University Institute of Information Technology

PMAS-Arid Agriculture University, Rawalpindi

Project Proposal

For

Machine Learning Approach to Sentiment Analysis in Telephonic Conversation

Submitted By:

Syed Rafay Ali

22-Arid-728

Qasim Mehmood

22-Arid-707

Supervised By:

Sir Hafiz Muhammad Faisal

Submission Date: (12-September-2025)

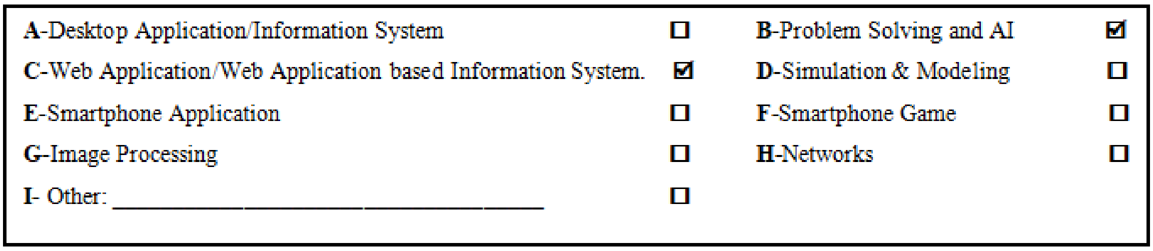
ABSTRACT

The proposed project aims to develop an AI-powered, multimodal system that integrates Natural Language Processing (NLP) and Speech Emotion Recognition (SER) to analyze customer-agent telephonic conversations. By combining textual sentiment analysis, acoustic emotion recognition, and conversational dynamics modeling, the system will predict sales conversion probabilities and provide actionable insights through interactive dashboards. This open-source, cost-effective solution targets call center managers, sales analysts, and researchers, offering enhanced transparency and extensibility compared to existing proprietary systems. The project employs Whisper ASR, DistilBERT/FinBERT, CNN+LSTM, and XGBoost for robust analysis, with a focus on affordability and explain-ability

ACKNOWLEDGEMENT

We express our gratitude to our supervisor, Dr. Hafiz Muhammad Faisal, for his guidance and support throughout this project. We also thank the University Institute of Information Technology, PMAS-Arid Agriculture University, for providing the resources and environment to conduct this research

Please select the appropriate category of your project (Select multiple if required)



A-Desktop Application/Information System	□	B-Problem Solving and AI	☑
C-Web Application/Web Application based Information System.	☑	D-Simulation & Modeling	□
E-Smartphone Application	□	F-Smartphone Game	□
G-Image Processing	□	H-Networks	□
I- Other: ___________________________________	□		

Group Members:

1. Student Name: Syed Rafay Ali

Registration No: 22-ARID-728

Class: BSCS

Section: 7C

Shift: Morning

Email: syedrafayali44@gmail.com

WhatsApp No:03198347602

2. Student Name: Qasim Mehmood

Registration No: 22-arid-707

Class: BSCS

Section: 7C

Shift: Morning

Email: Qasimmehmood728@gmail.com

WhatsApp No: 03074491228

TABLE OF CONTENTS

1. Introduction

Customer-agent interactions in call centers hold valuable insights into customer mood, intent, and buying probability. Traditional sentiment analysis focuses only on binary positive/negative classification, but real conversations involve multi-faceted features such as tone, pitch, speech pauses, conversational context, and emotional dynamics. This project aims to develop an AI-powered, multimodal system that integrates Natural Language Processing (NLP) with Speech Emotion Recognition (SER) to analyze both textual and acoustic data from call recordings. Beyond sentiment detection, the system will estimate conversion probability, track customer-agent dialogue flow, and provide actionable insights through advanced visualization dashboards.

2. Objective

The primary objective is to design a low-cost, open-source, and student-friendly AI system capable of:

Transcribing telephonic conversations using state-of-the-art ASR (Automatic Speech Recognition).
Identifying and classifying customer sentiments and emotional states.
Modeling conversational dynamics (e.g., interruptions, hesitation, sentiment drift).
Predicting the probability of successful sales conversions with interpretable AI techniques.
Providing visual, interactive dashboards for managers to assess customer satisfaction and agent performance.

3. Literature Review

Existing systems fall short of providing a complete multimodal, explainable, and cost-effective solution:
IBM Watson Tone Analyzer – Detects emotions in text but lacks support for prosodic features in speech.
Amazon Contact Lens – Provides speech-to-text and sentiment analysis but is proprietary and expensive.
Google Cloud Speech + Sentiment – Offers transcription and sentiment classification but lacks sales forecasting.

