Speech to text Community Application using Natural Language Processing

Rakesh Kumar
Department of Computer Science,
Chandigarh University,
Punjab, India
rakesh.e8623@cumail.in

Meenu Gupta
Department of Computer Science,
Chandigarh University,
Punjab, India
meenu.e8623@cumail.in

S.Rubina Sapra
Department of Computer Science,
Chandigarh University,
Punjab, India
Roseruby6006@gmail.com

Abstract – **Android-based mobile applications are on-demand these days. People in a community can communicate with each other through internet-based platforms using cloud firebase. Since typing messages is a cumbersome activity, a speech to text application would cater for a hassle-free mode of communication. A flutter based android application integrated with natural language processing can capture the spoken words and digitize them into textual format. The efficacy of the application depends on the vocal parameters of the speaker that includes- pitch, loudness, intonation etc. On the basis of the audio captured, the application processes the text and displays it on the screen. The application uses a cloud firebase for user authentication and sign-in. Once a user registers on this application, they can communicate with others using the speech to text feature. The performance of the application is evaluated using hidden Markov models that take variegated forms of audio input into consideration. The precision of the application is 91.5%, recall is 95.4%, f1 score is 86.8% and accuracy is 89%.**

Keywords: firebase, API, hidden Markov models

I. INTRODUCTION

Mobile-based transactions are one of the fastest-growing areas in Information Technology today. As people interact more with mobile applications, it has become the basis of human's daily life [1]. Online communication provides an easy medium to convey their current endeavours through chatting. A major problem faced by people is that there are fewer platforms available where people can get involved in a community where they can propel each other's progress. A community application would solve this problem. The user can connect with others and also post, chat and do a lot of activities to share their opinions using the community application. With the advent of technology, more features like chatting, video calling, sending audios has been made possible for everyone. There is always a positive global impact of mobile applications [2]. This facility can be used to nurture human beings on a professional front- to guide each other. For instance, a person wants to learn application development. There are a number of platforms available to learn but the main question is guidance on how to learn things so that we can complete them on time and also keep things on track. This issue is countered through this application in such a way that people can connect with each other irrespective of their geographic position.

Communication is essential for sharing our opinions, ideas and viewpoints [3]. It is a bridge that connects one person to another. The Community Application provides a medium of communication through real-time messaging. Chat facility gives a major advantage of communication. By using the speech to text API, users can easily chat with each other just by speech and it would be converted into the digital form. In this way, the people can have a personal conversation [4]. Whenever a person faces an issue in anything then they can schedule or fix a time to meet each other using this application. The facility of chatting gives an edge as in a social networking site. An internet-based application is preferred by everyone as in that case no data is stored in the mobile phone. All communication and posts are retained on the server. As soon as the user logs in, the data stored in the server is obtained on a real-time basis and is updated on the user's application. The Community Application gives the users the ability to speak out the words and it is converted into the textual format [5]. This feature serves as an easy method as compared to typing the content.

This paper is further classified into different subsections. Section 2 will discuss the different research views about social community platforms through a literature survey. The next section will explain the architecture of the proposed model and the fundamentals of the Android Application. Section 4 will discuss the working of the model and the metric analysis of the Android Application on the basis of its working. Finally, this paper has been concluded with its future scope in section 5.

II. LITERATURE SURVEY

This section provides an overview of the Android Applications conducted through researches and provides an idea of the methodologies implemented in developing the Applications.

In [1], S. Emmadi et al., assessed the existing chatting mediums like Hike, Telegram, Snapchat etc. and has highlighted the future prospects of the Android application development which include personal messaging groups, voice messaging and video calling. Various firebase services which include firebase authentication, firebase cloud, real-time database are analysed based on their functionality. In [2], R. Islam et al., presented the application and the utility of android application in the social and business areas and discussed the global impact of the android applications. In [3], M.S. Satu et al., showed the review of applications of the Chatbot which are developed using the AIML scripts. AIML based chatbot's ease of implementation and the efficiency of the chatbot are discussed. The detailed information about the different applications of the chatbots is evaluated considering a number of factors. In [4], N. Shakhovska et al., discussed the development of chatbot using Google APIs which is aimed at establishing a unified interface using available resources at hand. It implements natural language

processing. In [5], B.R. Reddy et al., explained the mode of operation of speech recognition using the HMM algorithm which uses a set of defined characteristics to convert the speech into digital format.

