A Voice Based Assistant Using Google Dialogflow and Machine Learning

Dr. Jaydeep Patil, Atharva Shewale, Ekta Bhushan, Alister Fernandes, Rucha Khartadkar

Information Technology, AISSMS's Institute of Information Technology, Pune, Maharashtra, India

ABSTRACT

Virtual Personal Assistant (VPA) is one of the most successful results of Artificial Intelligence, which has given a new way for the human to have its work done from a machine. This paper gives a brief survey on the methodologies and concepts used in making of an Virtual Personal Assistant (VPA) and thereby going on to use it in different software applications. Speech Recognition Systems, also known as Automatic Speech Recognition (ASR), plays An important role in virtual assistants in order to help user have a conversation with the system.

In this project, we are trying to make a Virtual Personal Assistant ERAA which will include the important features that could help in assisting ones' needs. Keeping in mind the user experience, we will make it as appealing as possible, just like other VPAs. Various Natural Language Understanding Platforms like IBM Watson and Google Dialogflow were studied for the same. In our project, we have used Google Dialogflow as the NLU Platform for the implementation of the software application. The User-Interface for the application is designed with the help of Flutter Software Platform. All the models used for this VPA will be designed in a way to work as efficient as possible. Some of the common features which are available in most of the VPAs will be added. We will be implementing ERAA via a smartphone application, and for future scope, our aim will be to implement it on the desktop environment.

The following Paper ensure to provide the methodologies used for development of the application. It provides the obtained outcomes of the features developed within the application. It shows how the available natural language understanding platforms can reduce the burden of the user, and therefore going on to develop a robust software application.

Keywords : Artificial Intelligence, Natural Language Understanding, IBM Watson, Google Dialogflow, Speech Recognition

I. INTRODUCTION

In this new age of technology and innovation, the use of artificial intelligence and machine learning has made our life much easier. These technologies have proved to be beneficial to the society in various fields such as education, industries, e-commerce, etc; and one of the most prominent is communication. Right from the 1960s, when IBM introduced the first digital speech recognition tool, i.e. IBM Shoebox, the idea of having a conversation with a computer seemed quite futuristic. An Intelligent Virtual Assistant is a digital life assistant made to contribute maximum convenience to the user. It is a highly developed software with a powerful speech recognition system which focuses on the processing of audio signal into the system, converting it to text and perform the required task.

Most of the Virtual Assistants work basically on voice as communication. It focuses on processing of audio signal into the system, converting them to text and performing the required task. In general, speech processing consists the following: A Speech-To-Text Module that converts speech signals to text, A Parser that extracts the semantic context, A Dialog Manager that determines system response through machine learning algorithms, An Answer Generator that provides the system response in text and A Speech Synthesizer that converts text to the speech signal [11].

When developed by a normal user, he/she may experience many issues, in terms of accuracy in recognition, robustness in performing operations, etc. and at times may not be able to understand the issues faced. Therefore, in this paper, we have tried to give an overview which will help user understand the methodologies and steps involved in the making of a Virtual Personal Assistant. We have taken into consideration the different methodologies, results and limitations published by different researchers.

II. LITERATURE REVIEW

The Speech Recognition Model is one of the most important part of a Virtual Assistant. Considering the various Neural Networks that are required for building up of a speech recognition system, it was necessary to survey the models that provided the insight by determining the accuracy and other factors of each Model. It was observed that High Accuracy and less Validation Accuracy was achieved for Convolutional Neural Network (CNN) model as compared to Basic Neural Network. Thus, proving that CNN is a better choice for speech recognition systems[1]. Considering the Limitations for the Model, other parameters such as Word Error rate, throughput of the system was not taken into consideration.

Various Machine Learning Algorithms are used for speech recognition. It was found that on application of Auto- WEKA on various algorithms, determined Random Forest as the best algorithm which is useful for learning the dataset based on the training set. However in this survey, Speech samples consisting of noise was not tested for determining the scalability and robustness of the models[3]. In the Survey of scaling speech recognition using CNN, following metrics were taken into consideration:
(i) throughput,
(ii) Real-Time Factor(RTF) and latency, and
(iii) Word Error Rate (WER)
for the overall framework, helped in achieving an efficient model. But due to the increase in the number of the layers the implementation of the same was difficult[4]. Other Algorithms such as the Long Short-Term Memory (LSTM) is very powerful in speech recognition and Hybrid model of Hidden Markov Model (HMM) and Gaussian Mixture Models (GMM) can give excellent results[5]. In various Projects of developing a Virtual Assistant it was observed that the platform failed to support various other languages of the countries including

China, Japan, India, etc.[6]. The survey paper provided with detailed study on the Recurrent Neural Networks (RNNs) that can be used for Speech Recognition System but with more research to be carried out on the same. However, the survey focused more on the Supervised Learning Models and less importance was given to the Unsupervised Learning Models.

