VOICE CONTROLLED HOME AUTOMATION USING AI AND NLP
MINI PROJECT REPORT

Submitted by

MOHANKUMAR.K **REGISTER NO: 21TH0164**

VISHAL.V **REGISTER NO: 21TN0024**

Under the guidance of

Mrs. S.Lavanya .M.E.,
Assistant Professor
DEPT OF AI&ML

in partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

MANAKULA VINAYAGAR INSTITUTE OF TECHNOLOGY,
KALITHEERTHALKUPPAM, PONDICHERRY
PONDICHERRY UNIVERSITY,
INDIA

MAY 2024

MANAKULA VINAYAGAR INSTITUTE OF TECHNOLOGY

PONDICHERRY UNIVERSITY

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

BONAFIDE CERTIFICATE

This is to certify that the project work entitled **"VOICE CONTROLLED HOME AUTOMATION USING AI AND NLP"** is a bonafide work done by **MOHANKUMAR.K [REGISTER NO: 21TH0164] ,VISHAL.V [REGISTER NO:21TN0024],** in partial fulfillment of the requirement for the award of B.Tech Degree in **DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** by Pondicherry University during the academic year 2023-24.

PROJECT GUIDE	**HEAD OF THE DEPARTMENT**
Mrs. S.LAVANYA	**Mr.R.RAJ BHARATH**
ASSISTANT PROFESSOR	**ASSOCIATE PROFESSOR & HOD**

Submitted for the University Examination held on............................

Internal Examiner **External Examiner**

ACKNOWLEDGEMENT

We express our deep sense of gratitude to **Theiva Thiru. N. Kesavan,** Founder, **Shri. M. Dhanasekaran,** Chairman & Managing Director**, Shri. S. V. Sugumaran,** ViceChairman, and **Dr. K. Narayanasamy** Secretary of **Sri Manakula Vinayagar Educational Trust, Puducherry** for providing the necessary facilities to successfully complete our mini project and report works.

We express our sincere thanks to our beloved Principal **Dr. S. Malarkkan** for having provided the necessary facilities and encouragement for the successful completion of this project work.

We express our sincere thanks to **Mr,R.Raj Bharath**, **Head of the Department**, Department of Artificial Intelligence and Machine Learning**,** for his support in making necessary arrangements for the conduction of the Project and for guiding us to execute our project successfully.

We express our sincere thanks to **Mrs.K.Anupriya**, Professor, Department of Artificial Intelligence and Machine Learning for her consistent reviews which motivated us in completing a project.

We thank all our department faculty members, non-teaching staff, and my friends for helping us to complete the document successfully on time.

We sincerely express our thanks to our mini project guide **Mrs.S.Lavanya,** Assistant Professor, Department of Artificial Intelligence and Machine Learning for continuously motivating and helping to develop our mini project.

We would like to express our eternal gratitude to our parents for the sacrifices they made for educating and preparing us for our future and their everlasting love and support..We thank the Almighty for blessing us with such wonderful people and for being with us always

SUSTAINABLE DEVELOPMENT GOALS (SGDs) MAPPING

Title : VOICE CONTROLLED HOME AUTOMATION USING AI AND NLP

SDG Goal : SDG Goal-7 (Affordable and Clean Energy)

SDG Goal-7:

Voice-controlled home automation utilizing AI and NLP plays a pivotal role in advancing Sustainable Development Goal 7 (SDG 7), which targets universal access to affordable, reliable, sustainable, and modern energy. By leveraging sophisticated technology, this innovation optimizes energy consumption within households, leading to enhanced energy efficiency and reduced overall energy usage. Through automated control of various devices such as lighting, heating, and cooling systems, tailored to user preferences and real-time environmental conditions, this technology facilitates a more judicious use of energy resources.

By intelligently managing energy consumption, voice-controlled home automation not only contributes to cost savings for households but also fosters a significant reduction in environmental impact. The seamless integration of AI and NLP allows for dynamic adjustments in energy usage, ensuring optimal efficiency without compromising comfort or convenience. Consequently, this technology supports the transition towards cleaner and more sustainable energy sources, aligning with global efforts to combat climate change and promote environmental stewardship.

