Sentiment and Emotion Classification Models Using Hybrid Textual and Numerical Features: A Case Study of Mental Health Counseling

Indri Ramayanti[1], Latius Hermawan[2,*], Rizma Adlia Syakurah[3], Deris Stiawan[4], Meilinda[5], Edi Surya Negara[6], Muhammad Fahmi[7], Ahmad Ghiffari[8], Muhammad Qurhanul Rizqie[9]

[1,8]*Department of Medical, Universitas Muhammadiyah Palembang, Indonesia*

[2]*Department of Doctoral Engineering Sciences, Universitas Sriwijaya, Indonesia*

[2]*Department of Informatics, Universitas Katolik Musi Charitas, Indonesia*

[3]*Department of Public Health, Universitas Sriwijaya, Indonesia*

[4,9]*Department of Computer Science, Universitas Sriwijaya, Indonesia*

[5]*Department of Teaching and Education Science, Universitas Sriwijaya, Indonesia*

[6]*Department of Informatics Engineering, Universitas Bina Darma, Indonesia*

[7]*Department of Accounting, Universitas Muhammadiyah Palembang, Indonesia*

(Received: February 1, 2025; Revised: April 1, 2025; Accepted: May 1, 2025; Available online: June 2, 2025)

Abstract

Understanding emotions during mental health counseling is important to deliver personalized and timely support. This study explores the use of machine learning to classify emotions in counseling conversations by combining two types of features: textual features (such as the words used) and numerical features (such as the number of characters or sentences in a response). The data includes 873 data, which were processed using common text-cleaning methods like stemming, stopword removal, and TF-IDF to capture important word patterns. At the same time, numerical features were added to reflect the structure of each response. Four machine learning models were tested: Bernoulli Naive Bayes, Decision Tree, Logistic Regression, and Random Forest. These models aimed to classify emotional categories: Depression, Happy, and Stress. After training and testing, Random Forest showed the best performance with 90.05% accuracy, followed closely by Logistic Regression (89.41%) and Bernoulli Naive Bayes (88.89%). The Decision Tree model had the lowest accuracy at 72.49%, which may be due to its limitations in dealing with complex data. These results show that combining text-based and numerical features helps improve the accuracy of emotion detection in mental health conversations. The findings also suggest that machine learning models, when properly designed, can support the development of AI-based tools for mental health support. This approach can be used as a foundation for building smarter chatbot systems that are able to recognize emotions and provide more empathetic responses in real-time. The study adds new insights into how artificial intelligence can be used to improve access to mental health services and make them more adaptive to users' emotional needs

Keywords: Sentiment Analysis, Emotion, Classification, Mental Health, Counseling

1. Introduction

Mental health challenges are a growing crisis, especially among university students. While offering opportunities for academic and personal growth, university life also brings unique stressors, including academic pressure, financial burdens, social dynamics, and career uncertainties. These stressors often lead to an increase in mental health issues such as depression, anxiety, and stress, which can significantly affect students' academic performance, social relationships, and overall well-being. According to the World Health Organization (WHO), untreated mental health disorders are among the leading causes of disability and poor quality of life globally, particularly among young adults. In academic settings, mental health issues can lead to absenteeism, academic underperformance, and higher dropout

*Corresponding author: Latius Hermawan (tiuz.hermawan@ukmc.ac.id)
DOI: https://doi.org/10.47738/jads.v6i3.632

rates, exacerbating the challenges faced by higher education institutions [1]. The increasing demand for mental health support on university campuses has strained existing resources, with counseling centers often overwhelmed by the volume of students seeking help. Many universities face significant barriers in delivering timely and effective support due to limited staff, financial constraints, and the stigma surrounding mental health, which discourages students from seeking help. These challenges underscore the urgent need for innovative, scalable solutions to enhance mental health services. One promising avenue is using artificial intelligence (AI) technologies, particularly sentiment analysis, to augment traditional mental health interventions.

Sentiment analysis, a subset of natural language processing (NLP), involves the computational assessment of text to identify the underlying emotions or sentiments expressed. This technology has shown great potential in mental health care, enabling practitioners to analyze large volumes of unstructured data, such as counseling transcripts, surveys, and social media posts. By identifying emotional patterns, sentiment analysis can help detect students experiencing distress and facilitate early intervention. For example, negative emotional trends in student communications can be flagged for follow-up, enabling counselors to focus on those most at risk [2]. Machine learning models are foundational to the success of sentiment analysis systems. Logistic Regression, a widely used algorithm, is particularly suitable for academic mental health applications due to its simplicity, interpretability, and effectiveness in binary and multiclass text classification tasks. Logistic Regression can efficiently identify emotional states, such as sadness or anger, from text data, making it a valuable tool for analyzing counseling records or surveys. On the other hand, ensemble models like Random Forest have proven highly effective in handling complex, imbalanced datasets, which are common in mental health contexts. By aggregating the decisions of multiple decision trees, Random Forest provides robust and accurate predictions, capturing subtle emotional nuances that simpler models may miss [3].

Beyond these models, hybrid feature integration has emerged as a powerful approach to improving the precision and utility of sentiment analysis in academic settings. Combining linguistic features, such as text and numeric datasets, with behavioral data, such as help-seeking frequency or participation in university activities, enhances the ability of these systems to detect emotional distress. Additionally, incorporating multimodal data—such as text from counseling sessions, audio recordings, and physiological signals—provides a more comprehensive understanding of a student's mental health. This multimodal approach enables tailored interventions that address mental health challenges' emotional and behavioral dimensions [4][5]. The potential applications of sentiment analysis extend far beyond traditional counseling. For instance, this technology can be used to monitor emotional trends across the student body, providing administrators with real-time insights into the collective mental health of the campus community. Such data can inform institutional policies, allowing universities to allocate resources strategically, prioritize interventions, and design programs that address the root causes of student stress. Additionally, automated sentiment analysis systems can serve as early-warning tools, identifying at-risk students who may not actively seek help due to stigma or other barriers [6].

This research seeks to explore the role of sentiment analysis in enhancing mental health care for university students. By leveraging AI technologies, this study aims to provide a framework for scalable, data-driven mental health interventions that address the unique challenges of academic environments. Integrating sentiment analysis into university counseling services can ultimately transform mental health care, fostering a culture of resilience, well-being, and educational success on campuses worldwide.

