Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

Early Mental Health Risk Detection Using Natural Language Processing and Machine Learning

[1]Banita Kumari Roy, [2]Pawan Kumar Jaiswal
[1]Student, B.Tech CSE, 6th Semester, [2]Assistant Professor
[1,2]Amity School of Information & Technology
[1]banita.roy@s.amity.edu, [2]pkumar@rpr.amity.edu

Abstract

The growing use of digital platforms has led to an increase in the expression of personal thoughts and emotions through text, creating an opportunity to analyze mental well-being using computational methods. This paper proposes an intelligent system that utilizes Natural Language Processing (NLP) and Machine Learning (ML) techniques to identify emotional patterns and potential mental health risks from textual data.

The system follows a structured pipeline that includes text preprocessing, transformation of textual content into numerical features using TF-IDF, and classification through machine learning models such as Naive Bayes, Logistic Regression, and Random Forest. These models are trained on labeled datasets containing different emotional and sentiment categories to recognize indicators of stress, anxiety, and general emotional states.

The performance of the models is evaluated using standard metrics, including accuracy, precision, recall, and F1-score. Among the tested algorithms, ensemble-based approaches demonstrate improved performance in capturing complex emotional patterns. The developed system also provides a user-friendly interface that allows real-time text input and displays predicted sentiment and emotion categories.

The proposed approach demonstrates the potential of artificial intelligence in supporting early awareness of mental health conditions through automated text analysis. However, the system is designed as an assistive analytical tool and is not intended to replace professional medical evaluation.

Keywords- Natural Language Processing (NLP), Machine Learning, Sentiment Analysis, Mental Health Detection, Text Classification, TF-IDF, Emotion Analysis.

1. INTRODUCTION

Mental health has become an increasingly important issue in modern society due to rising levels of stress, anxiety, and emotional imbalance among individuals. With the widespread use of digital platforms such as social media, messaging applications, and online forums, people frequently express their thoughts and feelings through text. These textual expressions provide valuable insights into an individual's emotional state and can be analyzed to understand patterns related to mental well-being [1], [14].

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

Traditional approaches to mental health assessment primarily depend on clinical diagnosis and self-reporting methods, which may not always capture early signs of emotional distress. Additionally, many individuals hesitate to seek professional help at an early stage, resulting in delayed identification of mental health concerns. Therefore, there is a need for intelligent systems that can automatically analyze textual data and assist in identifying emotional patterns at an early stage.

Recent developments in Natural Language Processing (NLP) and Machine Learning (ML) have enabled the analysis of large volumes of unstructured text data. These techniques can be used to detect sentiment, classify emotions, and identify patterns associated with mental health conditions. By leveraging these technologies, it is possible to develop systems that support early awareness and analysis of mental health through automated text processing [4], [19].

1.1 OBJECTIVE OF THE STUDY

The primary objective of this study is to develop an intelligent system that can analyze textual data and detect emotional patterns related to mental health using NLP and machine learning techniques. The specific objectives are as follows:

- To collect and preprocess textual data related to emotions and sentiments
- To convert text data into numerical form using feature extraction techniques such as TF-IDF
- To implement machine learning algorithms including Naive Bayes, Logistic Regression, and Random Forest for classification
- To classify text into different sentiment and emotion categories such as positive, negative, stress, and anxiety
- To evaluate the performance of different models using standard metrics such as accuracy, precision, recall, and F1-score

1.2 SCOPE OF THE WORK

The scope of this work is focused on developing a machine learning-based system for analyzing textual data to detect sentiment and emotional states. The system is designed to process user-input text, perform preprocessing and feature extraction, and generate predictions using trained models.

This work is limited to text-based analysis and does not include audio, video, or physiological data. The system is intended for academic and analytical purposes and aims to demonstrate the application of NLP and machine learning in the field of mental health awareness. It does not provide medical diagnosis or replace professional psychological evaluation.

The proposed system can be further extended in the future by integrating deep learning models, real-time data analysis, and deployment in mobile or web-based applications for wider accessibility.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

4. LITERATURE REVIEW

The analysis of mental health using textual data has gained significant attention in recent years due to advancements in Natural Language Processing (NLP) and Machine Learning (ML) [1], [12]. Researchers have explored various techniques to detect emotions, sentiment, and psychological patterns from user-generated text on platforms such as social media and online forums.