Moreover, by making energy management more accessible and intuitive, voice-controlled home automation enhances inclusivity and affordability. Its user-friendly interface caters to a diverse range of individuals, including those with disabilities or limited technical expertise, thereby promoting equitable access to energy-efficient solutions. By democratizing access to sustainable technologies, this innovation empowers individuals and communities to actively participate in the global transition towards a more sustainable future, where clean energy is readily accessible and affordable for all.

ABSTRACT

The aim of home automation is to make our lives easier and to improve the quality of life. The concept of Smart Homes builds on the progressing maturity of areas such as Artificial Intelligence and Natural Language Processing. Here, natural language processing (NLP) plays a vital role since it acts as an interface between human interaction and machines. Through NLP users can either command or control devices at home even though disabled persons command or request varies from presets. An application area of AI is Natural Language Processing (NLP). Voice assistants incorporate AI using cloud computing and can communicate with the users in natural language. Voice assistants are easy to use and thus there are millions of devices that incorporate them in households nowadays. Our project aims at providing a fully automated voice based solution that our users can rely on, to perform more than just switching on/off the appliances. The user sends a command through speech to the mobile device, which interprets the message and sends the appropriate command to the specific appliance. The primary objective is to construct a useful voice-based system that utilizes AI and NLP to control all domestic applications and services and also learn the user preferences over time using machine learning algorithms

TABLE OF CONTENT

CHAPTER 1
INTRODUCTION

Home Automation is the process of making access inside the house without anyone's help. Due to this, a physical or mental condition that restricts a person's need can be satisfied. Most people don't like to be dependent on others to do their daily work, especially the old people and the physically challenged people. So this technology will help these kinds of people to live their life without depending on others. Voice assistants use technologies like voice recognition, speech synthesis, and Natural Language Processing (NLP) to provide services to the users. A voice interface is essential for Internet of Things (IoT) devices that lack touch capabilities. Besides smartphones, voice assistants are now incorporated in devices that are equipped with a microphone and a speaker to communicate with the users, called smart speakers. Due to the aforementioned drawbacks of the current systems and with the rise of IoT technology, the proposed system is predicated on IoT which acts as an interface between physical home appliances and user, as well as having functionality kindred to a digital assistant.

1.1 OVERVIEW

Voice-controlled home automation leverages Artificial Intelligence (AI) and Natural LanguageProcessing (NLP) to enable users to interact with their smart home devices using voice commands. This project aims to develop a sophisticated system that enhances user experience and efficiency in controlling various aspects of their home environment through seamless voice interaction.

1.2 AIMS AND OBJECTIVES

- Develop a robust voice-controlled home automation system.
- Enhance speech recognition accuracy and NLP capabilities.
- Integrate seamlessly with a wide range of smart home devices.
- Provide intuitive user experiences through natural language interaction.

1.3 WORKING PRINCIPLE

Home Automation is the process of making access inside the house without anyone's help. Due to this, a physical or mental condition that restricts a person's need can be satisfied. Most people don't like to be dependent on others to do their daily work, especially the old people and the physically challenged people. So this technology will help these kinds of people to live their life without depending on others. Voice assistants use technologies like voice recognition, speech synthesis, and Natural Language Processing (NLP) to provide services to the users. A voice interface is essential for Internet of Things (IoT) devices that lack touch capabilities. Besides smartphones, voice assistants are now incorporated in devices that are equipped with a microphone and a speaker to communicate with the users, called smart speakers. Due to the aforementioned drawbacks of the current systems and with the rise of IoT technology, the proposed system is predicated on IoT which acts as an interface between physical home appliances and user, as well as having functionality kindred to a digital assistant.

1.4 IMPLEMENTATION

- **Platform Selection**: Chooses a suitable development platform such as Python or Node.js.
- **Speech Recognition**: Utilizes technologies like Google Cloud Speech-to-Text or Mozilla DeepSpeech.
- **Nlp**: Implements NLP algorithms using libraries like NLTK or spaCy.
- **Device Integration**: Establishes connections with smart home devices using APIs or IoT protocols.