2. Research Method

The dataset for this study was curated to capture textual data reflective of emotional expressions in university mental health counselling contexts. Sources include anonymized transcripts from counselling sessions, survey responses and ensuring a diverse representation of linguistic patterns and emotional nuances. To uphold ethical standards, participant consent was obtained before capture the data, and strict anonymization protocols were implemented to safeguard privacy. Measures such as removal of personally identifiable information and secure data storage were enforced to maintain confidentiality while preserving the integrity of the collected data.

2.1. Dataset Collection

This study utilizes a dataset of 873 responses from Indonesian university students participating in mental health counseling sessions. The dataset includes textual features such as response content, emotion labels, number of characters (num_of_characters), and number of sentences (num_of_sentences). The num_of_characters feature captures response length, where longer texts may indicate elaborate emotional expressions, while shorter ones may suggest strong or certain emotions. Similarly, num_of_sentences reflect response complexity, with higher counts often linked to distress or deep thought. These features were normalized and combined with TF-IDF vectorized text data, allowing machine learning models to better recognize patterns in emotional expression through text length and structure.

The dataset focuses on academic-related mental health challenges, including stress, happiness, and depression, with each response carefully annotated with emotional labels. A supervised learning approach was applied to classify emotions into stress, happiness, and depression leveraging pre-trained models and domain-specific feature engineering. Depression, happiness, and stress are commonly studied emotions in clinical psychology and psychopathology. Research indicates that depression is linked to difficulties in differentiating negative emotions, particularly sadness, guilt, and anxiety [7]. While emotions like disappointment and confusion are not explicitly categorized in diagnostic manuals such as the DSM-5, they are frequently included in broader psychological models of emotion recognition and psychopathology [8]. Additionally, emotions such as depression, stress, and pleasure are widely recognized in mental health research and are measured using psychological rating scales like the Beck Depression Inventory (BDI) and the Positive and Negative Affect Schedule (PANAS) [9]. This structured dataset plays a crucial role in advancing emotion detection research, particularly in academic settings, where early identification of emotional distress can facilitate timely mental health interventions and prevent long-term consequences. [10].

Figure 1 provides an illustration of preprocessed counseling conversation samples categorized under the emotional label "Confused." The original statements, initially in Indonesian, reflect participants' expressions of uncertainty, hesitation, and lack of clarity in various contexts. For instance, one statement, "Iya, aku paham. Aku harus belajar menerima..." translates to "Yes, I understand. I must learn to accept...," indicating a cognitive struggle with acceptance. Another example, "Kadang aku terlalu ragu dengan kemampuan..." ("Sometimes I am too hesitant with my abilities..."), highlights self-doubt regarding personal capabilities. This data structure ensures that emotional nuances are preserved while preparing the text for computational analysis.

	original_statement	emotion	num_of_characters	num_of_sentences	statement	tokens
0	Iya, aku paham. aku harus belajar menerima bah...	Confused	133	3	iya aku paham aku harus belajar menerima bahwa...	[iya, aku, paham, aku, harus, belajar, menerim...
1	Iya, benar. Kadang aku terlalu ragu dengan kem...	Confused	166	2	iya benar kadang aku terlalu ragu dengan kemam...	[iya, benar, kadang, aku, terlalu, ragu, denga...
2	Jarang ikut seminar, tapi pernah. Kalau di ana...	Confused	97	2	jarang ikut seminar tapi pernah kalau di anali...	[jarang, ikut, seminar, tapi, pernah, kalau, d...
3	Biasanya aku kasih tahu pematerinya kalau ada ...	Confused	65	1	biasanya aku kasih tahu pematerinya kalau ada ...	[biasanya, aku, kasih, tahu, pematerinya, kala...
4	Ya, ada yang ketawa, tapi biasanya efeknya cum...	Confused	102	2	ya ada yang ketawa tapi biasanya efeknya cuma ...	[ya, ada, yang, ketawa, tapi, biasanya, efekny...

Figure 1. Hybrid Textual and Numerical Dataset Sentiment Analysis

The data collection process adhered to ethical considerations, including anonymization and participant consent, ensuring compliance with data privacy regulations. Responses were derived from academic institutions' mental health programs, where participants voluntarily engaged in counseling to address their concerns. Similar studies, such as [11], have shown that such datasets can offer a robust foundation for developing AI-driven tools to enhance emotional support systems in educational settings [11]. The significance of emotion-labeled datasets is highlighted in research by [12], demonstrating how fine-grained emotional annotations can enhance the accuracy of predictive models in health-related contexts. Similarly, [13] underscores the importance of contextual emotion detection using supervised machine learning, noting its relevance in early mental health assessments [12], [13].

In academic mental health, datasets such as the one used in this research offer unique insights into students' emotional states. These insights can inform the design of interventions tailored to specific emotional patterns, ensuring more personalized and practical support. As noted by [14], using such datasets can contribute to developing advanced detection systems for psychological stress, ultimately enhancing the well-being of students in higher education environments.

2.2. Workflow Model Building

Figure 2 shows a diagram representing a systematic workflow for data preparation and model building in machine learning. The workflow is specifically tailored to ensure high-quality data and robust model performance. The pipeline comprises six main stages interconnected to optimize the data and modeling process.