A Survey included the detailed comparison of the Personal Voice-Based Assistants available in the market namely, Google Assistant, Cortana, Alexa and Siri. It concluded that Google assistant gave good results in VR and HFI by achieving 60% accuracy. Siri achieved 44% accuracy in VR and HFI. Cortana was observed with decrease in accuracy close to 30%. Other results included that Alexa wasn't suitable with simple questions whereas Cortana was poor in basic voice recognition[7]. The illustration to use AI-enabled content analysis has been discussed in one of the survey paper. The system can examine text of leadership speeches, content related to a specific organization. However, Only one type of content was analysed with limited samples and a Pre-defined Coding Scheme was used for the Project[9].

Since, the survey of this paper also included the comparison of IBM Watson and Google DialogFlow, various Projects carried on these platforms were studied before arriving at the conclusion for a better Platform for the Project. A Project was based on successful implementation of IBM Watson in developing an application for health care purpose[12]. This Project thus provided a base for building up of an AI Application with the help of IBM Watson. Various other Projects used IBM Watson as their platform for building the system which processed various queries with the help of its in-built Natural Language Processing (NLP) and Natural Language Understanding (NLU) Algorithms[8].

Projects based on successful implementation of the Google DialogFlow for an Organization were studied. It provided an insight to the various technologies like the Google Cloud Platform, Google Cloud Vision API for integrating detection features in the system and Firebase Real Time Database for developing the Application. The system ensures security of database with the help of OAuth Authentication for accessing the system. However, most of the actions carried out with the help of Google services required Internet Connectivity while accessing the system and thereby failed to service the queries offline[11]. Other Project, aimed to design a system for Educational purpose using Google DialogFlow. The proposed methodology consists of two main phases: Knowledge Abstraction and Response Generation. The methodology studied the deep learning model, The Decision Tree, that is been used in implementing the Dialogflow[10].

III. IBM WATSON VERSUS GOOGLE DIALOGFLOW

1. IBM Watson: Benefits and Limitations

In terms of software ranking, IBM Watson leads the software industry among the other Artificial Intelligence (AI) Platforms. IBM Watson proves to be a better choice for business looking for trusted computer program. It ensures focusing on behaviour of the customer, with the help of repository of data and analysis we can access more information that helps in future customer interactions and engagements. IBM Cloud also provides the user with IBM Discovery Services that includes services of Machine Learning Algorithms, Speech-To-Text and Text-To-Speech Modules, Artificial Intelligence (AI) Services, Cloud Functions (to be written in jQuery) for integration with the interface, Webhooks for connecting with the Web and so on. These services along with the Watson Assistant enables user to develop an interactive personal

assistant. The Chatbots developed with the help of IBM Watson makes the customer feel as if they are interacting with a real customer representative on the other line. It considers customer and employee behaviour for the benefit of the business. Apart from providing various advanced services for the development of an application system, IBM Watson fails to provide the facility of other languages other than English. This, therefore limits the usage of it within few locations in the world.

The Software also proves to be difficult to maintain and is not capable of processing structured data. The platform also takes time to integrate with services of a business and provides accessibility at higher prices as compared to other platforms available in the market, limiting the organizations that can afford to purchase their plans.

2. Google Dialogflow: Benefits and Limitations

The main benefit of Dialogflow is its connection with Google. The Machine Learning Algorithms inbuilt within the platform helps in understanding natural language that is the user's expression with the help of agents. Each Agent contains Intents which is matched with the user's expression and action is performed as a response to the query. It not only provides answers to customers but also enables agents to carry out small talks with the users. The Platform also provides pre-built templates that developers can use as foundations for their project.

