THE INFLUENCE OF CHATGPT ON UNIVERSITY STUDENTS SELF-LEARNING IN HO CHI MINH CITY

Dat Nguyen[1], Nhat Le[1], Duy Lam[1], Uyen Ngo[1], Hieu Le[2]
[1]Faculty of Information Technology, HUTECH, Ho Chi Minh City, Vietnam
[2] Faculty of Economics, HCMUTE, Ho Chi Minh City, Vietnam
Corresponding Author: (Phone: +84 868 008 379; Email: ngvdat.w@gmail.com)

Abstract

The use of ChatGPT in learning has created significant changes in the way students approach self-study. Many others appreciate the supportive value that this tool brings. However, this tool also causes many mixed opinions. Some students are concerned that overusing ChatGPT can reduce their ability to think independently and lead to dependence on technology.

Purpose: This study explores factors influencing ChatGPT usage decisions and frequency, its impact on learning outcomes, and students' expectations, analyzing its role in self-study.

Method: Data were collected from a survey of 858 students at universities in Ho Chi Minh City and analyzed using machine learning models such as XG-Boost, Random Forest, SVM, Logistic Regression, ANN and TF-IDF.

Results and Findings: The models achieved high accuracy (up to 90%). Students' use of ChatGPT for learning is the key factor across all three questions, emphasizing its central role in supporting education. Factors like prioritization, daily usage frequency, and positive perceptions of ChatGPT are also crucial in optimizing its effectiveness in education.

Originality: This study highlights ChatGPT's dual role as a learning support tool and a potential barrier to independent thinking, providing insights to optimize its integration into education.

Limitation: The study focuses on a specific geographic area and demographic (students in Ho Chi Minh City), which may limit the generalizability of findings.

Implication: The research proposes strategies to optimize ChatGPT use, balancing technology with self-learning development.

Keywords: ChatGPT, University Students, Data Preprocessing, Supervised Learning Models, TF-IDF.

1. INTRODUCTION

The influence of ChatGPT on university students' self-learning has become an increasingly popular research topic, particularly in Ho Chi Minh City. Recent studies have explored how ChatGPT can enhance various aspects of student learning. For instance, a study by (Dempere et al., 2023) examined the role of ChatGPT in improving self-learning, highlighting its potential to revolutionize students' learning experiences. Similarly, (Assegaf et al., 2024) explored ChatGPT's impact on student learning behaviors, focusing on its ability to enhance educational outcomes while addressing ethical concerns associated with AI. Another study by (Minh, 2024) investigated the use of ChatGPT to improve English writing skills and critical thinking among Ho Chi Minh City University of Industry and Trade freshmen. This research emphasized balancing technology use with critical engagement, particularly in language education.

The integration of ChatGPT into students' self-study has had positive impacts, particularly on self-directed learning (SDL) and academic writing. Research shows that AI tools promote personalized learning and independence, especially in research (Wini Indriani 2024). ChatGPT improves SDL by 31.2%, with learning motivation influencing SDL by 44.3%. Together, they account for 75.7% of SDL variation. Additionally, (Mekthanavanh Vinath, 2024) found that ChatGPT enhances programming proficiency, though students must still develop independent problem-solving skills. While beneficial, ChatGPT raises concerns about dependency, particularly in programming. (Khadija Hanifi 2023) found that 93% of software engineering students rely on it, which may affect their ability to work independently. This underscores the need for further study on ChatGPT's impact on students' self-study habits and attitudes.

As education undergoes a paradigm shift, AI tools like ChatGPT offer a dynamic, interactive approach to learning, with growing demand, especially in remote and online education. (Hasan et al., 2024) highlighted how these tools tackle educational challenges by providing real-time, personalized interactions that mimic a tutor's guidance. The importance of AI in education is increasingly vital as online learning continues to grow. ChatGPT has proven valuable in programming, offering code suggestions, detecting errors, and explaining complex concepts. (Paula De Barba 2024) noted its role in improving code review by providing diverse perspectives, benefiting both novice and experienced programmers. However, dependency remains a challenge, particularly in programming education, where students may rely too heavily on ChatGPT for solutions.