1.5 VERIFICATION

Verification for the topic "Voice-Controlled Home Automation Using AI and NLP" involves ensuring that the project is feasible, relevant, and aligns with current technological capabilities. Here's how verification can be approached:

CHAPTER 2
LITERATURE SURVEY

1.THE NLP COOKBOOK: MODERN RECIPES FOR TRANSFORMER BASED DEEP LEARNING ARCHITECTURE:

For higher performance, greater learning was required which resulted in larger data storage and model size. Due to the model's enormity and implicit knowledge storage, its learning ability had caveats in terms of efficient information access. Current Knowledge Retrieval models like ORQA, REALM , RAG , DPR attempt to alleviate implicit storage concerns of language models by providing external access to interpretable modular knowledge. This was achieved by supplementing the language model's pre-training with a 'knowledge retriever' that facilitated the model to effectively retrieve and attend over explicit target documents from a large corpus like Wikipedia. Further, the Transformer model's inability to handle input sequences beyond a fixed token span inhibited them to comprehend large textual bodies holistically. This was particularly evident when related words were farther apart than the input length. Hence, to enhance contextual understanding, architectures like Transformer-XL Longformer , Big Bird , were introduced with modified attention mechanisms to process longer sequences. Also, due to the surge in demand for NLP models to be economically viable and readily available on edge devices, innovative compressed models were launched based on generic techniques. These are apart from the Distillation, Pruning, and Quantization techniques described earlier. Such models deploy a wide range of computing optimization procedures ranging from hashing , sparse attention , factorized embedding parameterization , replaced token detection, inter-layer parameter sharing , or a combination of the above mentioned.The NLP Cookbook: Modern Recipes for Transformer Based Deep Learning Architecture

LIMITATIONS:

1.Limitations Increased Storage and Model Size: While achieving higher performance, the necessity for greater learning resulted in larger data storage and model size. This poses challenges for deployment, especially on resource-constrained devices or environments where storage capacity is limited.

2.Limited Efficient Information Access: The model's enormity and implicit knowledge storage can lead to limitations in efficiently accessing information. This hampers the model's learning ability, as it may struggle to effectively retrieve and utilize relevant information when needed.

2.PRIVACY-PRESERVING DEEP LEARNING NLP MODELS FOR CANCER REGISTRIES:

Among different AI approaches, Deep Learning (DL) has been successfully applied to classify and recognize complex features in images, speech, and text data. Recent studies have shown the potential of DL models in automatically extracting cancer key characteristics from cancer pathology reports by achieving accuracy superior to traditional machine learning NLP methods. Successfully applying DL in the specific domain requires a large training corpus that has similar characteristics as the prospective testing data. Furthermore, this success is proportional to the size of the training corpus. Obtaining a large enough corpus from a single cancer registry is challenging, particularly with respect to rare cancer anatomic location sites (i.e., body organs where cancer develops) and histologies (i.e., different cell types). This challenge can be overcome by aggregating cancer pathology reports from multiple cancer registries in a centralized hub which can serve as a neutral entity to train a generalized model on all the data. Upon completion of training, the trained model can be shared with the registries. However, data privacy and confidentiality concerns prevent cancer registries from sharing patient data and benefiting from each other's knowledge by leveraging DL.

LIMITATIONS:

Data Requirement and Size: DL models require large amounts of training data to perform optimally. The paragraph mentions the need for a training corpus with characteristics similar to the testing data, emphasizing that the success of DL applications is proportional to the size of the training data. This is a common challenge in machine learning where insufficient or non-representative data can significantly limit the performance and generalizability of models.

Data Availability and Rarity Issues: Gathering a sufficient corpus of data is particularly challenging for rare cancers due to the low occurrence of such cases. The rarity of certain cancer types means that the available data might not be enough to train robust DL models that can accurately classify or analyze these rarer conditions.

