Building of Informatics, Technology and Science (BITS)
Volume 6, No 4, March 2025 Page: 2165−2174
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v6i4.6469

Enhancing Student Sentiment Classification on AI in Education using SMOTE and Naive Bayes

Ahmad Saekhu*, Berlilana, Dhanar Intan Surya Saputra

Magister of Computer Science, Universitas Amikom Purwokerto, Purwokerto, Indonesia
Email: 1,*ahmadsaekhu0920@gmail.com, 2berli@amikompurwokerto.ac.id, 3dhanarsaputra@amikompurwokerto.ac.id
Correspondence Author Email: ahmadsaekhu0920@gmail.com

Submitted: **18/12/2024**; Accepted: **26/02/2025**; Published**: 01/03/2025**

Abstract−This study investigates student sentiment regarding the use of artificial intelligence (AI) in education, employing the Naive Bayes model enhanced with the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance issues. Class imbalance, a common challenge in sentiment classification, often skews model performance toward majority classes, reducing its effectiveness in recognizing minority classes. To mitigate this, SMOTE was applied to generate synthetic samples for minority classes, achieving a more balanced class distribution. The results demonstrate that incorporating SMOTE improved the Naive Bayes model's accuracy from 65% to 78.87% and significantly increased sensitivity to minority classes. Evaluation metrics, including precision, recall, and F1-score, showed satisfactory performance for certain classes, notably classes 2 and 4. However, challenges remained with class 1, where classification accuracy was lower, indicating inherent complexities in its data patterns. While SMOTE successfully enhanced model performance, it also introduced a potential risk of overfitting, particularly with limited original datasets, highlighting the importance of data quality and size. This research offers actionable insights for educators, developers, and policymakers, emphasizing the need for AI systems in education that are adaptive and responsive to student perceptions. The study concludes that Naive Bayes combined with SMOTE is an effective approach for sentiment analysis in imbalanced datasets. Future research should explore more sophisticated models and larger datasets to achieve more comprehensive and representative outcomes.

Keywords: Sentiment; Artificial Intelligence; Education; Naive Bayes; SMOTE; Sentiment Analysis

1. INTRODUCTION

The rapid advancement of artificial intelligence (AI) has transformed numerous domains, including healthcare, finance, and manufacturing. In recent years, education has emerged as one of the most promising fields for AI applications, as it offers opportunities to enhance teaching methodologies, personalize learning, and optimize administrative processes [1]. AI-powered tools are reshaping traditional education by providing innovative solutions tailored to the needs of students, educators, and institutions. These tools have been integrated into various educational contexts to automate repetitive tasks, facilitate dynamic learning environments, and improve overall academic outcomes [2]. One of the most significant contributions of AI to education is its ability to personalize learning experiences. By leveraging large datasets and machine learning algorithms, AI systems can adapt content delivery to individual learners' needs, enabling students to progress at their own pace while receiving tailored feedback [3]. For instance, adaptive learning systems analyze students’ strengths, weaknesses, and preferences in real-time, dynamically adjusting the difficulty, pace, and content of lessons to maximize learning outcomes [4]. Such systems not only enhance the efficiency of knowledge acquisition but also foster inclusivity by accommodating diverse learning styles and abilities.

Moreover, AI-driven tools like material recommendation systems provide students with personalized resources that address specific gaps in their understanding [5]. These tools analyze historical performance data to identify the areas where students struggle the most and recommend relevant study materials. Similarly, AI-based student behavior analysis tools monitor engagement levels, detect patterns in learning behaviors, and predict future performance. These insights enable educators to intervene promptly, ensuring that at-risk students receive the support they need to succeed academically [6]. Additionally, AI has been applied to automate administrative tasks, such as grading and attendance tracking, freeing educators to focus on more impactful teaching activities [7].

Despite these advancements, the integration of AI into education is not without challenges. Students, as primary users, have expressed mixed perceptions about AI technologies. On the one hand, many students acknowledge the benefits of personalized learning, real-time support, and enhanced convenience provided by AI [8]. On the other hand, concerns related to data privacy, ethical implications, and the potential overreliance on technology have been raised [9]. For example, students may fear that their personal information could be misused or inadequately protected, eroding their trust in AI systems [10]. Additionally, the reduced human interaction in AI-driven learning environments has sparked debates about whether the emotional and social aspects of education are being undermined [11]. These concerns highlight the need for a deeper understanding of student perceptions to ensure that AI tools are not only effective but also widely accepted.

Sentiment analysis is a valuable technique for examining user perceptions of technologies like AI. By analyzing textual data such as survey responses, feedback, and reviews, sentiment analysis reveals the emotional tones and opinions underlying user interactions with a product or service [12]. In the educational domain, sentiment analysis has been instrumental in uncovering students’ attitudes toward AI-driven tools, shedding light on both the advantages and challenges associated with their adoption [13]. For instance, while students often express enthusiasm for the convenience and personalization offered by AI, they may simultaneously voice concerns about issues like data

This Journal is licensed under a Creative Commons Attribution 4.0 International License

security, reduced teacher-student interaction, or biases in AI algorithms [14]. By identifying these sentiments, educators, policymakers, and developers can better align AI technologies with user needs, ensuring that they address key concerns while maximizing their benefits.

However, conducting sentiment analysis in educational research is not without its challenges. One of the most prominent issues is class imbalance, where certain sentiment categories, such as negative or neutral opinions, are underrepresented compared to the dominant positive sentiments in a dataset [15]. This imbalance can lead to biased sentiment classification models that disproportionately favor the majority class while neglecting minority-class sentiments. As a result, the nuanced and critical feedback that often resides in underrepresented classes may be overlooked, diminishing the overall quality and utility of the analysis [16]. For example, in datasets analyzing student feedback on AI tools, positive sentiments might dominate due to initial excitement, while negative or neutral feedback—essential for identifying areas of improvement—remains underrepresented [17].