3. Feature Comparison between IBM Watson and Google Dialogflow

i. Machine Learning:

Both IBM Watson and Google Dialogflow provides user with this feature to analyse the data. It includes the NLP and NLU Algorithms that are required for understanding natural language of the user's expression as input to the system. The Text Input is sent to the NLP Module where it is converted into structured data. The Speech input is handled using other Speech-To-Text Algorithms provided by the platforms for further processing.

ii. Chatbots:

Platforms ensure to provide chatbots for interaction of the user with the system. All the Question-Answer based queries by the user are handled here. These Chatbots are designed for providing interaction with the user just like humans, thus providing natural experience to the users using it.

iii. Maintenance:

Considering maintenance of software platform, Google Dialogflow wins the game. Various organizations and industries found maintaining IBM Watson difficult to maintain on the other hand dialogflow provided a platform that was easier to maintain.

iv. Handling of Structured Data:

Google Dialogflow has the capability of handling structured data as it converts the user's expression into structured data using Natural Language Processing Module. IBM Watson, however, fails to handle structured data directly which thereby limits the usage of the software in many organizations.

v. Services:

IBM provides various services through Watson Discovery Services, Watson Studio, Speech-to-Text and Text-to-Speech Services, Machine Learning and Artificial Language Modules, Deep Learning, Language Classifier and many more which caters the need for processing and analysing the data.

Google Dialogflow provides the services of two editions ES Edition and CX Edition, both having their capabilities to handle and help in building of an application system. Various NLP, ML and AI models, like in IBM, analyse the data for further processing and carrying out the required action to be performed.

vi. Integration and Overall Performance:

In terms of Integration, IBM Watson it takes time and efforts to integrate its services with the organization which therefore leads to delay of the Project. With the increase in the amount of data, the services in IBM are still limited to cater to their needs. Considering pricing for the IBM Software, it turns out to be more than any other software platforms in the market, thus enabling only those organizations that can afford Watson.

Google Dialogflow provides user with its two editions namely, Dialogflow ES and Dialogflow CX. ES being a Standard version provides facility of integrating application with dialogflow with the help of Fulfilment Feature or with the help of API Service as per the convenience of the user to use the Platform. Integration, therefore, has never been a problem with Dialogflow. Pricing in dialogflow is free for ES Model and is paid for CX Model whose charges varies according to the quotations and requests the user makes. Google Dialogflow promises to handle large amount of data and provide the services to them both in ES and CX Model. The Dialogflow platform also provides the facility of 14+ languages which on the other hand was not provided by IBM Watson as it avails the facility of only English Language on its platform, limiting the usage of the software platform in few regions across the world.

Thus, both IBM Watson and Google Dialogflow have their fair share of strengths and weaknesses. Considering factors and requirements for a particular project, one of the software platforms must be chosen. It also requires consideration of the size of business and the capability to access the plans that enables choosing the most suited software platform for the system application.

IV. METHODOLOGY

A. User-Interface

Flutter being an Open Source UI Software Development Platform is used for developing attractive UI for the Application. Various packages and graphics library provided by the Platform allow faster operations of the Application. It promises to give a stunning look to the Application Interface, irrespective of the operating platform. It enables the developer to create Cross-Platform Applications with ease. Thus the need of developing different Application to run on Android and iOS is eliminated.

B. Dialog Manager

The back-end includes the development of Dialog Manager of the Application. For the purpose of developing a Natural Language Understanding Platform for the Application we used the Google Dialogflow's ES (Standard) version which handles all the conversations and actions given by the user. It provides the following features:

Dialogflow Agent:

It is responsible for handling of the user conversations and converting the voice command given by the user into the text or the text command into Structured Data that is understandable to the application. Each of the Agent consists of Intents and Entities.

Intents:

It takes care of matching the user expressions obtained from the previous step to the best intent in the agent. This matching of intent is also known as Intent Classification. Figure 1 shows an example of an Agent Weather which contains an intent forecast.

Figure 1: Intent Classification

A basic Intent contains the following:

1. Training Phrases:

These are sample phrases for what end-users might say.

2. Action:

It triggers certain actions for each intent when activated.

3. Parameters:

When an intent is matched at runtime, Dialogflow provides the extracted values from the end-user expression as Parameters. Each parameter has a type, called the Entity Type, which dictates exactly how the data is extracted.

4. Response:

It provides the user with the responses for their queries.