In recent years, there has been a rise in the use of educational data mining and machine learning algorithms to predict student behaviors and outcomes. For instance, (Hasan et al., 2024) analyzed the impact of AI, specifically ChatGPT, using machine learning algorithms, with the

SVM model achieving the highest accuracy of 98% in analyzing ChatGPT data. Additionally, (Liu, 2024) discussed ChatGPT's role in online collaborative learning, noting its potential to enhance knowledge activation, critical thinking, and overall performance in collaborative learning settings. These insights contribute to a deeper understanding of how ChatGPT can be optimized to improve learning outcomes across different academic disciplines.

Overall, the growing body of research on ChatGPT and its role in self-learning reflects its significant potential to transform the educational landscape. However, it is essential to balance its use with independent learning skills to ensure long-term educational benefits.

2. METHOD

2.1 OBJECTIVE

In this research, we delve into how ChatGPT influences student learning by exploring four primary questions: What factors influence students' decision to use ChatGPT? What factors influence how often students use ChatGPT? How does ChatGPT affect student learning outcomes? What are students' perceptions of ChatGPT? These findings will provide valuable insights into ChatGPT's role in education, helping students better evaluate its usefulness in their studies. The results will also guide educators and institutions in crafting effective guidelines to optimize ChatGPT's implementation.

2.2 DATA COLLECTION

A total of 858 students from universities in Ho Chi Minh City took part in the study between June and August 2024. Data were collected using a 33-question Google Form survey, comprising linear scales, checkboxes, and paragraph prompts, including 5 questions on personal and academic information (Table 1), 5 questions on ChatGPT usage for learning and research (Table 2), 21 questions on personal evaluations and viewpoints (Table 3), and 2 open-ended questions about personal impressions of ChatGPT (Table 4).

Table 1: Personal and educational information

Features	Description
Gender	The gender of each student.
YearOfLearning	The number of years each student has been studying.
UniversityName	The name of the school the student is currently attending.
Major	The major the student is currently studying at school.
GradeRank	Academic classification of each student.

Table 2: Using ChatGPT in Study and Research

Features	Description
UseGPTorNot?	Whether the student uses ChatGPT in their studies?

GradeTrend(GPT)?	Whether the students' average grades tend to increase or decrease after using ChatGPT?
UsageInAcademics	How do you use ChatGPT for your learning and research?
UsageBeyondAcademics	How do you use ChatGPT for tasks other than learning and research?
AvgTimeUsage(1day)	On average, how much time do you spend using ChatGPT in a day?

Table 3: Personal evaluation and the impact of ChatGPT on learning and research

Features	Description
Priority	How do you prioritize ChatGPT compared to other educational research websites?
Reliability	The level of trust you place in ChatGPT for providing information for learning.
ResultofGPT	The effectiveness of ChatGPT in solving questions and academic problems.
InfluenceofGPT	The impact of ChatGPT on self-learning and research capabilities.
SatisfactionLevel	Your level of satisfaction with ChatGPT's assistance in answering academic questions.
Relevance	Does the content provided by ChatGPT align with your search queries?
Inspire&Motivate	Does ChatGPT have the ability to inspire and motivate you in learning/research?
PositiveInfluence	ChatGPT has a positive impact on your creative thinking ability in problem-solving.
NegativeInfluence	ChatGPT has a negative impact on your creative thinking ability in problem-solving.
Effective	ChatGPT can help you learn/research more effectively.
ReplaceTeacher	ChatGPT has the potential to completely replace the role of teachers in learning process.
SpeechDoExam	How many times faster do students using ChatGPT complete assignments compared to those who do not use ChatGPT?
Dependence&Constraint	ChatGPT creates dependency and constraints in learning and personal development.
LazyStudy&Think	Does using ChatGPT make you less inclined to study and think critically?
ImproveStudy	Does using ChatGPT improve your learning and research?
SatisfiedResult	Are you satisfied with the responses provided by ChatGPT to your questions?
MechanismOfAction	Do you understand the nature and working mechanism of ChatGPT?
EasyToUnderstand	Based on your experience using ChatGPT, are its answers generally easy to understand?
CompareGoogle	Do you prefer the answers provided by Google or ChatGPT?
FindJob	Will you use ChatGPT for your work in the future?
WaytoAsk	Are you interested in how to ask questions for chatGPT?