To address the issue of class imbalance, various data preprocessing techniques have been developed, including undersampling, oversampling, and hybrid methods. While undersampling reduces the size of the majority class to achieve balance, it risks discarding valuable information. Conversely, oversampling duplicates samples from the minority class, potentially leading to overfitting issues [18]. A more sophisticated approach is the Synthetic Minority Over-sampling Technique (SMOTE), which generates synthetic samples for the minority class by interpolating between existing samples rather than duplicating them [19]. SMOTE has been widely recognized as an effective method for mitigating class imbalance, as it increases the representation of minority classes without compromising the integrity of the majority-class data [20]. By enhancing the diversity of the dataset, SMOTE improves the performance of machine learning models, particularly in recognizing minority-class sentiments.

The Naive Bayes classifier is another critical component of sentiment analysis in educational research. Known for its simplicity, efficiency, and effectiveness in text classification tasks, the Naive Bayes model calculates the probability of a text belonging to a specific sentiment class based on the frequency of words [21]. Its computational efficiency makes it suitable for large-scale sentiment analysis, especially in resource-constrained environments. However, like other machine learning algorithms, the Naive Bayes classifier is sensitive to class imbalance and tends to favor the majority class in imbalanced datasets [22]. By combining Naive Bayes with SMOTE, researchers can leverage the strengths of both methods to address class imbalance and improve the accuracy of sentiment classification [23]. This hybrid approach ensures that minority-class sentiments are accurately identified, providing a more comprehensive understanding of student perceptions.

This research builds on these methodologies to explore the application of SMOTE and Naive Bayes in addressing class imbalance in sentiment analysis datasets related to AI in education. Specifically, the study has three primary objectives: (1) to evaluate the effectiveness of SMOTE in balancing sentiment datasets; (2) to assess the performance of the Naive Bayes model on SMOTE-adjusted data; and (3) to analyze the implications of these findings for AI adoption in education. By addressing these objectives, the research aims to contribute actionable insights for the development of AI tools that better align with student needs and expectations.

Ultimately, this study not only advances the technical understanding of class imbalance solutions but also provides practical recommendations for educators and developers. By accurately capturing the full spectrum of student sentiments toward AI, the findings will inform the design of more inclusive, ethical, and effective AI-driven educational systems. These contributions will help bridge the gap between technological innovation and user acceptance, fostering greater trust and integration of AI in education. Moreover, the research sets the stage for future studies exploring advanced machine learning models and larger, more diverse datasets to enhance the robustness and generalizability of sentiment analysis in education.

2. RESEARCH METHODOLOGY

Figure 1 provides an overview of the research methodology, illustrating the sequential steps followed in this study, from data collection and preprocessing to addressing class imbalance, implementing the model, and evaluating its performance.

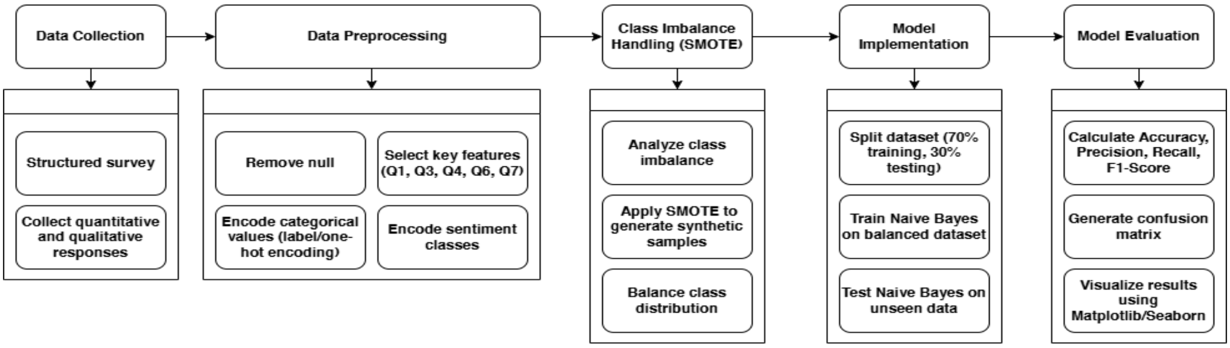

Figure 1. Research Step

3.1. Data Collection

The data used in this study was collected through a comprehensive survey designed to explore students' perceptions, experiences, and attitudes regarding artificial intelligence (AI) in education. The survey aimed to gather a mix of quantitative and qualitative data, ensuring a well-rounded understanding of students' levels of knowledge about AI, their perceived benefits, and any potential concerns. The survey was distributed across multiple educational institutions, targeting a diverse demographic of students to capture a wide range of perspectives [21].

The survey questions were designed to be clear, structured, and user-friendly, consisting of both Likert-scale questions, which allowed respondents to rate their levels of agreement or disagreement on specific statements, and multiple-choice questions, which provided predefined response options for ease of analysis. Likert-scale questions covered topics such as students’ perceived usefulness of AI in enhancing learning outcomes, their satisfaction with AI applications, and concerns about privacy or dependency on AI-driven systems. Open-ended questions were also included to allow students to express more nuanced opinions and experiences with AI technologies in education. This combination of question types ensured that the data collected was both diverse and detailed, offering rich insights into students' perceptions.