In [6] N. Chaudhari et al., elaborated the designing of a real-time chatting application that utilises translation on a real-time basis. This seeks to send instant messages in any language through the social medium. The receiver obtains the message and chooses the language to translate it to using the application. In [7] W. Wang et al., reviewed the functioning of an application that uses the speech to text feature that gives users the ability to operate on dual-mode which includes typing and giving input as speech. In [8], A.F. Muhammad et al., investigated a system that involves the Internet of Things and Android Application development integrated into a health monitoring system that uses the sensors that will diagnose and monitor the health of the patients and the users using the Android Application. The sensors can be worn by the users which would be constantly monitored by the medical staff. In [9], L. Xudong and M. Javed, reviewed a speech recognition module that can work on computers using mutual information matching and explains the SDSE architecture which is embedded in the device. The modules stated in the application include the admin module, guest module and user module. Based on the type of the user in the application, the functions and the access control of the application is discussed. In [10], D. Jurafsky evaluated the areas of speech recognition and explained how audio signals are analysed and the output is recovered as exact words as spoken by the user.

In [11], A. Farzindar and D. Inkpen reviewed the existing methods of implementing the data sets used in Natural Language Processing and summarize the new undertakings implemented to collect datasets that are collected from social media. In [12], A. Kumar summarized firebase as Backend as a Service (Baas) and explained the use cases of firebase as real-world software development. A detailed study of the features of Firebase functions and their implementation is provided. The Realtime Database rules are explained in context to the use case to be implemented. In [13], S. Maskey and J. Hirschberg [13] presented a method of evaluating speech without using textual data using a hidden Markov model framework. The expected summary is compared with the summarised output. In [14], A. Jadhav and A. Patil reviewed a smart text-based mobile system that utilises the voice/text option to send text messages. It recognizes the sentences built-in real-time and provides an easy method of entering data into the digital format using voice input. In [15], B.R. Reddy and E. Mahender have reviewed a speech to text recognition module implemented in an Android platform and elaborated the layers of the Android Application.

On the basis of the above research survey, a community application using speech to text API is studied in the upcoming section. It involves the architecture of the Android Application and the evaluation of the application on the basis of standard metrics.

III. METHOD AND MODEL

The application is developed to facilitate real-time conversation using speech to text features. The principal concept of the application is to bridge the gaps to give enough situations and opportunities to interact for common

benefit. For example, a person wants to learn application development. Although various online websites are available to learn application development, they might need guidance on how to learn in an efficient way. This issue is countered through this application where individuals can interact with one another and explain every one of the questions and get it resolved. So this application helps geeks to get insights from someone on any issue in an easy way using the speech to text feature.

A. Proposed Model

With the implementation of the speech to text API, people can easily use this Community Application to interact with each other. This provides the chance to share their present undertakings, considerations without any hitch. As shown in figure 1, the workflow of the Community Application is explained and it shows how data flows from the firebase for posting and chatting feature.
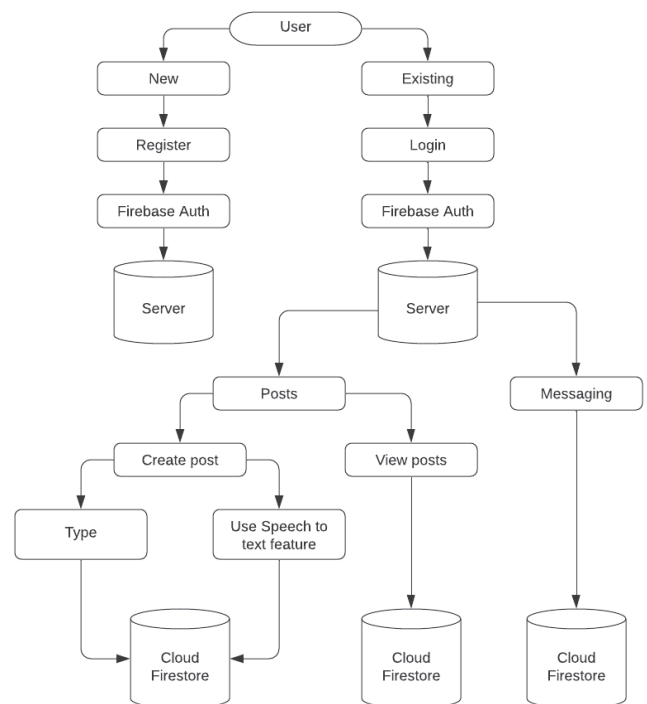

