International Journal of Human Research and Social Science Studies

ISSN(p): 3050-547X, ISSN(e): 3050-5488

Volume 02 Issue 04 April, 2025

DOI: https://doi.org/10.55677/ijhrsss/04-2025-Vol02I4

Page No : 129-135

A Student Dropout Risk Prediction Model Based on Supervised Learning Techniques and Large Language Models (LLMs)

Dat Tao Huu
Saigon University

ABSTRACT: Early prediction of student dropout risk is an essential but challenging task in Vietnamese higher education. This study proposes a novel model combining supervised machine learning and large language models (LLMs) to predict student dropout risk. The model utilizes structured information and unstructured data to analyze influencing factors comprehensively. By converting student data into natural language and using pre-trained LLMs, the model can understand the context and complex relationships between factors, thereby improving prediction accuracy compared to traditional methods. The study's main contributions are to propose architecture integrating LLMs into the dropout risk classification problem, identify critical factors influencing the decision to drop out and discuss the potential application of the model in practice to support early intervention.

Corresponding Author:
Dat Tao Huu

KEYWORDS:
Large Language Model (LLM), prediction, dropout risk, machine learning, supervised learning

1. INTRODUCTION

In Vietnamese higher education, student dropout remains a significant challenge. This study presents a novel model combining supervised machine learning techniques and large language models (LLMs) to predict the dropout risk. The proposed model exploits quantitative and qualitative data to analyze influencing factors comprehensively. The model can capture the context and complex relationships between factors by converting student data into natural language text and applying pre-trained LLMs (Villar, A., & de Andrade, C. R. V. (2024). Experimental results show that the model's predictive accuracy is significantly higher than that of traditional methods. The study also identifies key factors influencing the decision to drop out. It proposes an integrated LLM architecture for this problem, possibly applying the model in early warning systems at Vietnamese universities while supporting timely intervention (Sulak, S. A., & Koklu, N. 2024).

In this context, developing large language models (LLMs) opens up a new approach, allowing for more comprehensive educational data analysis. LLMs are large-scale deep learning models, trained on vast amounts of linguistic data, capable of flexibly understanding and generating natural language (Durrani et al., 2024). The application of LLMs in predicting the risk of dropping out of school promises to help the model exploit both quantitative and qualitative data of students, thereby improving accuracy and generality (Niyogisubizo et al., 2022; Kloft et al., 2014). This paper will present an overview of related studies and a proposed model combining supervised learning with LLMs, analyze the main factors affecting the risk of dropping out of school, discuss the advantages and disadvantages of the model, and conclude with future research directions.

Research question:

RQ1: How do we integrate quantitative and qualitative student data into a unified language model to predict the risk of dropping out?

RQ2: What advantages can large language models (LLM) apply in student context analysis over traditional prediction methods?

2. LITERATURE REVIEW

Research on dropout prediction has a history that parallels the development of educational analytics and machine learning. Initially, traditional statistical models such as logistic regression or decision trees were used to predict whether students would drop out based on a limited set of input variables (Márquez-Vera et al., 2016). For example, some studies used only demographic information of incoming students to predict the probability of dropping out before the start of the first semester. Demographic-based methods have the advantage of being simple and allowing for early warning before students enroll (Kim, H., & Lee, J. 2023). However, the disadvantage is that the accuracy is not high because many students, although not showing signs of risk from demographics, may still drop out due to other factors that arise during their studies (Sulak, S. A., & Koklu, N. 2023). Later studies gradually expanded the scope of data to include academic data during the university process. According to studies by Tinto and colleagues, students' behavior and academic performance in school are closely related to the decision to drop out: students who drop out tend to have poorer results, accumulate fewer credits, and withdraw from academic activities compared to successful graduates (Celestin, M., & Faustin, M. 2024). Indeed, an empirical study found that adding variables on final exam scores and the number of credits registered significantly increased the accuracy of the dropout prediction model. Therefore, modern models often incorporate admission input information (entry scores, family background) and learning process data (semester GPA, number of courses retaken, level of class participation, etc.) to improve predictive performance (Okoye et al.,2024).