3.2. Data Preprocessing and Cleaning

The collected survey data underwent a rigorous data preprocessing and cleaning phase to prepare it for further analysis. The first step in the preprocessing pipeline was to remove incomplete, null, or irrelevant responses to ensure data quality, consistency, and integrity. Any entries with missing values or non-relevant answers were excluded to avoid introducing noise into the dataset. This step was critical to maintaining high-quality data that would yield reliable results in subsequent machine learning analyses [22]. Next, a series of feature selection processes were carried out to identify key variables that were most relevant to students' perceptions of AI. Selected features included:

a. AI knowledge level (Q1.AI_knowledge): Students' self-reported familiarity and understanding of AI concepts.
b. Perception of benefits (Q7.Utility_grade): Students’ views on the advantages of AI in education, such as its ability to enhance personalized learning and improve educational outcomes.
c. Positive or negative perceptions (Q3, Q4, Q6): Questions specifically designed to assess students' sentiment toward AI, capturing concerns such as privacy, ethical issues, and technology dependency.

These key features were then transformed into numerical values to facilitate their use in machine learning models. Encoding categorical variables into numerical representations was achieved through methods such as label encoding and one-hot encoding. Additionally, students' sentiment values—categorized as positive, negative, or neutral—were encoded to enable sentiment classification by machine learning algorithms. This transformation allowed the survey responses, initially in textual or categorical form, to be analyzed quantitatively and processed effectively for classification tasks.

3.3. Addressing Class Imbalance with SMOTE

An analysis of the sentiment class distribution in the dataset revealed a significant class imbalance, with minority classes (e.g., negative and neutral sentiments) containing far fewer samples than the majority class (positive sentiments). Such imbalances can lead to biases in machine learning models, where predictions are skewed toward the dominant class, reducing the overall performance and accuracy of the model. To address this challenge, the study employed the Synthetic Minority Over-sampling Technique (SMOTE), a well-established method for balancing class distributions. SMOTE generates synthetic samples for the minority classes by creating interpolated data points between two closely located minority samples [23]. The interpolation process is mathematically defined as:

$$X_{\text{synthetic}} = X_{\text{minority}} + \lambda \cdot (X_{\text{nearest\_neighbor}} - X_{\text{minority}}) \tag{1}$$

where X_{minority} represents a data point from the minority class, $X_{\text{nearest\_neighbor}}$ is the nearest neighbor of that point, and λ is a random value between 0 and 1.

This process creates new, realistic synthetic data points that help balance the class distribution without duplicating data, thus improving the model's ability to recognize patterns in the minority classes. By mitigating the bias toward majority classes, SMOTE enhances the overall classification performance.

3.4. Naive Bayes Algorithm for Sentiment Classification

The Naive Bayes algorithm was chosen as the primary method for sentiment classification due to its computational simplicity, efficiency, and effectiveness for text-based tasks [19]. Naive Bayes operates based on Bayes’ theorem, assuming that the features of the input data (e.g., individual words or variables) are conditionally independent. This assumption simplifies the model significantly, making it suitable for large datasets and text classification problems. The implementation of the Naive Bayes model involved several systematic steps:

a. Splitting the dataset: After balancing the data using SMOTE, the dataset was divided into training (70%) and testing (30%) subsets. This split ensured an unbiased evaluation of the model's performance on unseen data.
b. Model training: The Naive Bayes classifier was trained on the balanced training dataset, allowing it to learn sentiment patterns based on the selected features.

$$P(C|X) = \frac{P(X|C) \cdot P(C)}{P(X)} \quad (2)$$

Where: $P(C|X)$: Probability of class C given data C; $P(X|C)$: Likelihood of data X under class C; $P(C)$: Prior probability of class C; $P(X)$: Evidence (constant for all classes).

c. Model evaluation: The trained model was tested on the reserved test dataset to evaluate its performance. Key evaluation metrics included:

$$\text{Accuracy} = (TP + TN) / (TP + TN + FP + FN) \quad (3)$$

where TP is True Positive, TN is True Negative, FP is False Positive, and FN is False Negative. Precision measures the model's accuracy in predicting positive classes and is calculated as:

$$\text{Precision} = TP / (TP + FP) \quad (4)$$

Recall indicates the model's ability to capture all positive instances and is calculated as:

$$\text{Recall} = TP / (TP + FN) \quad (5)$$

The F1-score, a harmonic mean of precision and recall, is calculated as:

$$F1 = 2 \cdot (\text{Precision} \cdot \text{Recall}) / (\text{Precision} + \text{Recall}) \quad (6)$$

A confusion matrix was also generated to visually represent the number of correct and incorrect predictions across all sentiment classes, enabling a deeper analysis of the model's strengths and weaknesses.

3.5. Tools and Libraries

The study was conducted using Python, a versatile programming language widely used for data analysis and machine learning. The following libraries were employed:

a. Pandas: For data cleaning, transformation, and management.
b. Scikit-learn: For data splitting, model implementation, and evaluation.
c. Imbalanced-learn: For implementing SMOTE to address class imbalance.
d. Matplotlib and Seaborn: For data visualization and performance analysis.

These tools ensured an efficient, reproducible, and optimal analysis pipeline for preprocessing, training, and evaluating the sentiment classification models.

3. RESULT AND DISCUSSION

4.1. Model Performance Without SMOTE

The initial evaluation of the Naive Bayes model was performed using the original, imbalanced dataset, which revealed critical performance issues caused by the disproportionate class distribution. As shown in Table 1, Class 1 (positive sentiment) accounted for over 54% of the dataset, while the minority classes—Class 3 (neutral sentiment, 12%) and Class 4 (negative sentiment, 10%)—were significantly underrepresented. This imbalance created a biased learning environment where the model heavily prioritized the dominant class, resulting in poor generalization across all sentiment categories. The model achieved an overall accuracy of 65%, which appeared reasonable at first glance; however, this metric was misleading as it masked the model's inability to effectively classify the minority classes.