Fig. 1. Proposed framework for online social community Platform

Basically, this application is divided into 4 main modules. First is the Login and Register module. The authentication process is tested and designed such that it is very secure and doesn't have a loophole. The second module is the Social feeds page which consists of posts that everyone can share. Third and the last module consist of the chat room. In this module, users can communicate with one another or in a group making the social bond much stronger. The facility of a global chat room is fascinating as all users can view the messages and the activities of all other users. In this way, physical barriers are no more a restriction in that case. They can share their thoughts through posts or even communicate privately through chats. The users can easily connect with all other users in the same community which can provide a platform to communicate for the common good.

The Application can connect both physical and mental barriers. People can associate with others through this

application. The usage of firebase further boosts the trait of facilitating global exposure as everyone across the globe can easily access this application. The feature of the API used is an efficient way of doing the chat activity very easily using instant messaging [6]. In today's world, everyone is busy and wants to make appropriate use of the time and the resources available. The API enables speech to text input. Messages which require 10 minutes to type can be easily created by a 2-minute speech.

B. Methodology Used

To suffice the need for the login and registration facility Google firebase using flutter is implemented. This provides the facility of authentication after the registration of the account. This is made possible using the Google JSON file which is imported into the android application. Once it is imported and the necessary details are filled, then log in and registration facility is implemented in the Android Application. Google firebase database is used for post and chat facility. The database is handled in such a way that each update of the user is stored in a blob and they are updated as soon as the user clicks the update button. The data sent is synced with the database on a real-time basis. The view post tab displays the previous posts by sorting them from the most recent ones to the old ones.

The striking feature of this application is the conversion of the audio into digital format. The user is given two options in this application to update posts- they can either type the messages or else click the speech to text button, then say the content to be typed and it is converted into the textual format. It is the discretion of the user to either choose to type or to use the speech to text feature [7]. This is made possible through the speech to text API using flutter which is integrated into the application that is used to simulate the conversation [8]. The API is implemented in such a way that whatever words the user says is updated on the screen at that instant. The Android Application works on a real-time basis, as soon as the user registers in the application, the details that are entered by the user are stored in the firebase and then, the user is redirected to the login page. As soon as the user enters the credentials, the entered data is checked with the data present in the firebase. If it matches, the user is authenticated to the main page which is the main page to upload posts. If the user wants to create a post, they can either type the content or else click on the mic icon. The mic icon is linked to the speech to the text API using the flutter which captures the words spoken by the user and converts them into the digital format using natural language processing which is real-time speech input analysis [9]. The output generated by the API is updated on a real-time basis in the application. The application uses natural language processing to parse output according to the given input using phonetics [10]. When the user clicks on the Update Post button, the post is updated with the time stamp. Simultaneously, the post details are updated at the firebase real-time database. On the post page, the posts uploaded by different users can be viewed by everyone. A similar feature is utilized in the chat application as well which has two methods of taking input – voice or text. The application is embedded with the speech recognition feature to provide dual functionality.

C. Application Fundamentals for the proposed model

This section discusses the fundamentals which constitute the Community Application by scrutinizing the features and highlighting their importance as a building block of the Android Application.

- *Enhance connectivity:* The application would provide a medium of interaction and help people to communicate and share things.

