International Journal of Technology and Applied Science (IJTAS)
E-ISSN: 2230-9004 ● Website: www.ijtas.com ● Email: editor@ijtas.com

AI-Based Analysis of Themes and Sentiments in Literary Texts

Thejovathi Murari[1], Charvitha Reddy Abbu[2], Alagala Bhavani[3], Chinnabbu Srinidhi[4], Tanishqa Bose Mazumdar[5]

[1]Associate Professor, [2,3,4,5]B. Tech 3rd year Student
[1,2,3,4,5]CSE (AI&ML), Vignan's Institute of Management and Technology for Women, Hyderabad, India.

Abstract:
This project proposes AI and NLP based system to automatically analyze and interpret the literary texts with objective, data driven evaluation of recurrent themes and changing sentiments in a narrative and then identifying the trends of emotions in literature. Traditional analysis of literary text depends upon the personal opinion of human readers, which is biased. Through the processing of text, it detects the hidden emotional arc and theme which cannot be perceived by only reading. This project links computing technology with creative arts and allow the user to quickly get some analysis of any long piece of text through our tool for research and study purposes.

I. INTRODUCTION:

Literary text such as novel, poem, short story, drama etc are filled with lots of emotion and other hidden meanings. Theme analysis and sentiment analysis of literary text is very important for the readers, researcher and student, because literatures reflect lot of human emotions such as love, happy, sad, fear, war, peace etc.[2] Analysis of text is done traditionally by reading the text and understanding the meaning, this will take lot of time and depends on the perspective of person who analyzes.[5] In the advent of Artificial Intelligence(AI) and natural language processing(NLP), machine can automatically interpret the texts with the help of natural language processing, which lets machines understand the human language by processing the text such as word, sentences, patterns etc.
Using NLP and ML methods this system will able to determine the topics or theme or in a text and able to detect its main sentiment (positive, negative or neutral) of the text.[3] This Project will use AI based techniques to help students/learners to gain better insights about a literary work.

II. PROPOSED SYSTEM:

The Proposed System is a text based literary analysis AI system which analyses literary text using most recent natural language processing techniques to provide meaningful information. The Proposed System will overcome the limitations of manual analysis.[6] The Proposed System supports Different type of literary text including Novels, poems etc. It extracts information such as topics/sentiments/emotions and characters etc. The proposed System pipeline consists of integration of many intelligent blocks. The Proposed System uses Topic modelling techniques such as Latent Dirichelet Allocation to extract topics from text. The proposed System uses BERT model to calculate Sentiment Analysis. The proposed System consists of Emotion Detection feature which detects deep emotions of text, such as joy, sadness, anger, fear.[7]

Named entity recognition is used to detect the character's name and extract information about the role played by them in the novel and about their actions and interaction with other characters. The proposed system does not only limit to sentiment and summarization, instead it also analyses the changing emotions in the text, analyses the change in behaviour of the character and analyzes the change in emotion of

character and reports them, by doing this providing a more structured interpretation of the book.[4] Overall, the proposed system helps automate and provide a more structured book analysis, while minimizing the efforts of humans. The reports can be generated to provide an in-depth analysis of the book automatically. So, the proposed system helps analyze the book in more structure and accurate way.

III. PROPOSED SYSTEM:

A. Overview of the Proposed System:

Proposed system takes raw literature as input from user, then processes this data using cleanings, tokenization, removing stop-words and applying lemmatization etc.[12] Then theme extraction is done using LDA, sentiment analysis using BERT and Emotion Analysis is done using Joy, Sadness and Anger Detection using ML models. NER is done to detect the characters and analysis of their action, behavior and their change in emotions.[8] Finally report generation is done and output is displayed to the user in a form of report or visualizations. System is Fast, structured way to analyse the Literature.

