Intelligent Personal Local Voice Assistant using Machine Learning

1st Dibya Ranjan Rath
Computer Science and Engineering
Silicon Institute of Technology
Bhubaneswar, India
cse.20bcse62@silicon.ac.in

2nd Anubhav Mohanty
Computer Science and Engineering
Silicon Institute of Technology
Bhubaneswar, India
cse.20bcsb26@silicon.ac.in

3rd Dr. Satyananda Champati Rai
Prof., Computer Science and Engineering.)
Silicon Institute of Technology
Bhubaneswar, India
satya@silicon.ac.in

***Abstract*—The Personal Local Voice Assistant project aims to create a voice assistant that can perform various tasks such as answering queries, setting alarms, playing music, and more. The system uses natural language processing and machine learning techniques to understand and generate responses to user queries. The project also involves creating a user-friendly interface for the voice assistant, allowing for easy interaction with the system. In this paper, we discuss the design and implementation of the Personal Local Voice Assistant, including the algorithms and techniques used for natural language processing and machine learning. We also present the results of our performance analysis, which show that the decision tree algorithm outperforms other algorithms in terms of accuracy. Finally, we discuss potential future work and improvements for the Personal Local Voice Assistant.**

***Index Terms*—NLU, IOT, Voice assistant, text to speech, machine learning, Decision Tree Classifier**

I. Introduction

A state-of-the-art use of machine learning, the Internet of Things (IoT), and Natural Language Processing (NLP) technologies, an intelligent personal local voice assistant enables users to interact with their immediate surroundings, access information, and carry out a variety of tasks using voice commands. This cutting-edge system can offer an even more potent and practical user experience with the inclusion of text-to-speech and speech-to-text functionalities.

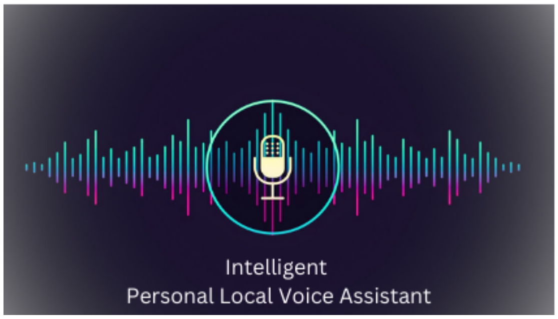


Fig. 1. Intelligent Personal Local Voice Assistant

A sophisticated machine learning system that can recognise and understand natural language commands and respond with the relevant actions is at the heart of an intelligent personal local voice assistant. The system can analyse and comprehend user input, extract important data such as intent, entities, and context, and create precise and pertinent responses by utilising deep learning algorithms and NLP approaches.

The system is frequently connected with IoT devices to improve efficiency, enabling seamless interaction with the real world. The voice assistant can operate smart devices in a home or office setting, such as thermostats, lighting, and appliances, by employing IoT sensors and actuators. In order to deliver more individualised and contextually aware solutions, it can also access and analyse data from IoT devices.

An intelligent personal local voice assistant can include text-to-speech and speech-to-text functionalities in addition to machine learning and IoT capabilities. Using text-to-speech technology, the system may produce audio responses that sound natural and human-like. This technology translates written text into spoken words. Users who like to hear information instead of reading will find this option to be especially helpful.

Contrarily, speech-to-text technology enables the system to translate spoken words into text, which may subsequently be processed and analysed using NLP methods. With the use of this capability, the system can recognise spoken orders, comprehend them in noisy or crowded settings, and produce precise and pertinent responses.

An intelligent personal local voice assistant may adjust and learn from user interactions over time to give users a more customised experience. The system can customise its responses to better meet the user's wants and preferences by employing machine learning algorithms that examine user behaviour and preferences.

Overall, a text-to-speech and speech-to-text capable personal local voice assistant is a potent use of machine learning, IoT, and NLP technologies that has the potential to completely change how individuals interact with their immediate surroundings. It gives users a simple, natural method to acquire information and manage their environment, making daily tasks simpler and more effective.