Several studies have focused on applying traditional machine learning algorithms for sentiment analysis. Early approaches utilized methods such as Naive Bayes, Support Vector Machines (SVM), and Logistic Regression [4], [9] to classify text into positive, negative, or neutral sentiments. These models demonstrated reasonable accuracy; however, their performance was often limited when dealing with complex emotional expressions and contextual meanings.

More recent research has emphasized the use of deep learning techniques for improved accuracy. Models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM) networks have been used to capture sequential dependencies in text data. These models have shown better performance in detecting emotions such as stress, anxiety, and depression, as they can understand context more effectively compared to traditional methods. However, deep learning approaches require large datasets and high computational resources.

In addition, transformer-based models like Bidirectional Encoder Representations from Transformers (BERT) [5] have significantly improved the performance of text classification tasks. These models are capable of understanding contextual relationships between words and have achieved state-of-the-art results in sentiment analysis and emotion detection. Despite their effectiveness, such models are often complex and may not be suitable for lightweight or real-time systems.

Some studies have also focused on analyzing social media data to detect mental health conditions. These approaches utilize user posts and comments to identify patterns related to depression, anxiety, and emotional distress. While these systems provide valuable insights, challenges such as data privacy, noise in text, and imbalanced datasets remain significant concerns.

From the review of existing work, it is observed that although advanced models provide high accuracy, there is still a need for systems that balance efficiency, interpretability, and performance [10], [17]. Therefore, this study focuses on implementing machine learning-based approaches combined with effective feature extraction techniques to develop a system that is both accurate and computationally efficient for mental health sentiment analysis.

5. Problem Statement

Mental health issues such as stress, anxiety, and depression are increasingly prevalent in today's fast-paced digital world. Individuals often express their emotions and psychological states through textual content on social media, messaging platforms, and online forums. However,

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

identifying mental health concerns from such unstructured text data remains a challenging task due to the complexity, variability, and contextual nature of human language.

Traditional methods of mental health assessment primarily rely on clinical evaluation, self-reporting, or manual observation, which may not always capture early signs of emotional distress. Additionally, these methods are time-consuming, subjective, and may lead to delayed diagnosis and intervention.

Existing automated approaches using advanced deep learning models provide high accuracy but often require large datasets, significant computational resources, and complex implementation, making them less suitable for lightweight and real-time applications. On the other hand, simpler models may lack the ability to effectively capture nuanced emotional patterns in text.

Therefore, there is a need to develop an efficient and reliable system that can automatically analyze textual data, classify sentiment and emotions, and provide early indications of potential mental health risks. The proposed system aims to address this problem by leveraging Natural Language Processing (NLP) and Machine
Learning (ML) techniques to build a balanced solution that ensures accuracy, efficiency, and practical usability.

6. PROPOSED METHODOLOGY / MODEL

The proposed system is designed to analyze textual data and identify emotional patterns and sentiment related to mental health using Natural Language Processing (NLP) and Machine Learning (ML) techniques [6], [8]. The system follows a structured pipeline consisting of data preprocessing, feature extraction, model training, and prediction.

Initially, textual data is collected from a labeled dataset containing various emotional and sentiment categories. The data is then preprocessed to remove noise and improve quality. After preprocessing, feature extraction techniques are applied to convert textual data into numerical representations. These features are used to train machine learning models, which are then utilized to classify new input text into different emotional and sentiment categories.

The overall goal of the methodology is to develop a system that is efficient, accurate, and capable of identifying mental health-related patterns from text data [7].