B. Overall System Architecture:

The Figure below illustrates the overall system architecture

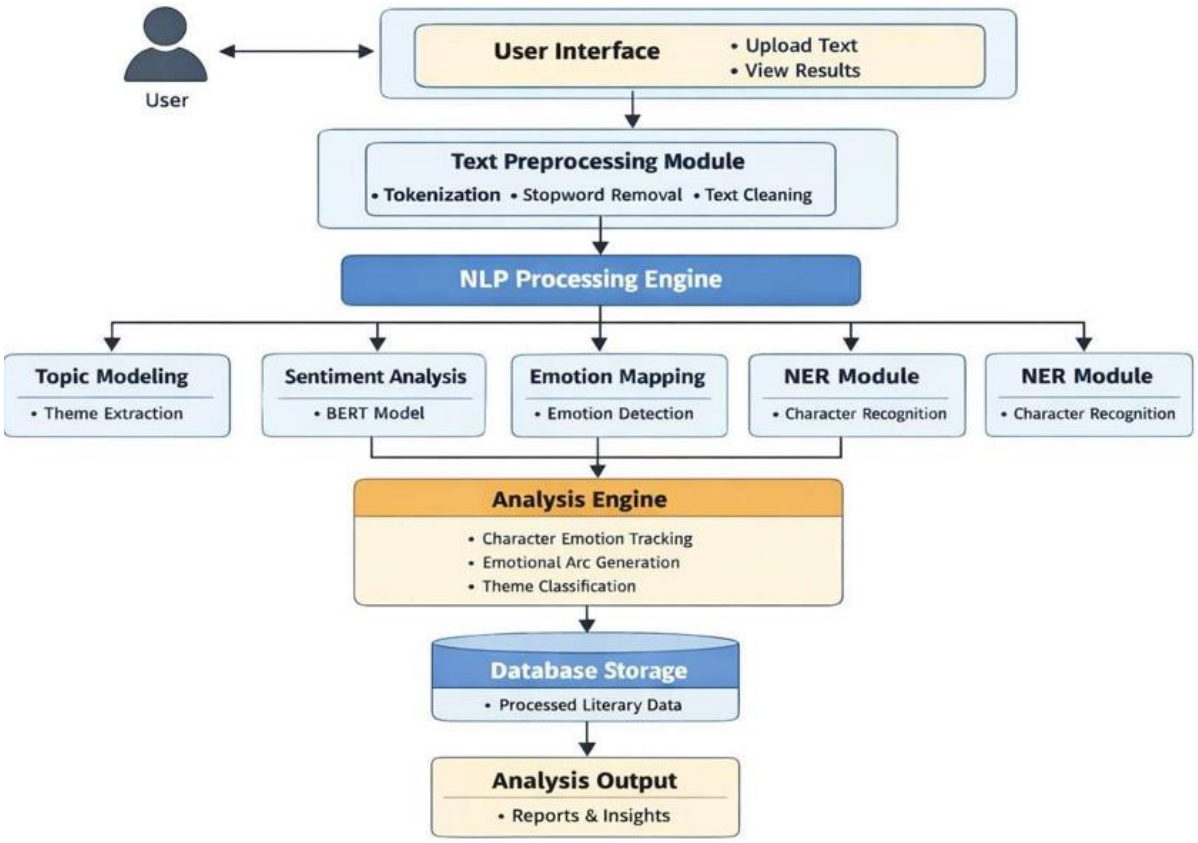

Figure: System Architecture

Navigation Flow:

Text is cleaned up in the Text Preprocessing Module. This module does a things to the Text. It breaks the Text into pieces and gets rid of extra words that are not needed.

The Text that has been cleaned and processed is then looked at by the NLP Processing Engine. The NLP Processing Engine does things to the Text. It looks at the text to see what it is about. It finds the topics

that're in the text. The system figures out how people feel about what they're reading. It detects the emotions that people have when they read the text.

IV. IMPLEMENTATION DETAILS:

The framework is built using Python and some popular libraries that are used for natural language processing. The system is made up of components that all do something important to help the system work properly.

Step 1: **Data Input**
Literary text is collected from users through manual input or from a dataset.

Step 2: **Text Preprocessing**
- Removal of punctuation and special characters
- Tokenization
- Stop-word removal
- Lemmatization

Step 3: **Theme Extraction Module**
This part of the process looks at the ideas and groups similar words into categories called themes. It uses something called Gensim LDA to do this.

Step 4: **Sentiment Analysis Module**
The Theme Extraction Module is followed by the Sentiment Analysis Module which uses a tool called BERT. The Sentiment Analysis Module sorts out how people feel into three groups: they like something they do not like something or they feel nothing about it.