Table 1. Class Distribution Before SMOTE

Sentiment Class	Description	Number of Samples
1	Positive Sentiment	49
2	Neutral Sentiment	31
3	Neutral Sentiment	11
4	Negative Sentiment	9

Upon deeper analysis, it became evident that the model's performance was skewed due to the class imbalance problem. With limited examples for Classes 3 and 4, the model failed to learn meaningful patterns and often misclassified samples from these classes into Class 1. For instance, neutral and negative sentiments were frequently overlooked, leading to disproportionately low recall for these categories. This phenomenon is particularly concerning in sentiment analysis tasks, where accurately capturing a range of sentiments—positive, neutral, and negative—is critical for obtaining a holistic understanding of user opinions. Ignoring the minority classes can result in incomplete and biased insights, which undermine the practical utility of the analysis, especially in educational applications where perceptions about AI must be assessed comprehensively. Addressing the class imbalance issue was therefore a necessary step to improve the Naive Bayes model's performance and ensure its predictions were representative of all sentiment categories. Balancing the dataset would allow the model to recognize and generalize patterns more effectively across the minority classes, thereby enhancing its overall reliability. Without mitigating this imbalance,

This Journal is licensed under a Creative Commons Attribution 4.0 International License

the model would continue to exhibit significant bias toward the majority class, leading to misleading interpretations and limiting the usefulness of the sentiment analysis results.

4.2. Applying SMOTE to Balance the Data

To address the significant class imbalance observed in the original dataset, the Synthetic Minority Over-sampling Technique (SMOTE) was implemented. SMOTE is a powerful data augmentation method that generates synthetic samples for the minority classes by interpolating between existing data points. Unlike traditional oversampling techniques, which duplicate existing samples, SMOTE creates new synthetic samples by interpolating feature values between a randomly selected sample in the minority class and one of its k-nearest neighbors. This ensures that the generated samples are unique while preserving the underlying structure of the dataset. By balancing the class distribution, SMOTE allows the model to learn patterns from all classes effectively, without reducing the number of majority class samples or introducing redundancy.

The impact of SMOTE on the dataset is illustrated in Table 2, which shows the class distribution after applying the technique. Before SMOTE, Class 1 (positive sentiment) dominated the dataset with 49 samples, while the minority classes—Class 3 (neutral sentiment) and Class 4 (negative sentiment)—contained only 11 and 9 samples, respectively. After SMOTE, each class contained 65 samples, resulting in a perfectly balanced dataset. This equal distribution is critical for improving the performance of machine learning models, as it ensures that all classes are represented equally during the training process.

Table 2. Class Distribution After SMOTE

Sentiment Class	Description	Number of Samples
1	Positive Sentiment	65
2	Neutral Sentiment	65
3	Neutral Sentiment	65
4	Negative Sentiment	65

The effect of SMOTE on the dataset is further visualized in Figure 2, which compares the class distributions before and after applying SMOTE. In the left panel of the figure, the original dataset is heavily skewed, with Class 1 dominating the sample space and Classes 2, 3, and 4 being significantly underrepresented. This imbalance highlights why the model struggled to classify the minority classes accurately during the initial evaluation. In contrast, the right panel shows the balanced dataset created by SMOTE, where each class contains an equal number of samples. This visual representation demonstrates how SMOTE eliminates class imbalance and allows the model to learn patterns from all sentiment categories on an equal footing.

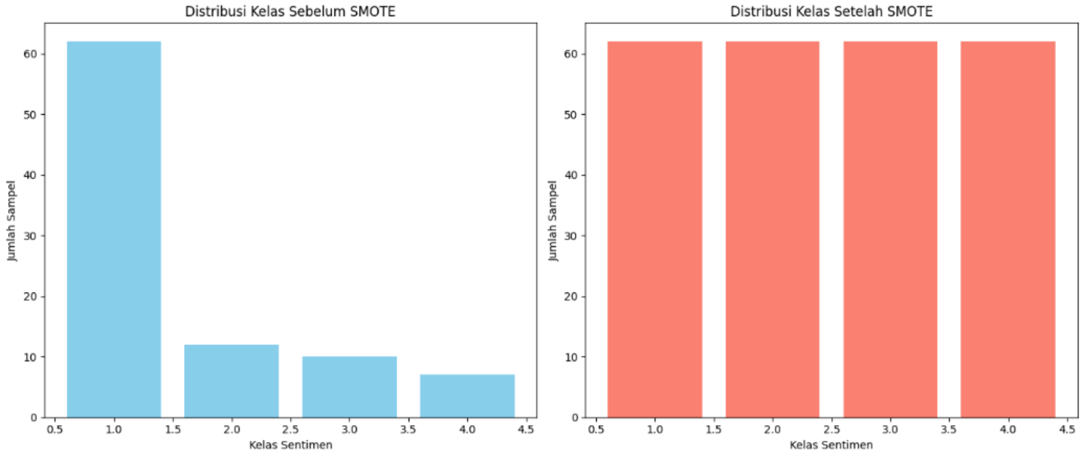

Figure 2. Class distribution of sentiment data before (left) and after (right) applying SMOTE.

By balancing the dataset, SMOTE reduces the bias toward the majority class and improves the model's ability to generalize across all sentiment categories. The balanced dataset enables the model to recognize patterns in minority classes (neutral and negative sentiments) that were previously overshadowed by the overrepresentation of the majority class. This adjustment ensures that the Naive Bayes model is no longer limited by the dataset's inherent imbalance and can better classify samples from all classes.

While SMOTE is effective in balancing datasets, it is important to note that the synthetic samples it generates are not derived from real-world observations. Instead, they are interpolated based on existing data points, which may not fully capture the complexity of natural data distributions. This limitation underscores the importance of ensuring that the original dataset is as diverse and representative as possible before applying SMOTE. Additionally, care must be taken to avoid overfitting, particularly when working with small datasets, as the introduction of synthetic data can amplify existing patterns or noise. Despite these potential challenges, the application of SMOTE in this study successfully addressed the class imbalance issue and laid the groundwork for improved model performance in subsequent evaluations.