- *Social Feeds:* The application would provide a medium for sharing the thoughts of different users for the development of the social life of the users within the college premises.

- *Speech to Text:* The user can give their audio input and the application converts the speech into a textual format using natural language processing using available datasets [11].

IV. RESULTS FORMULATION

The working of the application is analysed based on the standard metrics in this section. Various key factors like the total number of true positives, false positives, true negatives and false negatives during the working of the Android Application are utilized to calculate the precision, recall and f1 score.

A. Working Of Application Using Speech To Text

The Application is very easy to use interface as it has a global messaging system along with the timestamp. The user can register using basic details like email. Username and password in the community application. After registering, the login page appears. On authenticating, create a post page is displayed as shown in figure 2. This page contains a text box in which text can either be typed or the user can click on the mic button. As soon as the icon is pressed, the words spoken by the user is converted into digital form and is displayed in the textbox. If the user clicks on the update button, the post gets updated.
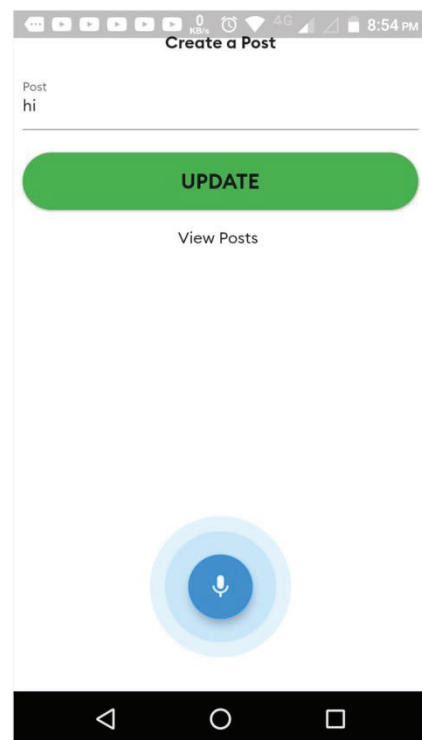

Fig. 2. Speech to Text page for updating posts

The firestore gives the facility of giving all users the ability to convey all their thoughts and opinions through the application. The Realtime Database rules [12] are modified according to the functionality to cater for the need. All users can provide all necessary details so that the other users can also know about their endeavours and then, they can connect which each other. All people using the application are given the benefit of posting all their updates in the feed. So, in an educational community when a number of people communicate with each other then they can help each other catalyse themselves on a professional front.

B. Metrics Used

Based on the work implemented, we calibrate the performance of the application to analyse the working of the Android Application adhering to the established features. The following are the standard metrics evaluated:

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (2)$$

$$\text{F1 Score} = 2 * \frac{(Recall*Precision)}{(Recall+Precision)} \qquad (3)$$

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+FN+TN} \qquad (4)$$

Here, TP refers to the total number of true positives, FP refers to the total number of false positives, TN refers to the number of true negatives and FN refers to the total number of false negatives.

True positive (TP) means that the application has generated the exact output same as the expected output. False positive (FP) means that the application has generated a wrong output. True negative (TN) means that the application has generated a correct output even though the input was inadequate to generate the correct output. False negative (FN) means that the application has generated an incorrect output since the input was inadequate.

C. Parameters for Metric Calculation

The system's ability to convert the spelt words into the digital format determines the efficiency. The system is able to capture 8 out of every 10 words maximum when the mic is on. The application can be used to convert the sentences spoken by the user into digital format on a real-time basis. The system has been tested against low pitch to high pitch and it has been accurate about 90% of the time. The accuracy of the system is proportional to the clarity of the words spoken by the user. The clearer is the expression of words, the better is the conversion of the speech into the text format using Hidden Markov Models [13].

The smart system[14] was subjected to the words spoken by reading from the blogs on medium, this was aimed at computing the precision up to which the system is accurate. The presence of a background noise decays the functionality of the system, it adversely affects the system output. The system is compatible with all android devices above the SDK version of 22.