Table 4: Personal feelings

Features	Description
OpinionOnImprovement	Students' views on ChatGPT's increasing intelligence.
ExperienceWithChatGPT	Students' perceptions of ChatGPT are based on their usage and interactions.

2.3 DATA PREPROCESSING

For data collection and cleansing, we downloaded survey responses in XLSX format, removed invalid entries, and handled duplicates to ensure the accuracy of the data. The cleaned data was then converted into numerical vectors to optimize the model's learning process. To address multicollinearity, we calculated the Variance Inflation Factor (VIF) (O'brien, 2007) for each feature and removed variables with a VIF greater than 5 to maintain model stability. We also analyzed the label distribution to check for class balance and applied the SMOTE technique

(Chawla et al., 2002) to address any imbalances, ensuring that the model could learn effectively from all classes. The data was split into training and testing sets, with 70% allocated for training and 30% for testing. We applied the MinMaxScaler (Raju et al., 2020) to rescale the feature values, improving model accuracy. Lastly, for text data cleaning and keyword extraction, we removed unnecessary noise, such as special characters, punctuation, and stopwords, to ensure cleaner data for analysis.

2.4 MODEL IMPLEMENT

This study explored five machine learning models XG-Boost (Ogunleye & Wang, 2020), Random Forest (Biau & Scornet, 2016), SVM (Salcedo-Sanz et al., 2014), Logistic Regression (Sperandei, 2014), and ANN (Agatonovic-Kustrin & Beresford, 2000). We relied on the confusion matrix, accuracy, precision, recall, and F1-score (Shekar & Dagnew, 2019) to gauge each model's classification accuracy and overall performance. To boost model reliability and effectiveness, we used GridSearch (Shekar & Dagnew, 2019) to identify optimal hyperparameters. This approach enhanced classification and prediction quality while minimizing risks of overfitting or underfitting. The machine learning models employed in this study were implemented using Python programming language and executed on Google Colab to perform data analysis.

3. RESULTS & DISCUSSION

3.1 RESULTS

Based on the analysis results from Table 5, Table 6 and Table 7, the XG-Boost model demonstrates outstanding performance in addressing all three target questions related to the use of ChatGPT among students in Ho Chi Minh City.

For the question about the decision to use ChatGPT, where label 0 indicates non-use and label 1 indicates use, the model achieved 88% accuracy, with Precision, Recall, and F1-score all at 88% (Table 5). The confusion matrix (Figure 1) confirms this stability, showing 166 True Positives and 160 True Negatives, with minimal misclassifications (20 FP and 24 FN). For the frequency of ChatGPT usage, where labels 0, 1, 2, and 3 represent <1 hour (including no usage), 1-2 hours, 2-3 hours, and >3 hours daily, XG-Boost performed best with 90% accuracy and consistent Precision, Recall, and F1-score (Table 6). While the model excelled in classifying higher usage (classes 2 and 3 with 174 and 168 True Positives), classes 0 and 1 showed more errors, with 29 FN and 12 FN, suggesting areas for improvement (Figure 2). For the impact on academic performance, with labels 0, 1, and 2 indicating decreased, unchanged, and increased performance, XG-Boost led with 87% accuracy, outperforming SVM (85%) and ANN (83%) (Table 7). The confusion matrix (Figure 3) highlights strong performance in

classes 0 and 2, with 168 and 154 True Positives, though class 1 faced challenges, with 19 FN and 13 FP.

