

# Business Analytics System

Abhinav Gupta<sup>1</sup>, Abhay Sharma<sup>2</sup>, Aks Saxena<sup>3</sup>, Abhay Kumar<sup>4</sup>, Ankit Kumar<sup>5</sup>

<sup>1</sup>Assistant Professor

<sup>2345</sup>Students, Department of Computer Science & Engineering

Dr. A.P.J. Abdul Kalam Technical University, India

Emails: abhinavguptamit@gmail.com,

main.abhaysharma@gmail.com,

akssaxena349@gmail.com,

abhaykumar1752005@gmail.com,

atulkumar12345678@gmail.com,

*Corresponding Author: <sup>2</sup>main.abhaysharma@gmail.com*

## ABSTRACT

Companies get large volume of audio call recordings from talking to customers in sales and support calls. Even if these recordings show how customers feel, checking audio calls manually is a slow and not steady way to do it which cannot be used much for making business decisions.

The proposed system integrates a speech-to text transcription module that converts speech to textual transcripts from audio recordings. The transcripts are preprocessed using Natural Language Processing including stop words removal and lemmatization techniques, where sentiment prediction is primarily performed using a TF-IDF based Logistic Regression model. A pretrained Distil BERT transformer model is used as a fallback mechanism to ensure reliable sentiment inference under varying operational conditions.

**Keywords:** Speech Recognition, Sentiment Analysis, Business Analytics, Whisper ASR, Logistic Regression, DistilBERT.

## I. INTRODUCTION

In the business world today, companies communicate with customers through sales and support calls to maintain relationships and improve service quality [1]. These conversations give us information about what customers think and how happy they are with the service. Even though we have recordings of these conversations it is still challenging to get useful information from them to make business decisions.

Recent advancements in Artificial Intelligence and Natural Language Processing provided automated speech recognition and contextual sentiment analysis with improved accuracy [2]. Understand how people feel from what they say. Transformer-based architecture and simple machine learning models are tools for looking at transcripts from audio data [3]. This research presents an AI-powered business analytics system developed to process customer call recordings through speech-to-text conversion and sentiment classification. The proposed system integrates automatic transcription, NLP-based preprocessing, hybrid sentiment prediction using Logistic Regression and DistilBERT, and visualization through an interactive dashboard.

## II. RELATED WORK AND RESEARCH GAP

Speech recognition and sentiment analysis have been intensively studied in the context of customer feedback and business interactions [4]-[6]. Transformer architectures, such as BERT and DistilBERT, have significantly advanced contextual understanding in various NLP tasks [7].

However, most studies reported to date are primarily concerned with sentiment classification accuracy and do not provide integrated analytics frameworks for decision support in a business environment [8], [9], [10].

### III. PROPOSED SYSTEM

To address the limitations identified in existing approaches, an AI-based business analytics system is proposed for automated analysis of audio call recordings of customer interactions. The system combines speech transcription, sentiment prediction, data storage, and visualization modules to transform conversational data into analytics dashboards.

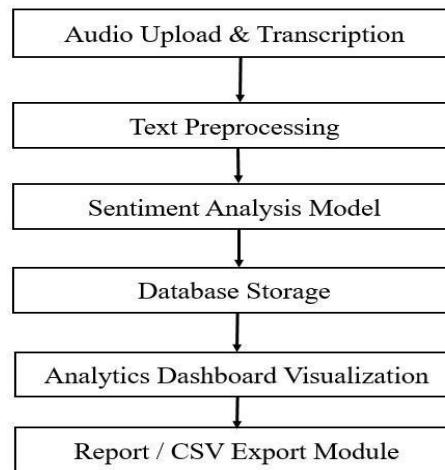


Figure 1: System Architecture of Business Analytics System

#### 1. Audio Input & Transcription

Customer audio files are uploaded and transcribed by Whisper ASR.

#### 2. Text Preprocessing

Stop-word removal and lemmatization are used to prepare transcripts for sentiment analysis.

#### 3. Sentiment Analysis

A TF-IDF Logistic Regression classifier is used for sentiment prediction, whereas DistilBERT is a fallback model.

#### 4. Data Storage

Transcripts and predictions are stored for analytical processing.