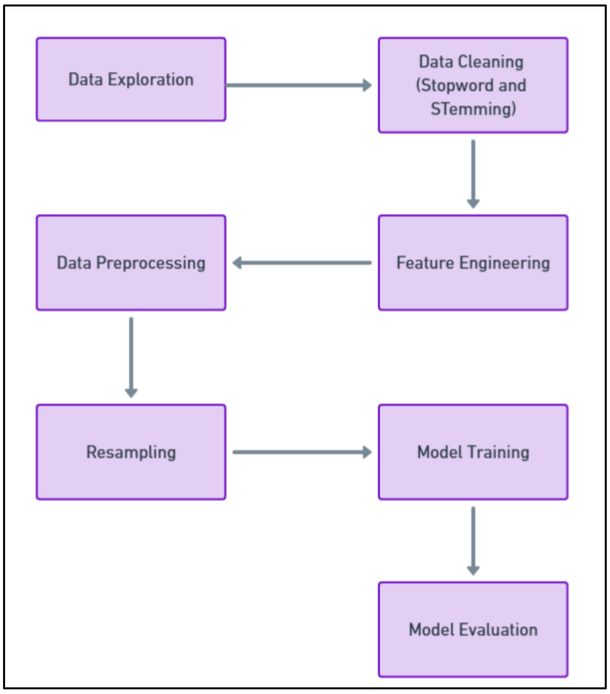

Figure 2. Workflow Model in Sentiment Analysis

Data Exploration: This initial step involves analyzing the dataset to understand its structure, distribution, and key characteristics. Exploratory Data Analysis (EDA) techniques, including visualization, summary statistics, and missing value analysis, help identify potential issues like outliers and data imbalances, which are crucial for planning subsequent preprocessing steps [15]. Data Cleaning: This step ensures data quality and suitability for modeling by removing irrelevant data, addressing missing values, and applying text preprocessing. Techniques such as stopword removal eliminate common but uninformative words (e.g., "the," "is") [16], while stemming reduces words to their root forms, ensuring consistency in text-based features. Feature Engineering: Raw data is transformed into meaningful features to enhance model performance. This involves creating, selecting, or encoding variables based on their relevance to the classification task. Effective feature engineering is especially critical for high-dimensional or sparse data, as it directly impacts model accuracy [17].

Data Preprocessing and Resampling: Standardization methods such as scaling, encoding categorical variables, and imputing missing values prepare the dataset for modeling. Resampling techniques, including oversampling minority classes or undersampling majority classes, help address class imbalance, ensuring that machine learning models generalize well and minimize bias [18]. Model Training: Machine learning algorithms such as Random Forest or Logistic Regression are applied to train predictive models. This phase includes parameter tuning to optimize performance based on the dataset’s characteristics [19]. Model Evaluation: The trained model is assessed using accuracy, precision, recall, and F1-score, ensuring reliability and identifying areas for improvement. Cross-validation is commonly used to evaluate robustness and prevent overfitting, making the model more generalizable to unseen data.

3. Machine Learning for Sentiment Analysis

3.1. Machine Learning Process

This study utilized a dataset of 873 records, split into 80% training (698 records) and 20% testing (175 records) to ensure a balanced evaluation of machine learning models. The training set was used for learning patterns, while the testing set provided an unbiased assessment of model performance [20], [21]. Four machine learning algorithms—Bernoulli Naive Bayes, Decision Tree, Logistic Regression, and Random Forest—were applied due to their effectiveness in text classification and emotion detection. The process included text preprocessing (stopword removal, stemming, and tokenization) and feature transformation using bag-of-words and TF-IDF vectorization [16]. Feature engineering was also employed to enhance sentiment classification by ensuring meaningful text representations, as shown in figure 3.

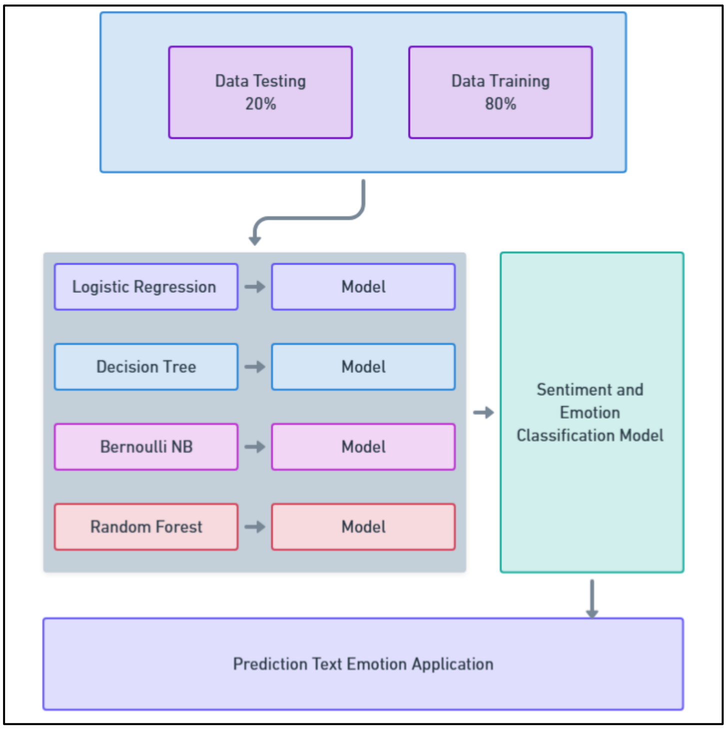

Figure 3. Sentiment Analysis Framework

This study employed a qualitative-quantitative hybrid research design to explore sentiment analysis and emotion detection in mental health counseling conversations. Data were collected from recorded counseling sessions, obtained with informed consent from counselors and clients, ensuring ethical compliance. The recorded audio was transcribed into text, focusing solely on the client's responses to preserve the context while maintaining participant confidentiality by anonymizing all identifiable information. The text data were then preprocessed through cleaning techniques such as stopword removal and stemming to enhance consistency and usability for analysis. Following this, the dataset underwent labeling, where emotional states such as happiness, sadness, anxiety, and anger were assigned to client responses to create a structured and annotated dataset. Machine learning models, including Random Forest and Logistic Regression, were trained and evaluated using this dataset to determine their effectiveness in identifying emotions.

Each algorithm was implemented and trained on the labeled dataset. BernoulliNB, a variant of the Naive Bayes classifier, was utilized for its ability to handle binary feature inputs effectively. It calculates probabilities based on the presence or absence of features, making it particularly well-suited for text data structured as binary vectors [22]. The Decision Tree algorithm was employed for its interpretability and capacity to handle categorical and numerical data. It uses a tree-like model of decisions and their possible consequences, including classifications of text data based on hierarchical rules derived from the features [23]. Logistic Regression was chosen for its simplicity and effectiveness in binary classification tasks. This algorithm estimates probabilities based on a linear combination of the input features, offering interpretable coefficients [24] that can provide insights into the significance of different textual features in determining sentiment. Random Forest, an ensemble learning algorithm, was included due to its robustness and high predictive accuracy. By aggregating the outputs of multiple decision trees, Random Forest reduces overfitting and improves generalization, making it ideal for datasets with complex or imbalanced class distributions [25].