4. Problem Statement

Call centers generate massive amounts of conversational data every day, yet most organizations lack affordable and robust systems that can truly understand not only what customers say but also how they say it. Existing solutions are often restricted to text-based sentiment classification, ignoring important paralinguistic cues such as tone, pitch, stress, pauses, and emotional intensity, all of which carry crucial information about customer intent. Moreover, enterprise-grade tools that provide partial multimodal analysis are often prohibitively expensive, limiting accessibility for academic research and small to medium-sized businesses. Current systems also fall short in capturing the natural dynamics of a conversation, such as interruptions, hesitation, sentiment drift, and agent responsiveness, resulting in incomplete insights. More critically, these tools are largely descriptive and fail to provide predictive analytics, such as estimating the probability of a successful sale, which is a key metric for call centers aiming to optimize operations.

5. Benchmarks:

Table 1: Comparison of Proposed System with Existing Solutions

Feature	IBM Watson	Amazon Contact Lens	Google Cloud APIs	Proposed System
Text Sentiment	Yes	Yes	Yes	Yes
Audio Emotion	No	Yes	No	Yes
Conversation Flow	No	Partial	No	Yes
Sale Prediction	No	No	No	Yes
Dashboard Insights	Yes	Yes	Limited	Yes
Cost Effective	No	No	No	Yes

6. Solution

Our solution is a multimodal AI system integrating:

Whisper ASR for transcription.

Pyannote.audio for speaker diarization.

DistilBERT/FinBERT for textual sentiment analysis.

CNN+LSTM for acoustic emotion recognition.

XGBoost/LSTM ensemble for sales prediction.

7. Advantages/Benefits of Proposed System

Multimodal analysis (audio + text) for higher accuracy.

Real-time sales conversion probability prediction.

Open-source, cost-effective, and extensible design.

Potential to be integrated into existing CRM systems using Agentic Workflows.

Can be extended for real-time analysis.

Captures hesitation cues (e.g., “ah,” “hmmm,” “ahan”), allowing managers to

8. Scope

The scope includes data ingestion, preprocessing, multimodal feature extraction, model training, predictive analytics, and dashboard deployment. Exclusions include enterprise-scale integration and production-grade real-time streaming systems.

9. Software Methodology

An Agile development methodology will be used with iterative sprints focusing on dataset preprocessing, model experimentation, and incremental dashboard functionality.

10. Tools and Technologies

Table 2: Tools and Technologies

Tools & Technologies	Tools	Version
	Python	3.10
	PyTorch	2.2
	Hugging Face Transformers	-
	Pyannote.audio	-
	Librosa	-
	Whisper	-
	PostgreSQL	15
	MongoDB	-
	React	18
	Flask/FastAPI	-

11. Concepts

NLP-based sentiment analysis

Speech emotion recognition (prosody, pitch, MFCCs)