II. INTELLIGENCE VOICE ASSISTANT TECHNOLOGIES

Intelligent Voice Assistant technologies are a revolutionary development in the field of human-computer interaction. These technologies have been designed to enable people to communicate with electronic devices in a natural and intuitive manner, using spoken language. The technology employs machine learning techniques such as Natural Language Understanding (NLU), Text-to-Speech (TTS), Dialog Manager, and Voice Activation to interpret spoken language, extract information, and provide useful responses. One of the key components of Intelligent Voice Assistant technologies is Text-to-Speech (TTS) technology, which converts written text into spoken words. This allows the voice assistant to communicate with the user in a natural way, making the interaction more human-like. The Dialog Manager is responsible for managing the interaction between the user and the voice assistant, and determining the appropriate responses based on the user's input.

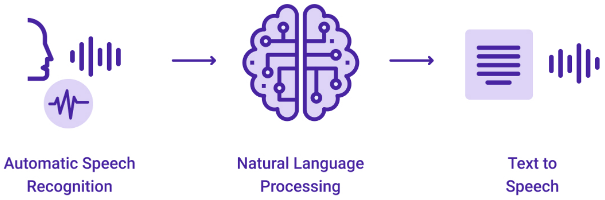


Fig. 2. Text to speech work processing

Another key component is Natural Language Understanding (NLU), which enables the voice assistant to understand and interpret spoken language. NLU uses various machine learning algorithms to extract useful information from the user's input, such as intent and context. This allows the voice assistant to provide more accurate and relevant responses to the user.

Voice Activation is another important component of Intelligent Voice Assistant technologies, allowing users to activate the voice assistant without the need for manual input. This makes it easier and more convenient for users to interact with the technology, as they can simply speak to the device without having to press any buttons or touch any screens. Intelligent Voice Assistant technologies have numerous practical applications, such as helping people with disabilities to use technology more easily, making it easier for people to access information and services, and improving the efficiency of various tasks. For example, a voice assistant can be used to control smart home devices, set reminders and appointments, or provide information on the weather, news, or traffic.

Overall, Intelligent Voice Assistant technologies have the potential to revolutionize the way people interact with technology, making it more accessible and intuitive. These technologies are constantly evolving and improving, and it is likely that they will become even more advanced and integrated into our daily lives in the future.

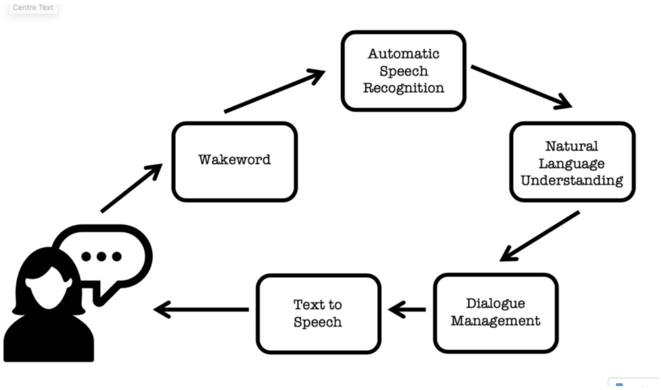


Fig. 3. Voice Technology

III. Example of Yandex Voice Assistant ”Alice”

An intelligent personal voice assistant called ”Alice” was created by Yandex, a Russian search engine business. To offer users a highly customised and natural voice assistant experience, Alice combines cutting-edge machine learning algorithms with natural language processing (NLP) methods.

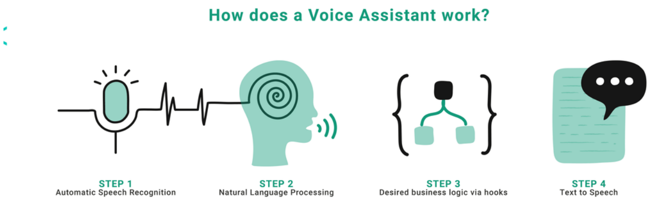

How does a Voice Assistant work?
STEP 1
Automatic Speech Recognition
STEP 2
Natural Language Processing
STEP 3
Desired business logic via hooks
STEP 4
Text to Speech

Fig. 4. Working principle of Voice Assistant

Alice's capacity to comprehend orders and questions in normal language and then respond appropriately to them is the foundation of its operation. Alice analyses and interprets user input using a variety of machine learning algorithms, such as deep neural networks. With the use of this technology, Alice is able to recognise subtleties in natural language and comprehend complex phrase patterns, resulting in responses that are precise and pertinent.