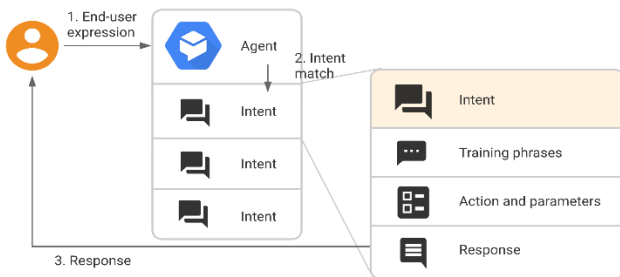Figure 2 shows the basic flow for intent matching and responding to the end-user.

Figure 2 : Intent Matching

Entities:

Pre-defined System Entities are provided by Dialogflow for matching dates, times, email addresses and so on. Entities can also be user-defined depending on the type of data handled by the system application

User Interactions with the API

For interacting with the Dialogflow API Service, a code must be written for direct interaction. Figure 3 shows the processing flow when interacting with the API Service.
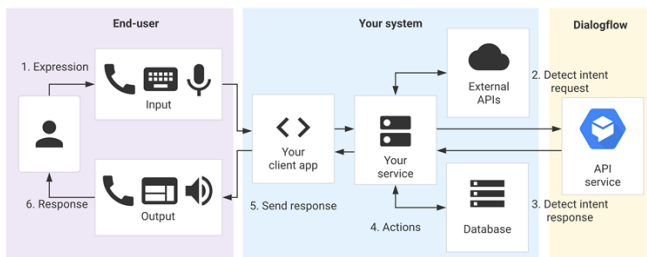
Figure 3: Interaction with the API

1. The end-user types or speaks an expression.

2. Your service then sends this end-user expression to Dialogflow in a detect intent request message.

3. Dialogflow sends a detect intent response message to your service. This message contains information about the matched intent, the action, the parameters, and the response defined for the intent.

4. Your service performs actions as needed, like database queries or external API calls.

5. Your service sends a response to the end-user.

6. The end-user sees or hears the response

The feature of Text-to-Speech of the system is developed by 'flutter_tts', a Text-To-Speech Package provided by Flutter. It helped in providing answers to the queries of the user in audio format. Thus, enhancing the usability of the Application.

Dialogflow Console

Dialogflow Console is a web user interface that enables us to create, build and test agents. Dialogflow APIs help in building agents for advanced scenarios.

V. IMPLEMENTATION

After the implementation of the user interface our main task was to develop a dialog manager that could handle all user commands either voice/text and perform the required task. Google Dialogflow, being an NLU Platform, was required for the same. The console of Dialogflow is as shown in the below Figure 4.
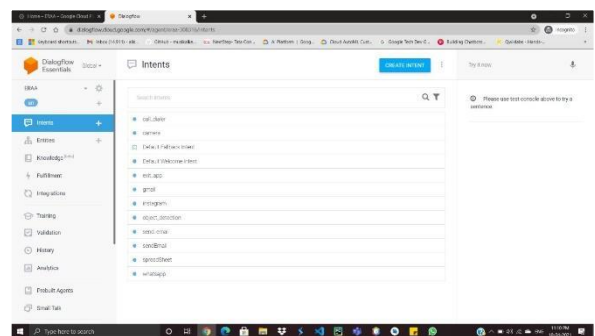

Figure 4: Google Dialogflow Console

With the help of Intents and Entities in each Dialog Agent the user expressions were matched and the task of responding to the same was carried out. The below Figure 5 and Figure 6 shows the various Intents and Entities included in our Application.
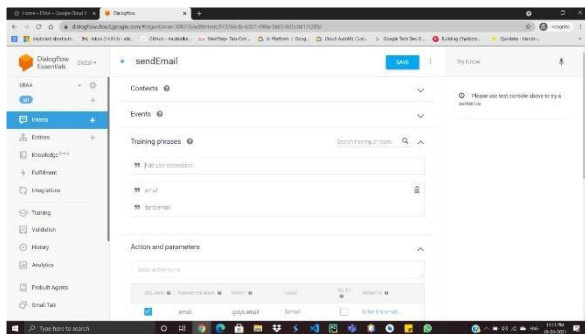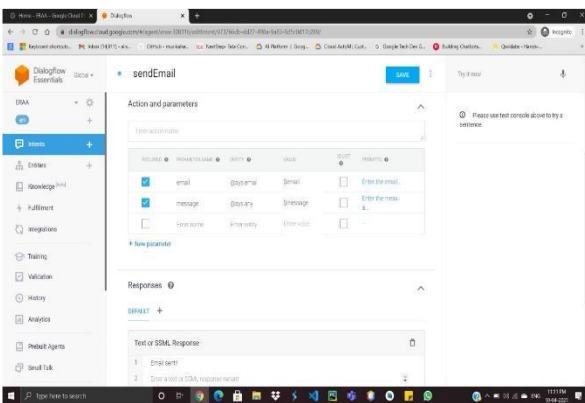

