DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

Major Project Report- [VIII Sem B.E]

on

RECORD.AI

Submitted by

Chinmay R C **1SI17CS027**

Mrinal Raj **1SI17CS062**

Sarthak Mishra **1SI17CS097**

Under the guidance of

Dr. Shobha K

Department of Computer Science and Engineering

Siddaganga Institute of Technology, Tumkur

An Autonomous Institute, Affiliated to Visvesvaraya Technological University Belagavi, Approved by AICTE B.H. road, Tumkur 572103, Karnataka, India

AY-2020-21

ABSTRACT

Communication over the phone has always been one of the most prone communication channels as it just requires a phone number to get connected and audio to communicate. Often the telecommunication companies use spam calls for promotions, offers and other services. Furthermore, the fraud calls, threat calls and blackmail calls, also fall under the broad division of non-legitimate calls. Through such type of calls, the caller stays behind the screen having malicious intents and pretends to be an authorized person.

We often come across the real-life incidents of such fraudulent activities including fund transfer on a call through illegal affairs or using the phone to make a threat or blackmail call to somebody. The solution for those fraudulent activity is to have software that can automatically categorize the call as non-legit or legit call and also store important points of conversations in the form of text-based notes for future reference, along with the audio recordings.

The solution, therefore, is an ML based mobile application that can run in the background, record the call and analyse the conversations by using the ML algorithms to build the classification model. The classification obtained would then lead to the call being categorized as legit/spam/threat/blackmail and all the conversations would also be converted and stored as notes along with the audio recording for future reference.

Contents

Chapter 1

Introduction

When I'm working on a problem, I never think about beauty. I think only how to solve the problem. But when I have finished, if the solution is not beautiful, I know it is wrong." This is the real beauty of software development with Machine Learning, Artificial Intelligence.

In daily life, the most widely used application 'Truecaller' exists in the market already which help the users to identify the type of call using phone number. However, the application works. The disadvantage of this is, that the application works on the basis of phone numbers but not on the basis of content of the call, so if any person's phone number is marked as non-legit then any type of call from him will be shown as non-legit. So, the application will not be more specific in suggesting the call type to the user.

So in this work, we introduce one such software called 'RECORD.AI'. This is an AI-powered solution to analyse and identify the non-legitimate call through the contents of the call by converting conversations between the two in the form of text notes, along with the audio recording, using NLP and Machine Learning.

1.1 Objectives

RECORD.AI software which is used to identify the call type based on the contents of call has several objectives.
The main objectives are as follows:

- *Record the on-going call*: To run in the background, record the on-going call, and produce an audio file having voice conversations.
- *Transcribe into Notes*: To transcript the recorded call into efficient text-based notes along with the audio, for future reference.
- *Detection of Non-Legitimate call*: To detect the non-legitimate call by considering many factors including speaker voice, age, emotion, etc.

- *Classification of Non-Legitimate call*: To precisely classify the type of non-legitimate call by categorizing it as spam or fraudulent or threat or blackmail call.
- *Track History*: To track the history of any record by saving the notes and classification in the software.
- *User Interface*: To develop an AI-powered mobile app with highly interactive and customized UI design.
- *Trustable system*: To build a trustable platform with a mobile app establishing integrity with the local mobile system and keeping users' data and privacy safe and secure.

1.2 Motivation

Real-life incidents of fraudulent happenings including fund transfer on a call through illegal affairs or using the phone to make a threat or blackmail call to somebody or the routine spam calls which serve as a call of a disturbance are the motivation behind this project.

We hear cases of threat, blackmail, and fraudulent calls happening on daily basis, and people getting trapped into this network. The seriousness of this can be understood from the fact that innocent people, who could not recognize such calls while being in conversation, often tend to fall into such criminal traps and incur huge loss leading sometimes to the danger of somebody's life especially in the case of a threat and blackmail call. So, the motto of this project is to save people from falling into spam, fraud, threat, or blackmail trap. Therefore, the safety of our people serves as the source of motivation for us.

Thus, the solution is to have software that can automatically categorize the call as to whether a legit or non-legit and also store important points of conversations in the form of text-based notes for future reference, along with the audio recordings and hence our proposed solution.

Chapter 2

Literature Survey

[1] Pala, Mahesh. (2016). A New Human Voice Recognition System. AJSAT. 5. 23-30.

The author proposes a new Human Voice Recognition system. Starting from the differentiation between isolated and continuous speech where it is easy to build such a system for isolated one, while words spoken in continuous speech may have coarticulation effect and so it becomes difficult for the system to get trained. Instead of two forms like Fourier Transform (FT) and Short-Term Fourier Transfer (STFT) the proposed idea is to use wavelets as a means of extracting features from a voice signal. Along with the existing Linear Predictive Coding method, a new algorithm RASTA (Relative Spectral Algorithm) is proposed. It is, hence, an improvisation in existing technology, by using the RASTA algorithm.

