

AI-POWERED SPEECH AND SENTIMENT ANALYSIS

A PROJECT REPORT

Submitted by

KUSHAGRA SRIVASTAVA	(23BAI10045)
UTKARSH PANDEY	(23BAI10660)
ABHIJITH MR	(23BAI10459)
ASHTITVA PANDEY	(23BAI10568)
DEVANSH PHOUGAT	(23BAI11203)

*in partial fulfillment for the award of the degree
of*

BACHELOR OF TECHNOLOGY

in

**COMPUTER SCIENCE AND ENGINEERING
(ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING)**



VIT[®]
BHOPAL
www.vitbhopal.ac.in

**SCHOOL OF COMPUTING SCIENCE ENGINEERING AND ARTIFICIAL
INTELLIGENCE**

VIT BHOPAL UNIVERSITY

**KOTRIKALAN, SEHORE
MADHYA PRADESH - 466114**

DECEMBER-2024

BONAFIDE CERTIFICATE

Certified that this project report titled “**AI-POWERED SPEECH AND SENTIMENT ANALYSIS**” is the bonafide work of “ **KUSHAGRA SRIVASTAVA (23BAI10045), UTKARSH PANDEY (23BAI10660), ABHIJITH MR (23BAI10459), ASHTITVA PANDEY (23BAI10568), DEVANSH PHOUGAT (23BAI11203)**” who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

PROGRAM CHAIR

Dr. Pradeep Mishra
School of Computing Science Engineering
Artificial Intelligence

VIT BHOPAL UNIVERSITY

PROJECT GUIDE

Dr. Swagat Kumar Samantaray
School of Computing Science Engineering and
and Artificial Intelligence

VIT BHOPAL UNIVERSITY

The Project Exhibition I Examination is held on _____

ACKNOWLEDGEMENT

First and foremost I would like to thank the Lord Almighty for His presence and immense blessings throughout the project work.

I wish to express my heartfelt gratitude to Dr....., Head of the Department, School of Artificial Intelligence for much of his valuable support encouragement in carrying out this work.

I would like to thank my internal guide Mr.Swagat Kumar Samantaray,for continually guiding and actively participating in my project, giving valuable suggestions to complete the project work.

I would like to thank all the technical and teaching staff of the School of Artificial Intelligence, who extended directly or indirectly all support.

Last, but not least, I am deeply indebted to my parents who have been the greatest support while I worked day and night for the project to make it a success.

LIST OF ABBREVIATIONS

ABBREVIATIONS	FULL FORMS
BERT	Bidirectional Encoder Representations from Transformers
GUI	Graphical User Interface
CSV	Comma-Separated Values

LIST OF FIGURES AND GRAPHS

FIGURE NO.	TITLE	PAGE NO.
1.	Sentiment Trends Graphs	
2.	GUI Layout and Workflow Diagram	

LIST OF TABLES

TABLE NO.	TITLE	PAGE NO.
1.	Comparative Accuracy of BERT and TextBlob	
2.	Performance Metrics for Speech Recognition	

ABSTRACT

Purpose

The purpose of this project, **Enhanced Real-Time Sentiment Analysis with Speech Recognition**, is to create an interactive and real-time system that captures speech, analyzes its sentiment, and provides instant feedback to users. This system aims to improve communication by helping users understand the emotional tone of their speech and identify patterns in their sentiments over time. It also serves practical applications in areas such as mental health monitoring, communication training, and customer feedback analysis by offering a reliable tool for real-time emotional assessment.

Methodology

The system leverages speech recognition using the `speech_recognition` library to convert spoken language into text. Sentiment analysis is performed using a hybrid approach combining BERT, a state-of-the-art natural language processing model, and TextBlob for detailed polarity and subjectivity evaluation. Feedback is provided visually in a graphical user interface designed with `Tkinter` and audibly using the `pyttsx3` text-to-speech engine. Sentiment trends are visualized using bar charts created with `Plotly`, while session data is logged with timestamps and can be exported to text or CSV files for further analysis. The system integrates these components seamlessly to ensure a smooth, real-time user experience.