#### 5. Analytics Dashboard

Visual dashboards provide sentiment trends and domain-wise insights.

#### 6. Export Module

Predictions can be exported as CSV reports.

#### 7. Real-Time Prediction

New recordings can be analyzed for instant sentiment classification.

#### 8. Rule-Based Recommendation

Business recommendations are generated when the negative sentiment exceeds predefined thresholds. In addition to the primary processing workflow shown in Figure 2, the system also provides real-time prediction and rule-based recommendation functionalities. These are the applications built on top of existing modules.

## IV. METHODOLOGY

The methodology section in this study defines the internal processing workflow used to transform transcripts into sentiment-based analytical outputs. The proposed workflow consists of text preprocessing, feature extraction, sentiment prediction, and output generation stages, as illustrated in Figure 2.

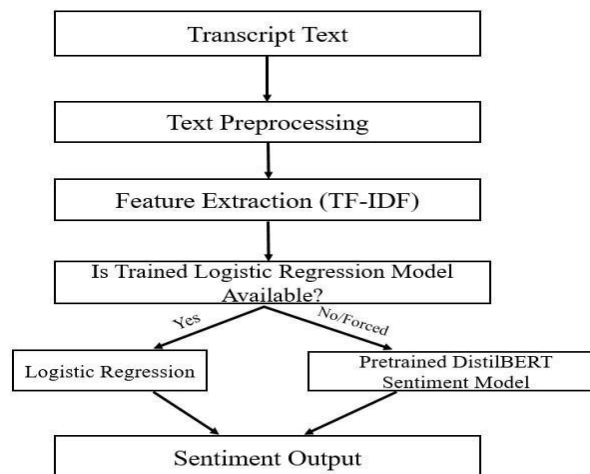


Figure 2: Methodology Flow of Business Analytics System

### A. Dataset Preparation

The dataset used for this system consists of simulated business interaction recordings representing customer inquiry conversations related to internship and course enrollment. Audio recordings were manually recorded to emulate real-world sales call environments. Sentiment labels were assigned through human annotation based on customer responses such as positive or negative, enabling supervised learning for sentiment classification.

### B. System Workflow

The initial step of the system workflow involves the transcription of customer call recordings into text through the Whisper Automatic Speech Recognition system. The generated transcripts are then preprocessed by removing stop words, lemmatizing, and converting to lower case to enhance the quality of the text for sentiment analysis. Preprocessed text is transformed into numerical representations using TF-IDF feature extraction technique. Sentiment prediction will be carried out with a Logistic Regression classifier since it is computationally efficient and performs well on medium-sized textual datasets; in case this primary model is not available, a pretrained DistilBERT transformer model will serve as a secondary option for keeping sentiment inference running. Outputs about predicted sentiments will be sent to an analytics module for visualization and extraction of insights relevant to business.

## V. RESULTS AND DISCUSSION

### 1. Model Performance Evaluation

The proposed sentiment classification model achieved a mean accuracy of **83.0% ± 5.1%** across 5-

fold cross-validation, as shown in Table 1. The low standard deviation observed across folds indicates consistent model behavior despite a limited dataset size.

Metric	Mean Score	Standard Deviation	Minimum	Maximum
Accuracy	0.830	0.051	0.75	0.900
Weighted Precision	0.821	0.13	0.562	0.912
Weighted Recall	0.830	0.051	0.75	0.900
Weighted F1-score	0.801	0.084	0.643	0.891

Table 1: Performance Evaluation of Sentiment Classification Model using 5-Fold Cross-Validation

## 2. Overall Sentiment Distribution Analysis

The sentiment distribution indicates a higher proportion of negative responses compared to positive interactions, as shown in Figure 3. This trend reflects realistic customer hesitation commonly observed during promotional or enquiry-based conversations.

