VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5967

ANALYSIS STUDENT EMOTIONS AND MENTAL HEALTH ON CUMULATIVE GPA USING MACHINE LEARNING AND SMOTE

Fadhil Muhammad Basysyar[1*]; Gifthera Dwilestari[2]; Ade Irma Purnamasari[3]

Information System[1,2,3]
STMIK IKMI Cirebon, Cirebon, Indonesia[1,2,3]
ikmi.ac.id[1,2,3]
fadhil.m.basysyar@gmail.com[1*], ggdwilestari@gmail.com[2], irma2974@yahoo.com[3]

(*) Corresponding Author
(Responsible for the Quality of Paper Content)

Abstract— *This research investigates the impact of emotions and mental health on students' cumulative grade point average (CGPA) using machine learning classification algorithms while addressing data imbalances with the Synthetic Minority Oversampling Technique (SMOTE). Emotional well-being and mental health are acknowledged as vital determinants of academic achievement. Data imbalance, particularly in mental health metrics such as anxiety and depression, frequently compromises forecast accuracy. This study improves the accuracy of CGPA prediction based on emotional and mental health factors by utilizing SMOTE in machine learning models such as logistic regression and random forest. A dataset including 226 university students, including academic records and self-reported mental health evaluations, was evaluated. The random forest model attained an accuracy of 87.63%, exceeding the logistic regression model's accuracy of 86.56%. These findings emphasize the significant role of emotions and mental health in academic outcomes and validate SMOTE's efficacy in addressing class imbalance. This work offers a fresh technique in educational data mining by revealing the possibility for improved academic achievement forecasts based on psychological characteristics, helping to the development of targeted therapies for students experiencing emotional issues. Implications for educational policy emphasize the significance of mental health support systems in promoting academic achievement. Subsequent research should investigate supplementary psychological variables and comprehensible models to improve predictive accuracy and facilitate evidence-based policymaking.*

Keywords: *cumulative grade point average, emotions, machine learning, mental health, synthetic minority oversampling technique.*

Intisari— *Studi ini mempelajari pengaruh emosi dan kesehatan mental mahasiswa terhadap Indeks Prestasi Kumulatif (IPK) mereka menggunakan algoritma klasifikasi pembelajaran mesin, dengan fokus khusus pada penyelesaian ketidakseimbangan data melalui Teknik Synthetic Minority Oversampling (SMOTE). Kesejahteraan emosional dan kesehatan mental semakin diakui sebagai prediktor utama kinerja akademis. Namun, kesulitan kumpulan data yang tidak seimbang, khususnya dalam mengumpulkan penanda kesehatan mental seperti kesedihan dan kecemasan, seringkali menghambat akurasi prediksi. Dengan mengintegrasikan SMOTE dengan beberapa model pembelajaran mesin, termasuk regresi logistik dan random forest, penelitian ini bertujuan untuk meningkatkan akurasi prediksi IPK berdasarkan data kesehatan emosional dan mental. Kumpulan data dari 226 mahasiswa, yang menggabungkan catatan akademis dan penilaian kesehatan mental yang dilaporkan sendiri, dievaluasi. Hasil penelitian mengungkapkan bahwa model random forest memperoleh akurasi 87,63%, mengalahkan regresi logistik pada 86,56%. Temuan ini menyiratkan bahwa emosi dan kesehatan mental memiliki dampak yang signifikan terhadap keberhasilan akademis, dan bahwa SMOTE merupakan teknik yang berguna untuk meminimalkan ketidakseimbangan kelas. Studi ini berkontribusi pada semakin banyaknya penelitian tentang penggalian data pendidikan dengan memberikan teknik baru untuk memprediksi keberhasilan akademis berdasarkan karakteristik psikologis. Konsekuensi dari temuan ini menggarisbawahi perlunya intervensi yang disesuaikan untuk mendukung siswa yang mengalami tekanan emosional, sehingga meningkatkan prestasi akademis mereka. Studi mendatang akan* *mengeksplorasi penyertaan indikator kesehatan mental tambahan dan model yang lebih dapat ditafsirkan untuk lebih meningkatkan prediksi dan membantu dalam pembuatan kebijakan pendidikan.*