3.IDENTIFYING SECURITY AND PRIVACY VIOLATION RULES IN TRIGGER–ACTION IOT PLATFORMS WITH NLP MODELS:

However, when defining an ECA rule, users may create several serious risks for their privacy and/or the security of the smart environment, mainly due to their general inexperience and lack of technical knowledge. For example, a user who wishes to play music from his/her smartphone as soon as his/her Bluetooth earphones connect may define the rule: “If a Bluetooth device is connected to the smartphone, then the smartphone plays the audio files contained in a folder.” This rule implies an important privacy disclosure as there may be a scenario where the selected folder contains personal audio files, and the smartphone connects to a Bluetooth speaker. Also, the rule in Fig. 1 could raise risks for the users. In fact, the opening of shutters is ruled only by the internal temperature of the smart house. In such a scenario, if the user forgets that the rule is active, shutters might open also on hot summer days when the house is empty, e.g., when the owners are on vacation, providing an entry point for thieves. Moreover, open windows represent a factor of risk for unsupervised children.

LIMITATIONS:

Privacy Risks: The example provided demonstrates how a seemingly simple automation rule can lead to unintended privacy breaches. By setting a rule that automatically plays audio files from a folder upon connection to any Bluetooth device, there is a risk that personal or sensitive audio files could be played where they can be heard by unintended parties, such as over a Bluetooth speaker in a public or shared space. This risk stems from the broad scope of the trigger ("any Bluetooth device") without specifying the type of device or additional conditions for security.

Security Vulnerabilities: The scenario also illustrates potential security issues in a smart home environment. For instance, automating the opening of house shutters based solely on internal temperature without considering other factors like whether the house is occupied can lead to security vulnerabilities. This could inadvertently signal to potential thieves that the house is empty, especially if such actions are predictable and occur at regular times (e.g., during a vacation).

4.NLP-BASED AUTOMATED COMPLIANCE CHECKING OF DATA PROCESSING AGREEMENTS AGAINST GDPR:

DPAs are legally binding agreements that regulate the data processing activities according to GDPR. To be deemed GDPRcompliant, the DPA must explicitly cover all the criteria imposed by the GDPR provisions concerning data processing. Establishing a DPA includes setting terms for how data is used, stored, protected, and accessed as well as defining the obligations and rights of the controller and processor. By signing a DPA, the processor is obliged to ensure that any software system deployed for processing personal data has to also comply with GDPR. To illustrate, consider the scenario of an educational institution (called as Sefer University) which shares personal data of its employees with a third-party service provider (e.g., accounting office) for a particular purpose (e.g., payroll administration). Examples of shared personal data include: first and last name, birth date, marital status, annual summary of leave or sickness absences and a scanned copy of a valid personal identification. The data controller in this scenario is Sefer University which collects the personal data of the employees (data subjects). The data processor is the service provider (in our example, the accounting office) performing financial services on behalf of the controller. Details concerning this service provider (e.g., the name and address of the hired accounting office) might not be shared with employees. To ensure that their personal data is sufficiently protected, GDPR requires Sefer university to sign a DPA with the accounting office.

LIMITATIONS:

Complexity of Compliance: The DPA must cover all GDPR criteria concerning data processing, which can be complex and exhaustive. This requires a deep understanding of GDPR provisions and how they apply specifically to the data handling practices of an entity. Misinterpretations or omissions in addressing these provisions in the DPA can lead to non-compliance.

Dependency on Third Parties: In scenarios like the one described with Sefer University and the accounting office, the university relies on third-party service providers to handle personal data. While the DPA obligates these third parties to comply with GDPR, the university has limited control over the actual day-to-day operations of these service providers. Ensuring continuous compliance across different third parties adds layers of complexity and risk.

5.UTILIZING A RAPIDLY EXPLORING RANDOM TREE FOR HAZARDOUS GAS EXPLORATION IN A LARGE UNKNOWN AREA:

In terms of gas distribution mapping, developed a method to drive a robot using an artificial potential field method, while used a particle filter algorithm to perform gas source localization. However, the problem was not complex, as the environment was assumed to be known and free from obstacles. Recent research by used optimal policies to perform gas distribution mapping in a cluttered environment, but the robot previously knew the occupancy map. Another researcher developed an integration between Simultaneous Localization And Mapping (SLAM) and gas distribution mapping in an unknown area, but the robot was controlled remotely. By aggregating all the problems above, we address the development of a fully autonomous robot to explore and exploit hazardous gas contamination in a cluttered and unknown large area.