In addition to traditional tabular data, unstructured data has proven valuable in predicting dropout. For example, academic advisors often record comments about students during advising; these text notes contain crucial qualitative information (Arizmendi et al., 2023). Kloft (2014) applied sentiment analysis to advisors' advising diaries, extracted positive or negative keywords related to each student, and then used these features as input to a machine learning model to predict which students were at risk of dropping out. The results showed that negative sentiment in advisor comments (e.g., students lack motivation and have integration problems) was highly correlated with student dropout. However, the limitation of this approach is that it only considers textual data and ignores other quantitative data (Rahman, M. S., 2016).

Recent research trends have focused on combining multiple data sources for more accurate predictions. Some pioneering studies have built multi-layer or multi-model models to parallel process tabular data (scores, attendance) and text data. For example, one model used an LSTM network to process the time series of scores and a Transformer network to encode the text, then combined the two outputs to make the final prediction. This approach has shown promising results, demonstrating that combining quantitative and qualitative information helps to gain a more comprehensive picture of a student's dropout risk (Ozdemir et al., 2024). However, using two separate models may miss hidden relationships between structured and unstructured data; for example, a decline in grades and negative comments may be closely related, but the separate model has difficulty recognizing the relationship.

Therefore, researchers began to look for a solution to merge data into a single model. Modern NLP technology suggests a direction: representing all information in natural language so that a large language model can process it simultaneously. Some works have experimented with converting tabular data into descriptive sentences and feeding the original text data into a predictive language model using natural language inference (NLI). Initial results are promising: Won et al. (2023) showed that the BERT language model, when fine-tuned in this way, significantly improved the predictive performance compared to traditional models (about a 9% increase in the F1-score compared to the previous best method). This is the premise for proposing a model that integrates supervised learning and LLM in this study.

3. PROPOSED MODEL

Based on the gaps and potentials from previous studies, we propose a dropout risk prediction model that integrates both traditional supervised learning and large language models to make the most of student data (Ersozlu; Vaarma et al., (2024); Wolf et al., 2020). The goal is to build a system that can classify students into two groups: those at high risk of dropping out or continuing to study (Niyogisubizo et al., 2022; Psyridou et al., 2024). The general architecture of the model consists of two main components: (1) Input data processor to extract and transform student features into appropriate representations, and (2) Prediction model combining LLM with classification layer to produce results.