Kata Kunci*: rata-rata nilai kumulatif, emosi, pembelajaran mesin, kesehatan mental, teknik oversampling minoritas sintetis.*

INTRODUCTION

The academic success of students, as assessed by their cumulative grade point averages (CGPA), is influenced by a range of factors, including emotional well-being and mental health[1]. In recent years, research has increasingly focused on how psychological states affect learning results as the pressures of modern education systems rise. Understanding the relationship between emotions, mental health, and academic achievement is vital for establishing focused treatments that can increase student success[2], [3].

With the rise of machine learning techniques, predictive models can now examine enormous datasets to identify and predict student performance based on emotional and mental health characteristics[4]. The Synthetic Minority Oversampling Technique (SMOTE) has demonstrated to be effective in dealing with imbalanced datasets, which is a prevalent difficulty in this sector, making it a good technique for this investigation [5], [6], [7], [8], [9].

The fundamental issue in studying the impact of emotions and mental health on students' academic performance is the imbalance in data distribution, particularly when discriminating between students with varying levels of emotional well-being and mental health[10], [11], [12]. This mismatch makes it challenging for typical machine learning models to anticipate outcomes effectively, often resulting in biased classifications. Recent studies have offered several solutions, such as applying resampling approaches and ensemble learning methods to address the class imbalance issue. However, the implementation of SMOTE in tandem with machine learning algorithms presents a viable approach, as it allows for better representation of minority classes without overfitting, thereby enhancing predictive accuracy.

Several studies have investigated specialized machine learning algorithms to solve the issues caused by imbalanced datasets in educational contexts. For example, introduced the use of SMOTE in educational data mining, which showed an increase in classification performance[13]. More recently, improved this work by adding adaptive boosting with SMOTE to enhance the accuracy of predicting student dropouts[14]. Additionally, employed deep learning techniques in conjunction with SMOTE to predict the emotional well-being of students based on their academic records and social media activity, producing noteworthy results in detecting at-risk students[9]. Despite these developments, there remains a void in the literature concerning the integration of these methodologies specifically for forecasting CGPA based on emotional and mental health aspects, underscoring the need for more study in this area[14], [15].

However, a fundamental problem in this discipline is the quality and balance of data. Many datasets, particularly those relating to mental health, tend to be imbalanced, with fewer cases of acute emotional suffering. This mismatch can bias predictions and diminish the performance of typical machine learning models. Techniques such as SMOTE have emerged as helpful tools to solve this problem, allowing researchers to supplement the minority class and produce more trustworthy predictions. In a recent study, showed that utilizing SMOTE with support vector machines (SVM) enhanced the prediction accuracy of student performance in connection to mental health parameters[16].

Additionally, the integration of mental health markers, such as anxiety and depression scores, into prediction models has become a key area of current research. That these characteristics, when added in machine learning models, considerably boost the capacity to predict academic outcomes. The combination of these markers with demographic and academic data has led to more holistic and accurate predictions; however, obstacles in data gathering and privacy concerns persist.

Despite these developments, there is still a notable void in the research concerning the integration of machine learning models, especially those geared to predict CGPA through mental health and emotional aspects, employing advanced resampling techniques like SMOTE[17]. Addressing this gap needs not only strengthening the models but also enhancing data gathering and integrating more complex emotional health variables[18].

In conclusion, this study emphasizes the significance of mental health and emotional well-being as critical determinants of academic success, specifically in predicting students' cumulative GPA. By applying machine learning algorithms improved with the Synthetic Minority Oversampling Technique (SMOTE), the research addresses a

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5967

prevalent difficulty in educational data analysis—imbalanced data—especially with regard to underrepresented mental health indicators like anxiety and depression. The findings reveal that SMOTE, when combined with algorithms such as random forest, greatly enhances the predicted accuracy of CGPA models, ensuring a more balanced appraisal of students' academic performance depending on their psychological status.