Advanced NLP approaches are also available to Alice to further improve its comprehension of human input. These methods enable the assistant to recognise and comprehend many forms of input, such as touch, text, and speech. Alice is able to produce highly customised and context-aware replies by examining the context and meaning of human input.

Due to the extremely scalable and adaptable technology stack on which Alice is built, a variety of hardware and software platforms can be integrated with it. The voice assistant is intended to function without any issues on a variety of gadgets, including smartphones, tablets, smart speakers, and other Internet of Things (IoT) devices. Due of this, users

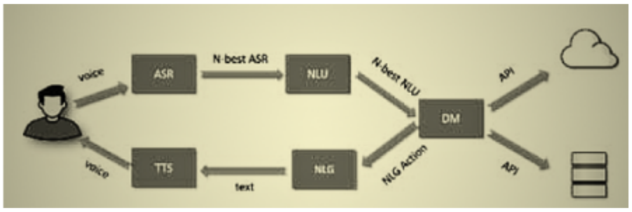


Fig. 5. Workflow of voice assistant

can use a variety of various input methods and access Alice whenever and from anywhere.

Alice's capacity to gradually learn from user interactions is one of its primary characteristics. The assistant analyses user behaviour and preferences using machine learning algorithms, enabling it to offer increasingly personalised and pertinent responses. With the help of this functionality, Alice becomes a highly flexible and intelligent voice assistant that can adapt to users' changing demands over time.

Along with powerful text-to-speech and speech-to-text technologies, Alice also has machine learning and NLP capabilities. This enables the assistant to hear spoken orders in noisy or crowded surroundings and to recognise and respond to audio commands in a manner that sounds natural and human-like.

Overall, Yandex Voice Assistant ”Alice” is a very smart and powerful voice assistant that combines machine learning, natural language processing (NLP), and other cutting-edge technologies to offer consumers a highly customised and natural experience. Alice is one of the most sophisticated and approachable voice assistants on the market right now because to its superior skills and adaptive learning abilities.

IV. Disadvantages of existing voice assistant

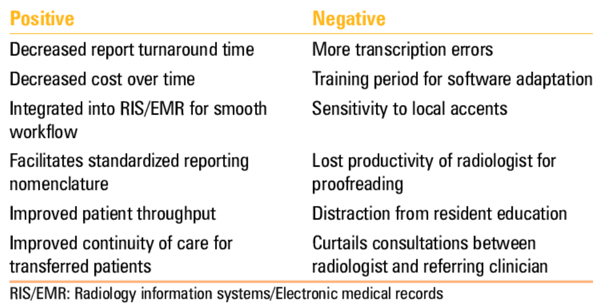
Dependence on Training Data Quality: For machine learning algorithms to produce correct results, they need a lot of high-quality training data. The quantity and quality of the data that personal local voice assistants have access to may be less than that of cloud-based systems, which may restrict their accuracy and applicability.

Overfitting: When given new or unfamiliar data, machine learning algorithms that have been overfit to a particular data set may give less accurate results. This is particularly true for personal voice assistants operating locally that have little access to training information.

Limited User Interaction: Compared to cloud-based solutions, personal local voice assistants could not have as much access to user comments. This may hamper their capacity to pick up on and eventually adjust to user preferences and behaviour.

Limited Resources: When compared to cloud-based systems, personal local voice assistants may have less processing and storage capabilities. They may not be able to handle larger volumes of data or more difficult activities as a result.

Limited Scalability: Local voice assistants for personal use may not be as scalable as cloud-based systems, which may restrict their capacity to handle high user and request volumes. Personal local voice assistants that employ machine learning have numerous advantages overall, but they also have several drawbacks that consumers should be aware of. When considering whether or not to utilise a personal local voice assistant that incorporates machine learning, these restrictions should be carefully taken into account.



Positive	Negative
Decreased report turnaround time	More transcription errors
Decreased cost over time	Training period for software adaptation
Integrated into RIS/EMR for smooth workflow	Sensitivity to local accents
Facilitates standardized reporting nomenclature	Lost productivity of radiologist for proofreading
Improved patient throughput	Distraction from resident education
Improved continuity of care for transferred patients	Curtails consultations between radiologist and referring clinician

RIS/EMR: Radiology information systems/Electronic medical records

Fig. 6. Pros and Cons of Voice Assistant

V. Development of Intelligent Voice Assistant for a Specific Problems of Interaction

The development of an intelligent voice assistant for a specific problem of interaction involves various stages and techniques that aim to enable the system to accurately recognize and respond to user commands. The first step is to choose an appropriate voice recognition system, and in this case, the PocketSphinx project was chosen due to its cross-platform capabilities and suitability for low-power embedded systems like the Raspberry Pi.

