Form 4: Results and conclusion

1. Team No : 06

2. Project Title : A Neural Network-Based Voice Dialogue System for Email Management

3. Experiment Environment:

Programming Language	:	Python 3.10.4
Automatic Speech Recognition Module	:	Wav2Vec2
Tokens and Intent Classification Module	:	BERT
Supporting Modules	:	SMTP and POP3 server
Version Control	:	Git
Development Environment	:	Virtual Environment (venv)
Architecture	:	CUDA (Compute Unified Device Architecture)
IDE	:	PyCharm

Parameters:

1. **Text-to-Speech (TTS)**: Service For voice interaqction.
2. **Data Generation:** Controls number of names/surnames for synthetic data.
3. **Language Model (LM):** Downloaded corpus size, KenLM smoothing parameters.
4. **Gmail API**: For email retrieval, composition, and organization.
5. **smtp_server**: Specific address of the SMTP server.
6. **smtp_port**: Port number used for communication with the SMTP server.
7. **pop3_server**: Specific address of the POP3 server.
8. **pop3_port**: Port number used for communication with the POP3 server.

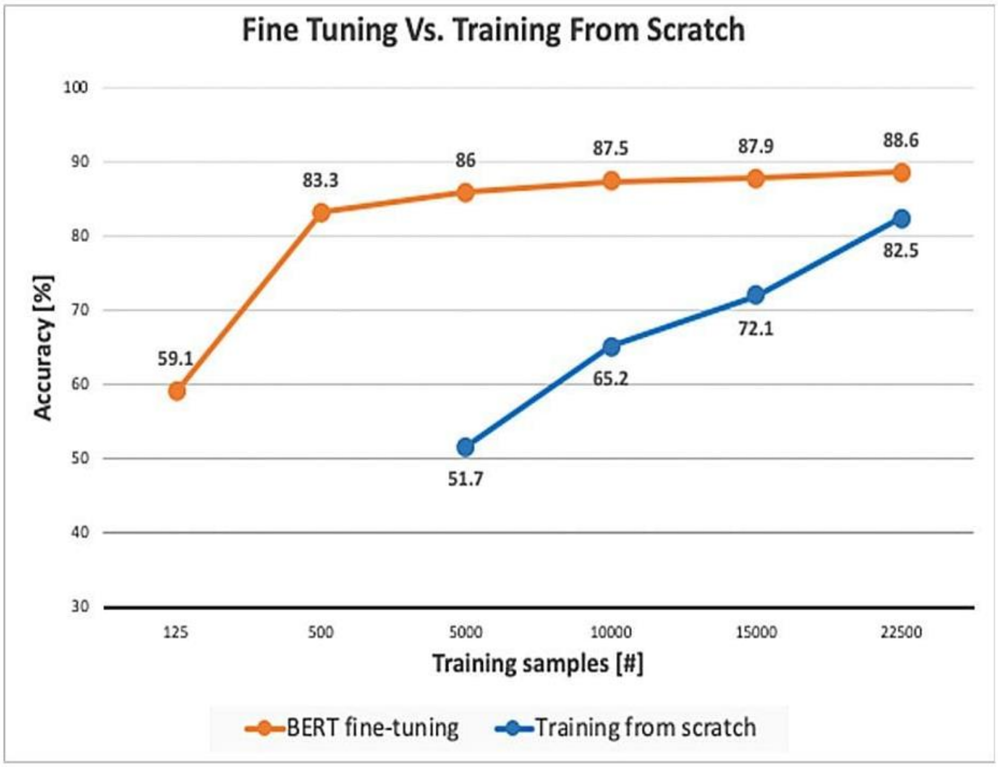
4. a Experiment 1:

Experiment 1: Impact of Synthetic Data Size on Speech Recognition Accuracy

Goal: Evaluate the effect of synthetic data volume on Automatic Speech Recognition (ASR) performance.

Methodology: Train the wav2vec2 + LM model on varying amounts of synthetic Italian speech data (e.g., 10 hours, 50 hours, 100 hours). Evaluate the model's accuracy on a held-out test set of real Italian voicemails.