LIMITATIONS:

Dependency on Pre-existing Data: The approach relies on the robot having prior knowledge of the occupancy map of the environment. This requirement for pre-known maps limits the robot's ability to operate effectively in environments that have not been mapped previously or in situations where the environment might change, such as in disaster zones or evolving industrial areas.

Lack of Autonomy: Although this method addresses the challenge of operating in unknown areas by using SLAM, the reliance on remote control suggests a lack of full autonomy. This reliance can be a significant drawback in situations where remote control might be impractical or impossible, such as in areas with poor communication infrastructure or where human operators cannot be continuously present.

CHAPTER 3
EXISTING SYSTEM

The existing technologies, in these regard are far outdated than what we have proposed. Most of the smart home automation systems that are existing only automate the basic process of changing the state of the appliances to ON/OFF. There are many smart home mechanization systems in the market that aim to automate the basic operations of these home appliances using various technologies such as GSM (Global System for Mobile), NFC (Near-Field Communication) and Wi-Fi. The existing smart home systems that have either been implemented or proposed have an elaborate procedure to interact with the home appliances. Some include pressing a button in a static location while some others include giving commands through a mobile device. Various technology companies have been trying to create amazing products in the department of home automation system since a decade ago. However, the Internet of things has become a briskly growing field only in the recent past.

CHAPTER 4
PROPOSED SYSTEM

Our project aims to provide the most easy and efficient way to interact with home appliances by giving voice commands in human (natural) language. We plan on eliminating the tedious process of clicking through various application screens with just one voice command. The natural language processing in the project provides a personal connection with our system. Primarily the user is authenticated by entering the specified username and password in the mobile device. The user sends a voice command to the mobile device, which interprets the message and sends the appropriate command to the specific appliance. The voice command given by the user is interpreted by the mobile device using Natural Language processing. The mobile device acts as a central console; it determines what operation must be done by which appliance to fulfill the user's request. The central console can also be either a desktop application, web application or a smart phone application as all of the data transferred can be processed by the cloud. However, for the convenience of the user and increased mobile capabilities we will be using a smart phone in this project.

CHAPTER 5
REQUIREMENTS

5.1 HARDWARE REQUIREMENT

- Microphones
- Processing Unit
- Smart Home Devices
- Networking Devices
- Speakers

5.2 SOFTWARE REQUIREMENT

- Operating System
- Voice Recognition Software
- Home Automation Platform
- Programming Languages
- APIs for Smart Devices
- Security Software
- Cloud Services (Optional

LIBRARIES

Tensorflow:

TensorFlow is an open-source machine learning (ML) framework developed by Google Brain and released in 2015. It provides a comprehensive ecosystem of tools, libraries, and resources for building and deploying ML models across a range of platforms, from desktops to mobile devices and cloud servers. TensorFlow is widely used in both research and production environments for various ML tasks, including

PyTorch:

PyTorch is an open-source machine learning framework developed by Facebook's AI Research lab (FAIR). It's known for its flexibility, ease of use, and dynamic computational graph, making it a popular choice among researchers and practitioners in the machine learning community. Here's an overview of PyTorch

Keras:

Keras is an open-source neural network library written in Python. It's designed to be user- friendly, modular, and extensible, making it an excellent choice for beginners and experts alike. Here's an overview of Keras:

CHAPTER 6
DESIGN

6.1 NATURAL LANGUAGE PROCESSING MODEL:

Natural Language Processing (NLP) is one of the hottest areas of artificial intelligence (AI) thanks to applications like text generators that compose coherent essays, chatbots that fool people into thinking they're sentient, and text-to-image programs that produce photorealistic images of anything you can describe. Recent years have brought a revolution in the ability of computers to understand human languages, programming languages, and even biological and chemical sequences, such as DNA and protein structures, that resemble language. The latest AI models are unlocking these areas to analyze the meanings of input text and generate meaningful, expressive output.