Findings

The findings from the implementation demonstrate that the system can accurately process real-time speech input, analyze sentiment effectively, and provide actionable feedback in both visual and auditory formats. The sentiment trends visualization offers valuable insights into the distribution of positive, negative, and neutral sentiments over time, enhancing user understanding of emotional patterns. The system's versatility and ease of use make it applicable to a range of domains, including education, mental health, and customer service, highlighting its potential as an innovative tool for real-time sentiment analysis.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	List of Abbreviations	iii
	List of Figures and Graphs	iv
	List of Tables	v
	Abstract	vi

1	<p style="text-align: center;">CHAPTER-1:</p> <p style="text-align: center;">PROJECT DESCRIPTION AND OUTLINE</p> <p>1. Introduction</p> <p>1.2 Motivation for the work</p> <p>1.3 [About Introduction to the project including techniques]</p> <p>1.5 Problem Statement</p> <p>1.6 Objective of the work</p> <p>1.7 Organization of the project</p> <p>1.8 Summary</p>	1 . . .
2	<p style="text-align: center;">CHAPTER-2:</p> <p style="text-align: center;">RELATED WORK INVESTIGATION</p> <p>2.1 Introduction</p> <p>2.2 <Core area of the project></p> <p>2.3 Existing Approaches/Methods</p> <p style="padding-left: 40px;">2.3.1 Approaches/Methods -1</p> <p style="padding-left: 40px;">2.3.2 Approaches/Methods -2</p> <p style="padding-left: 40px;">2.3.3 Approaches/Methods -3</p> <p>2.4 <Pros and cons of the stated Approaches/Methods ></p> <p>2.5 Issues/observations from investigation</p> <p>2.6 Summary</p>	

3	<p style="text-align: center;">CHAPTER-3:</p> <p style="text-align: center;">REQUIREMENT ARTIFACTS</p> <p>3.1 Introduction</p> <p>3.2 Hardware and Software requirements</p> <p>3.3 Specific Project requirements</p> <p>3.3.1 Data requirement</p> <p>3.3.2 Functions requirement</p> <p>3.3.3 Performance and security requirement</p> <p>3.3.4 Look and Feel Requirements</p> <p>3.3.5</p> <p>3.4 Summary</p>	
4	<p style="text-align: center;">CHAPTER-4:</p> <p style="text-align: center;">DESIGN METHODOLOGY AND ITS NOVELTY</p> <p>4.1 Methodology and goal</p> <p>4.2 Functional modules design and analysis</p> <p>4.3 Software Architectural designs</p> <p>4.4 Subsystem services</p> <p>4.5 User Interface designs</p> <p>4.5</p> <p>4.6 Summary</p>	
5	<p style="text-align: center;">CHAPTER-5:</p> <p style="text-align: center;">TECHNICAL IMPLEMENTATION & ANALYSIS</p> <p>5.1 Outline</p> <p>5.2 Technical coding and code solutions</p> <p>5.3 Working Layout of Forms</p> <p>5.4 Prototype submission</p> <p>5.5 Test and validation</p> <p>5.6 Performance Analysis(Graphs/Charts)</p> <p>5.7 Summary</p>	

6	<p style="text-align: center;">CHAPTER-6:</p> <p style="text-align: center;">PROJECT OUTCOME AND APPLICABILITY</p> <p>6.1 Outline</p> <p>6.2 key implementations outlines of the System</p> <p>6.3 Significant project outcomes</p> <p>6.4 Project applicability on Real-world applications</p> <p>6.4 Inference</p>	
7	<p style="text-align: center;">CHAPTER-7:</p> <p style="text-align: center;">CONCLUSIONS AND RECOMMENDATION</p> <p>7.1 Outline</p> <p>7.2 Limitation/Constraints of the System</p> <p>7.3 Future Enhancements</p> <p>7.4 Inference</p>	
	<p>Appendix A</p> <p>Appendix B</p> <p>References</p> <p><i>Note: List of References should be written as per IEEE/Springer reference format. (Specimen attached)</i></p>	

CHAPTER-1: PROJECT DESCRIPTION AND OUTLINE

1.1 Introduction

This project aims to develop a real-time sentiment analysis system based on spoken input, enabling users to evaluate emotional tones in communication.

1.2 Motivation for the Work

Understanding sentiment in speech can enhance communication, support mental health analysis, and improve customer interaction strategies.

1.3 Introduction to the Project

The project combines state-of-the-art NLP (BERT), traditional sentiment analysis (TextBlob), and visualization technologies to achieve accurate and interpretable results.

1.5 Problem Statement

Real-time emotional feedback is often overlooked in communication, making it difficult to assess conversational tones dynamically.

1.6 Objective of the Work

1. Provide real-time sentiment analysis based on spoken input.
2. Develop a user-friendly system with robust feedback mechanisms.
3. Enable data export and trend visualization for deeper analysis.