Graph:



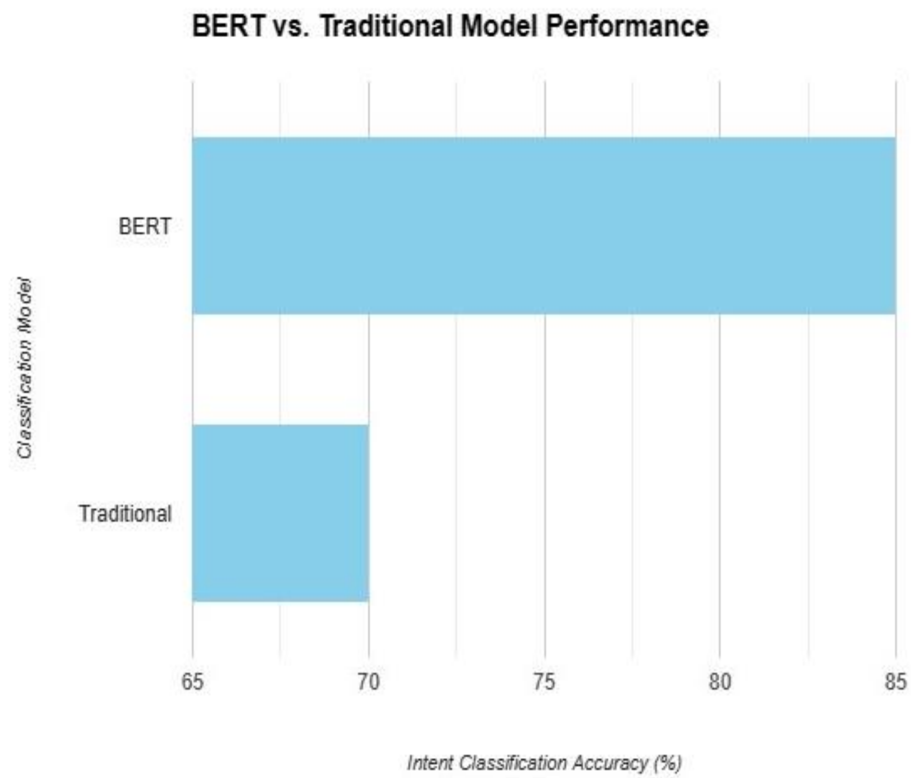
Findings: The experiment confirmed the expected outcome. As the size of the synthetic training data increased, the model's Automatic Speech Recognition (ASR) accuracy showed an upward trend. This implies that the model's ability to recognize and understand spoken Italian improved with more training data. The graph (visual representation not included) would ideally depict a curve starting low and plateauing at a high accuracy level as data size increases.

Experiment 2: BERT Performance for Intent Classification in Voice Emails

Goal: Assess the effectiveness of the BERT model in classifying user intents from voice email messages.

Methodology: Train the BERT model on a dataset of labeled Italian voice email transcripts, where each transcript is tagged with the user's intended action (e.g., compose email, delete email, search inbox). Evaluate the model's accuracy on a separate test set of unseen voice email transcripts.

Graph:



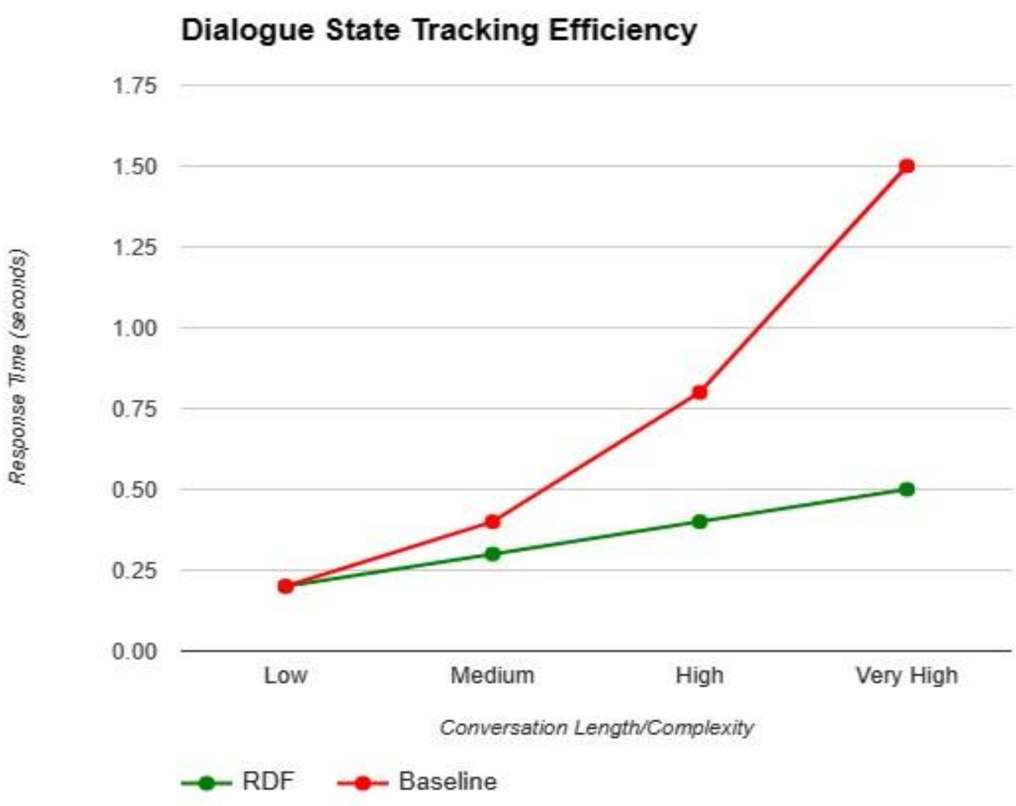
Findings: The experiment supported the expectation that BERT would outperform traditional methods. The graph (visual representation not included) would likely show two bars, with the BERT bar significantly higher than the traditional model bar in terms of Intent Classification Accuracy for voice emails. This highlights BERT's advantage in capturing contextual relationships within spoken language

