could measure the time that students spend on interacting with a virtual environment, in order to be able to assess its benefits with certainty [1]. And in Aydoğdu, Ş.’s work that concludes artificial neural networks was created according to the behaviors of the students in a semester and the predictions were made through this network. In the subsequent studies, it is planned to create a model for predicting the performance of students based on instant browsing data (number of clicks, what content is viewed) and developing a system making recommendations to students based on the model created [7].

# Prepare Your Paper Before Styling

Before you begin to format your paper, first write and save the content as a separate text file. Complete all content and organizational editing before formatting. Please note sections A-D below for more information on proofreading, spelling and grammar.

Keep your text and graphic files separate until after the text has been formatted and styled. Do not use hard tabs, and limit use of hard returns to only one return at the end of a paragraph. Do not add any kind of pagination anywhere in the paper. Do not number text heads-the template will do that for you.

## Abbreviations and Acronyms

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

## Units

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.
* Do not mix complete spellings and abbreviations of units: “Wb/m2” or “webers per square meter”, not “webers/m2”. Spell out units when they appear in text: “. . . a few henries”, not “. . . a few H”.

Identify applicable funding agency here. If none, delete this text box.

* Use a zero before decimal points: “0.25”, not “.25”. Use “cm3”, not “cc”. (*bullet list*)

## Equations

The equations are an exception to the prescribed specifications of this template. You will need to determine whether or not your equation should be typed using either the Times New Roman or the Symbol font (please no other font). To create multileveled equations, it may be necessary to treat the equation as a graphic and insert it into the text after your paper is styled.

Number equations consecutively. Equation numbers, within parentheses, are to position flush right, as in (1), using a right tab stop. To make your equations more compact, you may use the solidus ( / ), the exp function, or appropriate exponents. Italicize Roman symbols for quantities and variables, but not Greek symbols. Use a long dash rather than a hyphen for a minus sign. Punctuate equations with commas or periods when they are part of a sentence, as in:

*a**b* 

Note that the equation is centered using a center tab stop. Be sure that the symbols in your equation have been defined before or immediately following the equation. Use “(1)”, not “Eq. (1)” or “equation (1)”, except at the beginning of a sentence: “Equation (1) is . . .”

## Some Common Mistakes

* The word “data” is plural, not singular.
* The subscript for the permeability of vacuum *μ*0, and other common scientific constants, is zero with subscript formatting, not a lowercase letter “o”.
* In American English, commas, semicolons, periods, question and exclamation marks are located within quotation marks only when a complete thought or name is cited, such as a title or full quotation. When quotation marks are used, instead of a bold or italic typeface, to highlight a word or phrase, punctuation should appear outside of the quotation marks. A parenthetical phrase or statement at the end of a sentence is punctuated outside of the closing parenthesis (like this). (A parenthetical sentence is punctuated within the parentheses.)
* A graph within a graph is an “inset”, not an “insert”. The word alternatively is preferred to the word “alternately” (unless you really mean something that alternates).
* Do not use the word “essentially” to mean “approximately” or “effectively”.
* In your paper title, if the words “that uses” can accurately replace the word “using”, capitalize the “u”; if not, keep using lower-cased.
* Be aware of the different meanings of the homophones “affect” and “effect”, “complement” and “compliment”, “discreet” and “discrete”, “principal” and “principle”.
* Do not confuse “imply” and “infer”.
* The prefix “non” is not a word; it should be joined to the word it modifies, usually without a hyphen.
* There is no period after the “et” in the Latin abbreviation “et al.”.
* The abbreviation “i.e.” means “that is”, and the abbreviation “e.g.” means “for example”.

An excellent style manual for science writers is [7].

# Using the Template

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper. In this newly created file, highlight all of the contents and import your prepared text file. You are now ready to style your paper; use the scroll down window on the left of the MS Word Formatting toolbar.