The performance of these algorithms was evaluated using metrics such as accuracy, precision, recall, and F1-score to compare their effectiveness in detecting emotions from text. Random Forest and Logistic Regression were found to excel in accuracy and interpretability. At the same time, BernoulliNB and Decision Tree demonstrated strengths in specific scenarios, such as handling binary data or generating transparent decision paths. By implementing and analyzing these algorithms, this study highlights the potential of machine learning in sentiment analysis, providing valuable insights into emotion detection within mental health counseling contexts [16] [8]. The findings contribute to the growing research on AI-driven solutions for enhancing mental health interventions.

3.2. Data Exploration

The data exploration phase aimed to comprehensively understand and prepare the dataset for further analysis, ensuring its suitability for sentiment analysis and emotion detection tasks. This process began with an initial examination of the dataset to understand its structure, including the number of records (rows), features (columns), and labels representing different emotional states. This analysis identified key attributes and any potential challenges within the dataset. Handling missing data was a critical task during this phase. Any missing values were identified and removed by systematically scanning the dataset to maintain data integrity. This ensured that the dataset used in subsequent analysis was clean and free from inconsistencies. Removing missing data also reduced the likelihood of biases or errors during the modeling phase.

A key focus of this phase was analyzing the distribution of labels in the dataset, as class balance plays a significant role in machine learning model performance. The total number of occurrences for each emotional category, such as "Depression," "Happy," "Depression," was computed and visualized in the form of a pie chart (as shown in the figure 4). The visualization revealed the proportions of each category, providing critical insights into the dataset's composition as shown in figure 4. The pie chart illustrates that "Happy" is the most frequently occurring emotion, representing 34.9% of the dataset, followed by "Depression" (33.2%), and "Stress" (31.8%). This distribution demonstrates moderately balanced dataset, with all emotions represented in relatively equal proportions.

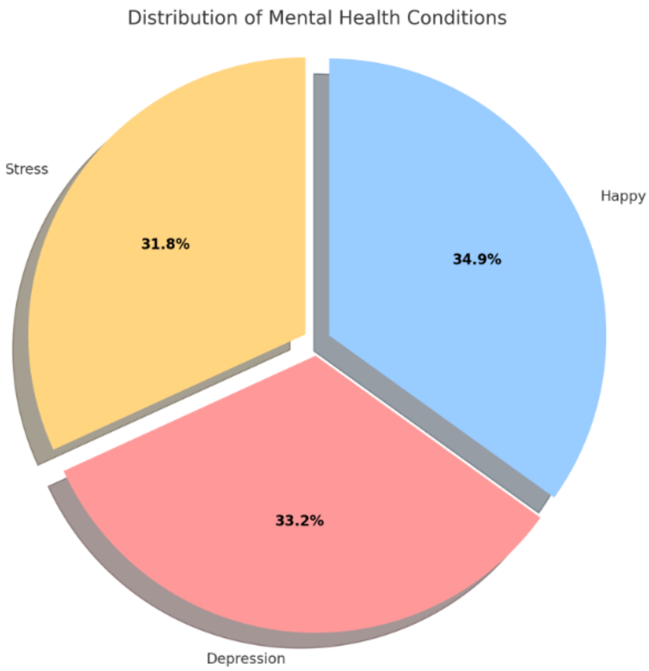

Figure 4. Data Exploration

Overall, the data exploration phase and the visualization of label distribution provided a detailed understanding of the dataset's structure and composition. These insights played a pivotal role in guiding preprocessing and model

ISSN 2723-6471

development, ensuring that the subsequent stages of the research were built on a solid foundation. The detailed analysis and visualization ensured transparency and informed decision-making throughout the process.

3.3. Data Cleaning

The data cleaning phase involved several critical preprocessing steps to prepare the dataset for sentiment analysis and emotion detection. This process began with stemming using the Sastrawi library, a widely used tool for Indonesian text processing [24]. Stemming reduced words to their root forms, helping to standardize the vocabulary by eliminating variations of the same word. For instance, words like "berlari (running)," and "pelari (runner)" were reduced to their root form "lari (run)," ensuring consistency across the dataset. In addition to stemming, stopword removal was performed using a predefined list of Indonesian stopwords from the Sastrawi library [26]. These stopwords, including "dan" (and), "yang" (which), "di" (at), and "adalah" (is), were removed as they carried little semantic meaning and were often irrelevant for sentiment analysis. The removal of stopwords helped focus on meaningful terms, improving the quality of feature extraction for machine learning models.

To assess the impact of preprocessing, a sample text before and after stemming and stopword removal is provided in figure 5. These steps were crucial in enhancing data quality by ensuring a more structured and meaningful representation of textual features, which ultimately improved the performance of emotion classification models.

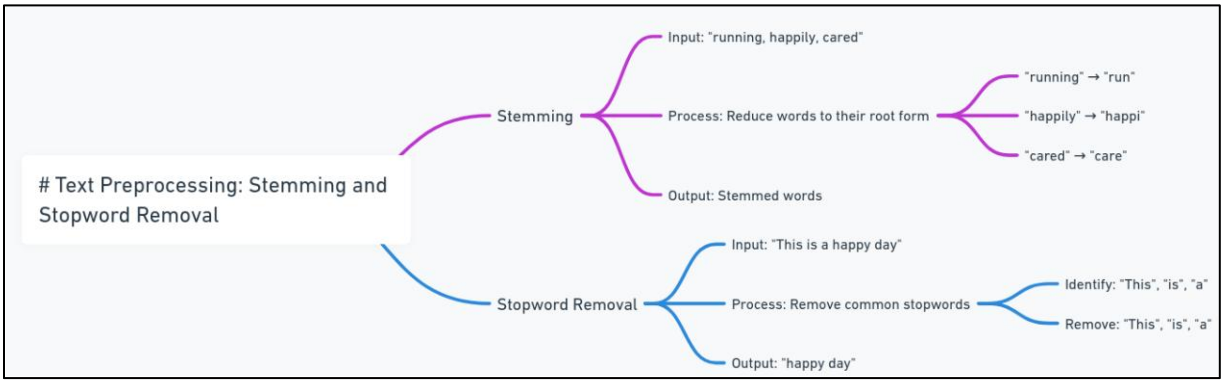

Figure 5. Stemming and Stopwords Process

A word cloud visualization was generated for each emotion label using the WordCloud library to gain further insights into the data. The word cloud provided a graphical representation of the most frequently occurring words for each emotional category, with the size of each word indicating its frequency. For example, in the "Anxious" category, common words such as "worry," "anxious," and "scared" were prominently displayed, reflecting the linguistic patterns associated with this emotion. Similarly, the "Depression" category, as shown in Figure 6, featured favorable terms like "scared" and "hard," emphasizing the words that frequently appeared in depressional contexts.