To verify the identity of individual based on text uses statistical computation, formant estimation, and wavelet energy on their speech signal. After performing tests using the LPC method accuracy is about 66.66% but after using RASTA, the accuracy rate is approximated to 90%. This helps us determine the right set of technology to be used to implement the project.

[2] Shet Shirodkar, Nilkanth. (2016). SPEECH TO TEXT RECOGNITION USING HIDDEN MARKOV MODEL TOOLKIT.

The author explains Hidden Markov Model (HMM) for speech-to-text conversion by presenting Mel Frequency Cepstral Coefficients (MFCC), an approach to extract features from the speech signals of isolated spoken words.

The author uses few methodologies including endpoint detection which could be a method supported the short- term log energy and short-term zero-crossing rate. Endpoint detection is employed to get rid of unwanted noise from the background. He also discusses the Mel Frequency Cepstral Coefficient (MFCC) where the steps like

framing, Mel frequency filtering, windowing, Discrete Fourier Transform (DFT), logarithmic function, and Discrete Cosine Transform (DCT) are used to extract features from speech signal. And lastly, he talks about Hidden Markov Model Recognizer which is one among the methodologies widely employed in various applications, HMM a tool for modelling a large range of time-series data, the accuracy is about 87.6%. This adds another layer of a deeper understanding of the speech-to-text conversion.

[3] A. Gelbukh, "Natural language processing," Fifth International Conference on Hybrid Intelligent Systems (HIS'05), Rio de Janeiro, Brazil, 2005.

The author elaborates on Natural Language Processing (NLP) which plays a significant role in artificial intelligence research, which is employed in fields of application and variety of other interactive artificial intelligence areas. The paper differentiates phases between NLP and components of Natural Language Generation (NLG). Natural Language Understanding and Natural Language Generation are two parts of Natural Language Processing which develops the task gradually to know and generate the text. The discussion about the method and terminologies used, make us understand how the conversion takes place. The terminologies like, Linguistics is the science of language which incorporates Phonology referring to sound, Morphology word formation, Syntax phrase structure, Semantics syntax, and Pragmatics which refers to understanding, helps us understand the concept better. For our point of interest, he also expands about the technology getting used to convert audio to short notes or summarized texts which are done by automatic summarization.

The paper mainly discusses the possible applications of traditional AI techniques and their combination during this fascinating area and thus validates the feasibility of our model.

[4] Zhao, Q., Chen, K., Li, T. et al. Detecting telecommunication fraud by understanding the contents of a call. Cybersecur 1, 8 (2018).

The author discusses the detection of fraud calls from the contents of the call instead of the callers' phone number. Here, data is analysed using Machine Learning which selects important descriptions from the data collected previously to construct datasets. Natural Language Processing is used to extract features from textual data and searches for similar contents for fraud analysis. The TF-IDF algorithm is used

to extract the keyword from all data. And these vectors are used to train using ML algorithms. Hence able to detect fraud call in the incoming call.

This enlightens one category of our non-legitimate calls i.e. fraud call cases and the way of its implementation helps us understand the way other type of calls like threat/spam/blackmail could be identified

[5] Ma, Li & Gu, Lei & Wang, Jin. (2014). Research and Development of Mobile Application for Android Platform. International Journal of Multimedia and Ubiquitous Engineering.

The author expands on the research and development of a mobile application for the Android platform. An android app is becoming more popular as the hardware of the mobile is improving. Three kinds of applications are developed based on Java and Android SDK they are video player, web Client, audio player. The authentication method is used by the system for user authorization to complete the login process. The specific functions of this system are developed based on the Android SDK. The interfaces of Android apps are pretty and smooth.

Furthermore, android is a linux-based operating environment. It consists of multiple layers of construction. The application layer is the home for all android applications and are written using the Java programming language running on a Java virtual machine called as Dalvik virtual machine. The application framework layer defined the Android application framework. All Android applications are based on this application framework. Libraries layers include a set of C/C++ libraries used by various components of the android system. An android service runs in the background to perform long-running operations. Content providers provide data share mechanisms among applications. An android is an open-source, free device platform with its powerful function.

This not only provides insight upon how can an app run in the background and recording of audio can be done but also validates the system integrity of android app with OS, the secured way of accessing implicit intents service from the system, the wide community of users using it, and thus the usability of the mobile app as the lifetime product solving the real-life problem.