6.2 WHAT IS NATURAL LANGUAGE PROCESSING (NLP)

Natural language processing (NLP) is the discipline of building machines that can manipulate human language — or data that resembles human language — in the way that it is written, spoken, and organized. It evolved from computational linguistics, which uses computer science to understand the principles of language, but rather than developing theoretical frameworks, NLP is an engineering discipline that seeks to build technology to accomplish useful tasks. NLP can be divided into two overlapping subfields: natural language understanding (NLU), which focuses on semantic analysis or determining the intended meaning of text, and natural language generation (NLG), which focuses on text generation by a machine. NLP is separate from — but often used in conjunction with — speech recognition, which seeks to parse spoken language into words, turning sound into text and vice versa.

6.3 WHY DOES NATURAL LANGUAGE PROCESSING (NLP) MATTER?

NLP is an integral part of everyday life and becoming more so as language technology is applied to diverse fields like retailing (for instance, in customer service chatbots) and medicine (interpreting or summarizing electronic health records). Conversational agents such as Amazon's and Apple's utilize NLP to listen to user queries and find answers. The most sophisticated such agents — such as GPT-3, which was recently opened for — can generate sophisticated prose on a wide variety of topics as well as power chatbots that are capable of holding coherent conversations. Google uses NLP to improve its search engine results, and social networks like Facebook use it to detect and filter hate speech.NLP is growing increasingly sophisticated, yet much work remains to be done. Current systems are prone to bias and incoherence, and occasionally behave erratically. Despite the challenges, machine learning engineers have many opportunities to apply NLP in ways that are ever more central to a functioning society

6.4 WHAT IS NATURAL LANGUAGE PROCESSING (NLP) USED FOR?

NLP models work by finding relationships between the constituent parts of language — for example, the letters, words, and sentences found in a text dataset. NLP architectures use various methods for data preprocessing, feature extraction, and modeling. Some of these processes are:

Data preprocessing: Before a model processes text for a specific task, the text often needs to be preprocessed to improve model performance or to turn words and characters into a format the model can understand. Data-centric AI is a growing movement that prioritizes data preprocessing. Various techniques may be used in this data preprocessing:

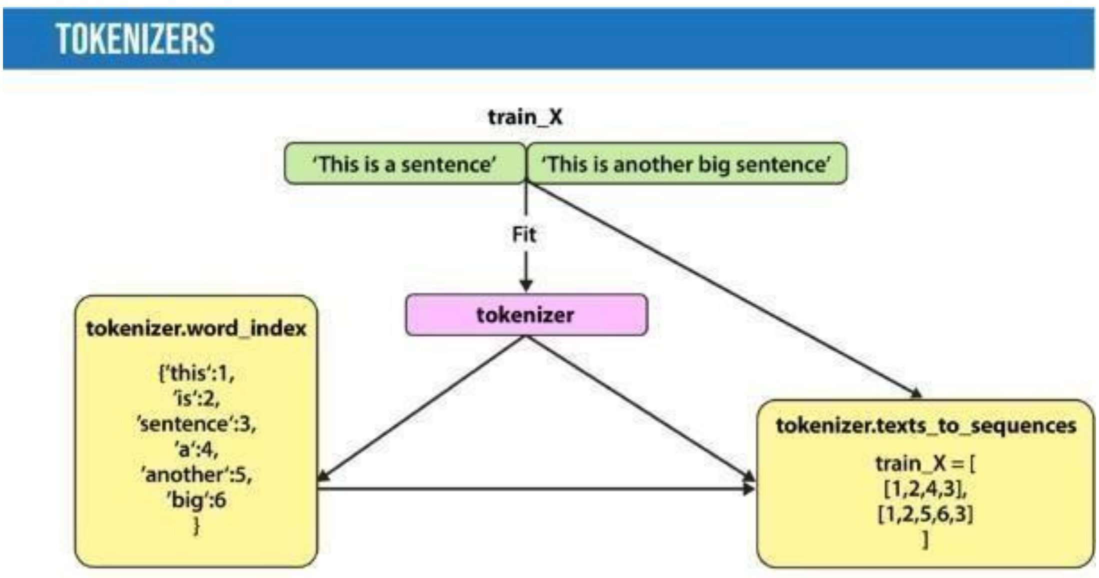
Stemming and lemmatization: Stemming is an informal process of converting words to their base forms using heuristic rules. For example, "university," "universities," and "university's" might all be mapped to the base univers. (One limitation in this approach is that "universe" may also be mapped to univers, even though universe and university don't have a close semantic relationship.) Lemmatization is a more formal way to find roots by analyzing a word's morphology using vocabulary from a dictionary. Stemming and lemmatization are provided by libraries like spaCy and NLTK.