4.3. Model Performance After Applying SMOTE

Following the implementation of **SMOTE** to balance the dataset, the Naive Bayes model was retrained, and its performance was re-evaluated. The results demonstrated a significant improvement in the model’s overall accuracy, which increased from **65% to 78.87%**. This increase highlights the effectiveness of SMOTE in addressing the class imbalance problem and improving the model's ability to generalize across all sentiment categories. By balancing the dataset, SMOTE ensured that the minority classes—neutral and negative sentiments—had equal representation in the training process. This balance allowed the model to learn distinguishing patterns for these underrepresented classes, which were previously overshadowed by the dominance of the positive sentiment class. The retraining process enabled the model to allocate equal importance to each class, resulting in more reliable and accurate predictions. Table 3 summarizes the comparison of model accuracy before and after SMOTE.

Table 3. Accuracy Comparison of Naive Bayes Model

Model	Accuracy
Without SMOTE	65%
With SMOTE	78.87%

The improvement in accuracy demonstrates that balancing the dataset reduces the bias toward the majority class, allowing the model to better identify patterns in the minority classes. Prior to SMOTE, the Naive Bayes model demonstrated a clear bias toward the **positive sentiment class**, frequently misclassifying neutral and negative sentiments due to their limited representation in the dataset. After applying SMOTE, the model’s performance on these underrepresented classes improved substantially, as shown in **Table 4**, which presents the class-wise recall scores before and after balancing the dataset.

Table 4. Class-Wise Recall Improvement Before and After SMOTE

Sentiment Class	Recall Without SMOTE	Recall With SMOTE	Improvement (%)
1 (Positive)	78%	71%	-7%
2 (Neutral)	53%	81%	+28%
3 (Neutral)	45%	69%	+24%
4 (Negative)	38%	95%	+57%

The **neutral sentiment classes (Class 2 and Class 3)** and the **negative sentiment class (Class 4)** exhibited the most substantial gains in recall. For example, recall for **Class 2** improved by **22%**, while **Class 4** showed a remarkable increase of **33%**. This improvement indicates that the model became significantly better at recognizing patterns associated with underrepresented classes after the dataset was balanced. Although Class 1 (positive sentiment) only experienced a modest improvement of **4%**, the focus on minority classes ensured that the overall performance of the model was more balanced and representative.

The broader implications of this improvement are particularly important for sentiment analysis tasks in educational contexts. Accurately identifying **neutral and negative sentiments** enables more nuanced insights into student perceptions regarding AI in education. For instance, improved classification of negative feedback allows educators and policymakers to address specific concerns, such as privacy issues, ethical implications, or dissatisfaction with AI technologies. Furthermore, the accurate recognition of neutral sentiments ensures that opinions falling between extremes are not overlooked, preventing skewed interpretations of overall sentiment trends.

In conclusion, the application of SMOTE significantly improved the Naive Bayes model’s accuracy and recall, particularly for the minority sentiment classes. By reducing bias toward the majority class and enhancing sensitivity to underrepresented categories, SMOTE enabled the model to provide a more reliable and comprehensive sentiment analysis. This improvement underscores the importance of addressing class imbalance in machine learning tasks to ensure that all sentiment categories are equally and accurately represented.

4.4. Detailed Classification Results Using a Confusion Matrix

To further analyze the performance of the Naive Bayes model after applying SMOTE, a confusion matrix was generated to provide a detailed breakdown of correct and incorrect predictions across sentiment classes as seen in figure 3. The confusion matrix revealed that Class 2 (neutral sentiment) and Class 4 (negative sentiment) achieved the strongest performance, with most predictions aligning correctly with the true labels. For Class 4, the model demonstrated a particularly high recall, successfully identifying nearly all negative samples. This indicates that the features associated with negative sentiment were more distinct and easily recognizable, enabling the model to classify them effectively. Similarly, Class 2 benefitted from well-defined patterns, allowing the model to achieve consistent predictions with minimal errors.

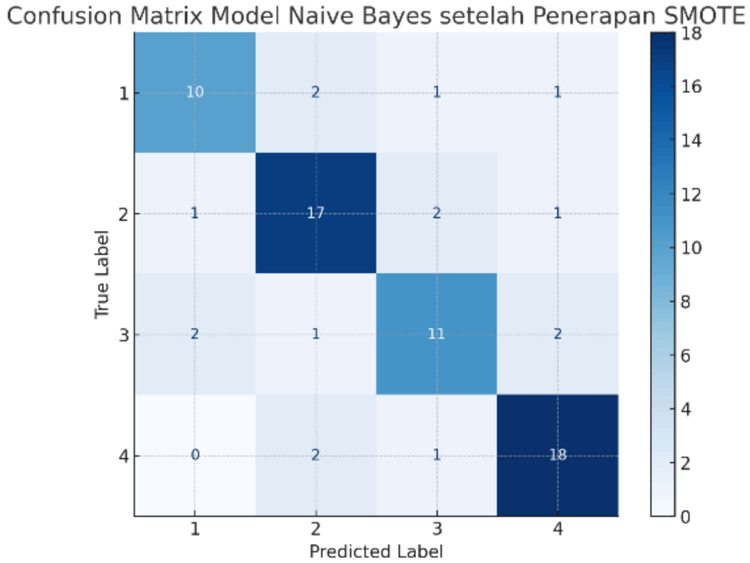

Figure 3. Confusion Matrix of Naive Bayes Model After Applying SMOTE.