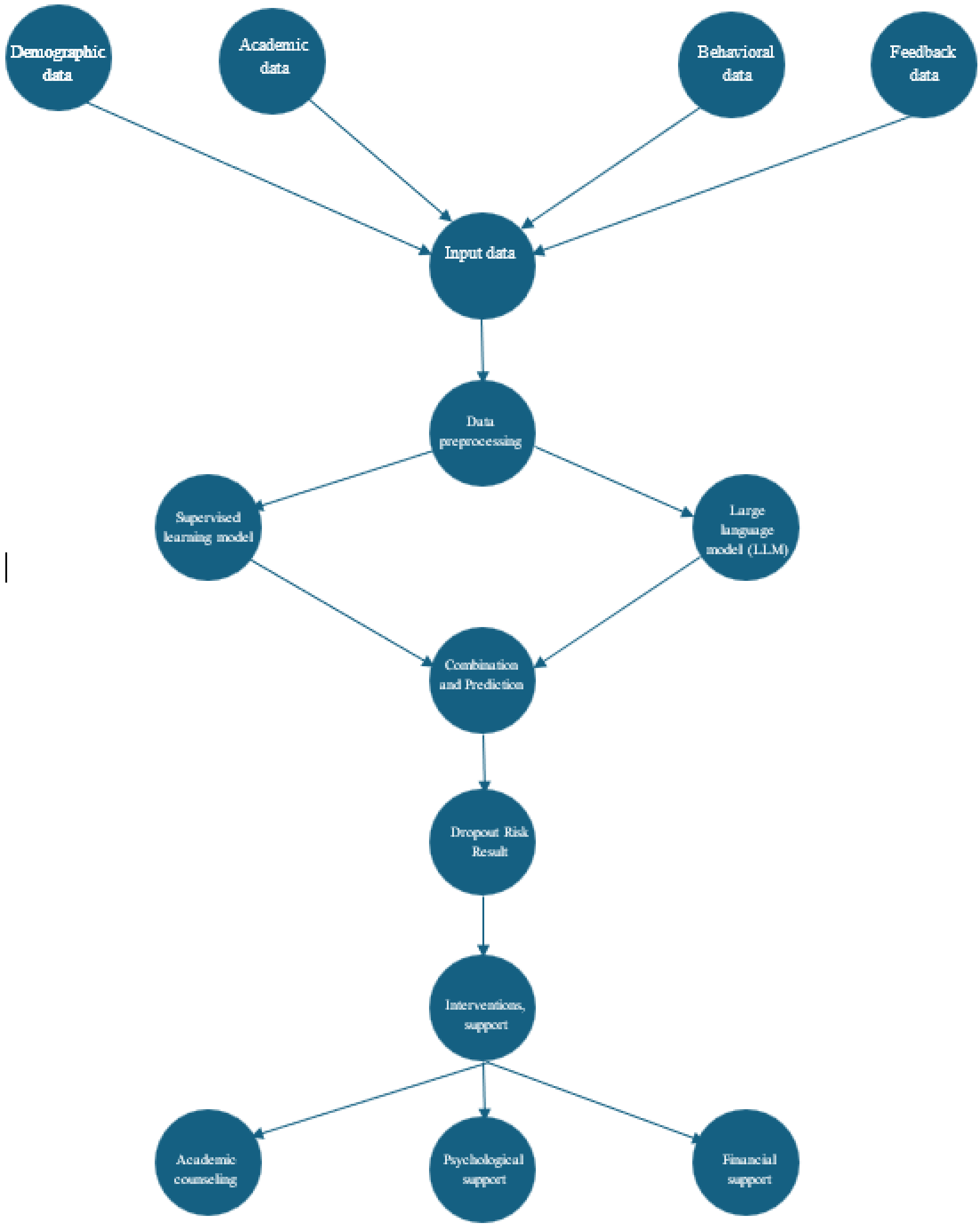

Figure 1. Research Model

(1) Input data:
- Demographic data (age, gender, region, family background, etc.)
- Academic data (GPA, course grades, credits, failed courses, etc.)
- Behavioral data (attendance, extracurricular participation, etc.)
- Feedback data (feedback from students, academic advisors, forums, social networks)

(2) Data preprocessing
- Normalize, encode, and transform text data to fit the model.

(3) Predictive model:
- Supervised learning model: Use algorithms such as Random Forest, XGBoost, or LSTM to predict based on structured data.
- Large language model (LLM): Analyze text data (comments, reviews, article content) to find factors related to the risk of dropping out.

(4) Combination and Prediction:
- Combine the output of the two models to decide on the student's dropout risk.

(5) Dropout Risk Result

- If the student is at high risk of dropping out, the system will trigger interventions such as academic support, psychological counseling, or financial aid.

(6) Interventions, support

- Academic counseling
- Psychological support
- Financial support

After collecting, the above data will be preprocessed to fit the language model. Numerical and categorical information (such as GPA, gender, major) is converted into short descriptive sentences. For example: "Male student, 20 years old, majoring in Computer Science, current GPA 2.5." or "Student has failed three subjects, regularly absent more than 20% of the sessions.". Existing texts, such as advisor comments, will be kept intact or summarized if too long. Next, the foundation will be used as a large pre-trained language model (such as BERT, Roberta, or XLM-R for Vietnamese). This LLM model can encode the aggregated text about students into semantically rich feature vectors. On top of the LLM, a classification layer (e.g., softmax neural network) is attached to predict the probability that students belong to two classes: dropout or non-dropout. The training process is supervised using historical data (which labels which students dropped out and which students graduated) to fine-tune the weights of the entire model (especially fine-tune LLM) to minimize prediction errors.