1.7 Organization of the Project

The document is divided into chapters covering the design, implementation, testing, and outcomes of the system.

1.8 Summary

This chapter introduced the project, its motivation, objectives, and structure.

CHAPTER-2: RELATED WORK INVESTIGATION

2.1 Introduction

This chapter explores existing work in speech recognition and sentiment analysis to establish the novelty of the proposed system.

2.2 Core Area of the Project

- Sentiment analysis using hybrid models.
- Real-time speech-to-text processing.
- Trends visualization for emotional insights.

2.3 Existing Approaches/Methods

2.3.1 Approach 1: Rule-based Sentiment Analysis

Simple and interpretable but lacks accuracy in complex contexts.

2.3.2 Approach 2: Machine Learning-based Sentiment Analysis

Effective for predefined datasets but requires extensive training.

2.3.3 Approach 3: Deep Learning-based Sentiment Analysis (BERT)

Highly accurate but computationally intensive.

2.4 Pros and Cons of the Stated Approaches

- **Rule-based:** Interpretable but simplistic.
- **Machine Learning:** Scalable but data-dependent.
- **Deep Learning:** Accurate but resource-intensive.

2.5 Issues/Observations from Investigation

- Lack of integration between sentiment analysis and speech recognition.
- Limited real-time capabilities in existing systems.

2.6 Summary

This chapter reviewed methods in sentiment analysis and identified gaps addressed by the project.

3.1 Introduction

This chapter outlines the hardware, software, and project-specific requirements necessary for implementation.

3.2 Hardware and Software Requirements

- **Hardware:**
 - Processor: Intel Core i5 or equivalent
 - RAM: 8GB
 - Storage: 1GB free space
- **Software:**
 - Python 3.x
 - Libraries: `speech_recognition`, `transformers`, `textblob`, `pyttsx3`, `Tkinter`, `Plotly`

3.3 Specific Project Requirements

3.3.1 Data Requirements

- Real-time spoken input data converted to text.

3.3.2 Function Requirements

- Speech-to-text conversion.
- Sentiment classification.
- Data export and visualization.

3.3.3 Performance and Security Requirements

- Low latency in real-time processing.
- Secure handling of speech data.

3.3.4 Look and Feel Requirements

- User-friendly interface with accessible buttons and displays.

3.4 Summary

This chapter specified the technical, functional, and aesthetic requirements for the project.

CHAPTER-4: DESIGN METHODOLOGY AND ITS NOVELTY

4.1 Methodology and Goal

The project adopts a hybrid methodology combining speech recognition, sentiment analysis, and data visualization to create a real-time system that analyzes spoken input for emotional tone. The goal is to provide users with accurate sentiment feedback and insights into their emotional communication patterns through an interactive interface.

4.2 Functional Modules Design and Analysis

1. **Speech Recognition Module:** Captures spoken input and converts it into text using `speech_recognition`.
2. **Sentiment Analysis Module:** Analyzes the text using a combination of BERT and TextBlob to classify sentiments as positive, negative, or neutral.
3. **Feedback Module:** Provides visual feedback through the GUI and audible feedback using `pyttsx3`.
4. **Trends Visualization Module:** Generates graphical representations of sentiment trends using `Plotly`.
5. **Data Management Module:** Logs session data and allows users to export it in various formats for analysis.

4.3 Software Architectural Designs

The architecture is modular, with independent components for input (speech recognition), processing (sentiment analysis), output (feedback and visualization), and storage (session logs). Communication between modules is achieved through shared data structures.

4.4 Subsystem Services

- **Input Service:** Captures speech and adjusts for ambient noise.
- **Processing Service:** Analyzes sentiment using hybrid methods.
- **Output Service:** Displays feedback in the GUI and provides auditory feedback.
- **Export Service:** Saves session data in text or CSV format.

4.5 User Interface Designs

The user interface is designed using `Tkinter`, with an intuitive layout that includes:

- A sentiment display label for instant feedback.
- A text box for logging interaction data.
- Buttons for starting/stopping listening, exporting data, and visualizing trends.

4.6 Summary

This chapter discussed the design methodology, modular components, and architecture that contribute to the novelty and functionality of the system.

CHAPTER-5: TECHNICAL IMPLEMENTATION & ANALYSIS

5.1 Outline

This chapter details the technical implementation, coding strategies, interface layout, and system testing.