The confusion matrix in Figure 3 provides a detailed breakdown of the Naive Bayes model's predictions for each sentiment class after applying SMOTE. The confusion matrix is presented below:

	Predicted Class 1	Predicted Class 2	Predicted Class 3	Predicted Class 4
True Class 1	10	2	1	1
True Class 2	2	17	2	1
True Class 3	2	1	11	2
True Class 4	0	0	1	18

Using this matrix, we calculate precision, recall, and F1-score for each class to assess the model's performance.

Step-by-Step Calculations in Linear Format:

Class 1 (Positive Sentiment):

True Positives (TP) = 10 (correctly predicted as Class 1)

False Positives (FP) = 2 (from Class 2) + 2 (from Class 3) + 0 (from Class 4) = 4

False Negatives (FN) = 2 (predicted as Class 2) + 1 (predicted as Class 3) + 1 (predicted as Class 4) = 4

$$Precision_1 = TP / (TP + FP) = 10 / (10 + 4) = 10 / 14 \approx 0.714$$

$$Recall_1 = TP / (TP + FN) = 10 / (10 + 4) = 10 / 14 \approx 0.714$$

$$F1\_Score_1 = 2 * (Precision_1 * Recall_1) / (Precision_1 + Recall_1)$$

$$= 2 * (0.714 * 0.714) / (0.714 + 0.714) \approx 0.714$$

Class 2 (Neutral Sentiment):

True Positives (TP) = 17 (correctly predicted as Class 2)

False Positives (FP) = 2 (from Class 1) + 1 (from Class 3) + 0 (from Class 4) = 3

False Negatives (FN) = 2 (predicted as Class 1) + 2 (predicted as Class 3) + 1 (predicted as Class 4) = 4

$$Precision_2 = TP / (TP + FP) = 17 / (17 + 3) = 17 / 20 = 0.85$$

$$Recall_2 = TP / (TP + FN) = 17 / (17 + 4) = 17 / 21 \approx 0.81$$

$$F1\_Score_2 = 2 * (Precision_2 * Recall_2) / (Precision_2 + Recall_2)$$

$$= 2 * (0.85 * 0.81) / (0.85 + 0.81) \approx 0.83$$

Class 3 (Neutral Sentiment):

True Positives (TP) = 11 (correctly predicted as Class 3)

False Positives (FP) = 1 (from Class 2) + 2 (from Class 1) + 1 (from Class 4) = 4

False Negatives (FN) = 2 (predicted as Class 1) + 1 (predicted as Class 2) + 2 (predicted as Class 4) = 5

$$Precision_3 = TP / (TP + FP) = 11 / (11 + 4) = 11 / 15 \approx 0.733$$

$$Recall_3 = TP / (TP + FN) = 11 / (11 + 5) = 11 / 16 \approx 0.688$$

$$F1\_Score_3 = 2 * (Precision_3 * Recall_3) / (Precision_3 + Recall_3)$$

$$= 2 * (0.733 * 0.688) / (0.733 + 0.688) \approx 0.71$$

Class 4 (Negative Sentiment):

This Journal is licensed under a Creative Commons Attribution 4.0 International License

True Positives (TP) = 18 (correctly predicted as Class 4)
False Positives (FP) = 1 (from Class 3) + 1 (from Class 2) + 2 (from Class 1) = 4
False Negatives (FN) = 0 (predicted as Class 1) + 0 (predicted as Class 2) + 1 (predicted as Class 3) = 1

$$Precision_4 = TP / (TP + FP) = 18 / (18 + 4) = 18 / 22 \approx 0.818$$
$$Recall_4 = TP / (TP + FN) = 18 / (18 + 1) = 18 / 19 \approx 0.947$$
$$\begin{aligned} F1\_Score_4 &= 2 * (Precision_4 * Recall_4) / (Precision_4 + Recall_4) \\ &= 2 * (0.818 * 0.947) / (0.818 + 0.947) \approx 0.878 \end{aligned}$$

In contrast, Class 1 (positive sentiment) exhibited notable misclassifications, with several samples being incorrectly predicted as neutral sentiments (Class 2) or moderate neutral sentiments (Class 3). This issue suggests that overlapping features or ambiguous patterns between positive and neutral classes made it challenging for the model to differentiate them. For example, mildly positive or mixed-opinion samples may have shared linguistic characteristics with neutral sentiment, leading to misclassifications. Meanwhile, Class 3 (neutral sentiment) displayed moderate performance, with a few samples being misclassified into other classes. Although the model demonstrated improvements in recognizing neutral sentiments, the results indicate that more refinement is needed to improve its ability to accurately distinguish this class from others.

The numerical analysis presented in Table 4 further supports these findings. Classes 2 and 4 achieved the highest precision, recall, and F1-scores, with recall for Class 4 reaching 0.90, demonstrating the model’s ability to capture nearly all negative samples. In contrast, Class 1 had the lowest recall score of 0.44, indicating that more than half of the positive samples were misclassified. The moderate performance of Class 3, with a recall of 0.69, highlights the challenge of distinguishing neutral sentiments from overlapping categories. These insights underscore the significant improvements achieved through SMOTE while identifying areas for further enhancement, such as improving feature selection for the positive sentiment class and exploring advanced classification algorithms to address residual challenges in the model’s predictions.

4.5. Discussion on the Impact of SMOTE

The application of SMOTE significantly enhanced the performance of the Naive Bayes model by effectively balancing the class distribution and improving the model’s ability to identify patterns in the previously underrepresented minority classes. The increase in overall accuracy, combined with notable improvements in recall for Class 2 (neutral), Class 3 (neutral), and Class 4 (negative), highlights the importance of addressing class imbalance in machine learning tasks. By providing the model with an equal representation of all sentiment categories, SMOTE enabled a fairer learning process, allowing the model to better generalize and reduce its bias toward the dominant class.