[6] Raahul, A & Sapthagiri, R & Pankaj, K & Vijayarajan, Vijayan. (2017). Voice based gender classification using machine learning. IOP Conference Series: Materials Science and Engineering.

According to the author, classification and recognition based on gender have been made for a long time. Here the main parameters considered are pitch and frequency, to differentiate male, female children's voice. Firstly, test data for the performance of the system is evaluated. Few algorithms namely Logistic Regression have the best accuracy of 92% while comparing to other algorithms like Random Forest which has the best accuracy of 93% with the voice data who speak the same language. Hence Random Forest is best for speech recognition based on frequency and pitch to classify female, male, and a child. Other algorithms used are Linear Discriminant algorithm and K-Nearest Neighbour (ML approach). Here the author mentions about dataset which has around 3169 records, out of which 1985 are female and 1584 are male records and for convenience labels of male and female are converted into 0 and 1.

This research shows how gender prediction can be made using voice data obtained through phone calls.

[7] C. Shan, J. Zhang, Y. Wang and L. Xie, "Attention-Based End-toEnd Speech Recognition on Voice Search," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Calgary, AB, 2018.

The research is about attention-based end-to-end speech recognition on voice search. IN this paper the author explores the usage of Mandarin language in which attention-based encoder-decoder model is used. Character embedding is used to deal with the large vocabulary. L2 regularization, Gaussian weight noise, and frame skipping is used for effective model training. With this mechanism, a sentence error rate (SER) of 7.43% and a character error rate (CER) of 3.58% on the Mi-TV voice search dataset is achieved. This explanation and understanding by this paper solve the problem of converting speech into short text-based notes using attention-based models.

[8] Jayatilleke, B.G., Ranawaka, G.R., Wijesekera, C. and Kumarasinha, M.C.B. (2018), "Development of mobile application through design-based research", Asian Association of Open Universities Journal, Vol. 13 No. 2.

The author discusses the event of the mobile application through design-based research. This paper explains about the rising use of digital method for education. Through iterative analysis, Design-based research aims to enhance educational practices throughout the planning, development, and implementation of the merchandise.

This paper validates use of mobile applications for better user engagement through the usage of high-end design practices. Hence, it's efficient to develop mobile apps with frameworks for solving a real-life problem.

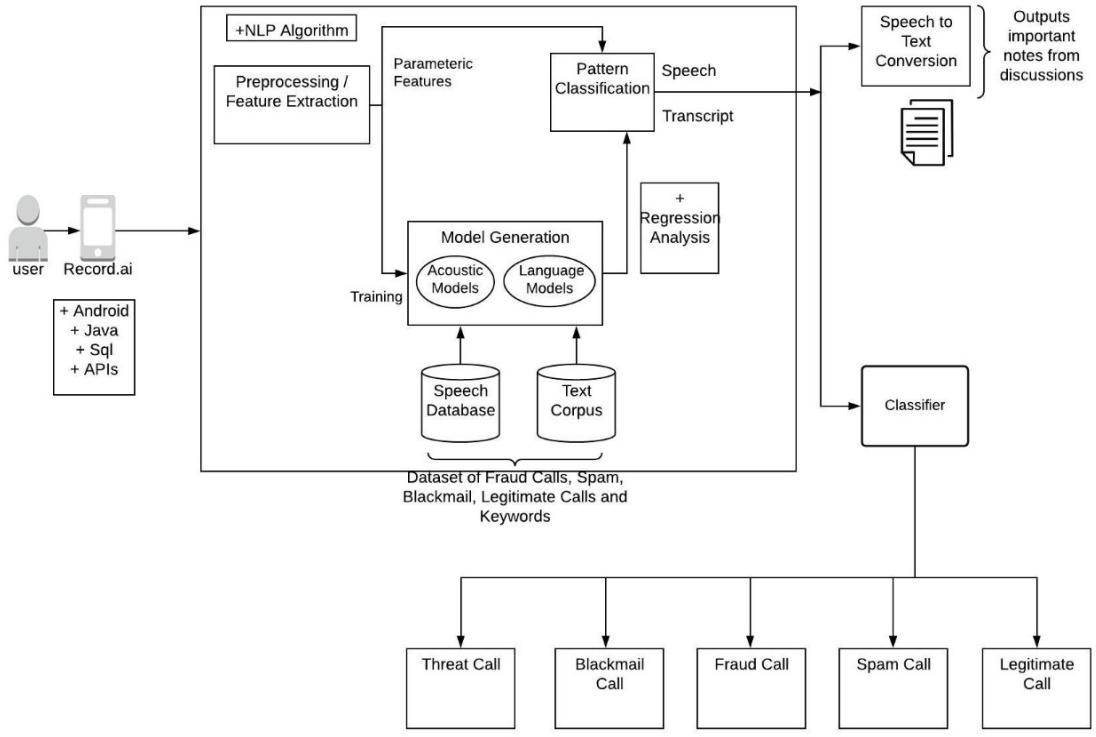
[9] Kabari, Ledisi & Nuka, Nanwin & Nquoh, Edikan. (2015). Telecmmunications Subscription Fraud Detection Using Artificial Nueral Networks. Transactions on Machine Learning and Artificial Intelligence.