Step 5: **Emotion Detection Module**
The Emotion Detection Module uses computer programs like DistilRoBERTa that have been trained already. The Emotion Detection Module can tell if the Theme Extraction Module and Sentiment Analysis Module are talking about feelings, sad feelings, angry feelings, scared feelings or surprised feelings.[11] The Emotion Detection Module also looks at how these feelings change throughout the text, about the Theme Extraction Module and Sentiment Analysis Module.

Step 6: **Character Analysis Module**
Uses spaCy NER to recognize characters and assess their actions and feelings.

Step 7: **Data Storage**
The results are saved in a MySQL database.

Step 8: **Visualization & Output**
Graphs and charts show theme distribution, sentiment trends, and emotional arcs. These are displayed through a web interface.

Technologies Used:
* Python
* NLTK
* spaCy
* Transformers
* Gensim

* PyTorch
* Scikit-learn
* MySQL

The system is made up of parts that work well together so it runs smoothly.

The system is easy to change when we need to make changes to the system.

This is because the system is made up of Python and other technologies like NLTK and spaCy that work together.

The system is made up of parts like Gensim and PyTorch and Scikit-learn that work well together so the system runs smoothly.

The system is easy to change when we need to make changes to the system because of technologies, like MySQL.

V. EXPERIMENTAL RESULTS AND ANALYSIS:

This section showcases the results/visuals of the website

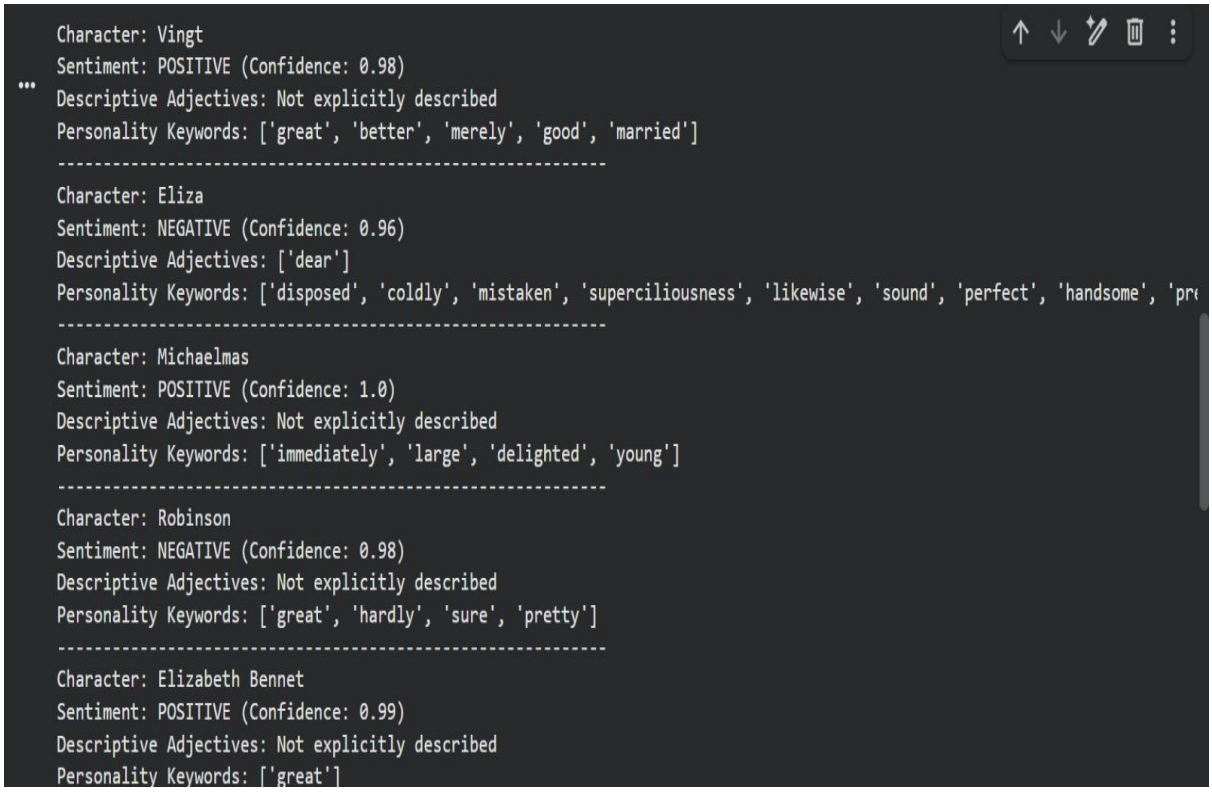

Figure: NLP Pipeline Output

The figure presents the textual output of a multi-stage Natural Language Processing (NLP) pipeline applied to an excerpt from Pride and Prejudice by Jane Austen. The results are organized into four major analytical stages: overall sentiment analysis, theme extraction, emotion detection, and character-level analysis.

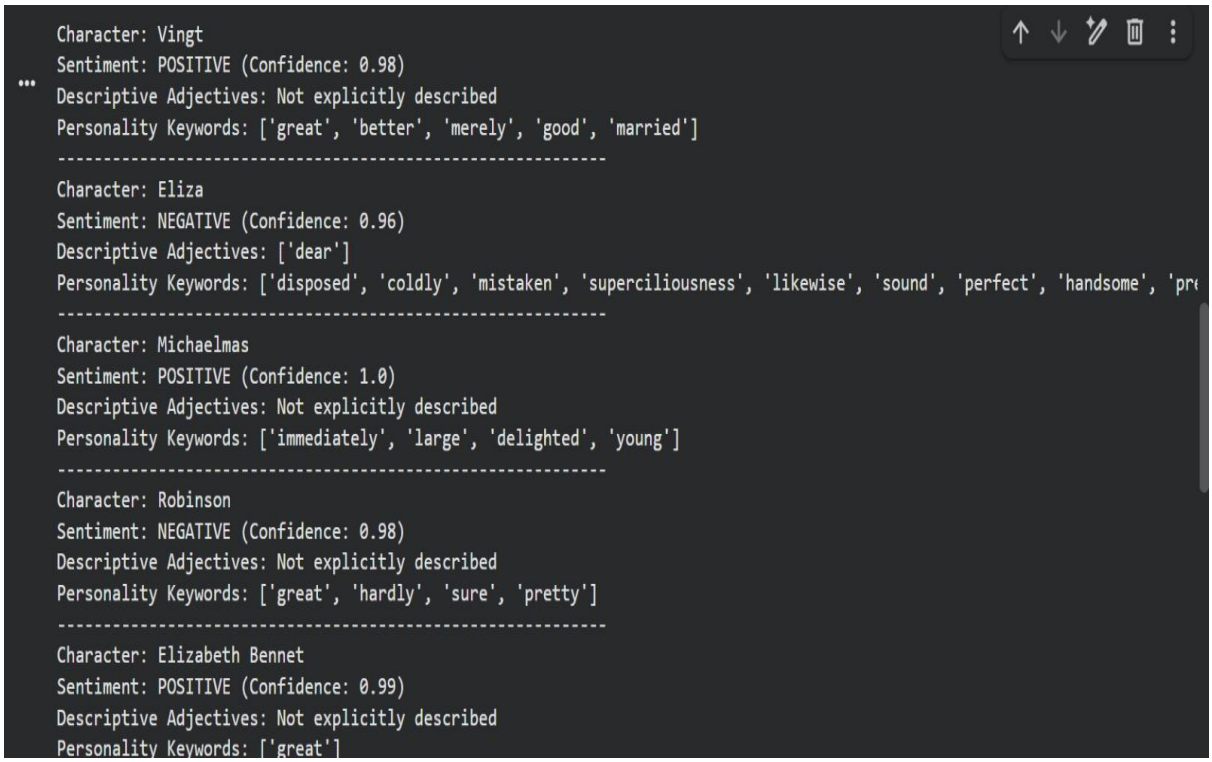

Figure: NLP Pipeline Output (2)