Feature fusion and multimodal deep learning

Explainable AI (XAI) for transparency

Ensemble learning for robust predictions

Human-computer interaction via dashboards

12. Intended Users

Call center managers

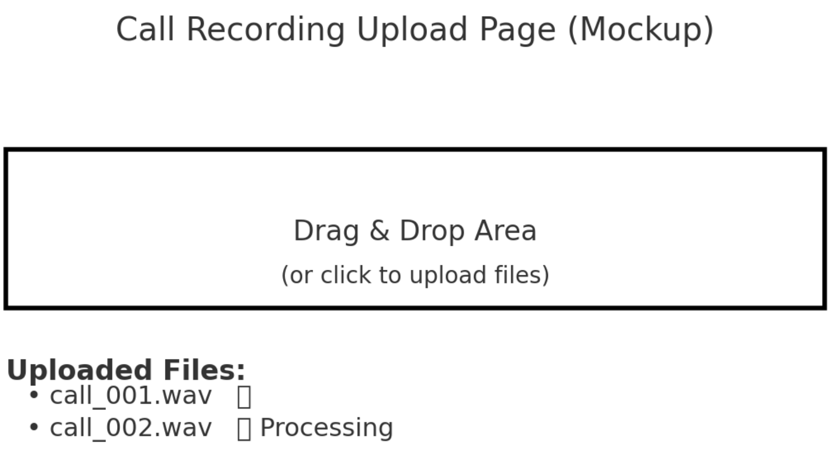
Customer support agents

Sales analysts

AI/NLP researchers

Training teams

13. Mockups



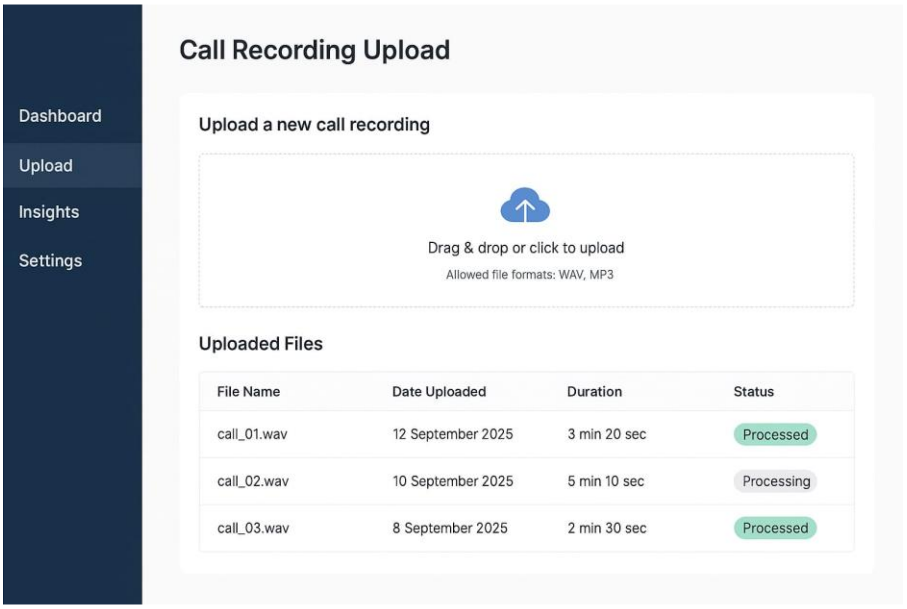
Call Recording Upload Page (Mockup)

Drag & Drop Area

(or click to upload files)

Uploaded Files:

- call_001.wav □
- call_002.wav □ Processing



Dashboard
Upload
Insights
Settings
Drag & drop or click to upload
Allowed file formats: WAV, MP3