## Authors and Affiliations

**The template is designed for, but not limited to, six authors.** A minimum of one author is required for all conference articles. Author names should be listed starting from left to right and then moving down to the next line. This is the author sequence that will be used in future citations and by indexing services. Names should not be listed in columns nor group by affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization).

### For papers with more than six authors: Add author names horizontally, moving to a third row if needed for more than 8 authors.

### For papers with less than six authors: To change the default, adjust the template as follows.

#### Selection: Highlight all author and affiliation lines.

#### Change number of columns: Select the Columns icon from the MS Word Standard toolbar and then select the correct number of columns from the selection palette.

#### Deletion: Delete the author and affiliation lines for the extra authors.

## Identify the Headings

Headings, or heads, are organizational devices that guide the reader through your paper. There are two types: component heads and text heads.

Component heads identify the different components of your paper and are not topically subordinate to each other. Examples include Acknowledgments and References and, for these, the correct style to use is “Heading 5”. Use “figure caption” for your Figure captions, and “table head” for your table title. Run-in heads, such as “Abstract”, will require you to apply a style (in this case, italic) in addition to the style provided by the drop down menu to differentiate the head from the text.

Text heads organize the topics on a relational, hierarchical basis. For example, the paper title is the primary text head because all subsequent material relates and elaborates on this one topic. If there are two or more sub-topics, the next level head (uppercase Roman numerals) should be used and, conversely, if there are not at least two sub-topics, then no subheads should be introduced. Styles named “Heading 1”, “Heading 2”, “Heading 3”, and “Heading 4” are prescribed.

## Figures and Tables

#### Positioning Figures and Tables: Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table heads should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1”, even at the beginning of a sentence.

1. Table Type Styles

|  |  |  |  |
| --- | --- | --- | --- |
| **Table Head** | **Table Column Head** | | |
| ***Table column subhead*** | ***Subhead*** | ***Subhead*** |
| copy | More table copya |  |  |

1. Sample of a Table footnote. (*Table footnote*)
2. Example of a figure caption. (*figure caption*)

Figure Labels: Use 8 point Times New Roman for Figure labels. Use words rather than symbols or abbreviations when writing Figure axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”. If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

##### Acknowledgment *(Heading 5)*

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression “one of us (R. B. G.) thanks ...”. Instead, try “R. B. G. thanks...”. Put sponsor acknowledgments in the unnumbered footnote on the first page.