This study's results underscore the value of integrating psychological data into academic predictive models, presenting a fresh perspective for educational institutions intending to assist student success through mental health efforts. Future studies can expand on these findings by integrating more mental health indicators and developing interpretable machine learning models, further increasing the precision of academic performance forecasts. This technique not only contributes to the field of educational data mining but also underlines the relevance of mental health interventions in increasing students' academic achievements.

MATERIALS AND METHODS

1. Data Collection

The dataset utilized in this study was acquired from a group of university students enrolled in several courses, including informatics engineering, information systems, and computer accounting. The sample consisted of 226 students, with relevant factors including gender, age, year of study, cumulative grade point average (CGPA), and self-reported mental health disorders such as depression, anxiety, and panic attacks. Additionally, information regarding expert treatment was also gathered at Table 1.

Table 1. Data Collection

No	Gender	Age	Course	Year	Cumulative GPA	Depression	Anxiety	Panic Attack
1	Female	18	Informatics Engineering	Year 1	3.00 - 3.49	Yes	No	Yes
2	Male	21	Information System	Year 2	3.00 - 3.49	No	Yes	No
3	Male	19	Informatics Engineering	Year 1	3.00 - 3.49	Yes	Yes	Yes
4	Female	22	Informatics Engineering	Year 1	3.00 - 3.49	Yes	No	No
5	Male	23	Informatics Engineering	Year 2	3.00 - 3.49	No	No	No
6	Male	19	Informatics Engineering	Year 1	3.50 - 4.00	No	No	Yes
...	...	...	...	...	...	...	...	...
225	Female	22	Informatics Engineering	Year 1	3.00 - 3.49	Yes	Yes	No
226	Male	24	Informatics Engineering	Year 1	3.50 - 4.00	No	No	Yes

Source: (Research Result, 2024)

The data was entered in the following fields: gender, age, course, year of study, cumulative grade point, depression, anxiety, panic attack, and specialized therapy. The majority of participants were between the ages of 18 and 24, coinciding with the normal age range for undergraduate students. Data columns (total 9 columns) at table 2.

Table 2. Data Columns

No	Column	Dtype
1	Gender	Object
2	Age	Int64
3	Course	Object
4	Year of Study	Object
5	Cumulative Grade Point	Object
6	Depression	Object
7	Anxienty	Object
8	Panic Attack	Object
9	Specialist Treatment	Object

Source: (Research Result, 2024)

2. Data Cleaning

To verify the correctness and integrity of the dataset, multiple preprocessing processes were conducted[19]:

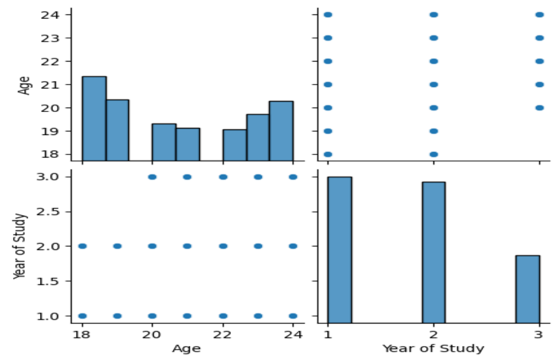
a. Handling Missing Data: All missing values were inspected and managed. Rows containing null values were either imputed or eliminated, depending on the severity and relevance of the missing data.
b. Normalization: The data columns, including those for cumulative grade points (e.g., 3.50 - 4.00), were normalized to remove extraneous gaps and ensure consistent formatting.
c. Duplicate Removal: Any duplicate entries were detected and eliminated to prevent skewed analysis.
d. Column Renaming: To ensure clarity and uniformity, column names were altered to match conventional naming practices. For example, columns such as "Year of Study" were renamed to represent numerical values rather than category descriptors like "Year 1" or "Year 2.".

3. Exploratory Data Analysis

Following data cleaning, exploratory data analysis was undertaken to analyze potential patterns in the dataset where there were no outlier data (figure 1).