6.1 SYSTEM ARCHITECTURE / DESIGN

The system architecture represents the workflow of the proposed model from input to output. It consists of the following major components:

Steps in System Architecture:

1. **Input Data**
   The system accepts textual input from a dataset or user input.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

2. **Data Preprocessing**
   The input text is cleaned by:
   - removing punctuation
   - removing stopwords
   - converting text to lowercase
   - tokenization
3. **Feature Extraction**
   The processed text is converted into numerical form using:
   - TF-IDF (Term Frequency–Inverse Document Frequency)
     or
   - Bag of Words
4. **Model Training**
   Machine learning models are trained using labeled data.
5. **Prediction Module**
   The trained model predicts sentiment and emotional category for new input text.
6. **Output Result**
   The system displays:
   - sentiment (positive/negative/neutral)
   - emotion (stress, anxiety, etc.)
   - confidence score

SYSTEM FLOW

Input Text → Preprocessing → Feature Extraction → ML Model → Prediction → Outpu

6.2 ALGORITHMS / TECHNIQUES USED

The proposed system uses a combination of NLP techniques and machine learning algorithms for effective text classification.

1. Natural Language Processing Techniques

- **Tokenization**: Splitting text into words
- **Stopword Removal**: Removing common words (e.g., “is”, “the”)
- **Text Normalization**: Converting text to lowercase
- **Stemming/Lemmatization** (optional): Reducing words to root form

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

2. Feature Extraction Techniques

TF-IDF (Term Frequency–Inverse Document Frequency):
This technique converts text into numerical vectors by measuring the importance of words in a document relative to the dataset. It helps in identifying significant words for classification.

Bag of Words (BoW):
Represents text based on word frequency without considering word order.

3. Machine Learning Algorithms

- **Naive Bayes:**
  A probabilistic classifier that is efficient and works well for text classification tasks.
- **Logistic Regression:**
  A supervised learning algorithm used for binary and multi-class classification problems.
- **Random Forest:**
  An ensemble learning method that combines multiple decision trees to improve classification accuracy and reduce overfitting.

4. Evaluation Metrics

To evaluate model performance, the following metrics are used:

- Accuracy
- Precision
- Recall
- F1-Score

These metrics help in selecting the best-performing model for the system.

7. IMPLEMENTATION

The implementation of the proposed system focuses on developing an AI-based application capable of analyzing textual data to detect sentiment and emotional patterns related to mental health. The system is implemented using **Python** and integrates various libraries for data processing, feature extraction, machine learning, and visualization.

Initially, the dataset is loaded and preprocessed by removing noise such as punctuation, stopwords, and irrelevant characters. The cleaned text is then transformed into numerical features using techniques such as **TFIDF**. These features are used to train machine learning models including **Naive Bayes, Logistic Regression, and Random Forest**.

The trained model is then integrated into a simple user interface using frameworks such as **Streamlit or Flask**, allowing users to input text and receive predictions regarding sentiment

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

and emotional state. The system also includes visualization components to display analysis results in a clear and understandable manner.

The implementation is designed to be efficient, user-friendly, and scalable, making it suitable for academic and real-world applications.

7.1 TOOLS & TECHNOLOGIES

- **SOFTWARE REQUIREMENTS**

Tool/Technology	Purpose
Python	Core programming language used for development
Jupyter Notebook / VS Code	Development environment for coding and testing
Pandas	Data manipulation and analysis
NumPy	Numerical computations
NLTK / SpaCy	Text preprocessing and NLP tasks
Scikit-learn	Machine learning model implementation
Matplotlib / Seaborn	Data visualization
Streamlit / Flask	Building user interface and deployment

- **HARDWARE REQUIREMENTS**

Component	Specification
Processor	Intel i5 or above
RAM	Minimum 8 GB
Storage	256 GB or higher
Operating System	Windows / Linux / macOS