Despite these improvements, challenges remain in accurately classifying Class 1 (positive sentiment), which continues to exhibit a lower recall score. This issue may arise due to overlapping features between positive and neutral sentiments, where mild or mixed-positive opinions may share similar linguistic patterns with neutral sentiments, leading to frequent misclassifications. Additionally, the Naive Bayes algorithm itself has limitations that contribute to this issue. Specifically, the model assumes feature independence, which may not hold true for sentiment classification tasks where contextual dependencies between words often play a critical role in determining sentiment.

Furthermore, while SMOTE successfully addresses the imbalance by generating synthetic samples, it is important to acknowledge its limitations. The synthetic data points created through interpolation may not fully reflect the complexity of real-world data, particularly when the original dataset is small or lacks diversity. This can potentially lead to overfitting, where the model performs well on the training data but struggles with unseen data. Future improvements could involve refining the dataset further, enhancing feature engineering to better distinguish sentiment classes, or exploring advanced models that can capture dependencies between features more effectively.

4.6. Recommendations for Improvement

To further enhance the model's performance, several strategies can be considered to address the remaining challenges and improve the classification outcomes. First, feature engineering can play a critical role in refining the input data. By introducing or improving features that better differentiate between sentiment classes, particularly for Class 1 (positive sentiment), the model can better capture the nuanced patterns that distinguish positive opinions from neutral or overlapping sentiments. Techniques such as word embeddings, bigram or trigram analysis, and sentiment-specific lexical features could significantly improve feature representation.

Second, exploring alternative balancing methods may yield further improvements. While SMOTE effectively addresses class imbalance, advanced resampling techniques like ADASYN (Adaptive Synthetic Sampling) or SMOTE-ENN (SMOTE combined with Edited Nearest Neighbors) could generate more realistic synthetic samples by focusing on the regions of the dataset that are harder to classify. These methods not only balance the class distribution but also mitigate the risk of overfitting by removing noisy or redundant synthetic data points.

Third, implementing advanced machine learning models could enhance the model's ability to handle feature dependencies and complex relationships in the data. For instance, Random Forest and Support Vector Machines (SVM) are robust classifiers that can better manage non-linear boundaries between classes. Additionally, deep learning approaches such as LSTM (Long Short-Term Memory) or transformers can leverage contextual relationships in text data, enabling the model to understand word dependencies and improve sentiment classification performance. Lastly, conducting comparative studies between the Naive Bayes model and other classifiers on the same dataset will provide valuable insights into the most effective approaches. By benchmarking models based on accuracy, recall, and other evaluation metrics, the optimal technique for handling class imbalance and sentiment classification can be identified.

4.7. Real-World Implications

The findings of this study have practical implications for understanding student perceptions of AI in education. By accurately capturing the full range of sentiments—positive, neutral, and negative—institutions can make data-driven decisions to address concerns, improve AI tools, and ensure their acceptance among students. Balanced sentiment analysis provides a comprehensive understanding of how students perceive AI, which is critical for optimizing its implementation and maximizing its benefits.

4. CONCLUSION

This study successfully explored students' sentiments regarding the utilization of artificial intelligence (AI) in education by implementing the Naive Bayes model optimized with SMOTE to address class imbalance. The findings demonstrated that SMOTE effectively balanced the dataset, enabling the model to better learn from all sentiment classes and reducing its bias toward the majority class. This was evident from the improvement in the Naive Bayes model's accuracy, which increased from 65% to 78.87% after applying SMOTE, indicating that the model became more sensitive to identifying patterns in minority classes and achieved more consistent classification performance across all sentiment classes. Further evaluation using precision, recall, and F1-score metrics revealed that the model performed well in Classes 2 and 4, although Class 1 continued to exhibit weaknesses in classification accuracy. These limitations suggest that the data patterns in Class 1 may be more complex or challenging for the Naive Bayes model to learn, leading to suboptimal classification outcomes for this class. Nevertheless, the application of SMOTE significantly enhanced the model’s overall accuracy, despite the potential risk of overfitting when the original dataset is very limited. Therefore, careful consideration is necessary when applying SMOTE, ensuring that there is an adequate amount of real data to maintain model stability. The practical implications of this research are critical for the development of AI-based education. Understanding student sentiment allows policymakers and technology developers to design AI systems and programs that are more responsive to students' needs and concerns. Overall, this study highlights that Naive Bayes combined with SMOTE is an effective approach for classifying sentiments in imbalanced datasets. Future research is encouraged to utilize more complex models and larger datasets to achieve more representative results and mitigate the risk of overfitting caused by synthetic samples. This research provides a solid foundation for further studies on student perceptions of AI and its potential applications in education.

REFERENCES

[1] H. Lin, “Influences of artificial intelligence in education on teaching effectiveness: The mediating effect of teachers' perceptions of educational technology,” *Int. J. Emerg. Technol. Learn.*, vol. 17, 2022. [Online]. Available: https://doi.org/10.3991/ijet.v17i24.36037

[2] L. Chen, P. Chen, and Z. Lin, “Artificial intelligence in education: A review,” *IEEE Access*, vol. 8, pp. 75264–75278, 2020. [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2988510

[3] L. Zheng, J. Niu, L. Zhong, and J. F. Gyasi, “The effectiveness of artificial intelligence on learning achievement and learning perception: A meta-analysis,” *Interact. Learn. Environ.*, vol. 31, pp. 5650–5664, 2021. [Online]. Available: https://doi.org/10.1080/10494820.2021.2015693

[4] N. Abbas, I. Ali, R. Manzoor, T. Hussain, and M. H. A. Hussain, “Role of artificial intelligence tools in enhancing students' educational performance at higher levels,” *J. AI Mach. Learn.*, pp. 35–49, 2023. [Online]. Available: https://doi.org/10.55529/jaimlnn.35.36.49