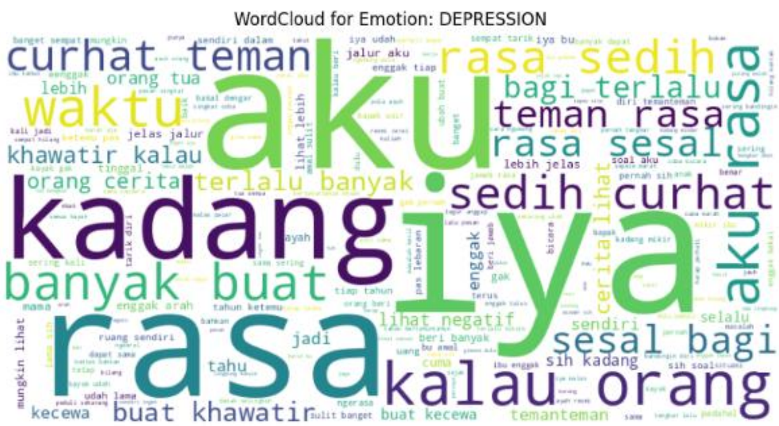

Figure 6. Depression WordCloud Example

The word cloud for "Depression" provides valuable insights into the common words and themes associated with depressive emotions in text data. Key words such as "aku" (I), "kadang" (sometimes), "rasa" (feeling), "iya" (yes),

ISSN 2723-6471

"sedih" (sad), "curhat" (venting), and "khawatir" (worried) highlight the core emotions and thoughts of individuals experiencing depression. This combination of stemming, stopword removal, and word cloud visualization provided a deeper understanding of the dataset and the unique linguistic characteristics of each emotional category. These steps cleaned and standardized the text and revealed valuable insights into the vocabulary associated with different emotions, supporting the subsequent feature engineering and model training phases. By leveraging these preprocessing techniques, the data cleaning process ensured that the dataset was high-quality and representative of the analyzed emotional states.

3.4. Data Encoding and Vectorizer

To prepare the cleaned text data for machine learning models, the textual data was converted into numerical features using the TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer. TF-IDF is a widely used technique in natural language processing (NLP) to represent text data in a way that highlights essential terms while reducing the impact of common words that appear across multiple documents [16]. This method ensures that the feature representation captures the relevance of terms to specific documents, making it particularly effective for tasks like sentiment analysis and emotion detection. For this study, the TF-IDF vectorizer transformed the preprocessed text into a numerical matrix, where each row corresponded to a document, as shown in Figure 7 (e.g., a client's response in counseling), and each column represented a term from the vocabulary.

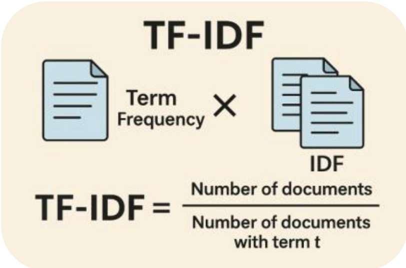

Figure 7. TF*IDF Algorithm

The values in the matrix indicated the TF-IDF score for each term in the respective document. This numerical representation allowed machine learning algorithms like Logistic Regression and Random Forest to process and analyze the data effectively. The text preprocessing phase involved stemming, stopword removal, and TF-IDF vectorization to enhance emotion classification. Stemming (using the Sastrawi library) standardized words to their root forms, reducing redundancy (e.g., "berlari," "lari," and "berlarian" → "lari") [27]. Stopword removal eliminated common words ("yang," "dan," "untuk") to improve feature extraction. Impact Analysis using word clouds confirmed that essential emotional indicators were preserved after preprocessing. TF-IDF Vectorization assigned higher weights to emotionally significant words while down-weighting frequent but uninformative words.

These steps improved model accuracy by ensuring that only relevant terms influenced classification. Future refinements may include lemmatization and domain-specific stopword lists for better feature selection. This approach captured the semantic meaning of the text, improving the model's ability to differentiate between emotional categories.
By applying TF-IDF vectorization, the textual data was effectively transformed into a structured format, paving the way for accurate and efficient emotion detection in the dataset. This process ensured that the models could capture the linguistic nuances of each emotion category, contributing to robust and reliable predictions.

4. Results and Discussion

The results of this study demonstrate the effectiveness of machine learning models in detecting emotions from counseling conversations, with varying degrees of performance across the four algorithms evaluated. Logistic Regression and Random Forest emerged as the most robust models, achieving the highest accuracy and F1 scores, attributed to their ability to handle complex patterns and imbalanced datasets effectively. Random Forest, in particular, excelled in capturing subtle emotional nuances due to its ensemble learning approach, while Logistic Regression offered interpretability, making it valuable for practical applications in counseling settings. Bernoulli Naive Bayes, while efficient for binary data, showed limitations in handling the complexity of emotional expressions, whereas the

Decision Tree model, though interpretable, faced challenges with overfitting on smaller datasets. The findings underscore the importance of feature engineering and dataset quality in emotion detection, as preprocessing steps like stemming and stopword removal significantly improved performance across all models. Furthermore, the discussion highlights the practical implications of deploying these models in mental health counseling, emphasizing their potential to support counselors by providing early insights into clients' emotional states. These results contribute to advancing sentiment analysis applications in mental health and open avenues for integrating multimodal data to enhance predictive accuracy in future research.