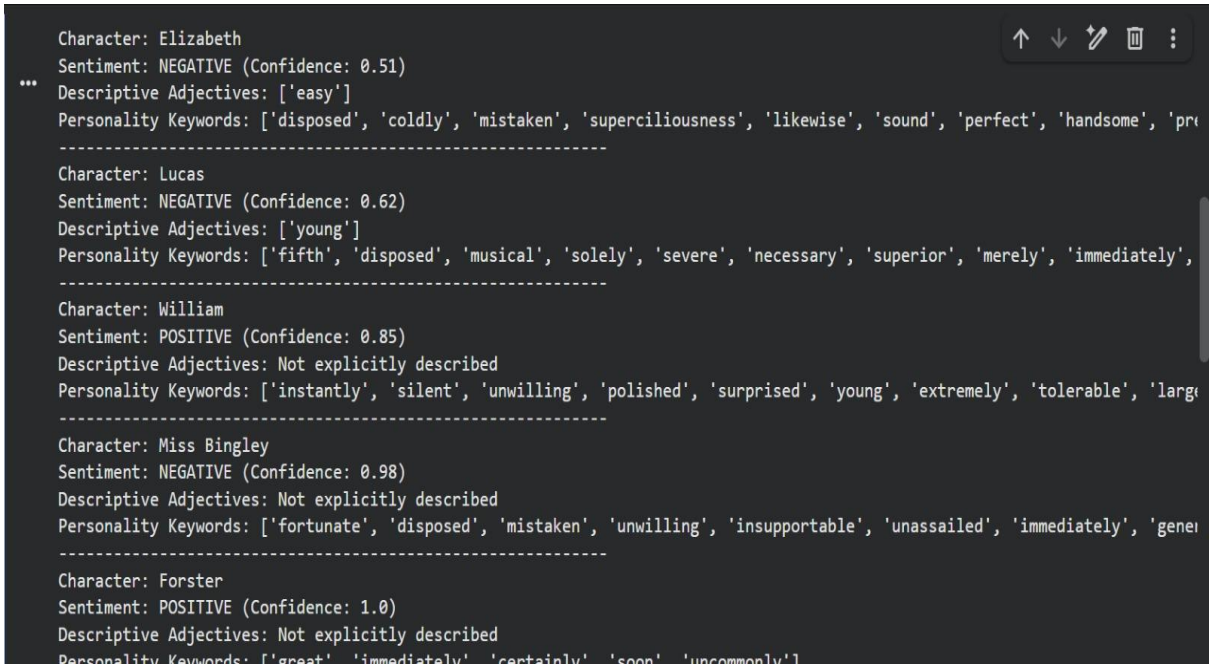

Figure: The image shows a text-based output screen.

Figure: Character Analysis Dashboard

The figure presents a multi-dimensional visualization of character-level analysis derived from the text of Pride and Prejudice by Jane Austen. It consists of three horizontally aligned bar charts, each capturing a distinct analytical dimension: sentiment score, character mentions, and personality richness.

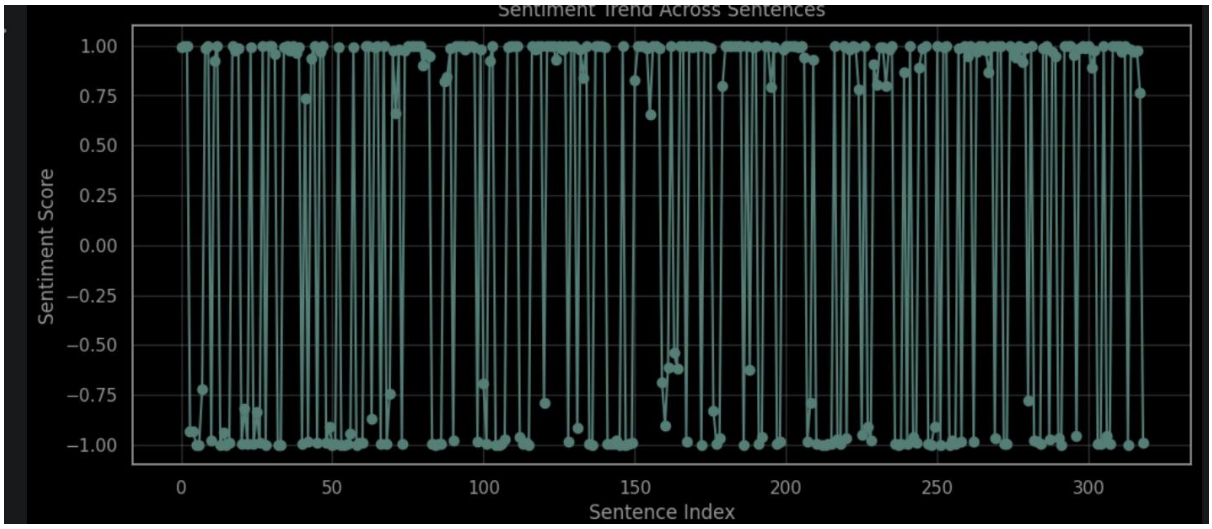

Figure: Sentiment Trend Across Sentences.