D. Metric Analysis

Attribute selection for speech to text recognition is aimed at reducing the redundant features which involve syntax, voice signals, structure, composition, grammar, intonation and speech flow. The attribute verifies the connection between the expected results and the system output.

The training set used for metric calculation contains variegated input data that contains various parameters of vocal quality that include- clarity, pitch, loudness, intensity, intonation, intensity etc. Since the performance of the application also varies with the background noise, the application has been evaluated with various levels of background noise. A training set on the basis of the above-mentioned parameters has been used to evaluate the performance of the system. Table 1 contains the values of TP, FP, TN and FN values for different values of input audio.

After getting the data for metric analysis, on the basis of the metrics as defined in 4.2 - precision, recall, f1 score and accuracy are calculated. Table 1 shows the observed values and Table 2 shows the values of the standard metrics for variegated input values on the basis of the attributes. Figure 1 shows the graphical representation of precision, recall and f1 score for different types of input values.

TABLE I.   OBSERVED VALUES FOR DIFFERENT PARAMETERS

	Parameter	Parameter Variation	TP	FP	TN	FN
1.	Clarity	Clear Voice	813	157	21	9
		Unclear Voice	842	88	34	36
2.	Pitch	High Pitch	797	132	54	24
		Low Pitch	787	55	64	94
3.	Age	Below 20	912	43	23	22
		Above 20	916	39	30	15
4.	Noise	Little background noise	904	53	12	31
		Much background noise	731	103	78	88
5.	Loudness	Loud voice	953	16	21	10
		Feeble voice	702	184	35	79
6.	Intonation	Rising intonation	942	43	9	6
		Falling intonation	897	29	32	42
7.	Intensity	High Intensity	903	49	20	28
		Low Intensity	874	53	71	2

	Parameter	Parameter Variation	Precision	Recall	F1 Score	Accuracy
1.	Clarity	Clear Voice	0.838	0.989	0.907	0.834
		Unclear Voice	0.905	0.959	0.931	0.876
2.	Pitch	High Pitch	0.857	0.970	0.910	0.851
		Low Pitch	0.934	0.893	0.913	0.851
3.	Age	Below 20	0.965	0.976	0.970	0.935
		Above 20	0.959	0.983	0.970	0.931
4.	Noise	Loud noise	0.944	0.967	0.955	0.916
		Feeble noise	0.876	0.892	0.883	0.809
5.	Loudness	Loud voice	0.983	0.989	0.986	0.974
		Feeble voice	0.792	0.887	0.836	0.737
6.	Intonation	Rising intonation	0.956	0.993	0.974	0.951
		Falling intonation	0.968	0.955	0.961	0.929
7.	Intensity	High Intensity	0.948	0.970	0.959	0.923
		Low Intensity	0.892	0.944	0.917	0.945
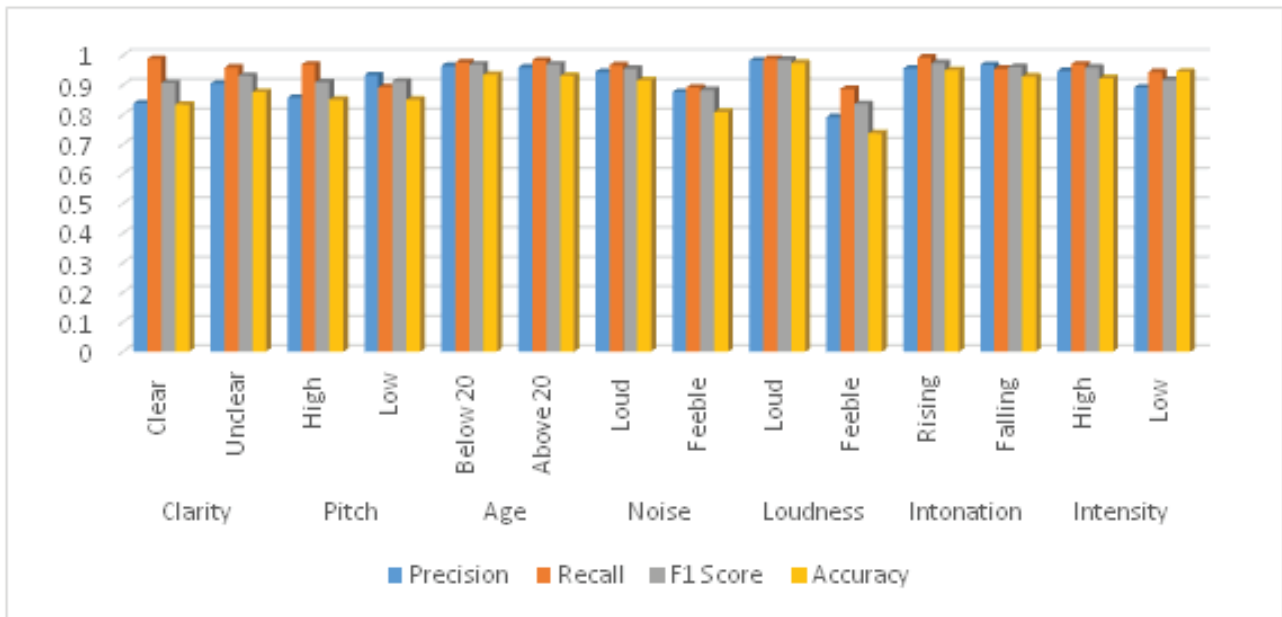
Fig. 3. Graphical Representation of Standard Metrics