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. Rivas, A., Fraile, J. M., Chamoso, P., González-Briones, A., Rodríguez, S., & Corchado, J. M. (2019). *Students Performance Analysis Based on Machine Learning Techniques. Learning Technology for Education Challenges, 428–438.* doi:10.1007/978-3-030-20798-4\_37
2. Sokkhey, P., & Okazaki, T. (2020). Hybrid machine learning algorithms for predicting academic performance. *Int. J. Adv. Comput. Sci. Appl*, *11*, 32-41.
3. Nieto, Y., García-Díaz, V., Montenegro, C., & Crespo, R. G. (2019). Supporting academic decision making at higher educational institutions using machine learning-based algorithms. *Soft Computing*, *23*(12), 4145-4153.
4. Xu, J., Moon, K. H., & Van Der Schaar, M. (2017). A machine learning approach for tracking and predicting student performance in degree programs. *IEEE Journal of Selected Topics in Signal Processing*, *11*(5), 742-753.
5. Rastrollo-Guerrero, J. L., Gomez-Pulido, J. A., & Duran-Dominguez, A. (2020). Analyzing and predicting students’ performance by means of machine learning: a review. *Applied sciences*, *10*(3), 1042.
6. Alshanqiti, A., & Namoun, A. (2020). Predicting student performance and its influential factors using hybrid regression and multi-label classification. *IEEE Access*, *8*, 203827-203844.
7. Aydoğdu, Ş. (2020). Predicting student final performance using artificial neural networks in online learning environments. *Education and Information Technologies*, *25*(3), 1913-1927.
8. Sekeroglu, B., Dimililer, K., & Tuncal, K. (2019, March). Student performance prediction and classification using machine learning algorithms. In *Proceedings of the 2019 8th International Conference on Educational and Information Technology* (pp. 7-11).
9. Alsheddy, A., & Habib, M. (2017). On the application of data mining algorithms for predicting student performance: A case study. *Int. J. Comput. Sci. Netw. Secur*, *17*(10), 189-197.
10. 1. Latrellis, O., Savvas, I. Κ., Fitsilis, P., & Gerogiannis, V. C. (2020). A two-phase machine learning approach for predicting student outcomes. Education and Information Technologies, 1-20.
11. Oku, A. Y. A., & Sato, J. R. (2021). Predicting student performance using machine learning in fNIRS data. Frontiers in human neuroscience, 15.
12. Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S., & Ragos, O. (2020). Transfer learning from deep neural networks for predicting student performance. Applied Sciences, 10(6), 2145.
13. Hamoud, A., Hashim, A. S., & Awadh, W. A. (2018). Predicting student performance in higher education institutions using decision tree analysis. International Journal of Interactive Multimedia and Artificial Intelligence, 5, 26-31.
14. Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. (2016). Predicting student performance using personalized analytics. Computer, 49(4), 61-69.
15. Daud, A., Aljohani, N. R., Abbasi, R. A., Lytras, M. D., Abbas, F., & Alowibdi, J. S. (2017, April). Predicting student performance using advanced learning analytics. In Proceedings of the 26th international conference on world wide web companion (pp. 415-421).
16. Liao, S. N., Zingaro, D., Alvarado, C., Griswold, W. G., & Porter, L. (2019, February). Exploring the value of different data sources for predicting student performance in multiple cs courses. In Proceedings of the 50th ACM Technical Symposium on Computer Science Education (pp. 112-118).
17. Soni, A., Kumar, V., Kaur, R., & Hemavath, D. (2018). Predicting student performance using data mining techniques. International Journal of Pure and applied Mathematics, 119(12), 221-227.
18. Kostopoulos, G., Lipitakis, A. D., Kotsiantis, S., & Gravvanis, G. (2017, August). Predicting student performance in distance higher education using active learning. In International Conference on Engineering Applications of Neural Networks (pp. 75-86). Springer, Cham.
19. Yang, S. J., Lu, O. H., Huang, A. Y., Huang, J. C., Ogata, H., & Lin, A. J. (2018). Predicting students' academic performance using multiple linear regression and principal component analysis. Journal of Information Processing, 26, 170-176.
20. Hamsa, H., Indiradevi, S., & Kizhakkethottam, J. J. (2016). Student academic performance prediction model using decision tree and fuzzy genetic algorithm. Procedia Technology, 25, 326-332.
21. Kaviyarasi, R., & Balasubramanian, T. (2020). Predicting the academic performance of college students through machine learning techniques.
22. Babić, I. Đ. (2017). Machine learning methods in predicting the student academic motivation. Croatian Operational Research Review, 443-461.
23. Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. SN Applied Sciences, 1(9), 1-10.
24. Pandey, M., & Taruna, S. (2016). Towards the integration of multiple classifier pertaining to the Student's performance prediction. Perspectives in Science, 8, 364-366.
25. Ha, D. T., Loan, P. T. T., Giap, C. N., & Huong, N. T. L. (2020). An Empirical Study for Student Academic Performance Prediction Using Machine Learning Techniques. International Journal of Computer Science and Information Security (IJCSIS), 18(3).
26. Satyanarayana, A., & Nuckowski, M. (2016). Data mining using ensemble classifiers for improved prediction of student academic performance.
27. Salal, Y. K., Abdullaev, S. M., & Kumar, M. (2019). Educational data mining: Student performance prediction in academic. IJ of Engineering and Advanced Tech, 8(4C), 54-59.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**