The author explains telecommunication subscription fraud, here he explains the types of services provided by the telecommunication system, they are post-paid services and the prepaid services. Telecommunication fraud is carried out through telecommunication services without the intention of paying. The types through which telecommunication frauds are made through subscription fraud, superimposed frauds, intrusion frauds, fraud based on technology, fraud based on the new regulation, social engineering, Masquerading as another user. From years, fraud has increased to the extent that losses to telephone companies are in terms of billions of American dollars. To overcome that the author uses various techniques for managing and detecting telephone fraud they are: Manual review of data, conventional analysis, adaptive flexible techniques and the methodologies used are Design of the Proposed System, knowledge base (for the database), Knowledge Base of Rule, Store of Nature of Fraud(storing all types of fraud available), Self- Learning system. The system is user friendly, effective and achieved a success rate of 85.7%.

Chapter 3

High Level Design

User uses the RECORD.AI app, where the app consists of models. The app runs in the background and records the telephonic conversation. These recorded audio clips are stored in the form of file. With the help of the file, text-based notes are made and these notes are used by the model for classification. The classifier is used to analyse the model and gives the call type.



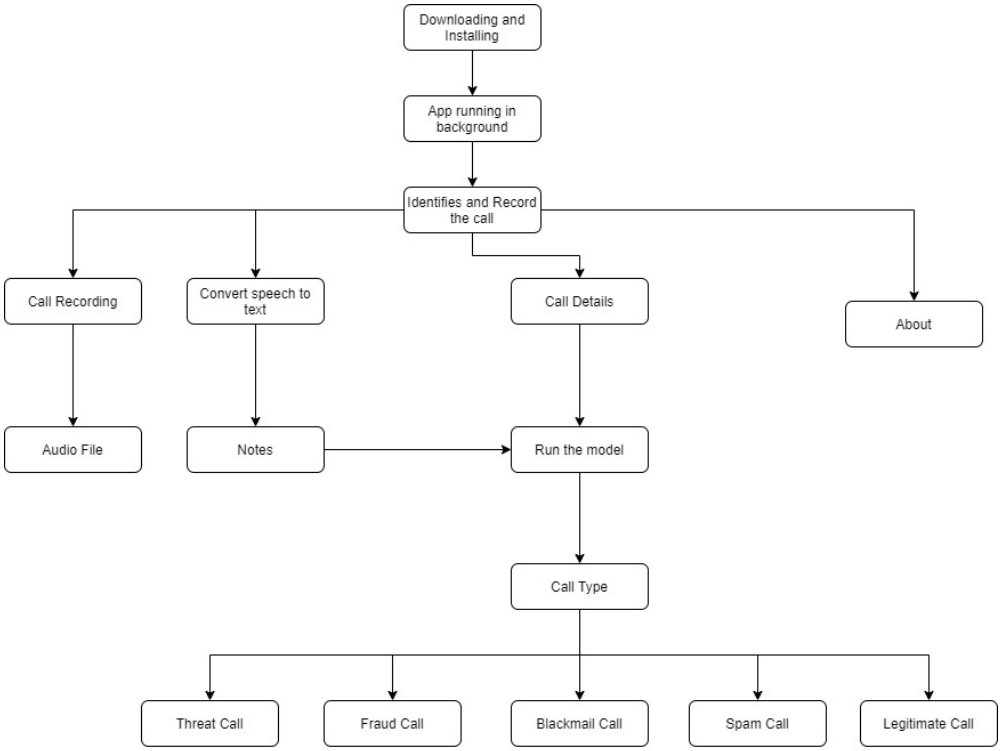


Figure 3.1: Flow chart of RECORD.AI app

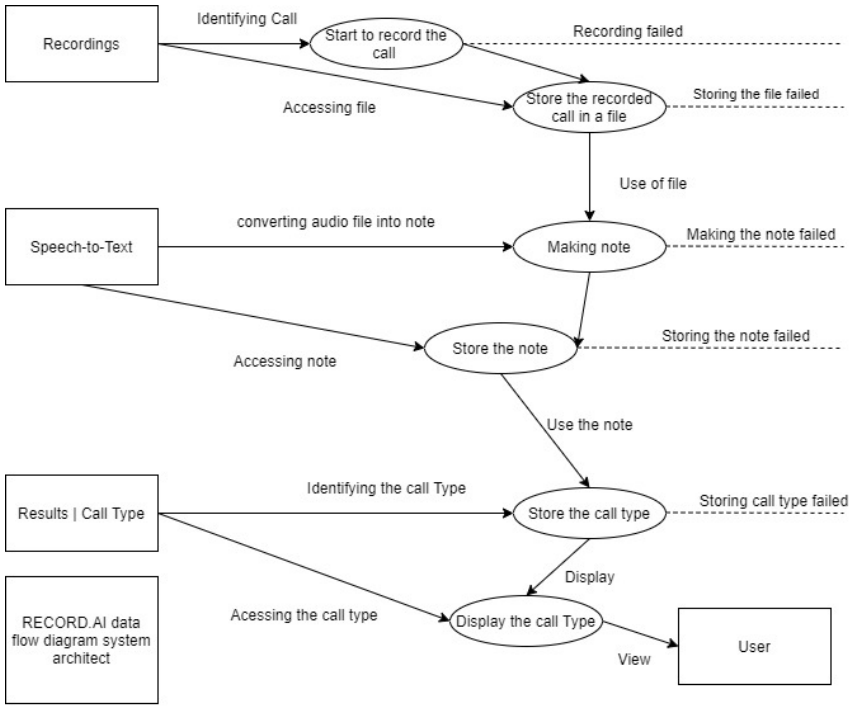


Figure 3.2: Data Flow Diagram of RECORD.AI

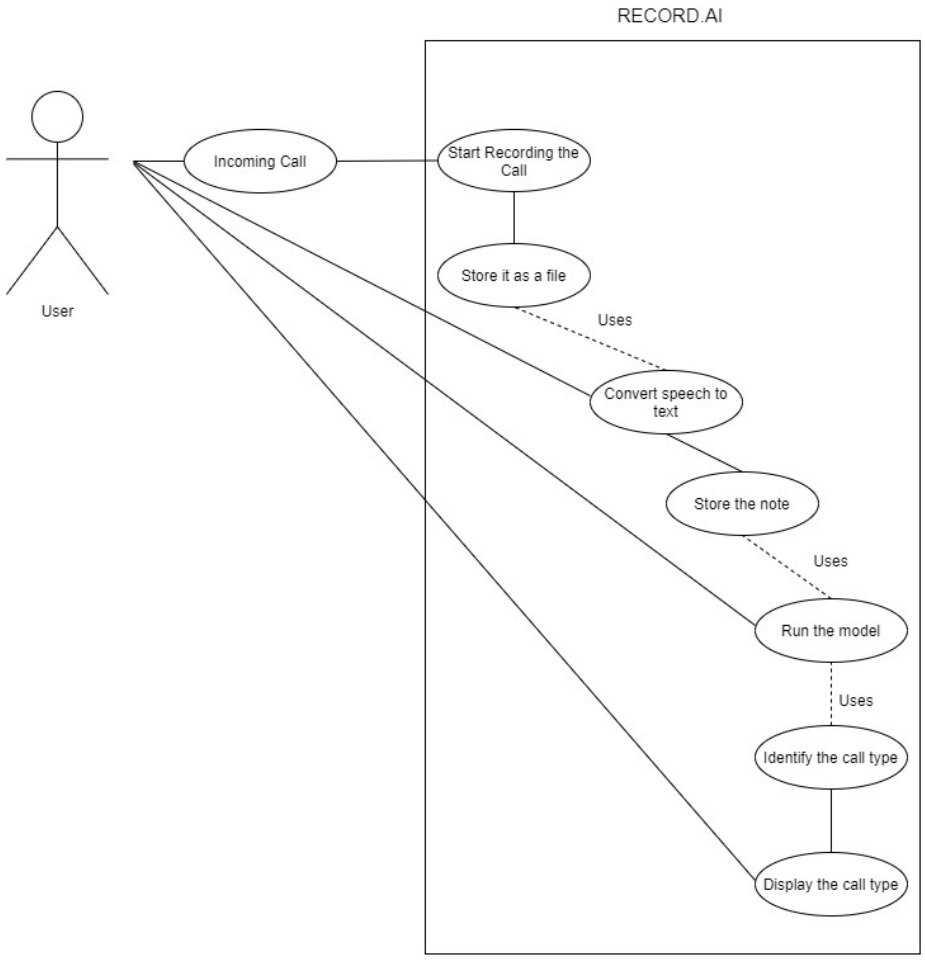


Figure 3.3: Use Case Diagram of RECORD.AI

Chapter 4

Implementation

The following are the tools, technologies and algorithms used:

Data structure used:

File: This app uses File data structure to store call conversations in the form of text files. One of the features of app i.e., 'Notes', reads from the content from the stored files and displays it. The files are stored locally on the users' device to protect the privacy of their data.