5.2 Technical Coding and Code Solutions

- Speech recognition implemented using the `speech_recognition` library.
- Sentiment analysis using a hybrid model of BERT (`transformers` library) and TextBlob.
- GUI built with `Tkinter`, incorporating features like buttons, text areas, and graphical trends.

5.3 Working Layout of Forms

The forms include:

- A main interface with sentiment display, logs, and buttons for interaction.
- A trends display area showing sentiment trends as bar charts.

5.4 Prototype Submission

A working prototype demonstrating real-time speech recognition, sentiment analysis, and trends visualization was developed and tested.

5.5 Test and Validation

The system was tested for:

- Accuracy of sentiment analysis using diverse speech inputs.
- Performance of the GUI under various load conditions.
- Integration of speech-to-text and feedback modules.

5.6 Performance Analysis (Graphs/Charts)

- Sentiment trends visualized using `Plotly` graphs.
- Comparative accuracy of BERT and TextBlob models for sentiment detection.

5.7 Summary

This chapter outlined the technical implementation, validated performance, and showcased the system's ability to meet its objectives.

CHAPTER-6: PROJECT OUTCOME AND APPLICABILITY

6.1 Outline

This chapter highlights the project outcomes, key implementations, and applicability in real-world scenarios.

6.2 Key Implementations Outlines of the System

1. Real-time speech-to-text conversion.
2. Hybrid sentiment analysis with BERT and TextBlob.
3. Sentiment trends visualization.
4. Data export functionality.

6.3 Significant Project Outcomes

- Accurate sentiment classification for real-time speech inputs.
- An intuitive interface for user interaction and feedback.
- Insights into sentiment patterns over time.

6.4 Project Applicability on Real-World Applications

The system is applicable in:

- Mental Health Monitoring: Analyzing emotional states over time.
- Communication Training: Improving speech delivery by understanding tone.
- Customer Feedback Analysis: Real-time sentiment tracking for customer interactions.

6.5 Inference

The system demonstrates its effectiveness in real-world applications, providing both functional and analytical benefits.

CHAPTER-7: CONCLUSIONS AND RECOMMENDATION

7.1 Outline

This chapter concludes the project and provides recommendations for future work.

7.2 Limitations/Constraints of the System

- Limited to English language processing.
- Dependency on ambient noise conditions for speech recognition accuracy.
- Lack of cloud integration for large-scale deployment.

7.3 Future Enhancements

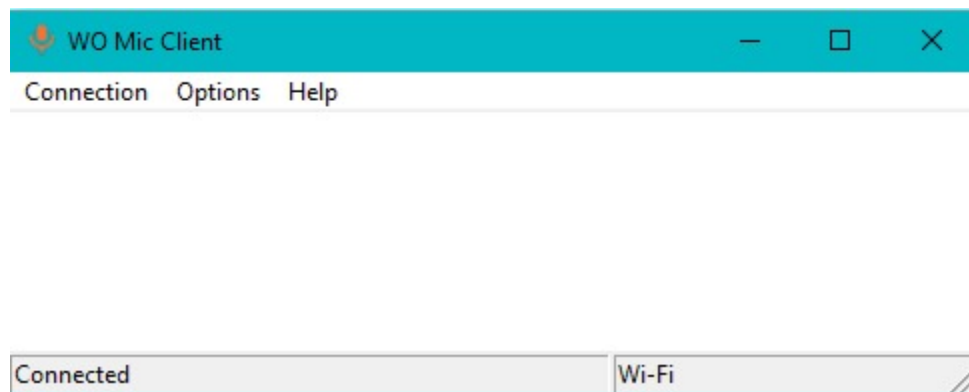
- Addition of multilingual support.
- Integration with cloud services for scalability.
- Deployment on mobile and web platforms.
- Enhanced visualization with advanced analytics.