Fig. 4. Values of standard metrics based on clarity
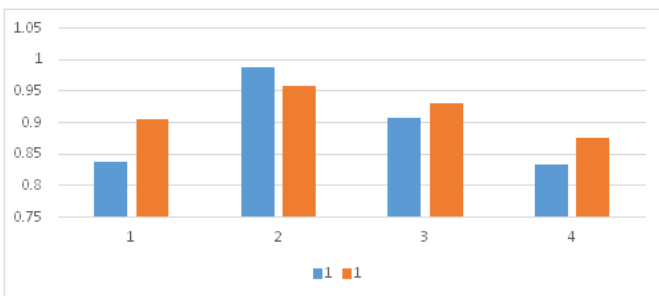

Fig. 5. Values of standard metrics based on pitch

Clarity refers to the quality to which the words are coherent. For clear words, the standard metrics of analysis are high and it is in contrast to the unclear ones. This indicates that precision, recall, f1 score and accuracy increases with clarity.
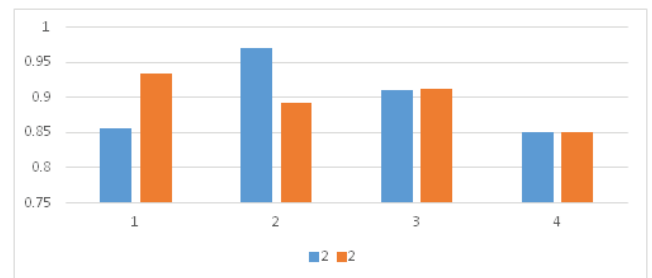
Pitch is the highness and the lowness of sound. High pitch calibrates high values of standard metrics since more pitch indicates that more easily the words are captured by the device and easier it is for the API to convert the speech into text.
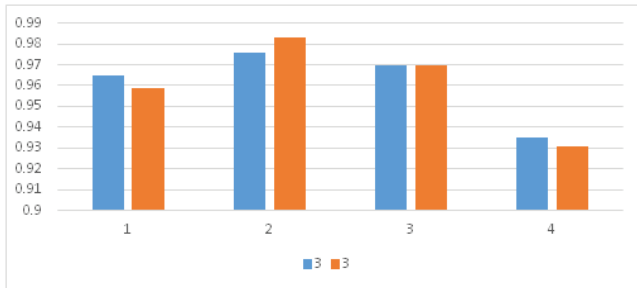
5

Fig. 6.   Values of standard metrics based on age

The values of the standard metrics are almost constant for the parameters above 20 and below 20. This implies that the functionality of the device is independent of age as it is dependent on other factors which include clarity, pitch, intonation etc. On the basis of the values of the standard metrics analysis, the precision is up to 91.5%, recall is 95.4%, the F1 score is 86.8% and Accuracy is 89%.

