# 

# Project Name:- UCO VoiceIQ: AI-Powered Analysis of Inbound Customer Calls

**Objective**

The objective of this project is to analyze inbound customer service calls received by UCO Bank using artificial intelligence. The system focuses on transcribing audio, identifying the language, classifying the topic, and analyzing sentiment. The goal is to extract actionable insights to improve customer experience, service efficiency, and operational responsiveness.

**Features Identified**  
Audio Processing:

* Enhanced transcription using Whisper large model with fallback to chunked transcription.
* Language detection improved via prompt-guided decoding (Hindi, Bengali, English).
* Audio is converted to mono, 16kHz for Whisper compatibility.

Noise Reduction and Enhancement:

* Uses noisereduce and pydub to convert, enhance, and denoise audio.
* Reduces background noise in chunks (30s) for long audio files.
* Ensures clearer transcription and speaker separation.

Transcription with Whisper (OpenAI):

* Uses Whisper 'large' model with chunked fallback to prevent hallucinations.
* Applies banking-specific prompts for Hindi and fallback for misdetected languages.
* Detects language using both auto and forced decoding strategy.

Speaker Diarization with Pyannote:

* Uses pyannote/speaker-diarization with Hugging Face authentication.
* Labels normalized to 'Agent' and 'Customer' using order of appearance.
* Fallback diarization applied when Whisper word timings are unavailable.

Sentiment Analysis with XLM-RoBERTa:

* Uses cardiffnlp/twitter-xlm-roberta-base-sentiment model.
* Summarized text is analyzed in chunks (512 tokens), scores are averaged.
* Multilingual support ensures consistent results across languages.

Topic Categorization using SentenceTransformers:

* Uses SentenceTransformer all-MiniLM-L6-v2 with cosine similarity.
* Applied on summaries generated by Mistral-7B.
* Embeddings averaged over text chunks, top-3 topics logged with confidence.
* Threshold for classification is adjustable (default: 0.55).

Report Generation:

* Summarization powered by mistralai/Mistral-7B-Instruct-v0.2.
* Summaries highlight actions, sentiment, resolutions, and timeline.
* Output includes diarized, translated, summarized transcripts and CSV summary

**Model Output**  
✅ Total Calls Analyzed: 2  
📊 Language Distribution: English  
😊 Sentiment Distribution:

* Negative: 2 calls (100%)

🏷️ Top Call Topics:

* Cheque Issues
* Pension / Benefit Credit Issues

🗂️ Result Files:

* english\_diarized\_summary\_analyzed.txt
* english\_2\_diarized\_summary\_analyzed.txt

**Technical Overview**

This section outlines the core technologies and models powering UCO VoiceIQ, an advanced system for analyzing UCO Bank’s inbound customer service calls.

**Audio Ingestion and Enhancement**

* Supports .wav, .mp3, .flac, and other formats.

Converts audio to mono and 16kHz using pydub.

* Applies chunk-wise noise reduction using noisereduce for improved clarity.
* Audio enhancement ensures cleaner transcriptions and better diarization.

**Transcription with Whisper (OpenAI)**

* Uses the Whisper large model.
* Automatically detects language or forces decoding for Indian scripts (Hindi, Bengali, English).
* Applies chunked fallback if hallucination or repetition is detected.
* Custom prompts improve banking-specific accuracy for Hindi/Bengali.

**Speaker Diarization with Pyannote**

* Uses pyannote/speaker-diarization via Hugging Face API.
* Separates customer and agent voices based on appearance and timing.
* Fallback diarization logic enabled if Whisper timestamps are missing.
* Labels standardized to "Customer" and "Agent" roles.

**Summarization with Mistral-7B**

* Uses mistralai/Mistral-7B-Instruct-v0.2 via Hugging Face text-generation pipeline.
* Generates detailed, paragraph-based summaries covering:
  + Call timeline
  + Customer emotions
  + Actions taken/resolution
* Token management ensures proper input chunking within 8192 limits.

**Sentiment Analysis with XLM-RoBERTa**

* Uses cardiffnlp/twitter-xlm-roberta-base-sentiment.
* Long summaries are chunked (512 tokens) and analyzed.
* Outputs are averaged for robust sentiment scoring (positive, neutral, negative).

**Topic Categorization with SentenceTransformers**

* Uses all-MiniLM-L6-v2 SentenceTransformer model.
* Computes semantic similarity to predefined banking complaint examples.
* Chunks are embedded, averaged, and compared using cosine similarity.
* Outputs include top-3 topic matches with confidence score; uses adjustable threshold.

**Models Used: Explained**

1. Whisper (OpenAI) – Speech-to-Text

* Model: Whisper large
* Function: Converts customer service calls to text.
* Highlights:
  + Chunked fallback to prevent repetition or hallucination.
  + Custom prompts improve banking context transcription.
  + Accurate across Hindi, Bengali, and English.

2. Pyannote – Speaker Diarization

* Model: pyannote/speaker-diarization
* Function: Distinguishes customer and agent segments.
* Highlights:
  + Uses Hugging Face token auth.
  + Fallback diarization when Whisper timestamps fail.
  + Normalized speaker roles for consistent labeling.

3. Mistral-7B – Summarization

* Model: mistralai/Mistral-7B-Instruct-v0.2
* Function: Generates summaries of call transcripts.
* Highlights:
  + Handles long-form transcripts with clear call flow.
  + Covers emotions, actions, resolutions, and complaints.

4. XLM-RoBERTa – Sentiment Analysis

* Model: cardiffnlp/twitter-xlm-roberta-base-sentiment
* Function: Detects sentiment (positive, neutral, negative) in summaries.
* Highlights:
  + Multilingual sentiment classification.
  + Uses chunk-wise averaging for long texts.

5. SentenceTransformers – Topic Categorization

* Model: all-MiniLM-L6-v2
* Function: Matches summaries to predefined complaint topics.
* Highlights:
  + Supports thresholding and top-N match logging.
  + Handles vague or fuzzy complaint expressions effectively.

**Why These Models Were Chosen**

|  |  |
| --- | --- |
| Model | Reason for Selection |
| Whisper (large) | Best-in-class multilingual ASR with fallback logic and domain-adaptive prompts. |
| Pyannote | Accurate diarization with speaker label normalization and Hugging Face integration. |
| Mistral-7B | Produces detailed, emotionally aware summaries optimized for customer service calls. |
| XLM-RoBERTa (XLM-R) | Multilingual, high-accuracy sentiment analysis; supports chunking and scoring. |
| SentenceTransformer | Lightweight yet powerful semantic similarity engine for topic classification. |
|  |  |

**Conclusion**  
The updated UCO VoiceIQ system demonstrates enhanced accuracy and efficiency in analyzing customer calls.  
The improved transcription and sentiment analysis enable the identification of critical banking issues.  
The insights help prioritize customer grievances and improve banking service quality.

The UCO VoiceIQ project successfully demonstrates the power of AI in automating customer voice interaction analysis. With just audio recordings, the system extracts meaningful insights like:

* Common customer issues
* Overall sentiment trends

These insights can be used to:

* Identify training needs for customer service agents
* Prioritize problem areas (like frequent ATM issues)
* Monitor customer satisfaction continuously