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480 /jitk.v10i2.5967



Source: (Research Result, 2024)

Figure 1. Exploratory Data Analysis

a. Demographic Overview

A pairplot was used to depict the correlations between several factors, such as gender, age, and year of study, and their probable correlation with mental health indicators (depression, anxiety, and panic attacks).

b. Distribution of Mental Health Conditions

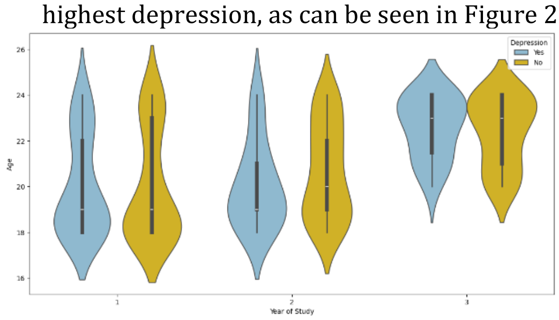
The incidence of depression, anxiety, and panic attacks was studied across different courses and years of study. For example, a comparative analysis was undertaken between students in informatics engineering and those in information systems to uncover potential variances in mental health results.

4. Statistical Analysis

An analysis of the correlation between CGPA and mental health disorders (depression, anxiety, panic attacks) by year was done using a series of statistical tests.

a. Depression, Year of Study, and Age

A chi-squared test was done to examine whether there was a significant link between depression, year of study, and age; 3rd year students did not experience depression or anxiety. 2nd year students had a diverse spectrum of classmates. 1st year students aged between 18 and 20 years old suffered the highest depression, as can be seen in Figure 2.

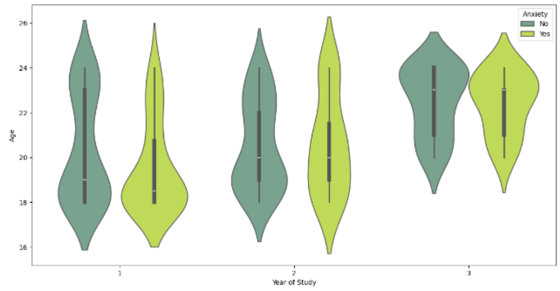


Source: (Research Result, 2024)

Figure 2. Depression, Year of Study, and Age

b. Anxiety, Year of Study, and Age

Cross-tabulation and chi-squared tests were undertaken to uncover potential correlations between the chosen course of study and the incidence of anxiety year of study, and age was found. Except for those aged 24 years, 3rd year students did not experience anxiety. 2nd year students had a diverse spectrum of classmates. 1st year students aged between 18 and 20 years reported the highest anxiety; 2nd year students were more prone to anxiety, as noted in figure 3.

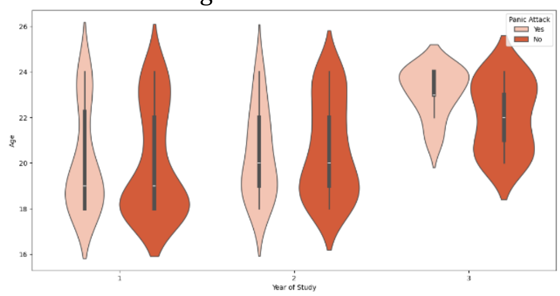


Source: (Research Result, 2024)

Figure 3. Anxiety, Year of Study, and Age

c. Panic Attacks, Year of Study, and Age

Statistical techniques were applied to evaluate if there was a significant difference in the prevalence of panic attacks across different panic attacks, year of study, and age. It was observed that 3rd year students did not have panic attacks except those aged 24 years, as shown in figure 4.



Source: (Research Result, 2024)

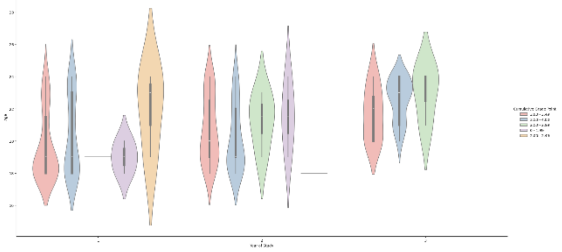
Figure 4. Panic Attacks, Year of Study, and Age

d. Cumulative Grade Point, Year of Study, and Age

Third-year students that are doing well academically and consequently have little or no mental health concerns. Many second-year students have a GPA below 2.0. Students in their first and second years perform better academically with a GPA above 2.5. First-year students aged 18-20 years old while having a

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5967

decent GPA who suffer mental problems can be noted in figure 5.