4.1. Model and Evaluation

The classification models were evaluated by comparing the accuracy of four machine learning algorithms: Bernoulli Naive Bayes (BernoulliNB), Decision Tree, Logistic Regression, and Random Forest, as visualized in the bar chart. The chart illustrates the accuracy scores achieved by each model after hyperparameter tuning to optimize their performance on the emotion detection task. Hyperparameter tuning was performed to optimize Bernoulli Naive Bayes, Decision Tree, Logistic Regression, and Random Forest models. Key adjustments included Laplace smoothing for Naive Bayes, depth and split constraints for Decision Tree and Random Forest, and L1 regularization for Logistic Regression. These optimizations aimed to balance bias and variance, enhancing model generalization and predictive accuracy. Model Performance and Simulation, as shown in table 1, figure 8 and figure 9

In the context of emotion detection, evaluation metrics such as accuracy, precision, recall, and F1-score provide deeper insights into model performance beyond overall correctness. Random Forest achieved the highest scores across all metrics, making it the most reliable model for capturing complex emotional patterns. Logistic Regression followed closely, offering competitive performance with the advantage of interpretability. Bernoulli Naive Bayes performed well, particularly in balancing precision and recall, making it a good choice for computational efficiency. Decision Tree had the lowest scores, indicating limitations in handling nuanced emotional distinctions. These findings suggest that Random Forest is the best model for high accuracy, while Logistic Regression remains a strong alternative for interpretable applications, depending on the trade-offs required for specific use cases as shown in table 1.

Table 1. Confusion Matrix Result

Model	Accuracy	Precision	Recall	F1-Score
BernoulliNB	0.88889	0.891727	0.88889	0.88832
DecisionTree	0.72486	0.734118	0.72486	0.72745
LogisticRegression	0.89418	0.898084	0.89418	0.89365
RandomForest	0.90047	0.900029	0.90047	0.90000

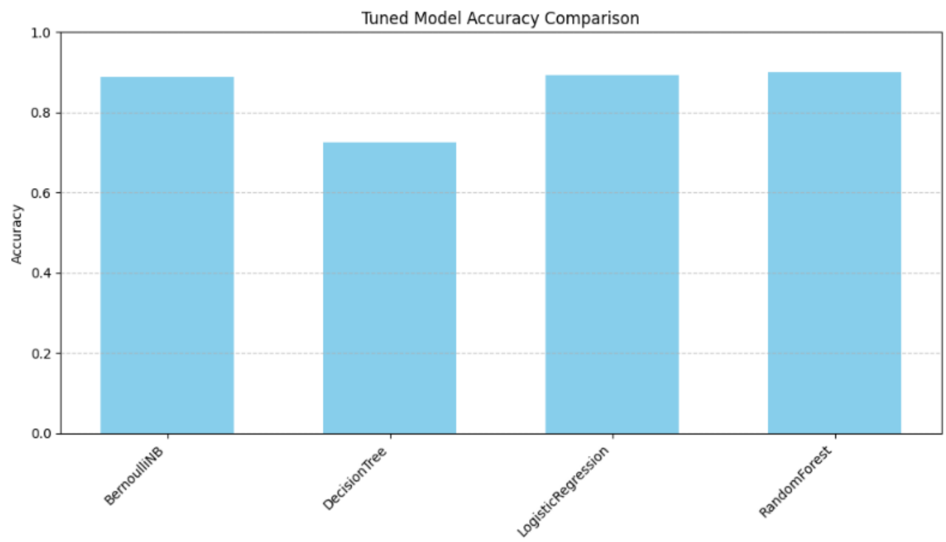

Figure 8. Model and Evaluation


```
••• Emotion Prediction System
    Type 'exit' to stop.
    Enter text: saya merasa tidak berguna
    Input Text: saya merasa tidak berguna
    Predicted Emotion: DEPRESSION
    Enter text: saya harus semangat
    Input Text: saya harus semangat
    Predicted Emotion: STRESS
    Enter text: Iya, aku akan berjuang
    Input Text: Iya, aku akan berjuang
    Predicted Emotion: HAPPY
    Enter text: nilaiku jelek
    Input Text: nilaiku jelek
    Predicted Emotion: STRESS
    Enter text:
```

Figure 9. Sentiment Analysis Testing

The Random Forest model outperformed the other models, achieving the highest accuracy of approximately 90%. Its strong performance is attributed to its ability to effectively capture complex relationships between features and output classes, particularly for datasets with non-linearly separable patterns. The ensemble learning approach, which aggregates predictions from multiple decision trees, enhances its robustness and reduces overfitting compared to a single Decision Tree. However, Random Forest can be computationally expensive, making it less suitable for real-time applications. The Logistic Regression model achieved an accuracy of approximately 89%, slightly lower than Random Forest but higher than Bernoulli Naive Bayes and Decision Tree.

Logistic Regression excels in interpretability and scalability, making it an ideal choice for datasets where linear relationships exist between independent and dependent variables. However, its performance may decline when handling non-linearly separable data or datasets with highly complex feature interactions. In contrast, Decision Tree (72%) had the lowest accuracy, likely due to overfitting and its limitations in handling high-dimensional text data. Decision Trees tend to create highly specific splits, making them less effective in distinguishing overlapping emotions.

The Emotion Prediction System output in figure 9 showcases how the model classifies different Indonesian text inputs into emotional categories. “saya merasa tidak berguna" (I feel useless) → Predicted Emotion: DEPRESSION. This phrase expresses self-worth issues and negative emotions, which are strongly associated with depression. The model correctly identifies it as DEPRESSION. "saya harus semangat" (I have to stay motivated) → Predicted Emotion: STRESS. This phrase suggests an attempt to stay positive, but the model interprets it as stress-related, possibly because such statements are often used when someone is under pressure or struggling. "Iya, aku akan berjuang" (Yes, I will fight/keep going) → Predicted Emotion: HAPPY. The phrase conveys determination and resilience, which could be seen as a positive or motivational statement. The model classifies it as HAPPY, assuming it reflects an optimistic outlook. "nilaiku jelek" (My grades are bad) → Predicted Emotion: STRESS. Academic performance concerns are a common source of stress. The model correctly associates this phrase with STRESS.