Figure 5: Intents and Entities

Figure 6: Intents and Entities

Then to check whether or not the conversations were handled we tested the entire dialog manager through the testing platform available in the Google Dialogflow. This is shown below Figure 7.
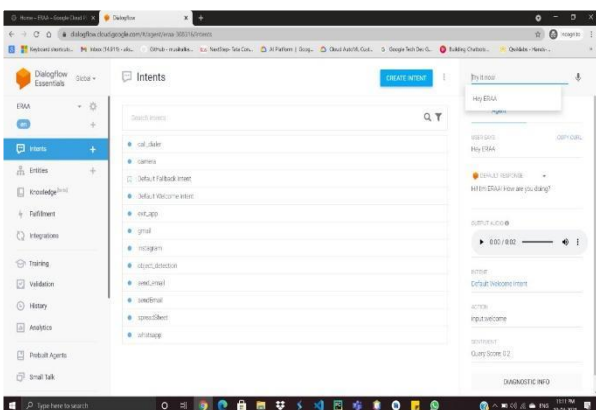

Figure 7: Testing

Speech Recognition

The input to the application could be either in text or in voice as per the convenience of the user. The text commands were automatically trained and tested with the help of in-built Natural Language Platform in the Google Dialogflow. This the Natural Language Processing(NLP) for the application was taken care of by the dialogflow. In terms of voice commands the Dialogflow also possessed an in-built feature of Speech-To-Text API that included various Machine Learning and Neural Network Algorithms to accomplish the extraction of text from speech even in noisy environments. Thus, the Speech Recognition was an added feature in ERAA, that made it more suitable to the users during their heavy workloads.

Other Features

The feature of opening device applications required accessing permission for the same which was handled by the 'permission_handler' plugin offered by Flutter. This plugin provided a cross-platform API to request and check permissions for the other applications present in the device. Thus, ERAA was then able to handle such requests from the user.

Object Detection Feature

The Object Detection was entirely based on the Convolutional Neural Network(CNN) Model that could help in detecting the objects with higher confidence levels. The Model was trained with various images and the network for the same was highly dense to achieve accurate results for the same. The below Figure 8 shows how CNN helps in detecting Objects. It includes various layers of the network namely Convolutional layer and the Max Pooling layer. Each layer extracts the most detectable feature from the image and converts them into a vector. A fully connected layer is developed within these vectors and then the image is detected based on its training set. This Neural

Network consisting of deep layers within itself is then able to detect the images on application of test set.
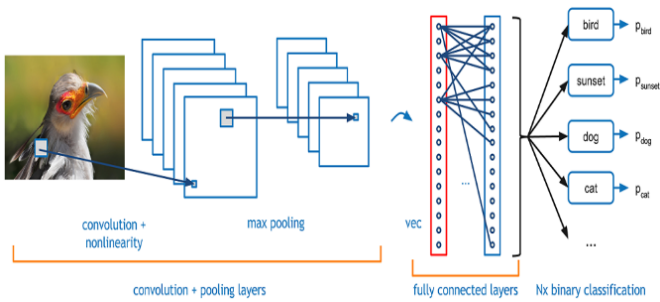

Figure 8: Object Detection with the help of CNN

VI. RESULTS

A. Sign-Up and Login-Page

The Login Page allows the user to access the application on providing the credentials i.e. user's mail id and password. The Sign-Up page is intended for the users whose information is not stored in the application. The required information is taken from the user and stored in Google's database with the help of Google Dialogflow.

The following section shows the pictures of the User Interface of the Application ERAA. It shows the Login page and Sign up module of the Application.