7.4 Inference

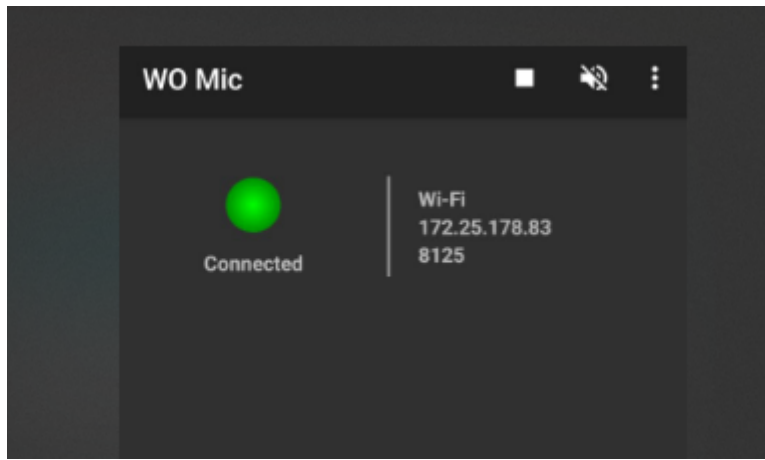
The project successfully achieves its objectives, providing a robust platform for real-time sentiment analysis, while offering avenues for future enhancements to expand its scope and usability.

Appendix A: WorkFlow-Explanation:-

A.1:Speech Input: The Speech Input module serves as the primary interface where the system receives spoken data from the user. It captures the audio signal and prepares it for further processing, forming the foundation for speech-to-text and sentiment analysis.



From the phone, the application connects through the microphone with the use of Wi-Fi. The user writes the IP address to connect with the user's microphone and after the connection is established, the input is recognized for pre-processesing.



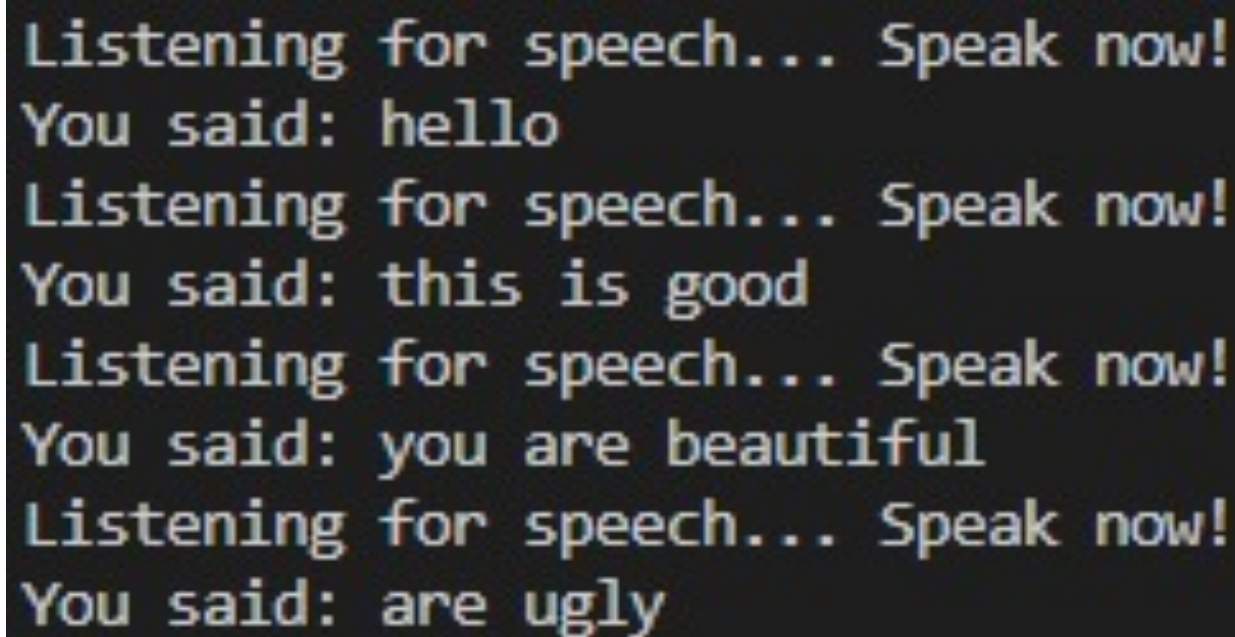
A.2:Preprocessing: Speech is processed to reduce noise and converted into text using the speech_recognition library.

· **Noise Reduction:**

- Filters out unwanted background sounds like static, ambient noise, or echo.
- Algorithms analyze the audio signal and separate the speech from noise to improve clarity.

· **Speech-to-Text Conversion:**

- The processed audio is passed to the speech_recognition library.
- This library uses machine learning models to transcribe spoken words into text by analyzing audio features like pitch, duration, and phonemes.