Source: (Research Result, 2024)
Figure 5. Cumulative Grade Point, Year of Study, and Age

In this study, hyperparameter tuning was a critical component in enhancing model performance, particularly given the imbalanced nature of the dataset, where mental health conditions such as depression, anxiety, and panic attacks appeared at variable frequencies across demographic subgroups.

Hyperparameter tuning was undertaken on multiple machine learning models, including logistic regression, random forest, and support vector machines, utilizing grid search and cross-validation approaches to increase predicted accuracy and stability. For instance, in the random forest model, parameters such as the number of trees, maximum depth, and minimum samples per leaf were rigorously changed to identify configurations that minimized prediction error and balanced accuracy across the minority and majority classes. For models like logistic regression and support vector machines, parameters like regularization strength (C) and kernel type were modified to handle potential overfitting and promote generalization.

Hyperparameter tuning was guided by k-fold cross-validation to limit the risk of model variation and provide robust performance across diverse subsets of data. In instances where tuning demonstrated small performance increases or computational efficiency was favored, default settings were preserved, as further tuning risked diminishing returns without considerable influence on model efficacy.In addressing the class imbalance inherent in the dataset, this study incorporates accuracy, recall, and F1-score as crucial evaluation measures to provide a rigorous assessment of model performance.

Precision is particularly important in assessing the fraction of accurately detected instances of mental health problems (depression, anxiety, and panic attacks) among all predictions of that condition. By focusing on precision, we strive to limit the possibility of false positives, which is critical for predicting mental health disorders since misclassification can lead to incorrect interventions or neglecting kids in need.

Conversely, Recall examines the model's capacity to properly detect all actual occurrences of mental health concerns, effectively rating the model's sensitivity. This statistic is especially significant for identifying pupils with specific mental health issues, as strong recall values lower the possibility of false negatives, ensuring that at-risk individuals are not neglected.

Finally, the F1-score harmonizes precision and recall, offering a single metric that balances both. Given the class imbalance in our data, the F1-score gives a more comprehensive evaluation than accuracy alone, particularly in circumstances where the costs of false positives and false negatives must be balanced to inform appropriate interventions.
Precision, recall, and F1-score collectively offer a nuanced assessment of the model's performance, particularly in discovering trends associated with students' emotional and mental health indicators, such as cumulative GPA and self-reported symptoms. This multi-metric method conforms with best practices in machine learning for healthcare and mental health investigations, as indicated in recent work on imbalanced datasets.

RESULTS AND DISCUSSION

The comparative analysis of classification models revealed notable differences in their performance metrics, each offering unique advantages and limitations. The logistic regression model, with an accuracy of 86.56%, provides an interpretable framework for studying the influence of mental health on academic success. Its simplicity and clarity make it excellent for educational environments, as logistic regression facilitates quicker detection of correlations between predictor factors and student outcomes. As noted in Table 4, this model obtained a precision of 0.88 for class 0 and 0.83 for class 1, alongside a recall rate of 0.83 and 0.89, respectively. Such balanced recall and precision indicate a reliable capacity for correct classification across both classes.

In contrast, the random forest model achieved a somewhat higher accuracy rate of 87.63%; however, its complexity limits interpretability. Although this model is well-suited to managing complicated, non-linear data structures, its decision processes are tough to trace, which can limit educational stakeholders' capacity to act on model outputs. However, random forest's ensemble mechanism makes it robust against overfitting, providing a broader capacity for capturing diverse student emotional and mental health attributes, which align with findings in educational data analytics literature, where complex algorithms like random forest often yield high accuracy on nuanced datasets. Finally, ensemble techniques, such as voting classifiers and stacking classifiers, further raised the accuracy to 85.2% and 85.6%, respectively, exhibiting enhanced performance through model combination. These ensemble techniques show the promise of hybrid approaches for managing imbalanced datasets effectively, as they blend the capabilities of numerous algorithms, boosting robustness. Prior studies affirm the efficacy of ensemble models in educational data contexts, notably for balancing predicted accuracy with the complexity inherent in student data features, can be seen at table 3.