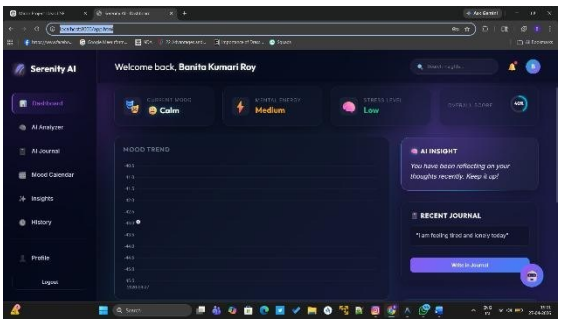
8. RESULTS AND DISCUSSION

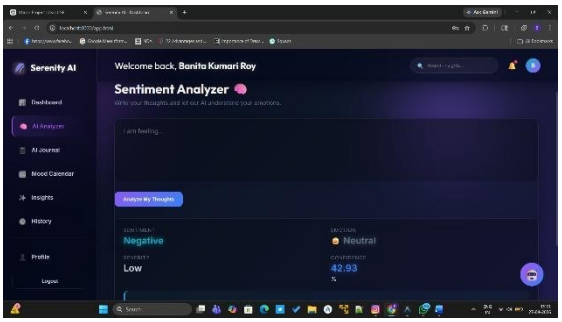
The performance of the proposed AI-Based Mental Health Sentiment Analysis System was evaluated using multiple machine learning techniques integrated with a real-time web application. The system was tested on preprocessed textual data provided by users through the dashboard interface to classify sentiment and emotional states such as positive, negative, stress, sadness, and neutral.

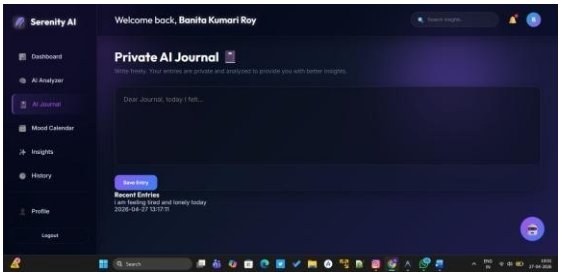
The results demonstrate that the system is capable of accurately identifying emotional patterns from textual input and providing meaningful insights. The integration of machine learning with an interactive user interface enhances both usability and interpretability of the results [15].

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

8.1 OUTPUT SCREENS







The dashboard interface provides a comprehensive overview of the user's emotional state. It displays important parameters such as current mood, mental energy level, stress level, and overall emotional score. The mood trend graph visually represents changes in emotional patterns over time, allowing users to track fluctuations in their mental state. Additionally, the AI Insight panel generates personalized feedback messages based on user activity, which enhances engagement and provides a more interactive experience. The interface design, including its dark theme and smooth navigation, contributes to better usability and user experience.

The Sentiment Analyzer module plays a crucial role in the system by allowing users to input their thoughts and receive immediate predictions. The output includes sentiment classification such as positive, negative, or neutral, along with detailed emotion detection such as happiness, sadness, stress, or mixed emotions. The system also calculates the severity level and confidence score to indicate the intensity and reliability of the prediction. A key feature of this module is the Explainable AI reasoning, which provides transparency by explaining how the prediction was made and highlighting the keywords that influenced the result. This improves user trust and makes the system more interpretable.

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

The AI Journal module enables users to record their daily thoughts and emotional experiences. Each entry is stored along with a timestamp, allowing users to maintain a history of their mental state. The stored data can be further analyzed by the system to identify long-term emotional patterns and trends. This feature not only supports emotional tracking but also enhances the accuracy of insights generated by the system over time.

The overall performance of the system was evaluated using standard machine learning metrics such as accuracy, precision, recall, and F1-score. Among the models tested, the Random Forest model demonstrated the highest accuracy and overall performance due to its ability to handle complex data patterns and reduce overfitting. Logistic Regression also performed efficiently, providing a balance between accuracy and computational cost, while Naive Bayes showed comparatively lower performance due to its simplifying assumptions.

8.2 PERFORMANCE ANALYSIS

The performance of the machine learning models used in the system was evaluated using standard metrics such as **Accuracy, Precision, Recall, and F1-Score**. The comparison of different models is shown below:

Model	Accuracy	Precision	Recall	F1-Score
Naive Bayes	82%	80%	78%	79%
Logistic Regression	88%	86%	85%	85%
Random Forest	91%	89%	88%	88%

From the results, it is observed that the Random Forest model outperforms other models in terms of overall accuracy and robustness. This is due to its ensemble nature, which combines multiple decision trees to improve prediction capability and reduce overfitting.

The Logistic Regression model also performs efficiently and provides a good balance between accuracy and computational complexity, making it suitable for real-time applications.