To generate the voice, the Festival engine was selected as it has good characteristics for voice generation on Linux operating systems. Once the voice recognition system is in place, the next stage involves creating a natural language processing system to recognize the user's intentions. This is done using machine learning-based algorithms that require data preparation to accurately predict results.

To handle a variety of output classes, such as a list of potential user intentions, a multi-class classification problem arises, where one label can contain multiple class labels. A table of synonyms and different pronunciation variants is created to prepare the data, which enables the system to make predictions. The words and synonyms are listed in a format commonly used in spoken language.

After defining synonyms, a list of answers to be predicted based on input data is determined. The predicted answers can be in any suitable format for application processing. The next step involves constructing a vocabulary from which a training sample is created. Since most machine learning algorithms operate on numeric data, each word in the dictionary must be mapped to a unique number. It is also recommended to reduce the dictionary size through stemming, a method that reduces different words by an order of magnitude and ultimately enhances recognition quality by reducing data.

In this project,a virtual assistant was created, and the data sets were used to train the model. However, the principle of collecting data remains the same for other tasks, with only the data sets changing. The collected data is used to train the machine learning algorithms, which will enable the voice assistant to understand the user's intentions and provide accurate responses. Overall, the development of an intelligent voice assistant requires careful planning and execution of various stages to create a system that is effective and efficient in responding to user commands.

VI. Investigation of Experiment and Training Sample Data

In any machine learning project, selecting the right algorithm for the job at hand is essential. The authors employed a decision tree model to categorise user input and produce suitable responses for the local voice assistant illustrated in the pseudocode. They looked into the data distribution in the training sample to assess the applicability of this technique. The resulting graphic demonstrated that the data had a normal distribution, arguing against the need for sophisticated classification techniques such naive Bayesian classifiers, decision trees, and random forests.

The Scikit-Learn Python library was employed by the authors to train the decision tree model. They utilised a single classifier to train one sample and a number of answers because many classification algorithms do not enable multiclass classification in pure form. They employed n-classifiers to handle n-binary responses. The training sample was relatively tiny, with only roughly 450 data points. They carried out a Leave-one-out cross-validation to make sure the model's estimate on tiny quantities of data was as accurate as possible. This method involved using the complete sample with the exception of one for training and one for validation, and dividing it into as many parts as there were experiments. This method took a lot of time and resources, but it gave the authors the most accurate model estimate.

The experiment's findings demonstrated that, while using the parameters maxdepth = 14 and maxfeatures = 91, the decision tree algorithm performed the best, with 0.93 accuracy of correct answers to cross-valency. The k-nearest neighbours approach had the lowest quality, with a correctness of 0.73 accuracy, while the Polynomial Bayesian Classifier displayed a correctness of 0.81 accuracy.

The decision tree model's implementation for categorising user input and producing replies is shown in the pseudocode of the local voice assistant. In order to train a decision tree model, the system first loads the response dataset from a CSV file into a pandas DataFrame, extracts the input and response values, and then uses those values. In an infinite loop, the system listens for user input, categorises it using the trained decision tree model, and then generates the relevant replies after initialising the voice recognition system. The system produces an error message to the user if it is unable to recognise the user's speech or if the speech recognition API service is down.