Table 3. Classification Model Accuracy

Classification Model	**Accuracy**
Logistic Regression	0.86549
Decision Tree	0.809807
Random Forest	0.876253
SVC	0.835158
XGBoost	0.865649
KNN	0.740877
Voting Classifier	0.852
Stacking Classifier	0.856

Source: (Research Result, 2024)

These results are in line with recent literature findings, which found that emotional and psychological factors have a significant influence on academic performance[20]. The application of the synthetic minority oversampling technique (SMOTE) also proved useful in this investigation by rectifying class imbalance, allowing each model to manage the dataset's skewed distribution with higher precision. This method corresponds with previous findings that underline the significance of balancing techniques like SMOTE in education-based machine learning applications, where minority class representation often correlates with crucial, underreported student behaviors or dangers.

The findings underline major implications for educational institutions attempting to serve students at risk owing to mental health difficulties. By employing machine learning models that account for emotional and psychological characteristics, schools and colleges can better detect kids potentially struggling academically due to mental health difficulties. The excellent accuracy rates reported, notably in logistic regression and random forest models, suggest these tools could reliably guide intervention procedures focused at pupils exhibiting patterns connected with academic performance reductions.

Model evaluation with logistic regression generated an accuracy score of 86.56%, which is relatively high when compared to other models such as Decision Tree (80.98%) and K-Nearest Neighbors (74.09%). This model has a precision of 0.88 for class 0 and 0.83 for class 1, with recalls of 0.83 and 0.89, respectively, according to the confusion matrix. This illustrates that the logistic regression model is able to classify the majority of data well but has a little trouble predicting some circumstances.

Table 4. Matrix Visualization

	precision	**recall**	**f1-score**	**support**
0	0.88	0.83	0.85	128
1	0.83	0.89	0.86	122
accuracy			0.86	250
macro avg	0.86	0.86	0.86	250
weighted avg	0.86	0.86	0.86	250

Source: (Research Result, 2024)

The findings underline major implications for educational institutions attempting to serve students at risk owing to mental health difficulties. By employing machine learning models that account for emotional and psychological characteristics, schools and colleges can better detect kids potentially struggling academically due to mental health difficulties. The excellent accuracy rates reported, notably in logistic regression and random forest models, suggest these tools could reliably guide intervention procedures focused at pupils exhibiting patterns connected with academic performance reductions.

Furthermore, this study underscores the need for employing interpretable models, such as logistic regression, in policy-making processes, where educators and administrators must grasp the underlying causes contributing to kids' academic issues. These models give actionable information that can influence resource allocation, such as the deployment of mental health support services or tailored academic tutoring.

Finally, incorporating ensemble models and methodologies like SMOTE into educational analytics boosts the predictive reliability of machine learning applications in education. This method not only aids in spotting at-risk pupils more precisely but also pushes for balanced and complete knowledge of varied student needs. This improvement in educational predictive analytics holds potential for future policy modifications aimed at enhancing academic outcomes across various student groups.

Other supporting literature, further suggests that Random Forest and ensemble models have

Accredited Rank 2 (Sinta 2) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No.225/E/KPT/2022, December 07, 2022. Published by LPPM Universitas Nusa Mandiri

VOL. 10. NO. 2 NOVEMBER 2024
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v10i2.5967

superior performance in multiclass classification, especially when used to complicated and imbalanced educational datasets[21]. The results of this study imply that the application of machine learning models can provide deeper insights into the factors that influence student academic attainment, with the potential to be employed in future educational policies and interventions.