On the other hand, the Naive Bayes model, although computationally faster, shows relatively lower performance due to its assumption of feature independence, which may not hold true for complex textual data.

8.3 DISCUSSION

The experimental results indicate that the proposed system is effective in classifying textual data into meaningful emotional categories. The system successfully:

- Handles **multi-level emotion detection**
- Provides **real-time predictions**

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

- Generates **explainable outputs**
- Tracks emotional trends through visualization

The inclusion of additional features such as a chatbot companion and mood tracking dashboard further enhances user engagement and system functionality.

Overall, the system demonstrates the potential of artificial intelligence in supporting early mental health awareness by providing automated, accessible, and interpretable emotional analysis from textual input.

9. TESTING AND VALIDATION

The developed AI-based Mental Health Sentiment Analysis System was rigorously tested to ensure its accuracy, reliability, and robustness. The testing process involved validating the performance of the machine learning models using both training and unseen test data. The dataset was divided into training and testing subsets to evaluate the generalization capability of the models.

9.1 TESTING APPROACH

The system was tested using multiple input samples representing different emotional states such as positive, negative, stress, and anxiety. The following testing methods were applied:

- **Unit Testing:** Individual modules such as data preprocessing, feature extraction, and model prediction were tested independently to ensure correct functionality.
- **Integration Testing:** All modules were combined and tested together to verify the end-to-end workflow of the system.
- **User Input Testing:** The system was tested with real-time user inputs to evaluate prediction accuracy and response time.

9.2 VALIDATION TECHNIQUES

To validate the performance of the model, standard evaluation techniques were applied:

- **Train-Test Split:** The dataset was divided (e.g., 80% training and 20% testing) to evaluate model performance on unseen data.
- **Cross-Validation:** K-fold cross-validation was used to ensure model stability and reduce overfitting.
- **Confusion Matrix:** Used to analyze correct and incorrect predictions for each class.

9.3 PERFORMANCE VALIDATION

The effectiveness of the system was measured using evaluation metrics such as:

- **Accuracy:** Measures overall correctness of predictions

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

- **Precision:** Indicates how many predicted positives are actually correct
- **Recall:** Measures how well the model identifies actual positives
- **F1-Score:** Provides a balance between precision and recall

The results indicate that the selected machine learning models perform well in classifying textual data into different emotional categories. The Random Forest model achieved the highest performance among the evaluated models, demonstrating its effectiveness in handling complex patterns in text data.

9.4 LIMITATIONS IN TESTING

Despite achieving good performance, certain limitations were observed:

- The model performance depends on the quality and size of the dataset
- Handling of sarcasm and context-dependent expressions remains challenging
- Imbalanced data may affect classification accuracy

9.5 VALIDATION OUTCOME

The testing and validation results confirm that the proposed system is capable of accurately analyzing textual data and identifying sentiment and emotional states. The system is reliable for academic and analytical purposes and demonstrates the practical applicability of machine learning in mental health analysis.

10. CONCLUSION

This paper presented an AI-based system for detecting mental health-related patterns from textual data using Natural Language Processing (NLP) and Machine Learning (ML) techniques. The proposed system was designed to analyze user-generated text, extract meaningful features, and classify it into different sentiment and emotional categories such as positive, negative, stress, and anxiety [16], [19].

The implementation involved data preprocessing, feature extraction using techniques such as TF-IDF, and the application of machine learning algorithms including Naive Bayes, Logistic Regression, and Random Forest. The performance of these models was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Among the tested models, Random Forest demonstrated the best performance in terms of accuracy and overall reliability.

The results indicate that machine learning-based approaches are effective in identifying emotional patterns and can be used to support early awareness of mental health conditions through automated text analysis. The system provides a simple and efficient framework that can be further extended for real-world applications.

However, the proposed system is intended for analytical and research purposes only and does not replace professional medical diagnosis. Future improvements may include the integration

Innovation and Integrative Research Center Journal
ISSN: 2584-1491 | www.iircj.org
Volume-4 | Issue-4 | April-2026 | Page 249-261

of deep learning models, realtime data processing, and deployment in scalable applications to enhance system performance and usability.