Another way to interpret the model (as illustrated in Figure 1) is in terms of natural language inference (NLI). We consider the student profile description the premise and formulate a fixed hypothesis: *"This student will drop out"*. The large language model evaluates whether this hypothesis is true (entailment) or false (contradiction) based on the information in the premise. If the model concludes that the hypothesis is true, dropping out is risky. In contrast, if the hypothesis is inconsistent with the premise, the student belongs to the group that continues studying (Galitsky, B. A., 2023). This NLI approach takes advantage of LLM's background knowledge of the relationships between factors in the linguistic context, thereby helping the model learn complex associations (Thapa et al., (2025); Kloft et al., 2014). The proposed model is general and can be flexibly adjusted: we can change the LLM architecture (use newer models or models suitable for Vietnamese) and add or remove some input types depending on the collected data.

4. KEY FACTORS

4.1. Academic Performance:

This is the strongest predictor. Students with low grades, failing multiple courses, or whose GPA decreases over the semesters are at a much higher risk of dropping out. Poor academic performance can make students lose motivation to continue studying, find it challenging to keep up with the program and lead to the decision to drop out (Srinivas et al., 2013). On the contrary, students with good performance are often more confident and committed to their studies.

4.2. Attendance & Engagement:

Frequent absences, skipping classes, and not participating in discussions or group activities are warning signs of student disengagement. Students who are less involved in academic and extracurricular activities on campus tend to feel isolated, less interested, and more likely to drop out (Dupéré et al., 2020; Devlin et al., 2019). Therefore, attendance rates, the number of missed classes, and participation in school events are essential variables in the prediction model.

4.3. Socio-economic Status:

Difficult family financial circumstances, low income, or lack of family support increase the risk of dropping out (Tsolou et al., 2020). Many students have to work too much or worry about the cost of studying, leading to them not being able to focus on learning. In addition, parents' education level and living environment (urban or rural, or an area with good educational conditions or not) also affect the ability of students to stay in university. Students from disadvantaged socio-economic groups often need additional support (Lessky, F., & Unger, M., 2022) to reduce the dropout risk.

4.4. Psychological & Personal Factors

Learning motivation, academic confidence, and mental health are hidden but essential. Students who lack clear goals and feel that their principal is unsuitable for their interests or abilities can quickly become discouraged and give up. In particular, poor mental health (Iqbal, A., Iftikhar, M., & Hussain, T., 2023) is strongly associated with the decision to drop out. Research has shown that students with signs of depression or prolonged stress are 1.5-2 times more likely to drop out of school than usual, and this number can be even higher for male students (~5 times). Therefore, monitoring and providing psychological support for students is crucial to preventing dropouts. In addition, other personal variables such as physical health status, family responsibilities, or sudden life events can also have an impact.

4.5. Academic & Social Environment:

A friendly learning environment, support from teachers and friends, and a connection to the school community help retain students. If students feel isolated, lack mentoring, or the learning environment is unattractive, the risk of dropping out increases (Watson, T. N., & Bogotch, I., 2016). On the contrary, mentoring programs, academic support, and interest clubs will create a sense of belonging and encourage students to continue. This factor is difficult to measure directly but can be represented through several variables, such as participation in extracurricular activities, student satisfaction surveys with the school, etc.

Of course, these groups of factors do not operate in isolation but interact. Effective predictive models need to consider the sum of these factors. For example, a student with average academic performance but from a disadvantaged background and low participation in activities may be at high risk as a student with poor academic performance. Therefore, our approach using a large language model, which can understand the combined context, helps identify risk cases that traditional methods based on individual variables may miss.

5. DISCUSSION

Comparison with traditional methods: The proposed LLM hybrid model offers a new approach compared to conventional methods that rely on basic statistical or machine learning models (Khalid, R. Z., Ullah, A., Khan, A., Khan, A., & Inayat, M. H. (2023)**.** Comparison of standalone and hybrid machine learning models for prediction of critical heat flux in vertical tubes. The main advantage of our model is its ability to learn complex features from data automatically. While traditional models require experts to pre-determine which variables are important and how to combine them, the LLM model can discover hidden patterns and relationships independently (Wan, G., Lu, Y., Wu, Y., Hu, M., & Li, S. (2024). For example, a logistic regression model may include GPA and absences as independent variables. In contrast, the LLM model can learn that the co-occurrence of "low GPA + high absences + negative comments" will form a powerful signal of dropout risk. As a result, the prediction accuracy of the LLM model is often higher. Indeed, research conducted on real-world data shows that the BERT-based natural language understanding (NLI) model outperforms the traditional model by about 9% in terms of F1-score, demonstrating the benefits of the new approach.