CONCLUSION

The results of this study underscore the considerable influence of emotional and mental health factors on students' academic performance, measured through cumulative GPA. With the Random Forest model achieving an accuracy of 87.63% and logistic regression reaching 86.56%, the findings underscore the potential of machine learning models in identifying kids at risk of academic issues due to mental health challenges.

The adoption of the Synthetic Minority Oversampling Technique (SMOTE) to manage data imbalances has considerably boosted prediction accuracy, particularly for minority classes, thereby permitting a more nuanced understanding of these implications. This advancement resonates with current work that underlines the crucial impact of mental health on school attainment and the relevance of accommodating minority student profiles.

Educational institutions can employ these predictive algorithms to proactively establish support systems for students showing indicators of psychological distress. By implementing such data-driven solutions, institutions can introduce more targeted resources, such as counseling services and academic help tailored to students' emotional and mental health requirements. For instance, early identification of students likely to have academic difficulty due to mental health issues could enable the provision of preventative care, lowering the risk of academic deterioration and increasing well-being.

The addition of explainable AI (XAI) techniques in future models can significantly boost these systems' practical utility. XAI promotes transparency by making model outputs and predictions more comprehensible for stakeholders, such as educators and counselors, who may require clear reasons to lead focused actions. By identifying particular aspects contributing to academic performance, XAI may deliver actionable insights, permitting individualized educational support and creating a more holistic approach to student success.

However, it is necessary to address this study's weaknesses, including the confined dataset that may not adequately represent various student populations. A future study could expand dataset inclusivity by integrating additional mental health variables, such as stress and burnout levels, while also widening the sample to represent students from various academic disciplines and geographic regions. This extension will ensure a more complete model that better reflects the intricacies of student mental health and academic achievement on a worldwide scale.

This study contributes significantly to educational predictive analytics, revealing the potential for machine learning to increase understanding of mental health's impact on academic achievement in increasingly complicated educational environments. By enhancing prediction accuracy and interpretability, our work lays the framework for future applications that enhance student well-being and academic performance, ultimately promoting a more supportive and responsive educational system.

REFERENCE

[1] I. E. Fattoh, F. A. Mousa, and S. Safwat, "Converting cumulative grade point average to an equivalent percentage value based on a fuzzy logic," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 24, no. 3, 2021, doi: 10.11591/ijeecs.v24.i3.pp1823-1831.

[2] X. Wen, Y. Cai, K. Li, Z. Wang, W. Zhang, and M. Qin, "A Cross-Sectional Association Between Screen-Based Sedentary Behavior and Anxiety in Academic College Students: Mediating Role of Negative Emotions and Moderating Role of Emotion Regulation," *Psychol Res Behav Manag*, vol. 16, 2023, doi: 10.2147/PRBM.S430928.

[3] Y. Tang and W. He, "Meta-analysis of the relationship between university students' anxiety and academic performance during the coronavirus disease 2019 pandemic," 2023. doi: 10.3389/fpsyg.2023.1018558.

[4] N. Nordin, Z. Zainol, M. H. Mohd Noor, and C. Lai Fong, "A comparative study of machine learning techniques for suicide attempts predictive model," *Health Informatics J*, vol. 27, no. 1, 2021, doi: 10.1177/1460458221989395.

[5] D. Gonzalez-Cuautle *et al.*, "Synthetic minority oversampling technique for optimizing classification tasks in botnet and intrusion-detection-system datasets,"

Applied Sciences (Switzerland), vol. 10, no. 3, 2020, doi: 10.3390/app10030794.

[6] Z. Xu, D. Shen, Y. Kou, and T. Nie, "A Synthetic Minority Oversampling Technique Based on Gaussian Mixture Model Filtering for Imbalanced Data Classification," *IEEE Trans Neural Netw Learn Syst*, vol. 35, no. 3, 2024, doi: 10.1109/TNNLS.2022.3197156.

[7] W. Suci and S. Samsudin, "Algoritma K-Nearest Neighbors dan Synthetic Minority Oversampling Technique dalam Prediksi Pemesanan Tiket Pesawat," *Jurnal Media Informatika Budidarma*, vol. 6, no. 3, 2022, doi: 10.30865/mib.v6i3.4374.