Figure 9: Application Logo
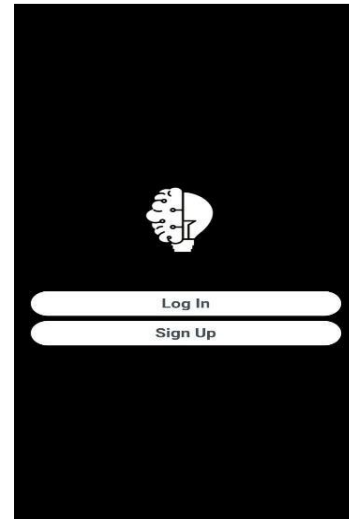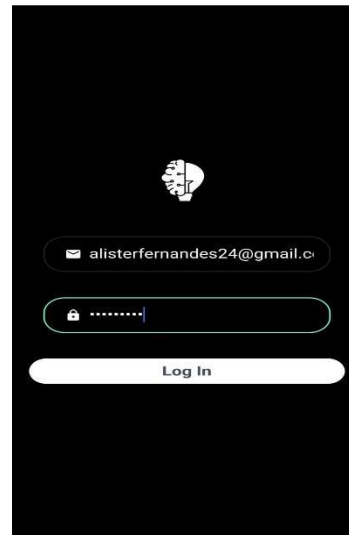

Figure 10: Sign-up/Login Page
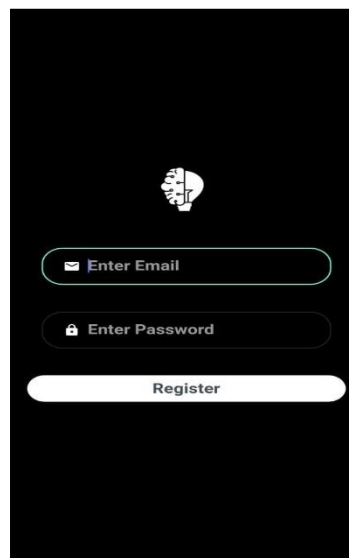

Figure 11: Login Page

Figure 12: Sign-Up Page

B. ERAA's Launch Page

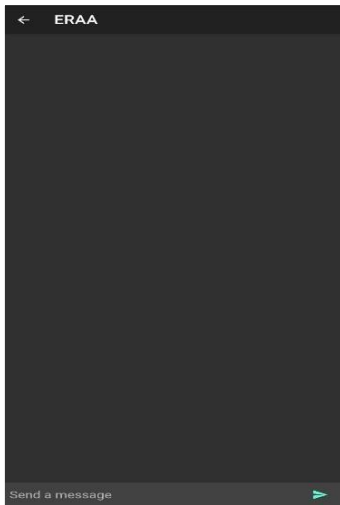After the Login Page the launch page is activated by the Application, the figure is shown as below.

Figure 13: ERAA's Launch Page

C. Opening of Various Applications through ERAA

ERAA is successfully able to open the Applications installed in the Device as the Application is provided with permission to full access to all the applications with help of Flutter Software Platform. The working for the same is as shown in the following images.
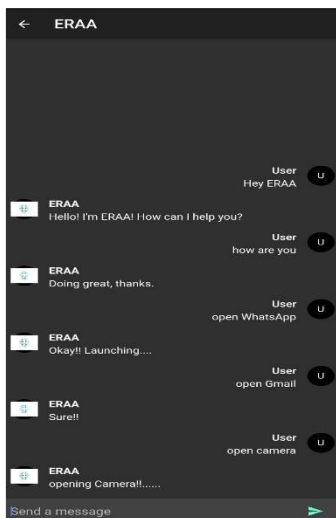

Figure 14: Accessing of various applications through ERAA

D. Object Detection Feature

The feature of Object Detection in the Application is an added feature apart from performing the basic tasks like accessing other applications. It enables the user to use Device Camera or the Images present in the Gallery of the device for the purpose of Object Detection. The feature provides the Confidence level of each label it matches to. Thus, this helps the user to identify the exact identity of the object.
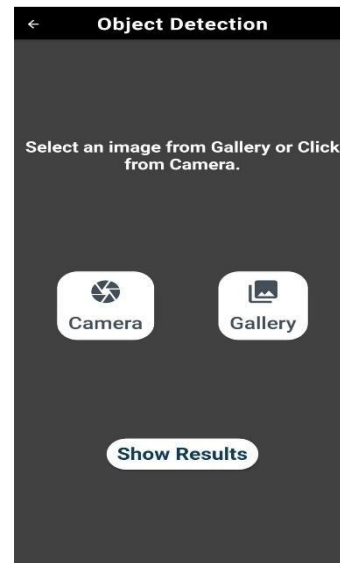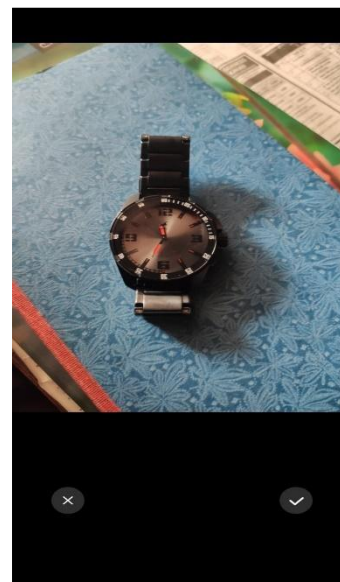