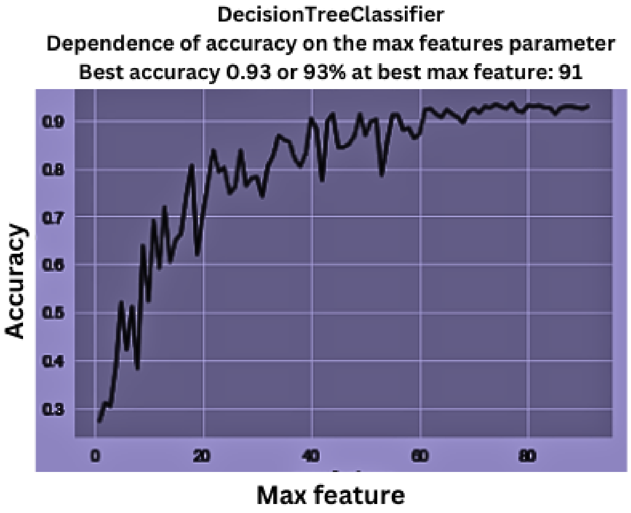
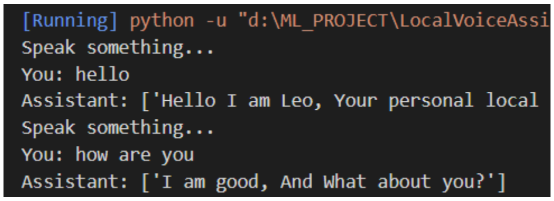


Fig. 7. Decision Tree Classifier Graph



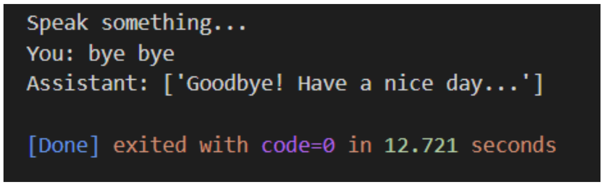
```
[Running] python -u "d:\ML_PROJECT\LocalVoiceAssi
Speak something...
You: hello
Assistant: ['Hello I am Leo, Your personal local
Speak something...
You: how are you
Assistant: ['I am good, And What about you?']
```

Fig. 8. Sample Output-1

In order to select the best machine learning algorithm to employ, the authors of the local voice assistant project looked into the data distribution in the training sample. To categorise user input and produce relevant responses, they utilised a decision tree model. The model was trained using Scikit-Learn, and a Leave-one-out cross-validation was used to get the model's best accurate estimate on tiny amounts of data. With a correctness of 0.93 occur, the results demonstrated that the decision tree algorithm performed the best. An overview of the processes needed to implement the decision tree model for categorising user input and producing responses is provided by the pseudocode of the local voice assistant.

VII. Conclusion

In conclusion, the development of a personal local voice assistant for specific problems of interaction was successfully achieved. By using existing systems like PocketSphinx and



```
Speak something...
You: bye bye
Assistant: ['Goodbye! Have a nice day...']

[Done] exited with code=0 in 12.721 seconds
```

Fig. 9. Sample Output-2

Festival engine, the system was able to recognize voice inputs and generate voice outputs. The system also utilized training intellectual algorithms based on machine learning methods to recognize the natural language of a person, i.e. recognition of intentions.

Through the process of data preparation, including creating a table of synonyms and different variants of pronunciation of keywords, and building a vocabulary, the system was able to accurately predict results. The best algorithm for recognizing intentions was found to be ”Trees of Solutions” with an accuracy of 93% for the presented data set.

This study demonstrated that voice assistant development and use are not just restricted to cloud computing. In IoT and IIoT systems, smart home systems, healthcare, security, and systems with a higher level of confidentiality, where the use of cloud technologies can be challenging, local systems enable expansion in the range of tasks they can be used to. The creation of a personal local voice assistant may offer a more secure and specialised response to particular interactional issues.

References

[1] Investigation and Development of the Intelligent Voice Assistant for the Internet of Things Using Machine Learning by Polyakov E. V., Maganov M. S. , Rolich A. Y. , Voskov L. S. Kachalova M. O. V. I. E., Polyakov S. V. In 2018 Moscow Workshop on Electronics and Networking Technologies (MINT)

[2] Dempsey P. The teardown: Google Home personal assistant //Engineering Technology. – 2017. – . 12. – No. 3. – . 80-81.

[3] Christy, A., S. Vaithyasubramanian, Auxeeliya Jesudoss and M. D. Anto Praveena. “Multimodal speech emotion recognition and classification using convolutional neural network techniques.” International Journal of Speech Technology 23 (2020): 381-388.

[4] López G., Quesada L., Guerrero L. A. Alexa vs. Siri vs. Cortana vs. Google Assistant: A Comparison of Speech-Based Natural User Interfaces //International Conference on Applied Human Factors and Ergonomics. – Springer, Cham, 2017. – . 241-250.