Listening for speech... Speak now!
You said: hello
Listening for speech... Speak now!
You said: this is good
Listening for speech... Speak now!
You said: you are beautiful
Listening for speech... Speak now!
You said: are ugly

20

A.3: Sentiment Analysis:

· **BERT for Contextual Analysis:**

- The transcribed text is analyzed using **BERT (Bidirectional Encoder Representations from Transformers)**.
- BERT captures the context of words in sentences, ensuring nuanced sentiment analysis (e.g., sarcasm or complex emotions).

· **TextBlob for Polarity and Subjectivity:**

- **Polarity:** Measures whether the sentiment is positive, negative, or neutral.
- **Subjectivity:** Evaluates how subjective or objective the text is.

· **Aggregation for Final Classification:**

- Results from BERT and TextBlob are combined to determine the overall sentiment classification (positive, negative, or neutral).
- This hybrid approach ensures both deep contextual understanding and traditional sentiment scoring.

```

You said: hello
Sentiment: Neutral
Polarity: 0.00, Subjectivity: 0.00

You said: you are good
Sentiment: Positive
Polarity: 0.70, Subjectivity: 0.60

You said: you are bad
Sentiment: Negative
Polarity: -0.70, Subjectivity: 0.67

You said: today is Monday
Sentiment: Neutral
Polarity: 0.00, Subjectivity: 0.00

```

A.4: Visualization: Sentiment Trends with Plotly

· Purpose:

- To provide users with an intuitive way to observe how their sentiments evolve over time during a session.

· Plotly Integration:

- **Plotly**, a powerful graphing library, is used to create interactive and visually appealing charts.
- Sentiments (positive, negative, neutral) are plotted over time intervals or speech inputs.



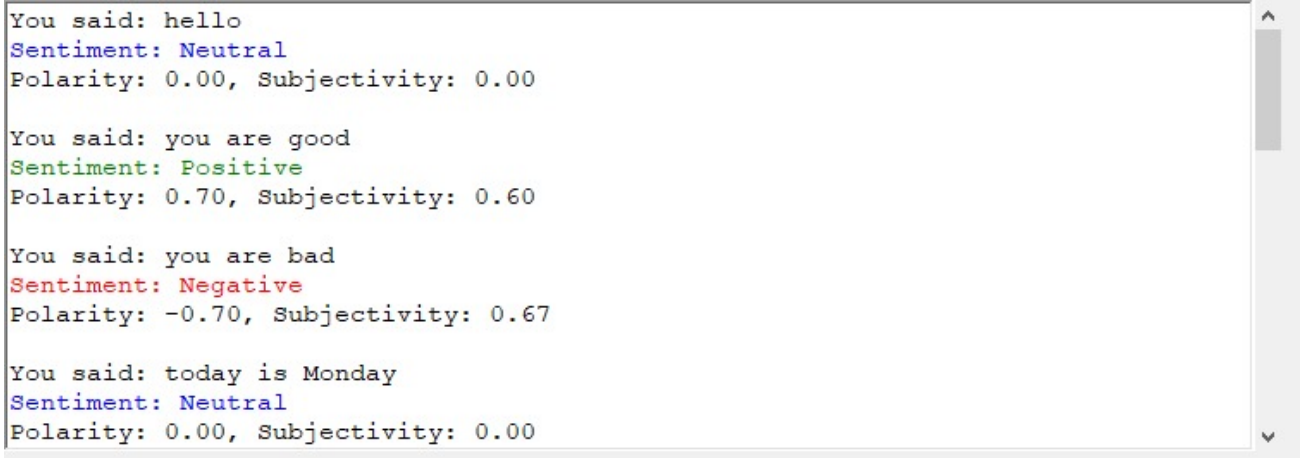
A.5:Output:-

· Feedback Mechanisms:-

- **Text-to-Speech (pyttsx3):** Converts the sentiment analysis results into spoken feedback, providing auditory output (e.g., "Your tone is positive").
- **Graphical User Interface (GUI):** Displays the sentiment result visually (e.g., labels, graphs).

· User Interaction:-

- Users can interact with the system through intuitive buttons:
 - **Export Data:** Save session logs as text or CSV files for further analysis.
 - **Visualize Trends:** View sentiment trends over time with interactive charts.