Figure 15(a): Object Detection Feature

Figure 15(b): Object Detection Feature

Results

Label: Jewellery Confidence: 61.19 %

Label: Metal Confidence: 59.34 %

Label: Clock Confidence: 56.44 %

Label: Wheel Confidence: 52.13 %

Label: Leisure Confidence: 51.85 %

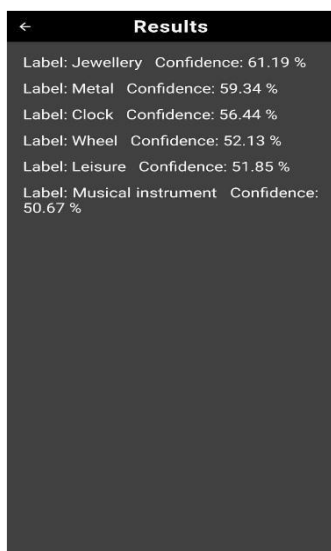Label: Musical instrument Confidence: 50.67 %

Figure 15(c): Object Detection Feature

Thus, ERAA is able to perform all the tasks as required for being a Virtual Personal Assistant

VII. CONCLUSION

The proposed research will help in understanding the basic methodologies which are used in developing their own Virtual Personal Assistant. The survey given in this paper will help in gaining a clear understanding of the difference between the Natural Language Understanding Platforms that is; IBM Watson and Google Dialog Flow. This will help in choosing the most appropriate amongst them for future projects. The survey also provided the research on various projects carried by using Google Dialog Flow and IBM Watson thus determining their features.

In our Project the application, ERAA, developed with the help of Google Dialog Flow was able to perform various tasks like accessing the other applications like WhatsApp, Instagram, Gmail that are installed in the device. Its User-friendly Platform developed with the help of Flutter provided ease in accessing the Application. With the help of Graphic packages in Flutter we were able to provide an attractive user-interface. It was able to

perform the basic features as required in an ideal Personal Assistant. The Speech Recognition feature in the application allowed users to perform the tasks by giving Voice Commands. The Application was also capable of handling small talk with the user. With the development of the Application, we were able to gain enough knowledge on Natural Language Understanding Platforms and Machine Learning Models which are the foundations for developing future Artificial Intelligence Models.

VIII. FUTURE SCOPE

In future, the proposed system can be used in making a software application which could be used in various sections of the society; be it healthcare services, educational institutions, etc. Many services which are externally available to the users could be incorporated in a single application with the help of Google Dialogflow and various other NLU Platforms available in the market and thereby making the application a multi-functioned software. We would encourage the readers to gain a deeper insight of Natural Language Platforms and use to develop applications which will cater the different needs of the society.

IX. ACKNOWLEDGEMENT

X. REFERENCES

[1]. Mohit Bansal, Dr. T. K. Thivakaran, "Analysis of Speech Recognition using Convolutional Neural Network", Journal of Engineering Sciences, Vol 11, Issue 1, 2020, Page 285-291.

[2]. J. Huang, J. Li and Y. Gong, "An analysis of convolutional neural networks for speech

recognition," 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), South Brisbane, QLD, Australia, 2015, pp 4989-4993, doi: 10.1109/ICASSP.2015.7178920

[3]. T. B. Mokgonyane, T. J. Sefara, T. I. Modipa, M. M. Mogale, M. J. Manamela and P. J. Manamela, "Automatic Speaker Recognition System based on Machine Learning Algorithms," 2019 Southern African Universities Power Engineering Conference/Robotics and Mechatronics/Pattern Recognition Association of South Africa (SAUPEC/RobMech/PRASA), Bloemfontein, South Africa, 2019, pp. 141-146, doi: 10.1109/RoboMech.2019.8704837.