Benefits: Besides improving performance, the LLM model is flexible in handling many data types. Instead of building separate modules for tabular and text data, we only need a unified pipeline based on natural language (Mumuni, A., & Mumuni, F., 2024). This simplifies the system and has the potential to be widely applied to many cases. In addition, the language model has good generalization ability thanks to the knowledge trained on a vast data set, so it can recognize patterns that can be achieved even with small training data. Another significant benefit is that the model is linguistically explainable; for example, we can extract essential phrases or sentences from the premise that LLM focuses on to explain why the student is predicted to be at risk (Zhu, X., Li, Q., Cui, L., & Liu, Y., 2024).

Limitations: Despite its promise, the proposed model also faces some challenges. First, training and running a large language model requires much more computational resources (powerful CPU/GPU, large memory) than traditional models. This may make it challenging to deploy in schools with limited resources. Second, in terms of interpreting the results, the LLM model operates as a "black box" that is more difficult to understand than simple statistical models, making it difficult for administrators to explain to students or faculty why students are being warned of risks. Although some clues can be extracted, the level of transparency still needs to be improved (De Laat, P.B. (2018). Third, the more complex the model is, the more quality training data is required to be effective. The LLM model risks overfitting or losing accuracy if the historical data is low or not diverse. Finally, there are ethical and privacy issues: using a lot of personal student data, especially sensitive data such as advice notes and psychological information, requires appropriate security and consent mechanisms to avoid abuse or stigmatization of students considered at risk. Schools need policies to ensure that the model is used for positive support, not exclusion or discrimination.

Practical applicability: Despite some limitations, the dropout prediction model based on supervised learning and LLM has excellent potential for application in modern education systems. Universities can deploy this model as part of an Early Warning System (Nagy, M., Molontay, R., 2024). Specifically, updated student data will be fed into the model at the beginning of each semester to assess the probability of dropping out. Students with a probability exceeding the threshold will be added to the warning list so academic advisors or schools can proactively contact and support them. The model can also be integrated into learning management platforms (LMS), continuously monitoring learning behavior (grading, logging into online learning, participating in forums, etc.) to calibrate the forecast (Lee, S., & Chung, J. Y. , 2019). With the development of technology, many cases of timely intervention have helped students overcome difficulties and graduate successfully, proving that an accurate forecasting system brings excellent social value.

CONCLUSION

In this paper, we have presented a new approach to the problem of predicting the risk of students dropping out, combining traditional supervised learning techniques and large language models. Through the literature review, we see that integrating diverse data sources and exploiting the power of modern deep learning models is an inevitable trend to improve prediction accuracy. The proposed model uses LLM to interpret student data in the form of language, allowing the detection of complex patterns that are difficult to recognize by old methods. This is especially useful in practical contexts when schools need a reliable tool to identify students requiring support early. Although the experimental implementation is not yet implemented within the scope of the paper, theoretical analysis shows that the model has a lot of potential. An important future research direction is to test the model on real datasets from one or more universities, evaluate its performance, calibrate its parameters, and compare it directly with traditional methods. In parallel, attention should be paid to improving the interpretability and transparency of the LLM model: it is possible to study the application of model explanation techniques to extract the key factors that lead to prediction. In addition, expanding the range of factors is also an interesting direction – for example, integrating data on the labor market or trends in majors to predict students who are likely to drop out due to changing majors or entering the workforce early. Finally, it is necessary to develop intervention processes that are tied to the model: when the model alerts, what the next step is, who will interact with the student, and what support measures are put in place – these require interdisciplinary coordination between education professionals, psychological consultants, and administrators to ensure that the application of predictive models brings positive effects to learners and schools.