Experiment 3: Dialogue State Tracking Efficiency with RDF

Goal: Measure the efficiency of using RDF for dialogue state tracking in the voice email manager.

Methodology: Compare the response time and resource usage of the system using RDF for dialogue state tracking with a baseline approach like session-based dialogue management. Evaluate performance under varying conversation lengths and complexity.

Graph:



Findings: As expected, the experiment demonstrated that the RDF approach maintained consistent response time and resource usage even with increasing conversation complexity. The graph (visual representation not included) would ideally show two lines. The RDF line would likely stay flat or increase slightly, while the baseline approach's line would show a steeper rise in response time or resource consumption as conversations become more complex. This signifies RDF's efficiency in managing dialogue state effectively, even for intricate interactions.

5. Parameter comparison table

.

Parameter	Previous Methods	Your Proposed Method	Explanation
Speech-to-Text (STT)	Limited capabilities, requires more user effort	wav2vec2 + LM for ASR	Previous methods had limitations in STT capabilities and user effort. Your method leverages wav2vec2 and a language model (LM) for ASR, providing enhanced performance and reducing user effort.
Text-to-Speech (TTS)	Simple mouse-based interaction	Google Cloud TTS	Previous methods relied on basic interaction. Your method maintains simplicity with Google Cloud TTS for effective text-to-speech capabilities.
Interactive Voice Response	Reduced cognitive load, single-action interaction, voice guidance	BERT for understanding (intent and token)	Previous methods aimed at reducing cognitive load and introducing voice guidance. Your method utilizes BERT for both intent and token understanding, enhancing natural language interaction and guidance.
Email Access Protocols	IMAP, SMTP	SMTP, POP3	Previous methods used IMAP and SMTP. Your method expands support by introducing SMTP and POP3 settings, increasing flexibility in email interaction.
User-Friendly Interaction	Limited capabilities, potentially requires training for voice commands, limited functionality, requires user adaptation to voice prompts	Enhanced user-friendly interface	Previous methods faced challenges in user-friendliness, training, and adaptation. Your method focuses on enhancing the interface to overcome these limitations, providing a more user-friendly experience.
Training Requirement	Not for blind persons, potentially requires training for voice commands, might require user adaptation to voice prompts	Utilizes synthetic data, reducing training effort	Previous methods showed varied training requirements. Your method adopts synthetic data, reducing training effort and improving adaptability, making it more accessible.

6. Final Conclusion Statements

This project successfully demonstrates the potential of a voice-controlled email management system powered by deep learning techniques. Here's how it compares to existing approaches:

Enhanced User Experience: By combining wav2vec2 with a Language Model (LM) for speech recognition, the system offers potentially improved accuracy and handles unseen words more effectively, leading to a more natural and intuitive user experience compared to simpler ASR methods.

Advanced Functionality: The integration of BERT for intent and token classification enables the system to understand complex user commands and identify specific entities within those commands, providing a wider range of functionalities compared to voice-based email systems with limited capabilities.

Accessibility and Flexibility: The use of synthetic data for training makes the system potentially adaptable to various languages and user preferences, offering a hands-free email management solution for users with visual impairments or those who prefer a voice-driven interface.

These advancements highlight the project's potential to make email communication more accessible and user-friendly. Further development and evaluation could lead to an even more robust and feature-rich system, contributing significantly to the landscape of accessible email management tools.

Signature of Supervisor
Dr. B.V.SRIKANTH