Detection of Fraud Communication and Phone Scams

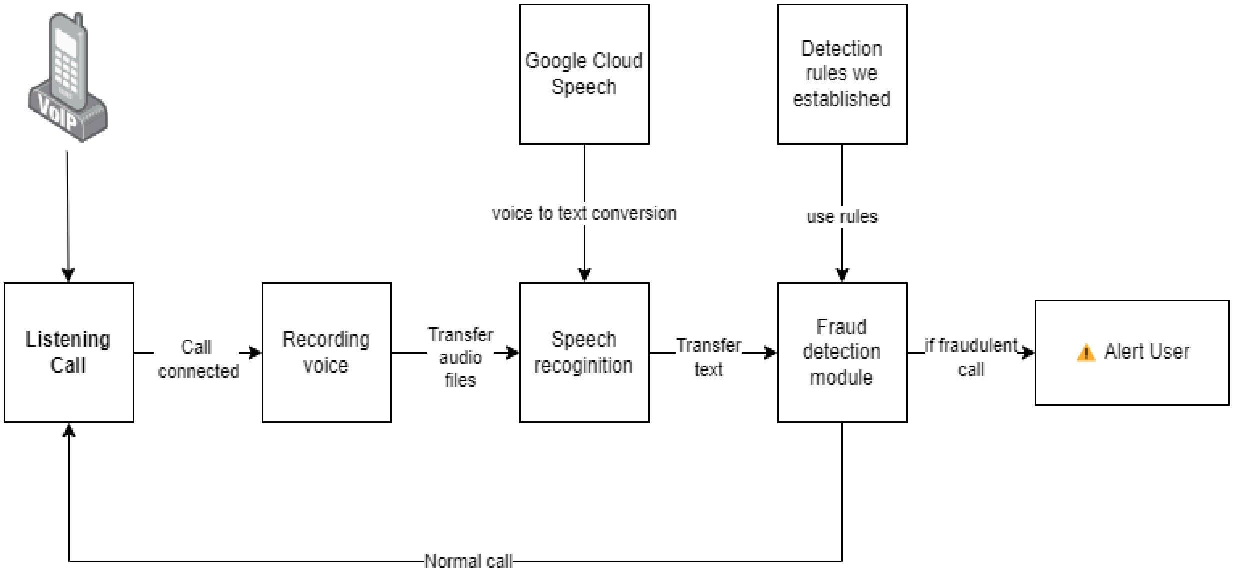
Objective:

Fraudulent communication and phone scams are growing threats, requiring more advanced detection methods. This mini project presents an ML-based approach to detect fraud by analyzing real-time call transcriptions and voice data. Using supervised machine learning models, the system identifies suspicious patterns, keywords, and behaviors linked to scams. It also incorporates sentiment analysis to detect pressure tactics commonly used in fraud.

Team Members:

1. Shefaoudeen Z - 21EC1096
2. Gowtham Balaji R - 21EC1076
3. Shruthi D - 21EC1097
4. Thejashri R - 21EC11103
5. Reena MD - 21EC1080
6. Rithish S - 21EC1081

Flow Diagram:



Abstract:

Phone scams and fraudulent communications have become significant threats in today’s increasingly digital world, impacting individuals and organizations alike. Detecting these scams early and efficiently is crucial to mitigate potential financial losses and security risks. This paper introduces a Python-based system utilizing Google Speech-to-Text and machine learning (ML) techniques to detect fraudulent activity in voice communications. The system operates by capturing incoming phone calls and transcribing them into text in real-time using Google’s Speech-to-Text API, a robust tool for converting speech into accurate text data.

Once the voice data is transcribed, natural language processing (NLP) techniques are applied to clean and preprocess the text, extracting essential features such as keywords, phrases, and patterns commonly associated with scams. The system specifically identifies words and phrases indicative of fraud, such as requests for sensitive information, offers that are too good to be true, or urgency, which are common characteristics of scam calls.

To enhance detection accuracy, machine learning models are trained on a labeled dataset containing both legitimate and fraudulent communications. Additionally, a rule-based layer complements the ML approach by flagging specific scam-related keywords, phrases, or behaviors commonly associated with fraud calls. The rule-based component allows for faster identification of known fraud patterns, such as the use of phrases like “urgent action required” or requests for sensitive financial information.

This Python-based approach, leveraging Google Speech-to-Text and ML, offers an advanced and scalable solution for detecting fraudulent communications, providing proactive protection against the growing threat of phone scams.