From these analyses, it can be concluded that the XG-Boost model is the optimal choice for analyzing and predicting factors related to the use of ChatGPT among students in Ho Chi Minh City, thanks to its superior performance and stability compared to other models.

Table 5: The machine learning model results for factors influencing students' decision to use ChatGPT.

Models	Accuracy	Precision	Recall	F1-score
SVM	0.87	0.87	0.87	0.87
Logistic Regression	0.86	0.86	0.86	0.86
Random Forest	0.86	0.86	0.86	0.86
XG-Boost	0.88	0.88	0.88	0.88
ANN	0.87	0.87	0.87	0.87

Table 6: The machine learning model results for factors influencing how often students use ChatGPT.

Models	Accuracy	Precision	Recall	F1-score
SVM	0.88	0.89	0.89	0.88
Logistic Regression	0.58	0.59	0.58	0.58
Random Forest	0.84	0.85	0.84	0.83
XG-Boost	0.90	0.90	0.90	0.90
ANN	0.87	0.87	0.87	0.87

Table 7: The machine learning model results for how ChatGPT affects student learning outcomes.

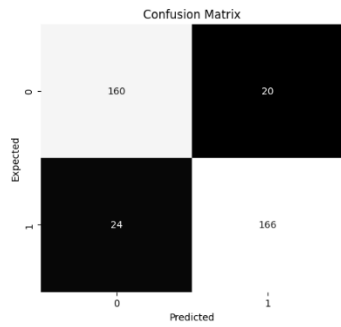| Models	Accuracy	Precision	Recall	F1-score
SVM	0.85	0.85	0.85	0.85
Logistic Regression	0.61	0.61	0.62	0.61
Random Forest	0.81	0.81	0.81	0.80
XG-Boost	0.87	0.87	0.87	0.87
ANN	0.83	0.83	0.83	0.83

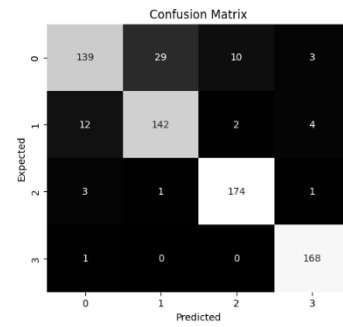Figure 1: Confusion matrix for factors influencing students' decision to use ChatGPT.

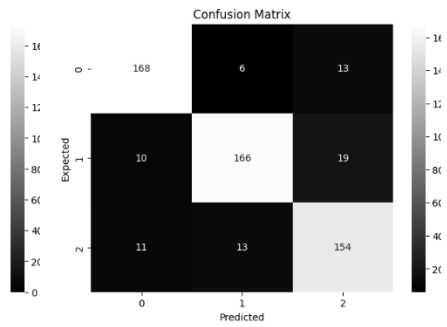Figure 2: Confusion matrix for factors influencing how often students use ChatGPT.

Figure 3: Confusion matrix for how ChatGPT affects student learning outcomes.

Based on top 10 features importance charts of the XG-Boost model, the key factors influencing the decision to use, the frequency of use, and the impact of ChatGPT on academic performance are clarified.

For the decision to use ChatGPT, UsageInAcademics (using ChatGPT for academic purposes) emerges as the most critical factor, indicating that students primarily choose ChatGPT to support their studies. Additionally, UsageBeyondAcademics (usage beyond academics), Priority (priority level), and WaytoAsk (ways of asking questions) also play significant roles, highlighting how students interact with and use ChatGPT both inside and outside academic contexts (Figure 4).

When analyzing the frequency of ChatGPT usage, Priority leads as the most influential factor, suggesting that students who prioritize ChatGPT tend to use the tool more frequently. UsageInAcademics and Dependence&Constraint (dependence and constraints) also have substantial impacts, indicating that the way ChatGPT is utilized in studies and the sense of reliance on the tool are key factors. Personal factors, such as SpeechDoExam (exam speed) and Gender, also contribute to shaping the frequency of use (Figure 5).