The figure illustrates the variation of sentiment across individual sentences in the selected excerpt from Pride and Prejudice by Jane Austen. It is represented as a line and scatter plot, where each point corresponds to the sentiment score of a sentence.

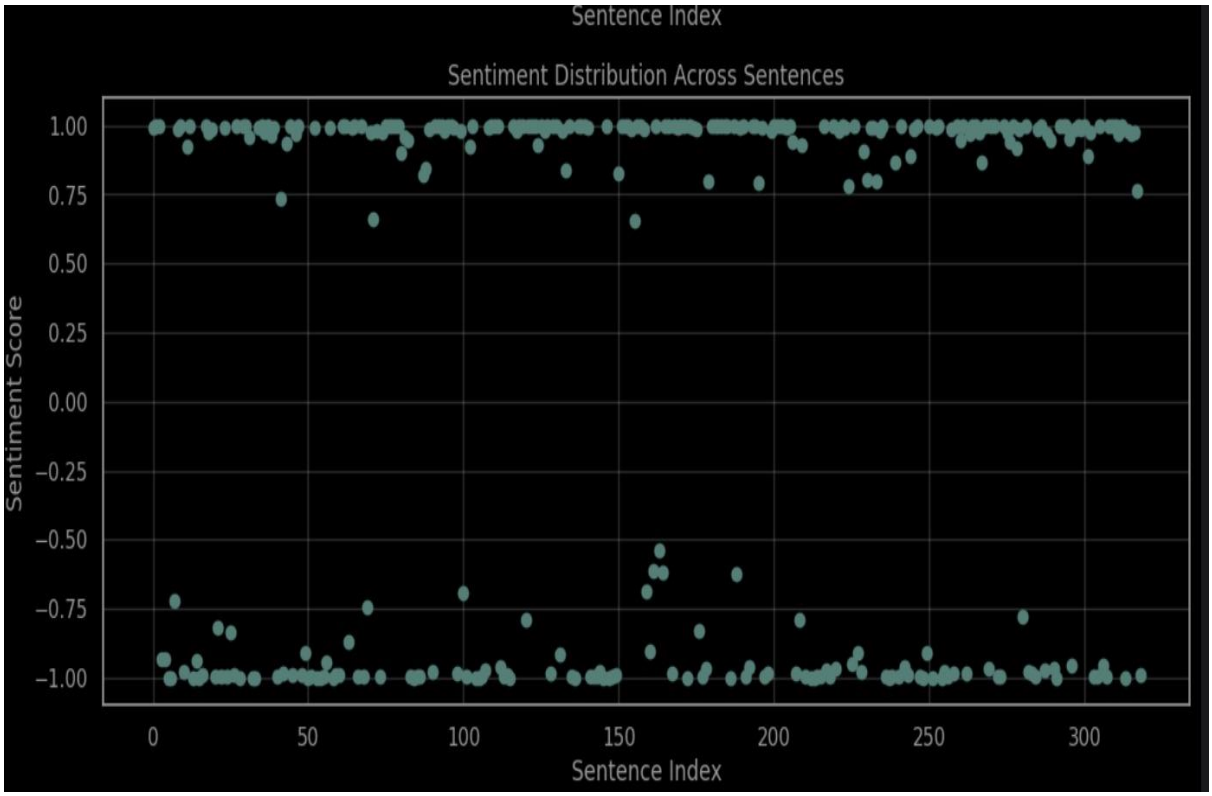

Figure: Sentiment Distribution Across Sentences.

The figure presents a scatter plot illustrating the distribution of sentiment scores across individual sentences in the selected excerpt from Pride and Prejudice by Jane Austen.