V.   CONCLUSION AND FUTURE SCOPE

The android application is thus a smart system in an android device that increases the efficacy. Since the model is trained with adequate data sets, it can recognize the input voice and generate output accordingly. The application functions smoothly as compared to existing speech recognition systems in terms of capturing the audio and generating the output. The accuracy rate of the existing speech recognition system is 75-85%. The speech recognition system using NLP presented here is 89% since the training data set embedded in this application has been upgraded with the latest phonetic dataset. The present android application can be enhanced to various degrees and in different languages. The future enhancements of the application are explained as follows: 1. A video calling feature can be added. In this way, despite being in a different physical location they can contact each other. Using this feature, they can have live video conferencing and can coordinate with each other just like a face to face communication. 2. Rich text-based chatting can be added that involves giving suggestions to all the words we type so that users can find it very time-saving. 3. Privacy features can be implemented by which users can hide their posts from the public and make them visible to the people they choose. 4. Granting admin and moderator privileges to users for groups. Admin privileges give the facility to the user to lead the pages and add members to the group and remove them if they are unfit for the group. 5. Artificial intelligence and instant keras like platforms can be used to provide chat assistants and guides which can use algorithms to make the application up to date with the current ongoing technology.

REFERENCES

[1] Emmadi, S. S. R., & Potluri, S. (2019). Android based instant messaging application using firebase. International Journal Recent Technology and Engneering, 7(5), 352-355.

[2] Islam, R., Islam, R., & Mazumder, T. (2010). Mobile application and its global impact. International Journal of Engineering & Technology (IJEST), 10(6), 72-78.

[3] Satu, M. S., & Parvez, M. H. (2015, November). Review of integrated applications with aiml based chatbot. In 2015 International Conference on Computer and Information Engineering (ICCIE) (pp. 87-90). IEEE.

[4] Shakhovska, N., Basystiuk, O., & Shakhovska, K. (2019). Development of the Speech-to-Text Chatbot Interface Based on Google API. In MoMLeT (pp. 212-221).

[5] Reddy, B. R., & Mahender, E. (2013). Speech to text conversion using android platform. International Journal of Engineering Research and Applications (IJERA), 3(1), 253-258.

[6] Chaudhari, N., Shinkar, S., & Pagare, P. (2018). Chatting Application with Real Time Translation.

[7] Wang, W. (2019). Using Speech. In Pro iPhone Development with Swift 5 (pp. 471-490). Apress, Berkeley, CA.

[8] Muhammad, A. F., Susanto, D., Alimudin, A., Rochman, Z. A., Assidiqi, M. H., & Nabhan, S. (2020, September). Development of English Conversation Practice App with Artificial Intelligence & Speech Recognition. In 2020 International Electronics Symposium (IES) (pp. 442-449). IEEE.

[9] Xudong, Y. Y. Z. H. Z. (2002). Speech Recognition Engine SDSE Based on Mutual Information Matching and App lication [J]. Computer Engineering and Applications, 11.

[10] Jurafsky, D. (2000). Speech & language processing. Pearson Education India.

[11] Farzindar, A., & Inkpen, D. (2015). Natural language processing for social media. Synthesis Lectures on Human Language Technologies, 8(2), 1-166.

[12] Kumar, A. (2018). Mastering Firebase for Android Development: Build real-time, scalable, and cloud-enabled Android apps with Firebase. Packt Publishing Ltd.

[13] Maskey, S., & Hirschberg, J. (2006, June). Summarizing speech without text using hidden markov models. In Proceedings of the Human Language Technology Conference of the NAACL, Companion Volume: Short Papers (pp. 89-92).

[14] Jadhav, A., & Patil, A. (2012). A Smart Texting System for Android Mobile Users. International Journal of Engineering Research and Applications, 2(2), 1126-1128.