Regarding ChatGPT's impact on academic performance, UseGPT (whether ChatGPT is used or not) stands out as the most prominent factor, affirming the critical role of accessing ChatGPT in improving academic results. Furthermore, UsageInAcademics, AvgTimeUsage(1day) (average daily usage time), and factors such as Effective (effectiveness) and SatisfactionLevel (satisfaction level) indicate that positive usage patterns and perceptions of ChatGPT significantly enhance students' academic abilities (Figure 6).

In summary, UsageInAcademics emerges as the core factor across all three questions, emphasizing the central role of ChatGPT in academic support. Factors like Priority, AvgTimeUsage(1day), and positive perceptions of ChatGPT also contribute significantly to optimizing the tool's effectiveness in education.
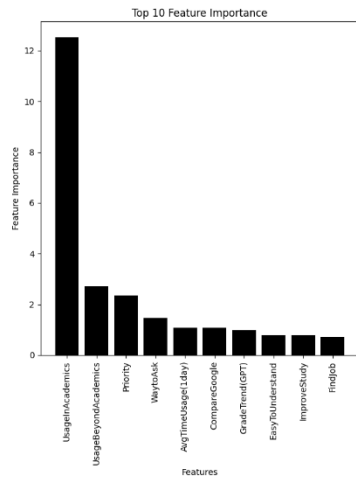
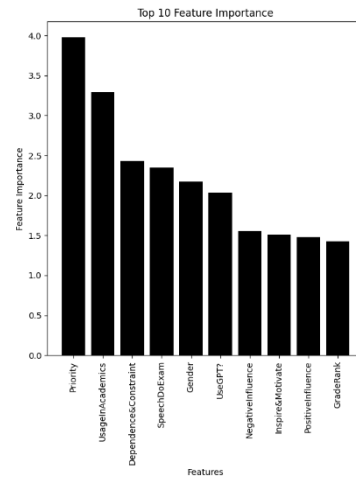*Figure 4: Features importance for factors influencing students' decision to use ChatGPT.*

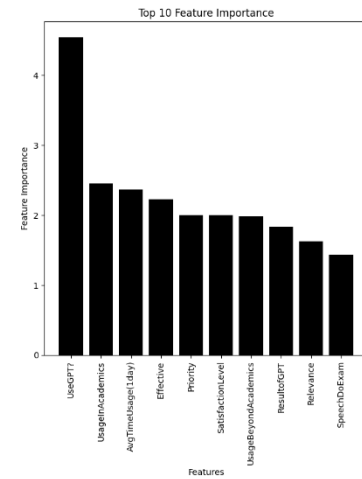*Figure 5: Features Importance for factors influencing how often students use ChatGPT.*

Figure 6: Features importance for how ChatGPT affects student learning outcomes.

To answer the question: "*What are students' perceptions of ChatGPT?*", TF-IDF (Ganiger & Rajashekharaiah, 2018) identified frequently used and distinctive terms, providing deeper insights into students' perspectives on ChatGPT while unveiling opportunities for improving its services.

Top 20 TF-IDF terms for the question "Students' perceptions of ChatGPT's increasing intelligence" indicate that the most frequent and distinctive terms are mostly positive or neutral, such as "good," "correct," "agree," "okay," "opinion," "normal," and "think." These terms suggest that students highly appreciate the intelligence and development of ChatGPT. Terms like "good" and "correct" highlight students' satisfaction with the quality and accuracy of this tool. Additionally, "opinion" and "agree" reflect the consensus among students regarding the usefulness and continuous development of ChatGPT. Words like "think" and "normal" suggest that students are comparing ChatGPT's intelligence to their own thinking abilities, showing their interest in how the tool could replace or assist humans.