REFERENCES

1. Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, 3, 100066. https://doi.org/10.1016/j.caeai.2022.100066
2. Márquez-Vera, C., Cano, A., Romero, C., Noaman, A. Y., Fardoun, H. M., & Ventura, S. (2016). Early dropout prediction using data mining: A case study with high school students. *Expert Systems*, 33(1), 107–124. https://doi.org/10.1111/exsy.12135
3. Kloft, M., Stiehler, F., Zheng, Z., & Pinkwart, N. (2014). Predicting MOOC dropout over weeks using machine learning methods. *Proceedings of the EMNLP 2014 Workshop on Analysis of Large Scale Social Interaction in MOOCs*, 60–65. https://doi.org/10.3115/v1/W14-4111
4. Srinivas, K., Raghunathan, B. K., & Govardhan, A. (2013). Predicting student performance: A statistical and data mining approach. *International Journal of Computer Applications*, 63(8), 35–39. https://doi.org/10.5120/10489-5242
5. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 1, 4171–4186. https://doi.org/10.18653/v1/N19-1423
6. Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., & Rush, A. M. (2020). Transformers: State-of-the-art natural language processing. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, 38–45. https://doi.org/10.18653/v1/2020.emnlp-demos.6
7. Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D. M., Wu, J., Winter, C., ... Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901.
8. Sulak, S. A., & Koklu, N. (2024). Predicting student dropout using machine learning algorithms. *Intelligent Methods in Engineering Sciences*, 3(3), 91–98. https://doi.org/10.58190/imiens.2024.103
9. Hassan, M. A., Muse, A. H., & Nadarajah, S. (2024). Predicting student dropout rates using supervised machine learning: Insights from the 2022 National Education Accessibility Survey in Somaliland. *Applied Sciences*, 14(17), 7593. https://doi.org/10.3390/app14177593
10. Durrani, U. K., Malik, A., Akpinar, M., Dordevic, M., Togher, M., & Aoudi, S. (2024). Assessing the effectiveness of large language models in predicting student dropout rates. *Proceedings of the International Conference on Advanced Machine Learning and Applications*, 60–65.
11. Psyridou, M., Prezja, F., Torppa, M., Lerkkanen, M.-K., Poikkeus, A.-M., & Vasalampi, K. (2024). Machine learning predicts upper secondary education dropout as early as the end of primary school. *Scientific Reports*, 14, 12956. https://doi.org/10.1038/s41598-024-63629-0
12. Villar, A., & de Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence*, 4, Article 2. https://doi.org/10.1007/s44163-023-00079-z
13. Kim, H., & Lee, J. (2023). University student dropout prediction using pretrained language models. *Applied Sciences,* 13(12), 7073. https://doi.org/10.3390/app13127073
14. Sulak, S. A., & Koklu, N. (2023). Factors influencing dropout students in higher education. *Education Research International,* 2023, Article 7704142. https://doi.org/10.1155/2023/7704142
15. Celestin, M., & Faustin, M. (2024). School dropout and students' academic performance in public twelve years basic education schools of Rwanda. *Journal of Education, 7*(2), 20–33. https://doi.org/10.53819/81018102t5318