[8] Z. Qu, H. Li, Y. Wang, J. Zhang, A. Abu-Siada, and Y. Yao, "Detection of electricity theft behavior based on improved synthetic minority oversampling technique and random forest classifier," *Energies (Basel)*, vol. 13, no. 8, 2020, doi: 10.3390/en13082039.

[9] D. Dablain, B. Krawczyk, and N. V. Chawla, "DeepSMOTE: Fusing Deep Learning and SMOTE for Imbalanced Data," *IEEE Trans Neural Netw Learn Syst*, vol. 34, no. 9, 2023, doi: 10.1109/TNNLS.2021.3136503.

[10] K. D. Hinkson *et al.*, "An examination of the mental health and academic performance of student veterans," *Journal of American College Health*, vol. 70, no. 8, 2022, doi: 10.1080/07448481.2020.1858837.

[11] F. Bolinski, N. Boumparis, A. Kleiboer, P. Cuijpers, D. D. Ebert, and H. Riper, "The effect of e-mental health interventions on academic performance in university and college students: A meta-analysis of randomized controlled trials," 2020. doi: 10.1016/j.invent.2020.100321.

[12] S. Zada, Y. Wang, M. Zada, and F. Gul, "Effect of mental health problems on academic performance among university students in Pakistan," *International Journal of Mental Health Promotion*, vol. 23, no. 3, 2021, doi: 10.32604/IJMHP.2021.015903.

[13] T. Wongvorachan, S. He, and O. Bulut, "A Comparison of Undersampling, Oversampling, and SMOTE Methods for Dealing with Imbalanced Classification in Educational Data Mining," *Information (Switzerland)*, vol. 14, no. 1, 2023, doi: 10.3390/info14010054.

[14] A. Anggrawan, H. Hairani, and C. Satria, "Improving SVM Classification Performance on Unbalanced Student Graduation Time Data Using SMOTE," *International Journal of Information and Education Technology*, vol. 13, no. 2, 2023, doi: 10.18178/ijiet.2023.13.2.1806.

[15] A. M. Sowjanya and O. Mrudula, "Effective treatment of imbalanced datasets in health care using modified SMOTE coupled with stacked deep learning algorithms," *Applied Nanoscience (Switzerland)*, vol. 13, no. 3, 2023, doi: 10.1007/s13204-021-02063-4.

[16] F. Syah, H. Fajrin, A. N. Afif, M. R. Saeputra, D. Mirranty, and D. D. Saputra, "Analisa Sentimen Terhadap Twitter IndihomeCare Menggunakan Perbandingan Algoritma Smote, Support Vector Machine, AdaBoost dan Particle Swarm Optimization," *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, vol. 7, no. 1, 2023, doi: 10.35870/jtik.v7i1.686.

[17] S. Akuma and H. Abakpa, "Predicting Undergraduate Level Students' Performance Using Regression," *Nigerian Annals Of Pure And Applied Sciences*, vol. 4, no. 1, 2021, doi: 10.46912/napas.224.

[18] Y. Baashar *et al.*, "Evaluation of postgraduate academic performance using artificial intelligence models," *Alexandria Engineering Journal*, vol. 61, no. 12, 2022, doi: 10.1016/j.aej.2022.03.021.

[19] R. Jain, M. Gupta, S. Taneja, and D. J. Hemanth, "Deep learning based detection and analysis of COVID-19 on chest X-ray images," *Applied Intelligence*, vol. 51, no. 3, 2021, doi: 10.1007/s10489-020-01902-1.

[20] M. M. Porras and F. H. Ortega, "Procrastination, test anxiety and academic performance on university students," *Interdisciplinaria*, vol. 38, no. 2, 2021, doi: 10.16888/INTERD.2021.38.2.16.

[21] A. Rácz, D. Bajusz, and K. Héberger, "Effect of dataset size and train/test split ratios in qsar/qspr multiclass classification," *Molecules*, vol. 26, no. 4, 2021, doi: 10.3390/molecules26041111.