The evaluation of model performance revealed several key insights and trends. Random Forest emerged as the most accurate model, followed closely by Logistic Regression. This highlights the effectiveness of ensemble methods in text-based classification tasks, as Random Forest leverages multiple decision trees to enhance robustness and reduce overfitting. Logistic Regression also performed well, demonstrating its capability to handle complex feature interactions effectively. In contrast, the Decision Tree model exhibited relatively lower accuracy, emphasizing its limitations when dealing with datasets containing overlapping features or high-dimensional spaces. Despite being a simpler algorithm, Bernoulli Naive Bayes performed competitively, reflecting its effectiveness in text classification tasks where features are represented in a binary format. This makes it a strong candidate for scenarios requiring efficient computations with minimal resource usage. Based on these results, Random Forest is the best-performing model for this emotion detection task, offering a strong balance between accuracy and robustness. Logistic Regression also demonstrated high accuracy, making it a reliable alternative, especially for tasks that prioritize scalability and interpretability. While Bernoulli Naive Bayes may be preferred for real-time applications due to its speed and efficiency, Decision Tree’s simplicity and transparency can be beneficial when interpretability is a key requirement. These findings underscore the importance of selecting models based on task-specific needs and dataset characteristics to achieve optimal results.

ISSN 2723-6471

5. Conclusion

This study evaluated four machine learning models—Bernoulli Naive Bayes, Decision Tree, Logistic Regression, and Random Forest—for emotion detection in mental health counseling conversations. The Random Forest model achieved the highest accuracy at 90%, demonstrating its robustness and effectiveness in handling complex feature interactions. Logistic Regression followed closely with an accuracy of 89.4%, showcasing its reliability, scalability, and suitability for text classification tasks. Bernoulli Naive Bayes performed competitively with an accuracy of 88.89%, excelling in computational efficiency and binary feature representation. In contrast, the Decision Tree model had the lowest accuracy at 72.49%, highlighting its limitations in handling overlapping features and high-dimensional data. These findings align with the challenges discussed in the introduction, emphasizing the need for effective and scalable tools in mental health emotion analysis. Random Forest and Logistic Regression demonstrated superior performance in accurately classifying key emotions like anxiety, stress, and motivation, which are critical in identifying mental health states and guiding interventions. Future studies should focus on integrating machine learning models into chatbots or automated counseling systems such as audio recording and transcribe it to improve early emotion detection and intervention also exploring cross-lingual embeddings, transfer learning, and language-specific fine-tuning to enhance adaptability across diverse languages in mental health counseling and automated interventions. Enhancing models like Random Forest and Logistic Regression for real-time applications could increase their effectiveness and accessibility in mental health support.

6. Declarations

6.1. Author Contributions

Conceptualization: I.R., L.H., R.A.S., D.S., M., E.S.N., M.F., A.G., and M.Q.R.; Methodology: L.H., D.S., and A.G.; Software: I.R., M.F., and E.S.N.; Validation: I.R., L.H., R.A.S., D.S., and M.Q.R.; Formal Analysis: I.R., L.H., R.A.S., and M.Q.R.; Investigation: I.R., M., and M.F.; Resources: L.H. and D.S.; Data Curation: L.H. and M.F.; Writing—Original Draft Preparation: I.R., L.H., R.A.S., and A.G.; Writing—Review and Editing: L.H., I.R., R.A.S., M., and D.S.; Visualization: I.R. and M.Q.R. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

This research was supported by the DRTPM KEMENDIKBUD KATALIS 2024 Funding with contract number 044/E5/PG.02.00/PL.BATCH.2/2024, dated August 1, 2024, Universitas Sriwijaya (UNSRI), Universitas Bina Darma (UBD), and Universitas Muhammadiyah Palembang (UMP). We sincerely appreciate their place, team, and financial support, which made this study and research possible.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] M. Srivastava, I. Singh, Er. S. Khanna, and Dr. A. K. Srivastava, “Depression Detection And Sentiments Analysis,” *International Journal Of Scientific Research In Engineering And Management*, vol. 07, no. 04, pp. 1-8, Indospace Publications, Apr. 09, 2023. doi: 10.55041/ijsrem18726.

[2] Mr. J. Dias and Dr. P. D. Tawde, “Predicting Mental Health Disorders based on Sentiment Analysis,” *International Journal Of Scientific Research In Engineering And Management*, vol. 08, no. 09. Indospace Publications, pp. 1–4, Sep. 18, 2024. doi: 10.55041/ijsrem37524.

[3] M. Deshpande and V. Rao, “Depression detection using emotion artificial intelligence,” *2017 International Conference on Intelligent Sustainable Systems (ICISS)*. IEEE, vol. 2017, no. 12, pp. 858-862, Dec. 2017. doi: 10.1109/iss1.2017.8389299.

[4] B. Kathole, S. Lonare, G. Dharmale, J. Katti, K. Vhatkar, and V. V. Kimbahune, “Sentiment Analysis-Based Automatic Stress and Emotion Recognition using Weighted Fused Fusion-Based Cascaded DTCN with Attention Mechanism from EEG Signal,” *Journal of Information andamp; Knowledge Management*, vol. 23, no. 05, pp. 1-12, World Scientific Pub Co Pte Ltd, Jun. 07, 2024.

[5] J. Alanya-Beltran, A. Gehlot, U. Gurav, C. Valderrama-Zapata, H. Anandaram, and A. Jain, “Emotional Detection Based on Bio Information and Data Analytics,” *2022 International Conference on Innovative Computing, Intelligent Communication and Smart Electrical Systems (ICSES)*. IEEE, vol. 2022, no. 07, pp. 1–6, Jul. 15, 2022. doi: 10.1109/icses55317.2022.9914197.

[6] D. K. Choudhary, A. Gupta, S. Singh, T. Abhinav, and T. Agrawal, “Sentiment Analysis for Depression Detection Using Artificial Intelligence,” *2024 3rd International Conference on Artificial Intelligence For Internet of Things (AIIoT)*. IEEE, vol. 2024, no. 05, pp. 1–5, May 03, 2024. doi: 10.1109/aiiot58432.2024.10574749.