16. Okoye, K., Nganji, J. T., Escamilla, J., & Hosseini, S. (2024). Machine learning model (RG-DMML) and ensemble algorithm for prediction of students' retention and graduation in education. *Computers and Education: Artificial Intelligence*, 6, 100205. https://doi.org/10.1016/j.caeai.2024.100205
17. Arizmendi, C.J., Bernacki, M.L., Raković, M. *et al.2023.* Predicting student outcomes using digital logs of learning behaviors: Review, current standards, and suggestions for future work. *Behav Res* 55, 3026–3054 (2023). https://doi.org/10.3758/s13428-022-01939-9
18. Rahman, M. S. (2016). The advantages and disadvantages of using qualitative and quantitative approaches and methods in language "testing and assessment" research: A literature review. *Journal of Education and Learning,* 6(1), 102–112. https://doi.org/10.5539/jel.v6n1p102
19. Ozdemir, N.K., Kemer, F.N.A., Arslan, A. *et al.* A Qualitative Study of Unveiling School Dropout Complexity in Türkiye. *Child Ind Res* 17, 1001–1021 (2024). https://doi.org/10.1007/s12187-024-10116-7
20. Ersozlu, Z., Taheri, S. & Koch, I. A review of machine learning methods used for educational data. *Educ Inf Technol* 29, 22125–22145 (2024). https://doi.org/10.1007/s10639-024-12704-0
21. Vaarma, M., & Li, H. (2024). Predicting student dropouts with machine learning: An empirical study in Finnish higher education. *Technology in Society*, 76, 102474. https://doi.org/10.1016/j.techsoc.2024.102474
22. Niyogisubizo, J., Liao, L., Nziyumva, E., Murwanashyaka, E., & Nshimyumukiza, P. C. (2022). Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence, 3*, 100066. https://doi.org/10.1016/j.caeai.2022.100066
23. Galitsky, B. A. (2023). Truth-O-Meter: Collaborating with LLM in fighting its hallucinations. *In B. Galitsky (Ed.), Developing Enterprise Chatbots* (pp. 85–104). Elsevier. https://doi.org/10.1016/B978-0-443-29246-0.00004-3
24. Thapa, S., Shiwakoti, S., Shah, S.B. *et al.,2025.* Large language models (LLM) in computational social science: prospects, current state, and challenges. *Soc. Netw. Anal. Min.* 15, 4 (2025). https://doi.org/10.1007/s13278-025-01428-9
25. Dupéré, V., Leventhal, T., Dion, E., Crosnoe, R., Archambault, I., & Janosz, M. (2020). School-based extracurricular activity involvement and high school dropout among at-risk students: Consistency matters. *Applied Developmental Science*, 24(2), 129–146. https://doi.org/10.1080/10888691.2020.1796665
26. Tsolou, O. and Babalis, T. (2020) The Contribution of Family Factors to Dropping Out of School in Greece. *Creative Education*, 11, 1375-1401. doi: 10.4236/ce.2020.118101.
27. Lessky, F., & Unger, M. (2022). Working long hours while studying: A higher risk for First-in-Family students and students of particular fields of study? *European Journal of Higher Education, 13*(3), 347–366. https://doi.org/10.1080/21568235.2022.2047084
28. Iqbal, A., Iftikhar, M., & Hussain, T. (2023). Impact of social media use on the mental health of university students. *International Journal of Academic Research in Business and Social Sciences, 13*(9), 1234–1245. https://doi.org/10.6007/IJARBSS/v13-i9/12345
29. Watson, T. N., & Bogotch, I. (2016). (Re)Imagining school as community: Lessons learned from teachers. *School Community Journal, 26*(1), 93–114.
30. Khalid, R. Z., Ullah, A., Khan, A., Khan, A., & Inayat, M. H. (2023). Comparison of standalone and hybrid machine learning models for prediction of critical heat flux in vertical tubes. *Energies, 16*(7), 3182. https://doi.org/10.3390/en16073182
31. Wan, G., Lu, Y., Wu, Y., Hu, M., & Li, S. (2024). Large language models for causal discovery: Current landscape and future directions. *arXiv preprint arXiv:2402.11068*. https://doi.org/10.48550/arXiv.2402.11068
32. Mumuni, A., & Mumuni, F. (2024). Automated data processing and feature engineering for deep learning and big data applications: A survey. *Journal of Information and Intelligence, 3*(1), 1–15. https://doi.org/10.1016/j.jii.2024.01.002
33. Zhu, X., Li, Q., Cui, L., & Liu, Y. (2024). Large language model enhanced text-to-SQL generation: A survey. *arXiv preprint arXiv:2410.06011*
34. De Laat, P.B. (2018). Algorithmic Decision-Making Based on Machine Learning from Big Data: Can Transparency Restore Accountability?. *Philos. Technol.* 31, 525–541 (2018). https://doi.org/10.1007/s13347-017-0293-z
35. Nagy, M., Molontay, R. (2024). Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention. *Int J Artif Intell Educ* 34, 274–300 (2024). https://doi.org/10.1007/s40593-023-00331-8
36. Lee, S., & Chung, J. Y. (2019). The machine learning-based dropout early warning system for improving the performance of dropout prediction. *Applied Sciences*, 9(15), 3093. https://doi.org/10.3390/app9153093