On the other hand, top 20 TF-IDF terms for the question "Students' perceptions of ChatGPT based on their usage and interactions" show that terms such as "good," "useful," "okay," "convenient," "used," and "chatgpt" appear frequently, reflecting students' positive views on the tool's usefulness and convenience. Words like "normal" and "information" indicate that students expect ChatGPT to provide accurate and standard information. Moreover, terms like "intelligent" and "fast" suggest that students are satisfied with ChatGPT's intelligence and response speed. The word "tool" emphasizes ChatGPT's role as an important learning support tool.

Overall, the results from both questions indicate that students view ChatGPT positively, appreciating its usefulness, accuracy, and convenience.
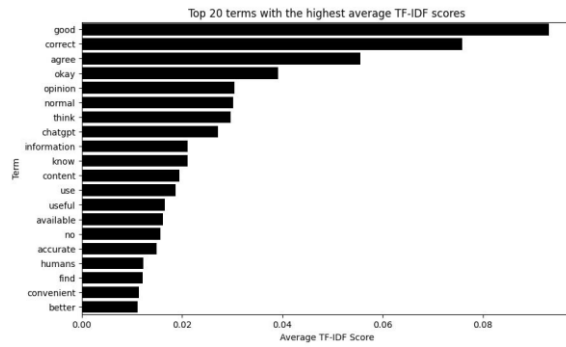
Figure 7: TF-IDF results for students' views on ChatGPT's increasing intelligence.
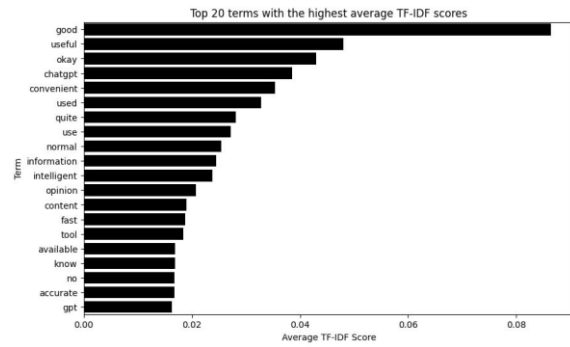
Figure 8: TF-IDF results for students' perceptions of ChatGPT based on their usage and interactions.

3.2 DISCUSSION

The study highlights the dual influence of ChatGPT on university students' self-learning in Ho Chi Minh City. ChatGPT enhances academic performance, creativity, and task efficiency, as demonstrated by the high accuracy of the XG-Boost model in analyzing usage factors like prioritization, frequency, and reliability. TF-IDF analysis revealed positive perceptions, with students appreciating ChatGPT's intelligence, speed, and user-friendliness.

However, concerns about over-dependence on ChatGPT persist, with reports indicating reduced motivation for traditional learning and critical thinking. The study emphasizes the need for a balanced approach, as reliance on ChatGPT without fostering independent problem-solving skills can hinder long-term learning development. The findings suggest that while ChatGPT benefits students academically, challenges remain in optimizing its use and mitigating dependency risks.

These insights underline the importance of structured guidelines to promote effective ChatGPT use. Integrating ChatGPT into formal learning environments, raising awareness of its limitations, and combining its use with traditional methods can maximize its benefits while addressing potential drawbacks.

4. CONCLUSION

This study confirms that ChatGPT is a powerful tool for supporting self-directed learning among university students, particularly in enhancing academic performance and creativity. Its role in providing accessible and efficient learning support is undeniable. However, reliance on ChatGPT without fostering independent thinking can pose significant challenges to students' long-term educational development.

To harness the potential of ChatGPT effectively, educational institutions should emphasize its role as a supplementary tool rather than a replacement for traditional learning methods. This involves promoting critical thinking, guiding students on effective usage, and addressing

dependency concerns through structured curricula. By leveraging ChatGPT responsibly, students can enjoy its advantages while cultivating essential lifelong learning skills.