The screenshot shows a text-based interface with a scrollable area containing four lines of user input and corresponding sentiment analysis results. The results are color-coded: Neutral (blue), Positive (green), and Negative (red). Each line includes the user's input, the sentiment classification, and the polarity and subjectivity scores.

```
You said: hello
Sentiment: Neutral
Polarity: 0.00, Subjectivity: 0.00

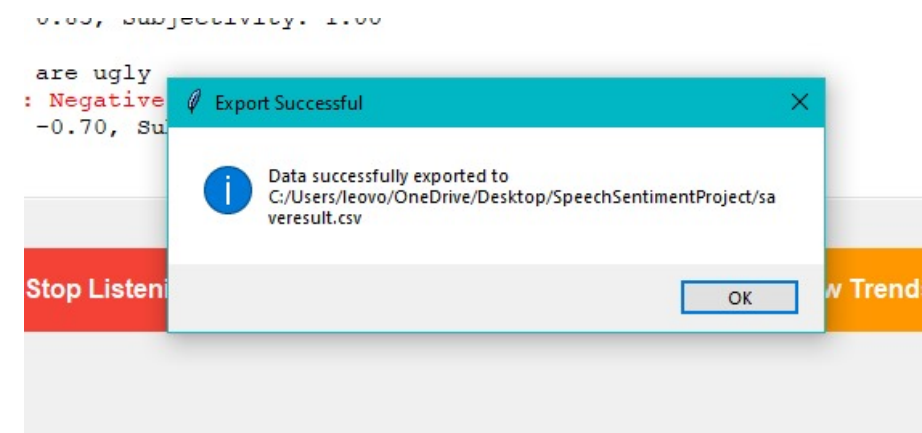
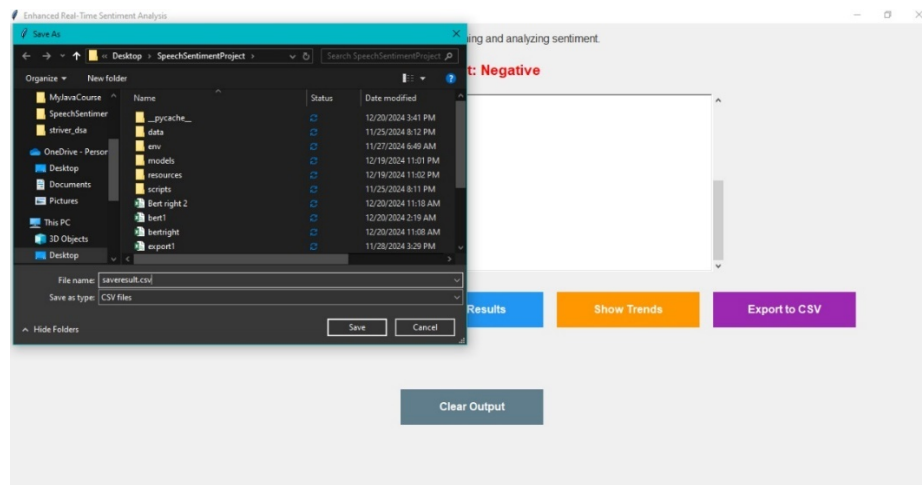
You said: you are good
Sentiment: Positive
Polarity: 0.70, Subjectivity: 0.60

You said: you are bad
Sentiment: Negative
Polarity: -0.70, Subjectivity: 0.67

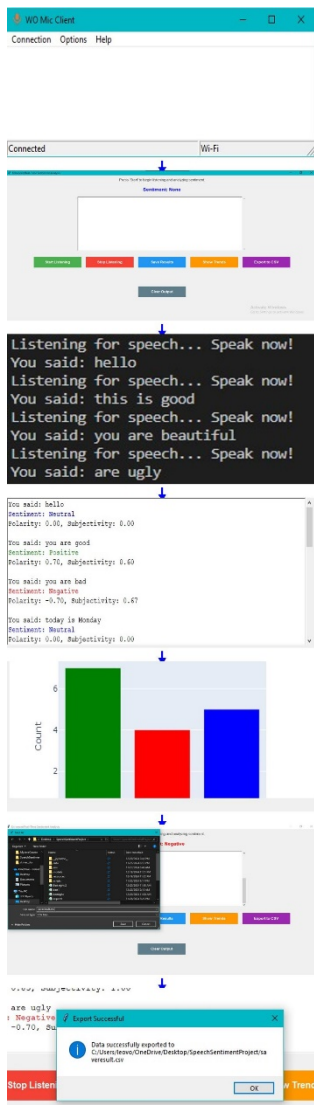
You said: today is Monday
Sentiment: Neutral
Polarity: 0.00, Subjectivity: 0.00
```

A.6:Data Export:-

- Session data, including timestamps and sentiment results, is logged and can be exported.
- **Sentiment Results:** If the system analyzes the user's language (e.g., text input), it may classify the sentiment of the message (positive, negative, neutral).