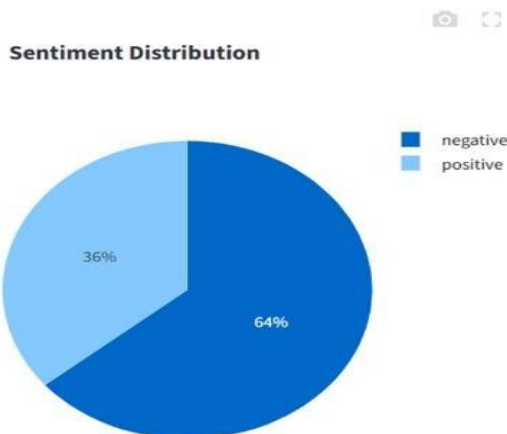


Figure 3: Overall Sentiment Distribution obtained from customer interaction transcripts

## 3.

## 4. Domain-wise Sentiment Analysis

Domain-wise analysis reveals variations in customer response across different courses, as shown in Figure 4. Such insights enable organizations to identify the most trending domains and help with further decision-making.

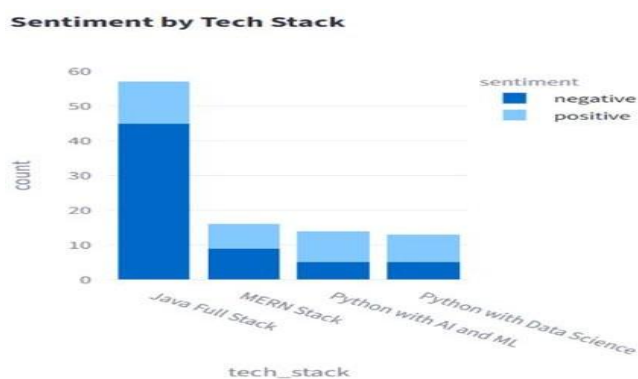


Figure 4: Domain-wise sentiment analysis across different stacks

## VI. LIMITATIONS and FUTURE SCOPE

The proposed system was tested with a simulated conversational dataset, thus limiting generalization to real-world environments. Future work will include testing on large-scale real-world call center datasets, domain-specific transformer fine-tuning, and integration with live CRM platforms for real-time sentiment monitoring and Automatic NER Extraction.

## VII. CONCLUSION

This study presents an AI-integrated business analytics system that combines speech-to-text transcription, hybrid sentiment analysis, and visualization-based insights. The proposed approach provides a automated mechanism for transforming recorded customer calls into valuable intelligence to support business decisions. Experimental results demonstrate consistent performance in sentiment classification as well as significant analytical insights derived from conversational data.

## REFERENCES

- [1] T. Scheerschmidt and D. R. Metzler, "Voice Analytics Applications and Corporate Communication: Current State and Future Research Directions," Proceedings of the European Conference on Information Systems (ECIS), 2024.
- [2] H. Kheddara, M. Hemis and Y. Himeurc, "Automatic Speech Recognition using Advanced Deep Learning Approaches: A survey" in *arXiv preprint arXiv:2403.01255*, 2024.
- [3] U. Narayan, and D. Kumar, "Sentiment Analysis Using Transformer-Based Model", *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)*, March 2024.
- [4] X. Liao, "Sentiment analysis based on machine learning models," in *Proc. 4th Int. Conf. Signal Processing and Machine Learning*, 2024.
- [5] H. Ahlawat, N. Aggarwal, and D. Gupta, "Automatic speech recognition: A survey of deep learning techniques and approaches," *Int. J. Cognitive Computing in Engineering, Elsevier*, 2023.
- [6] A. Riswawan, "Implementing TF-IDF and logistic regression for sentiment analysis of YouTube comments on the iPhone 16," *Jurnal Teknologi dan Open Source*, Dec. 2024.
- [7] A. Joshy and S. Sundar, "Analyzing the performance of sentiment analysis using BERT, DistilBERT, and RoBERTa," in *Proc. IEEE Int. Power and Renewable Energy Conf. (IPRECON)*, 2022.
- [8] M. A. Khaldy and A. Shaheen, "NLP systems for intelligent report generation in business," in *Proc. 1st Int. Conf. Computational Intelligence Approaches and Applications (ICCIAA)*, Amman, Jordan, 2025.
- [9] Bhavya V, "Machine Learning Model for Speech Sentiment Analysis" in *International Journal of Research and Analytical Reviews (IJRAR)* , Jan 2024.
- [10] H. Takci, "Converting Call Center Recordings into Valuable Insights Using Sentiment Analysis" *In 5th International Conference on Data Science and Applications (ICONDATA'22)*, December 2025.