REFERENCES

Agatonovic-Kustrin, S., & Beresford, R. (2000). Basic concepts of artificial neural network (ANN) modeling and its application in pharmaceutical research. *Journal of Pharmaceutical and Biomedical Analysis*, *22*(5), 717–727. https://doi.org/10.1016/S0731-7085(99)00272-1

Assegaf, S., Qonitatillah, M., Prameswari, N. N., Kurniawan, B., & Ratnawati, N. (2024). THE IMPACT OF USE OF CHATGPT ON CHANGES IN STUDENTS' LEARNING BEHAVIOR. *International Journal of Geography, Social, and Multicultural Education*, *2*(1), 49–58. https://doi.org/10.26740/ijgsme.v2n1.p49-58

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, *25*(2), 197–227. https://doi.org/10.1007/s11749-016-0481-7

Challenges and opportunities in using ChatGPT as a team member to promote code review education and self-regulated learning. (2024). *ASCILITE Conference Proceedings*, 108–117. https://doi.org/10.14742/apubs.2024.1152

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Dempere, J., Modugu, K., Hesham, A., & Ramasamy, L. K. (2023). The impact of ChatGPT on higher education. *Frontiers in Education*, *8*. https://doi.org/10.3389/feduc.2023.1206936

From Conversation To Competence: Analysis Of The Influence Of Using ChatGPT And Learning Motivation In Increasing Self-Directed Learning. (2024). *Academic Journal of Psychology and Counseling*, *5*(2). https://doi.org/10.22515/ajpc.v5i2.8971

Ganiger, S., & Rajashekharaiah, K. M. M. (2018). Comparative Study on Keyword Extraction Algorithms for Single Extractive Document. *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 1284–1287. https://doi.org/10.1109/ICCONS.2018.8663040

Hasan, N., Polin, J. A., Ahmmed, M. R., Sakib, M. M., Jahin, M. F., & Rahman, M. M. (2024). A novel approach to analyzing the impact of AI, ChatGPT, and chatbot on education using machine learning algorithms. *Bulletin of Electrical Engineering and Informatics*, *13*(4), Article 4. https://doi.org/10.11591/eei.v13i4.7158

Liu, H. (2024). Applicability of ChatGPT in Online Collaborative Learning: Evidence Based on Learning Outcomes. *Proceedings of the International Academic Conference on Education*, *1*(1), 33–43. https://doi.org/10.33422/iaceducation.v1i1.656

Minh, A. N. (2024). Leveraging ChatGPT for Enhancing English Writing Skills and Critical Thinking in University Freshmen. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (Online)*, *3*(2), Article 2. https://doi.org/10.60087/jklst.vol3.n2.p62

O'brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Ogunleye, A., & Wang, Q.-G. (2020). XGBoost Model for Chronic Kidney Disease Diagnosis. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *17*(6), 2131–2140. https://doi.org/10.1109/TCBB.2019.2911071

On ChatGPT: Perspectives from Software Engineering Students. (2023, October 22). SciSpace - Paper. https://doi.org/10.1109/qrs60937.2023.00028

Raju, V. N. G., Lakshmi, K. P., Jain, V. M., Kalidindi, A., & Padma, V. (2020). Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 729–735. https://doi.org/10.1109/ICSSIT48917.2020.9214160

Salcedo-Sanz, S., Rojo-Álvarez, J. L., Martínez-Ramón, M., & Camps-Valls, G. (2014). Support vector machines in engineering: An overview. *WIREs Data Mining and Knowledge Discovery*, *4*(3), 234–267. https://doi.org/10.1002/widm.1125

Shekar, B. H., & Dagnew, G. (2019). Grid Search-Based Hyperparameter Tuning and Classification of Microarray Cancer Data. *2019 Second International Conference on Advanced Computational and Communication Paradigms (ICACCP)*, 1–8. https://doi.org/10.1109/ICACCP.2019.8882943

Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia Medica*, 12–18. https://doi.org/10.11613/BM.2014.003

Vinath, M. (2024). ChatGPT as a Learning Assistant: Its Impact on Students Learning and Experiences. *International Journal of Education in Mathematics, Science and Technology*, 1603–1619. https://doi.org/10.46328/ijemst.4471