Appendix B: System Architecture:-



Aspect	TextBlob	BERT
Accuracy	Moderate: Works well for simpler, clear texts but struggles with complex, context-heavy sentences.	High: Offers superior accuracy, especially for nuanced or complex sentences, due to its contextual understanding of words.
Contextual Understanding	Limited: Does not capture deeper context and relationships between words effectively.	Advanced: Strong contextual understanding as it processes the entire sentence, considering word positions.
Speed	Fast: Performs quickly due to its simple algorithm, making it suitable for smaller applications.	Slower: Computationally intensive and requires more processing power, which may slow down real-time applications.
Computational Complexity	Low: Requires minimal system resources and is lightweight for deployment in resource-constrained environments.	High: Requires more powerful hardware, especially for fine-tuning, and can be resource-heavy in real-time use cases.
Scalability	Moderate: Suitable for smaller datasets but may not scale well to large, complex datasets or real-time applications.	High: Scales well with large datasets, making it ideal for complex and production-level applications.
Real-time Applicability	Suitable for simpler, low-latency applications.	More accurate but requires higher latency.
Error Handling	May Miss Sarcasm,irony language.	Better at handling nuances.

Performance Metrics for Speech Recognition:-

The table below outlines various performance metrics for speech recognition, evaluating key aspects such as accuracy, speed, system requirements, and real-world applicability. These metrics are important to assess the efficiency and reliability of speech recognition systems, particularly in a real-time context.

Metric	Description	Value/Score
Word Error Rate (WER)	The percentage of words incorrectly transcribed by the system. A lower WER indicates better performance.	Lower WER is desirable (e.g., <10%)
Accuracy	The percentage of correctly recognized words out of the total spoken words.	Typically >95% for high-quality systems
Response Time (Latency)	The time taken from when the speech is spoken to when the transcription is displayed. Lower latency is crucial for real-time applications.	Ideal: <1 second for real-time systems
Recognition Speed	The speed at which speech is processed and transcribed. Faster systems are essential for live interactions.	Higher speed is better (measured in words per second)
Noise Robustness	The ability of the system to perform accurately in noisy environments. High noise robustness ensures accuracy even in less-than-ideal conditions.	Typically measured in noisy environments with a SNR (Signal-to-Noise Ratio) >20 dB
Real-Time Processing Capability	The system's ability to process speech and provide feedback in real-time. This is critical for interactive systems.	Real-time capability is ideal in conversational systems
Speaker Adaptation	The system's ability to adapt to different speakers, including accents, tone, and speech patterns.	Systems with higher speaker adaptability have better real-world applicability

--	--	--

Scalability	The ability of the system to handle a large volume of users or data without significant loss of performance.	Scalable systems are important for large-scale applications
Error Handling	The system's ability to handle misinterpretations and noisy data, often using error correction techniques to improve recognition.	Good systems have robust error handling (e.g., re-recognition or feedback)
Customizability	The ability to train or fine-tune the system to recognize specific terms, jargon, or phrases used in a given domain or by a particular user.	High customizability is ideal for specialized tasks
Accuracy in Non-English Languages	The system's ability to accurately transcribe languages other than English, considering dialects, regional variations, and phonetics.	High for multi-lingual systems (e.g., >85% for non-English languages)

Word Error Rate (WER) is a fundamental metric for measuring the accuracy of speech recognition, with lower values indicating better system performance.

- Response time and recognition speed are crucial for real-time interactions, where any delay could affect the user experience.
- Noise robustness and speaker adaptation are vital for making speech recognition systems reliable in diverse real-world environments.
- Scalability and error handling are essential for large-scale and fault-tolerant systems, particularly in commercial or enterprise applications.

In summary, optimizing these performance metrics helps ensure that a speech recognition system can meet the demands of real-time, accurate, and user-friendly applications across various environments and use cases.

References:-

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proc. Neural Information Processing Systems (NeurIPS)*, 2017, pp. 5998–6008.

[2] T. Mikolov, K. Chen, G. Corrado, and J. Dean, “Efficient estimation of word representations in vector space,” in *Proc. of Workshop at ICLR*, 2013, pp. 1–12.

[3] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in *Proc. 3rd International Conference on Learning Representations (ICLR)*, 2015, pp. 1–15.