[4]. Vineel Pratap, Qiantong Xu, Jacob Kahn, Gilad Avidov, Tatiana Likhomaneko, Awni Hannun, Vitaliy Liptchinsky, Gabriel Synnaeve, Ronan Collobert, " Scaling up Online Speech Recognition Systems using ConvNets", 27th January 2020.

[5]. A. B. Nassif, I. Shahin, I. Attili, M. Azzeh and K. Shaalan, "Speech Recognition Using Deep Neural Networks: A Systematic Review," in IEEE Access, vol. 7, pp. 19143-19165, 2019, doi: 10.1109/ACCESS.2019.2896880.

[6]. M. A. Khan, A. Tripathi, A. Dixit and M. Dixit, "Correlative Analysis and Impact of Intelligent Virtual Assistants on Machine Learning," 2019 11th International Conference on Computational Intelligence and Communication Networks (CICN), Honolulu, HI, USA, 2019, pp. 133-139, doi: 10.1109/CICN.2019.8902424.

[7]. Tulshan A.S., Dhage S.N. (2019) Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa. In: Thampi S., Marques O., Krishnan S., Li KC., Ciuonzo D., Kolekar M. (eds) Advances in Signal Processing and Intelligent Recognition Systems. SIRS 2018. Communications in Computer and Information Science, vol 968. Springer, Singapore.

[8]. N. A. Godse, S. Deodhar, S. Raut and P. Jagdale, "Implementation of Chatbot for ITSM Application Using IBM Watson," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-5, doi: 10.1109/ICCUBEA.2018.8697411.

[9]. Linda W. Lee, Amir Dabirian, Iran Paul McCarthy, Jan Kietzmann. (2020), " Making sense of text: artificial intelligence-enabled content analysis", European Journal of Marketing, Vol.54 No.3, pp 615-644.

[10]. Roberto Reyes, David Garza, Leonardo Garrido, Victor De la Cueva and Jorge Ramirez, "Methodology for the Implementation of Virtual Assistants for Education Using Google Dialogflow.", Advances in Soft Computing (pp.440-451).

[11]. Chinnapa Reddy Kanakanti and Sabitha R., "Ai and Ml Based Google Assistant for an Organization using Google Cloud Platform and Dialogflow", International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-5, January 2020, Page 2722-2727

[12]. Mayank Aggarwal and Mani Madhukar, "IBM's Watson Analytics for Health Care: A Miracle Made True.", Cloud Computing Systems and Applications in Healthcare. DOI: 10.4018/978-1-5225-1002-4.ch007.

[13]. G. E. Dahl, D. Yu, L. Deng, and A. Acero, "Contextdependent pre-trained deep neural networks for largevocabulary speech recognition," IEEE Trans. on Audio, Speech and Language Processing, vol. 20, no. 1, pp. 30– 42, 2012.

[14]. Sánchez-Díaz X., Ayala-Bastidas G., Fonseca-Ortiz P., Garrido L. (2018) A Knowledge-Based Methodology for Building a Conversational Chatbot as an Intelligent Tutor. In: Batyrshin I., Martínez-Villaseñor M., Ponce Espinosa H. (eds) Advances in Computational Intelligence. MICAI 2018. Lecture Notes in Computer Science, vol

11289. Springer, Cham. https://doi.org/10.1007/978-3-030-04497-8_14.

[15]. Winkler, Rainer & Söllner, Matthias. (2018), "Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis", Academy of Management Proceedings. 2018. DOI: 10.5465/AMBPP.2018.15903abstract

[16]. A. P. Singh, R. Nath and S. Kumar, "A Survey: Speech Recognition Approaches and Techniques," 2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), Gorakhpur, India, 2018, pp. 1-4, doi: 10.1109/UPCON.2018.8596954.

[17]. Ossama Abdel-Hamid, Abdelrahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu, "Convolutional Neural Networks for Speech Recognition", IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol.22,2010.

[18]. Ying Zhang, Mohammad Pezeshki, Philemon Brakel, Saizheng Zhang, Cesar Laurent Yoshua Bengio, Aaron Courville, "Towards End-to-End Speech Recognition with Deep Convolutional Neural Networks", arXiv:1701.02720v1,2017.

Cite this article as :