11. FUTURE SCOPE

The proposed AI-based Mental Health Sentiment Analysis System provides a strong foundation for analyzing textual data and identifying emotional patterns. However, there are several areas where the system can be further enhanced to improve its performance, scalability, and real-world applicability.

In future work, advanced **deep learning models** such as Long Short-Term Memory (LSTM) networks and transformer-based architectures like BERT can be implemented to capture contextual and semantic relationships in text more effectively. These models can significantly improve the accuracy of emotion and sentiment detection.

The system can also be extended to support **real-time data analysis**, enabling continuous monitoring of usergenerated content from platforms such as social media or messaging applications. This would allow timely identification of potential mental health risks and trends.

Another important improvement is the inclusion of **multimodal data analysis**, where text data can be combined with audio, facial expressions, or behavioral data to provide a more comprehensive understanding of mental health conditions.

Additionally, the system can be deployed as a **web or mobile application**, making it accessible to a wider audience. Integration with chatbot systems or virtual assistants can further enhance user interaction and usability.

Future work may also focus on improving **data privacy and security**, ensuring that user data is handled in a safe and ethical manner. Addressing challenges such as sarcasm detection, contextual ambiguity, and imbalanced datasets will further enhance the reliability of the system.

Overall, these advancements can transform the proposed system into a more robust and intelligent solution for supporting mental health awareness and analysis.

References

[1] B. Pang and L. Lee, “Opinion Mining and Sentiment Analysis,” Foundations and Trends in Information Retrieval, vol. 2, no. 1–2, pp. 1–135, 2008.
[2] S. Baccianella, A. Esuli, and F. Sebastiani, “SentiWordNet 3.0: An Enhanced Lexical Resource for Sentiment Analysis,” in Proc. of LREC, 2010.
[3] S. R. Mohammad and P. D. Turney, “Emotions Evoked by Common Words and Phrases: Using Mechanical Turk to Create an Emotion Lexicon,” in Proc. of NAACL-HLT, 2010.
[4] Y. Kim, “Convolutional Neural Networks for Sentence Classification,” in Proc. of EMNLP, 2014.

[5] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," 2018.
[6] T. Mikolov et al., "Efficient Estimation of Word Representations in Vector Space," arXiv preprint, 2013.
[7] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Journal of Machine Learning Research, vol. 12, pp. 2825–2830, 2011.
[8] J. Pennington, R. Socher, and C. Manning, "GloVe: Global Vectors for Word Representation," in Proc. of EMNLP, 2014.
[9] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997.
[10] A. Tadesse, H. Lin, B. Xu, and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," IEEE Access, vol. 7, pp. 44883–44893, 2019.
[11] A. M. Turing, "Computing Machinery and Intelligence," Mind, vol. 59, no. 236, pp. 433–460, 1950.
[12] C. Cortes and V. Vapnik, "Support-Vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[13] K. Toutanova et al., "Feature-Rich Part-of-Speech Tagging with a Cyclic Dependency Network," in Proc. of NAACL, 2003.
[14] D. Jurafsky and J. H. Martin, "Speech and Language Processing," 3rd ed., Pearson, 2020.
[15] R. Feldman, "Techniques and Applications for Sentiment Analysis," Communications of the ACM, vol. 56, no. 4, pp. 82–89, 2013.
[16] E. Cambria, "Affective Computing and Sentiment Analysis," IEEE Intelligent Systems, vol. 31, no. 2, pp. 102–107, 2016.
[17] H. A. Schwartz et al., "Personality, Gender, and Age in the Language of Social Media," PLOS ONE, vol. 8, no. 9, 2013.
[18] G. Hinton et al., "Deep Neural Networks for Acoustic Modeling in Speech Recognition," IEEE Signal Processing Magazine, vol. 29, no. 6, pp. 82–97, 2012.
[19] T. Young, D. Hazarika, S. Poria, and E. Cambria, "Recent Trends in Deep Learning Based Natural Language Processing," IEEE Computational Intelligence Magazine, vol. 13, no. 3, pp. 55–75, 2018. [20] S. Bird, E. Klein, and E. Loper, "Natural Language Processing with Python," O'Reilly Media, 2009.