Technology used:

- **Tensorflow:** It is an open-source symbolic math library that uses dataflow and differentiable programming to perform several tasks for Machine Learning applications focused on training and inference of deep neural networks. It allows developers to create machine learning applications using various tools, libraries, and community resources.
- **Tensorflowlite:** It is the technology that we used to integrate our Machine Learning model with the android app. This library is supported by both Tensorflow and the Android Studio IDE. The working of this library includes the conversion of model into tflite format which basically is the file to be added in Android studio. Then through the tftlite file, we read the model, configure Interpreter and give input in the format of model and get the output in form of 2D array.
- **Jupyter Notebook:** It is the tool used to make and train the Machine Learning model using python.

Dataset

Dataset is a collection of related sets of data that will be manipulated as a unit by a computer. For building the model for RECORD.AI we've referred the datasets that are available on the Kaggle. we've combined both threat and blackmail dataset as they both have almost similar dataset.

In the look for the datasets, we got 3950 of spam dataset, 5212 of fraud dataset, 8048 of blackmail and threat dataset, 10,871 of legit dataset.

Input Data Preparation

The datasets of all spam, fraud, blackmail, threat, legit are combined and stored in a file. With the help of the notes, strings of conversation are sent to the stop-words remover block where all the stop-words are removed.

After removing the stop-words the remaining words are sent to tokenizer where, each word is mapped as key value pair. Here each word is key and has a numeric number as value. Each word value is added in linear fashion and rest are appended with zero's until it reaches the specified length.

Now the value of specific length is passed to the model, where it contains many classifiers. Here we are taking each percentage of value available then, based on the higher percentage the type of call is decided whether it is a spam or fraud or blackmail and threat or legit call-type.

Algorithm

We extracted data from different reliable sources including Kaggle and Machine Learning dataset repositories, for each type of classification i.e., spam, fraud, legit, blackmail and threat. Then the dataset is fed to the Machine Learning model.

- **Machine Learning Algorithm** The total dataset gets converted into 2D matrix to be fed as input. To define the model, we have used Keras model defining input in the form of input layer, and using classifiers to train it using the prepared dataset. A sigmoid function is used at the output to get it in the range 0 and 1 i.e., probabilities for each label. Hence, the output layer has only 1 node.
  After defining the model, we compiled it using TensorFlow. Next, we fit the model on dataset.
  We then evaluate the model accuracy and test it on test data.

After the model is trained, we converted and stored it in the tflite format, using TFLite converter function, for us to integrate it with the android app. From the interface of the app, we take the conversation texts as the input, prepare the data again using Stop words removal, Tokenizer and string to integer conversion process. It is then fed to the model stored in tflite format and we get output in the form of 2D array having 1 row and 5 columns, one for each type of classification result.

The output is processed to get probabilities and show the overall call type of one having maximum probability.

The App

The RECORD.AI app which helps the user to know that the call made by anonymous person is a legitimate call or any threat, spam, fraud, blackmail calls by analysing the content of the call.

This app is made by keeping user privacy by not uploading the recorded call or content of the call anywhere. The interface of the app is user friendly.

RECORD.AI app offers the following features:

4.1 Recording

The app records the telephonic conversation by running in background as a service. These recorded audio clips are stored in the form of file.

Whenever there is a call, app performs the following:

- Identifies the incoming and outgoing call.
- Start recording the call.
- Store the recorded audio clip in a file.
- Stored file can be played/deleted.

This file is stored only in user's mobile and can be accessed by the user at anytime.

4.2 Speech-to-Text

As soon as it finishes the recording and storing of the file, the next stage of the app starts functioning which is Speech-to-Text which is in user side. Here the recorded audio file is taken and with the help of the model notes have been made.

The notes are made by excluding the normal conversation keywords and considering all the keywords of non-normal conversation keywords. This note can be accessed by the user at anytime.

4.3 Call Type

With user's permission, the app records the telephonic conversation by running in background as a service, and then uses a pre-trained Machine Learning model to classify the call into the categories namely Spam, Fraud, Threat, Blackmail and Legit. As a result, every call gets mapped to its call type regardless of the phone number.