[7] Rutten, I., Voorspoels, W., Koster, E.H., and Vanpaemel, W., “Emotional Expressions as an Implicit Dimension of Categorization”. *Cognitive Science*. vol. 40, pp. 2370-2375, 2018

[8] N. S. Eckland, S. H. Sperry, A. A. Castro, and H. Berenbaum, “Intensity, frequency, and differentiation of discrete emotion categories in daily life and their associations with depression, worry, and rumination.,” *Emotion, American Psychological Association (APA)*, vol. 22, no. 2, pp. 305–317, Mar. 2022. doi: 10.1037/emo0001038.

[9] Bhawna and S. D. Bhambri, "Effect of Positive Emotions on Mental Health," *International Journal for Research Trends and Innovation (IJRTI)*, vol. 7, no. 11, pp. 340-347, Nov. 2022, ISSN: 2456-3315.

[10] R. Cobo-Rendón, M. V. Pérez-Villalobos, D. Páez-Rovira, and M. Gracia-Leiva, “A longitudinal study: Affective wellbeing, psychological wellbeing, self-efficacy, and academic performance among first-year undergraduate students,” *Scandinavian Journal of Psychology*, vol. 61, no. 4. Wiley, pp. 518–526, Feb. 05, 2020. doi: 10.1111/sjop.12618.

[11] S. S. Sethi and K. Jain, “AI technologies for social-emotional learning: recent research and future directions,” *Journal of Research in Innovative Teaching and Learning*, vol. 17, no. 2. Emerald, pp. 213–225, May 27, 2024. doi: 10.1108/jrit-03-2024-0073.

[12] de León Languré and M. Zareei, “Evaluating the Effect of Emotion Models on the Generalizability of Text Emotion Detection Systems,” IEEE Access, vol. 12. Institute of Electrical and Electronics Engineers (IEEE), vol. 12, no. 05, pp. 70489–70500, 2024. doi: 10.1109/access.2024.3401203.

[13] Waheed Awan, I. Taj, S. Khalid, S. Muhammad Usman, A. S. Imran, and M. Usman Akram, “Advancing Emotional Health Assessments: A Hybrid Deep Learning Approach Using Physiological Signals for Robust Emotion Recognition,” IEEE Access, vol. 12. Institute of Electrical and Electronics Engineers (IEEE), vol. 12, no. 09, pp. 141890–141904, 2024. doi: 10.1109/access.2024.3463746.

[14] L. Hermawan, D. Stiawan, R. A. Syakurah, Meilinda, and D. S. Ikhsan, “Mental Health with Machine Learning: A Prediction-Based Intervention Chatbot for Mental Health Conversations,” *2024 11th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. IEEE, vol. 2024, no. 09, pp. 590–595. doi: 10.1109/eecsi63442.2024.10776244.

[15] Malakar and B. Nepal, “Conceptualizing Explorative Data Analysis in Applied Statistics,” Patan Gyansagar, vol. 6, no. 1. Nepal Journals Online (JOL), pp. 46–63, Jul. 09, 2024. doi: 10.3126/pg.v6i1.67406.

[16] D. Manning, H. Schütze, and P. Raghavan, Introduction to Information Retrieval. Cambridge, UK: Cambridge University Press, 2009.

[17] X. Li, Y. Wang, and R. Ruiz, “A Survey on Sparse Learning Models for Feature Selection,” IEEE Transactions on Cybernetics, vol. 52, no. 3. Institute of Electrical and Electronics Engineers (IEEE), vol. 52, no. 03, pp. 1642–1660, Mar. 2022. doi: 10.1109/tcyb.2020.2982445.

[18] L. Sha, M. Rakovic, A. Das, D. Gasevic, and G. Chen, “Leveraging Class Balancing Techniques to Alleviate Algorithmic Bias for Predictive Tasks in Education,” IEEE Transactions on Learning Technologies, vol. 15, no. 4. Institute of Electrical and Electronics Engineers (IEEE), vol. 15, no. 04, pp. 481–492, Aug. 01, 2022. doi: 10.1109/tlt.2022.3196278

[19] M. Bishop, Pattern Recognition and Machine Learning. New York, NY, USA: Springer, 2006.

[20] M. L. F. Martanto and W. Istiono, "Sentiment Analysis of M-Paspor App Reviews Using Multinomial Naive Bayes," *Journal of Logistics, Informatics and Service Science*, vol. 11, no. 10, pp. 311–326, 2024, doi: 10.33168/JLISS.2024.1017. ISSN: 2409-2665.

[21] O. K. Afshin Gholamy, Vladik Kreinovich, "Why 70/30 or 80/20 relation between training and testing sets: A pedagogical explanation," Computer Sciences Commons, 2018.

[22] S. Gan, S. Shao, L. Chen, L. Yu, and L. Jiang, "Adapting Hidden Naive Bayes for Text Classification," *Mathematics*, vol. 9, no. 19. MDPI AG, pp. 1-14, Sep. 25, 2021. doi: 10.3390/math9192378.

[23] D. Mienye and N. Jere, "A Survey of Decision Trees: Concepts, Algorithms, and Applications," IEEE Access, vol. 12. Institute of Electrical and Electronics Engineers (IEEE), vol. 12, pp. 86716–86727, 2024. doi: 10.1109/access.2024.3416838.

[24] W. Hosmer Jr., S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression, Wiley Series in Probability and Statistics. Wiley, Mar. 22, 2013.

[25] L. Breiman, Machine Learning, vol. 45, no. 1. Springer Science and Business Media LLC, pp. 5–32, 2001. doi: 10.1023/a:1010933404324.

[26] M. A. Rosid, A. S. Fitrani, I. R. I. Astutik, N. I. Mulloh, and H. A. Gozali, "Improving Text Preprocessing For Student Complaint Document Classification Using Sastrawi," *IOP Conference Series: Materials Science and Engineering*, vol. 874, no. 1. IOP Publishing, pp. 1-6, Jun. 01, 2020. doi: 10.1088/1757-899x/874/1/012017.

[27] A. Siswanto and Y. Dani, "Sentiment Analysis about Oximeter as Covid-19 Detection Tools on Twitter Using Sastrawi Library," *2021 8th International Conference on Information Technology, Computer and Electrical Engineering (ICITACEE)*. IEEE, vol. 2021, no. 12, pp. 161–164. doi: 10.1109/icitacee53184.2021.9617216.