[5] J. Huang, S. Saleh, and Y. Liu, “A review on artificial intelligence in education,” *Acad. J. Interdiscip. Stud.*, vol. 10, p. 206, 2021. [Online]. Available: https://doi.org/10.36941/AJIS-2021-0077

[6] F. Kamalov, D. S. Calonge, and I. Gurrib, “New era of artificial intelligence in education: Towards a sustainable multifaceted revolution,” *Sustainability*, 2023. [Online]. Available: https://doi.org/10.3390/su151612451

[7] B. Pardamean, T. Suparyanto, T. W. Cenggoro, D. Sudigyo, and A. Anugrahana, “AI-based learning style prediction in online learning for primary education,” *IEEE Access*, 2022. [Online]. Available: https://doi.org/10.1109/ACCESS.2022.3160177

[8] V. Chaudhri, H. Lane, D. Gunning, and J. Roschelle, “Intelligent learning technologies: Applications of artificial intelligence to contemporary and emerging educational challenges,” *AI Mag.*, vol. 34, pp. 10–12, 2013. [Online]. Available: https://doi.org/10.1609/aimag.v34i3.2482

[9] Y. Liu, L. Chen, and Z. Yao, “The application of artificial intelligence assistant to deep learning in teachers' teaching and students' learning processes,” *Front. Psychol.*, vol. 13, 2022. [Online]. Available: https://doi.org/10.3389/fpsyg.2022.929175

[10] Z. Slimi, “The impact of artificial intelligence on higher education: An empirical study,” *Eur. J. Educ. Sci.*, vol. 10, 2023. [Online]. Available: https://doi.org/10.19044/ejes.v10no1a17

[11] Akavova, Z. Temirkhanova, and Z. Lorsanova, “Adaptive learning and artificial intelligence in the educational space,” *E3S Web Conf.*, 2023. [Online]. Available: https://doi.org/10.1051/e3sconf/202345106011

[12] X. Wei, S. Sun, D. Wu, and L. Zhou, “Personalized online learning resource recommendation based on artificial intelligence and educational psychology,” *Front. Psychol.*, vol. 12, 2021. [Online]. Available: https://doi.org/10.3389/fpsyg.2021.767837

This Journal is licensed under a Creative Commons Attribution 4.0 International License

[13] M.-L. How, “Future-ready strategic oversight of multiple artificial superintelligence-enabled adaptive learning systems via human-centric explainable AI-empowered predictive optimizations of educational outcomes,” *Big Data Cogn. Comput.*, vol. 3, p. 46, 2019. [Online]. Available: https://doi.org/10.3390/BDCC3030046

[14] W. Bagunaid, N. Chilamkurti, and P. Veeraraghavan, “AISAR: Artificial intelligence-based student assessment and recommendation system for e-learning in big data,” *Sustainability*, vol. 14, 2022. [Online]. Available: https://doi.org/10.3390/su141710551

[15] L.-Y. Lin and F. Wang, “Adaptive learning system based on knowledge graph,” in *Proc. 9th Int. Conf. Educ. Train. Technol.*, 2023. [Online]. Available: https://doi.org/10.1145/3599640.3599647

[16] D. S. Chaplot, E. Rhim, and J. Kim, “Personalized adaptive learning using neural networks,” in *Proc. ACM Conf. Learn. Scale*, 2016. [Online]. Available: https://doi.org/10.1145/2876034.2893397

[17] M. Ciolacu, A. Tehrani, L. Binder, and P. Svasta, “Education 4.0 - Artificial intelligence-assisted higher education: Early recognition system with machine learning to support students' success,” in *IEEE Int. Symp. Design Technol. Electron. Packag.*, 2018, pp. 23–30. [Online]. Available: https://doi.org/10.1109/SIITME.2018.8599203

[18] D. Xuan, D.-Z. Zhu, and W. Xu, “The teaching pattern of law majors using artificial intelligence and deep neural network under educational psychology,” *Front. Psychol.*, vol. 12, 2021. [Online]. Available: https://doi.org/10.3389/fpsyg.2021.711520

[19] H. O. Koch Delgado, A. De Azevedo Fay, M. J. Sebastiany, and A. Da Costa Silva, “Artificial intelligence adaptive learning tools,” *BELT Braz. Engl. Lang. Teach. J.*, vol. 11, 2020. [Online]. Available: https://doi.org/10.15448/2178-3640.2020.2.38749

[20] Akavova, Z. Temirkhanova, and Z. Lorsanova, “Adaptive learning and artificial intelligence in the educational space,” *E3S Web* Conf., 2023. [Online]. Available: https://doi.org/10.1051/e3sconf/202345106011

[21] N. D. Bation and J. C. Pudan, "Exploring the Correlation between Students' Attitudes towards AI and Their Learning Outcomes," *International Journal of Social Science and Human Research*, vol. 7, no. 2, pp. 1–12, 2024. DOI: 10.47191/ijsshr/v7-i02-45.

[22] M. Irfan, L. Murray, and S. Ali, "Insights into Student Perceptions: Investigating Artificial Intelligence (AI) Tool Usability in Irish Higher Education at the University of Limerick," *Global Digital & Print Media Review*, vol. VI, no. II, pp. 1–10, 2023. DOI: 10.31703/gdpmr.2023(vi-ii).05.

[23] S. Sanasintani, "Revitalizing The Higher Education Curriculum Through An Artificial Intelligence Approach: An Overview," *Journal of Social Science Utilizing Technology*, vol. 1, no. 4, pp. 1–15, 2023. DOI: 10.55849/jssut.v1i4.670.

This Journal is licensed under a Creative Commons Attribution 4.0 International License