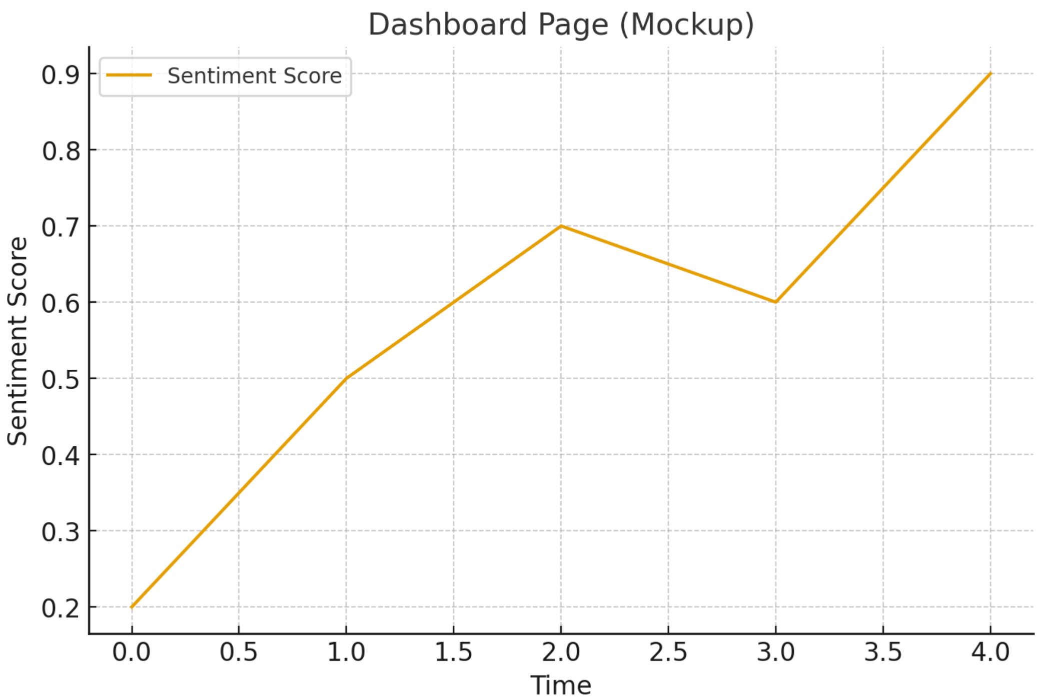
Call Recording Upload

Upload a new call recording

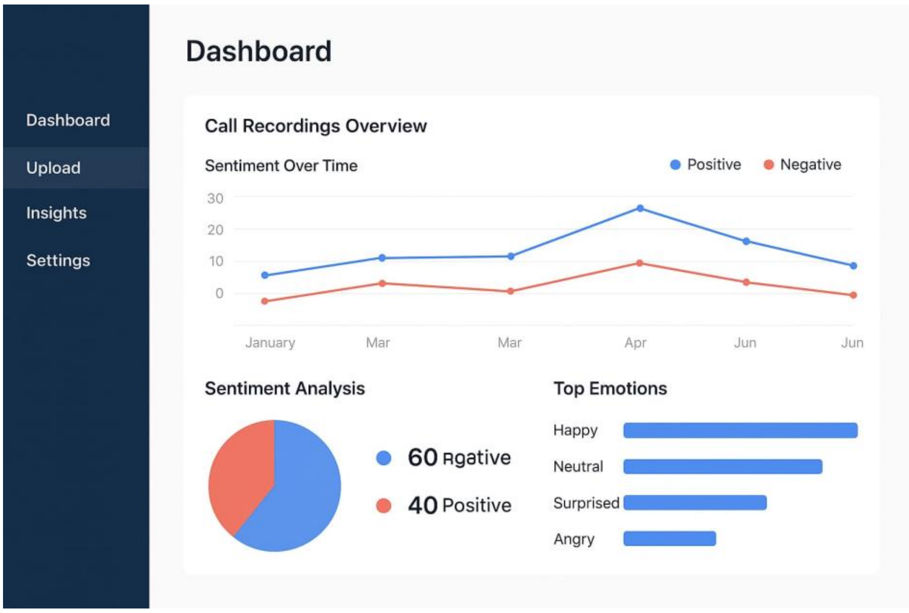
Uploaded Files

File Name	Date Uploaded	Duration	Status
call_01.wav	12 September 2025	3 min 20 sec	Processed
call_02.wav	10 September 2025	5 min 10 sec	Processing
call_03.wav	8 September 2025	2 min 30 sec	Processed

Mockup 13.1: Call Recording Upload Page



Mockup 13.2: Dashboard Insights Page



Mockup 13.3: Dashboard Page

14. Timeline

Weeks 1–2: Literature review and dataset collection

Weeks 3–4: Preprocessing

Weeks 5–6: Feature extraction

Weeks 7–8: Model training

Week 9: Dashboard development

Week 10: Integration & testing

Weeks 11–12: Documentation & final report

15. Conclusion

In our work, we aim to design an AI-powered system that not only analyzes customer sentiment from text but also captures emotional cues from speech and conversational dynamics. By integrating Automatic Speech Recognition (ASR), Natural Language Processing (NLP), speech emotion recognition, and predictive modeling, we expect to generate deeper insights into customer intent and sales conversion probability. This dual focus on both multimodal sentiment analysis and predictive forecasting will enhance the practical value of our system, making it suitable for real-world applications in call centers, customer experience management, and sales optimization..

References

IBM, “Watson Tone Analyzer,” [Online]. Available: https://www.ibm.com/watson/services/tone-analyzer/. [Accessed: Sep. 2025].

Amazon Web Services, “Amazon Connect Contact Lens,” [Online]. Available: https://aws.amazon.com/connect/contact-lens/. [Accessed: Sep. 2025].

Google Cloud, “Cloud Speech-to-Text and Natural Language Sentiment Analysis,” [Online]. Available: https://cloud.google.com/. [Accessed: Sep. 2025].

Hugging Face, “Transformers: State-of-the-Art Natural Language Processing,” [Online]. Available: https://huggingface.co/transformers. [Accessed: Sep. 2025].

H. Bredin et al., “pyannote.audio: Neural building blocks for speaker diarization,” [Online]. Available: https://github.com/pyannote/pyannote-audio. [Accessed: Sep. 2025].

OpenAI, “Whisper: Robust Speech Recognition via Large-Scale Weak Supervision,” [Online]. Available: https://github.com/openai/whisper. [Accessed: Sep. 2025].

T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (KDD), 2016, pp. 785–794.

B. McFee et al., “librosa: Audio and music signal analysis in Python,” in Proc. 14th Python in Science Conf. (SciPy), 2015, pp. 18–25.