4.4 User confidentiality

In the view of maintaining confidentiality recording of the audio clip, storing the audio file, storing of the notes, result of the call type is all stored in the user side so that there is no risk of revealing the confidential information to others.

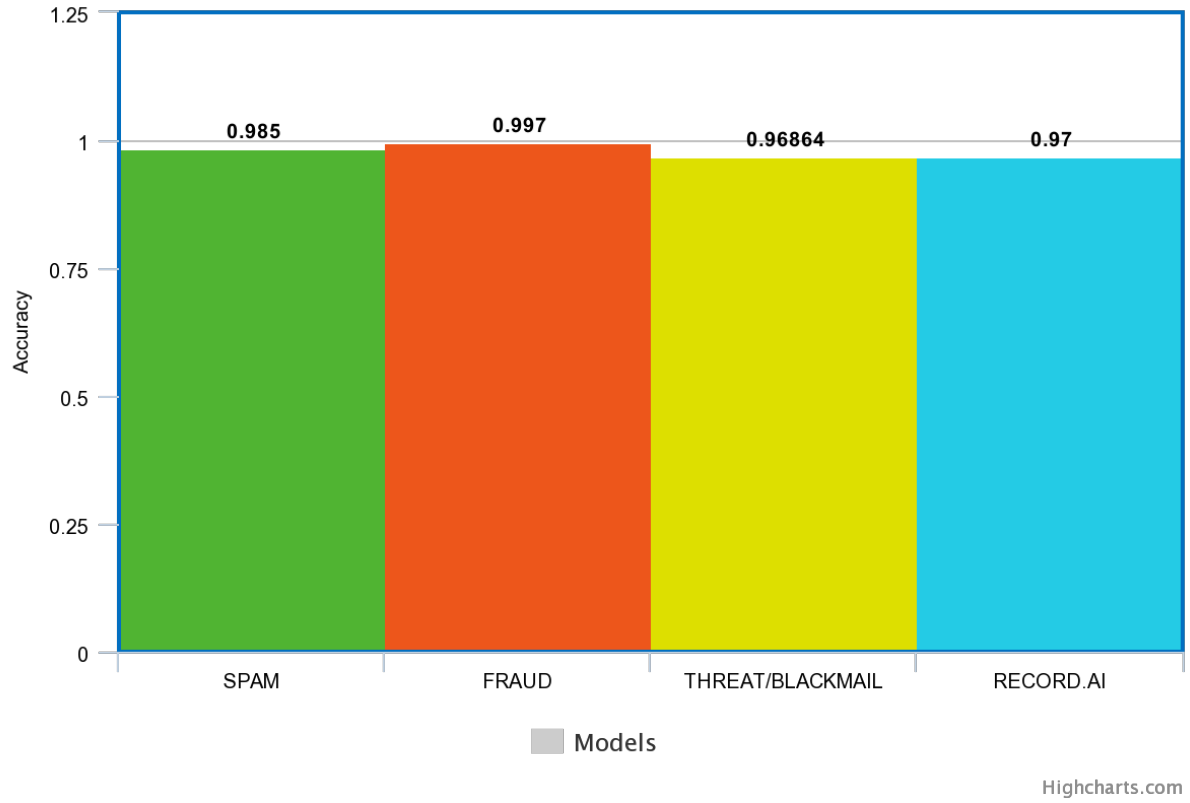
Chapter 5

Unit Testing and Results

Performance

The following figure compares the accuracy of Record.AI model with existing classifier models. The accuracies of the models are as follows:

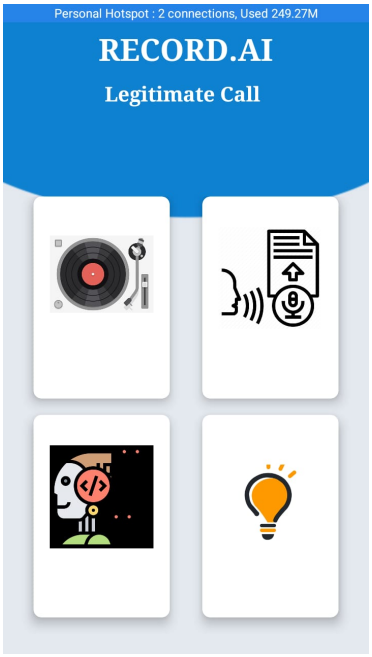
- Spam classifier: 98.5%
- Fraud classifier: 99.7%
- Threat/Blackmail classifier: 96.864%
- Record.AI: 97%



The following are the screenshots:

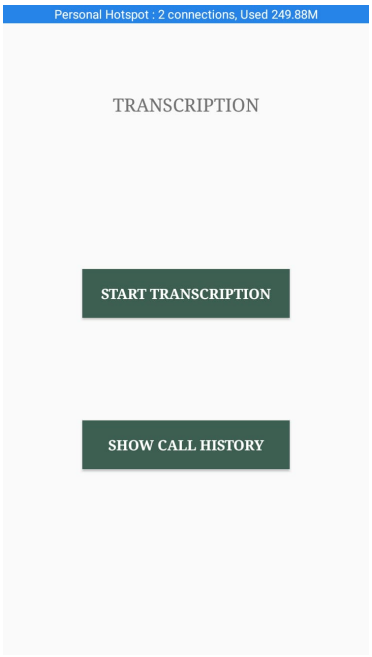
5.1 Home Page

This is the home page of our app. It involves 3 features namely Recording for recording the call, Speech-to-Text for converting audio to text and Call type for the display the identified call type.



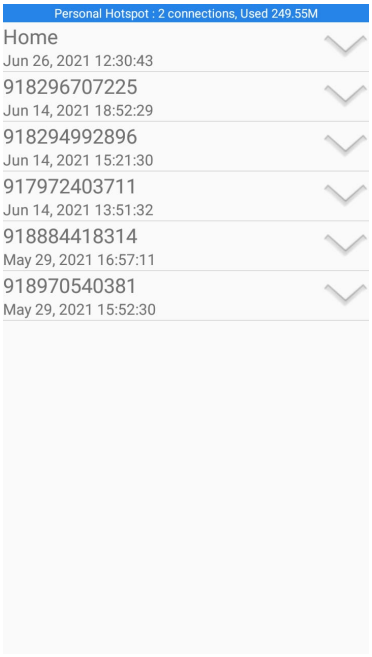
5.2 Transcription

Transcription consists of two features namely Start Transcription for transcription of the audio file to text file i.e., to make notes and Show Call History is for displaying the stored transcription of the call in a file.



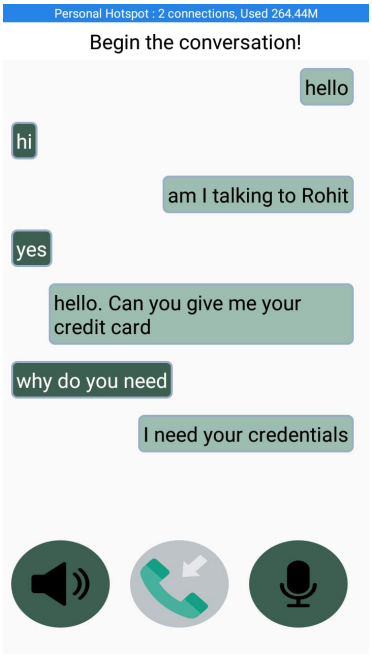
5.3 List of Call File

This page will list the all the call file, which consist of all conversions which are stored in a txt format.



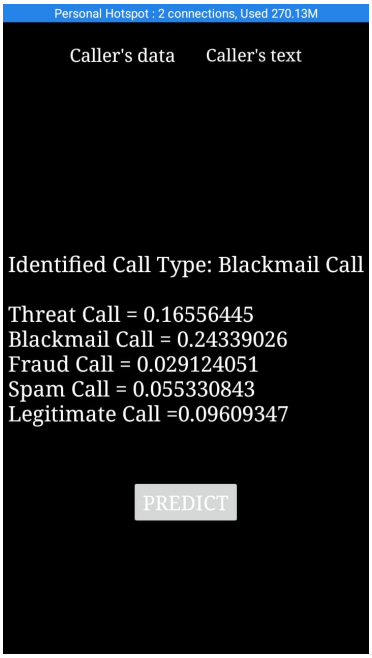
5.4 Conversations

This page will display the conversations between the two on a phone call in text format



5.5 Call Type

This page is used the display the identified call type by displaying the percentage of each spam, blackmail, fraud, thread, legit calls.



Chapter 6

Conclusion

"Is really there any application of technology better than saving one's life?"

The proposed AI-powered mobile app which records the voice call, identifies the non-legitimate phone call by classifying them precisely as spam/fraud/threat/black-mail, taking voice as the input and other important factors like emotion, age into the consideration and transcripts the conversations into notes along with the audio recording for future references, using NLP and Machine Learning, is RECORD.AI.

The easy-to-use and highly customized UI of the app with background running features and the integrity of the software system along with the security concern of users' data and privacy make it more reliable, trustable and feasible to become a real time product.

This software development makes a realistic and strong move in the direction of identification and thus prevention from not only spam or fraudulent happenings but also from real life endangering threat or blackmail calls, with the use of the right set of technologies and hence, *saving one's life*.