VI. CONCLUSION:

The Literary Text Analysis System effectively demonstrates how modern NLP techniques can simplify and enhance the understanding of complex literary texts. The results show that automated analysis can capture themes, emotions, and character dynamics with meaningful accuracy, reducing the need for time-consuming manual interpretation. While the system performs well, it can be further improved with more advanced models and larger datasets for deeper contextual understanding. Overall, the project highlights the potential of technology in transforming literary studies into a more interactive and insightful experience. It also opens opportunities for real-time analysis and integration with educational tools. In the future, the system can be expanded to support multiple languages and more nuanced literary interpretations.

REFERENCES:

[1].Annadurai, T., A. Mohanraj, C. Sakthimurugan, G. Padmapriya, D. Vedamanickam, and P. Kavinkumar. "Sentiment Analysis Meets Literary Studies: AI for Interpreting Tone and Mood in Fictional Works." In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG), pp. 1-7. IEEE, 2025.

[2]. Yu, Jie, and Chunhong Qi. "Machine learning-based sentiment analysis in English literature: Using deep learning models to analyze emotional and thematic content in texts." IEEE Access (2025).

[3]. Setiyoningsih, Titi, Chafit Ulya, Sarwiji Suwandi, Nugraheni Eko Wardani, Sugit Zulianto, and Kundharu Saddhono. "Using AI for Thematic Analysis in Poetry and Prose." In 2025 3rd International Conference on IoT, Communication and Automation Technology (ICICAT), pp. 1-8. IEEE, 2025.

[4]. Saddhono, Kundharu, Nanda Saputra, Erikson Saragih, Elina Lulu Bimawati Rumapea, Sri Ninta Tarigan, and Cahyo Hasanudin. "AI-Powered Automated Criticism Design Tool Based on Texts and its Themes." In 2024 4th International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE), pp. 1528-1533. IEEE, 2024.

[5].Srivarshini, S., and V. Lizy. "Literary Analysis of Anthropophobia with Natural Language Processing (NLP): A Generative AI Approach." In 2025 10th International Conference on Smart Structures and Systems (ICSSS), pp. 1-6. IEEE, 2025.

[6].Biswas, Priyanka. "An AI-Driven Framework for Computational Literary Analysis: Bridging English Literature and Technology." In 2025 IEEE 5th International Conference on ICT in Business Industry & Government (ICTBIG), pp. 1-6. IEEE, 2025.

[7].Al Maruf, Abdullah, Fahima Khanam, Md Mahmudul Haque, Zakaria Masud Jiyad, Muhammad Firoz Mridha, and Zeyar Aung. "Challenges and opportunities of text-based emotion detection: A survey." IEEE access 12 (2024): 18416-18450.

[8].Gal, Zoltan, and Erzsébet Tóth. "Deep learning-based analysis of ancient Greek literary texts in English version: A statistical model based on word frequency and noise probability for the classification of texts." Infocommunications Journal 16, no. Specia (2024): 2-11.

[9].Wang, Ming, Ruiyang Ma, Geoffrey Qiping Shen, and Jin Xue. "How large language models empower the analysis of online public engagement for mega infrastructure projects: Cases in Hong Kong." IEEE Transactions on Engineering Management (2025)

[10].Ali, Muhammad Hamza, Mahbuba Afrin, Redowan Mahmud, and Aneesh Krishna. "Cross-Layered Sentiment Analysis for Identifying Learner Intent in AI Chatbots." IEEE Transactions on Technology and Society (2025).

[11].Lescano, Germán, Juan Manuel Rodriguez, and Rosanna Costaguta. "Deep Learning Techniques for Text-based Emotional Response Generation: A Systematic Review." IEEE Transactions on Affective Computing (2025).

[12].Sharma, Riya, Balraj Singh, and Aditya Khamparia. "Machine learning and generative AI techniques for sentiment analysis with applications." Generative Artificial Intelligence for Biomedical and Smart Health Informatics (2025): 183-208.

[13].Bravo, Luis, Ciro Rodriguez, Pedro Hidalgo, and Cesar Angulo. "A systematic review on artificial intelligence-based multimodal dialogue systems capable of emotion recognition." Multimodal Technologies and Interaction 9, no. 3 (2025): 28.