Sentence segmentation breaks a large piece of text into linguistically meaningful sentence units. This is obvious in languages like English, where the end of a sentence is marked by a period, but it is still not trivial. A period can be used to mark an abbreviation as well as to terminate a sentence, and in

this case, the period should be part of the abbreviation token itself. The process becomes even more complex in languages, such as ancient Chinese, that don't have a delimiter that marks the end of a sentence.

Stop word removal aims to remove the most commonly occurring words that don't add much information to the text. For example, "the," "a," "an," and so on.

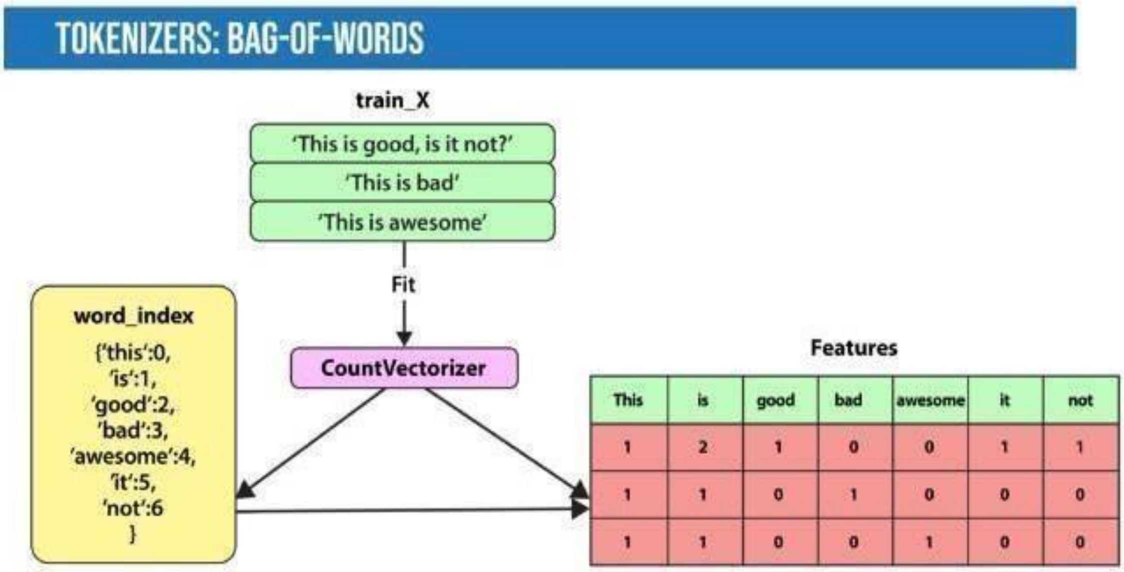
Tokenization splits text into individual words and word fragments. The result generally consists of a word index and tokenized text in which words may be represented as numerical tokens for use in various deep learning methods. A method that instructs language models to can improve efficiency.



Given a corpus of documents, a tokenizer maps every word to an index. Then it can translate any document into a sequence of numbers.

Feature extraction: Most conventional machine-learning techniques work on the features – generally numbers that describe a document in relation to the corpus that contains it – created by either Bag-of-Words, TF-IDF, or generic feature engineering such as document length, word polarity, and metadata (for instance, if the text has associated tags or scores). More recent techniques include Word2Vec, GLoVE, and learning the features during the training process of a neural network.

Bag-of-Words: Bag-of-Words counts the number of times each word or n-gram (combination of n words) appears in a document. For example, below, the Bag-of-Words model creates a numerical representation of the dataset based on how many of each word in the word_index occur in the document.



Bag-of-Words (through the CountVectorizer method) encodes the total number of times a document uses each word in the associated corpus.

TF-IDF:In Bag-of-Words, we count the occurrence of each word or n-gram in a document. In contrast, with TF-IDF, we weight each word by its importance. To evaluate a word's significance, we consider two things:

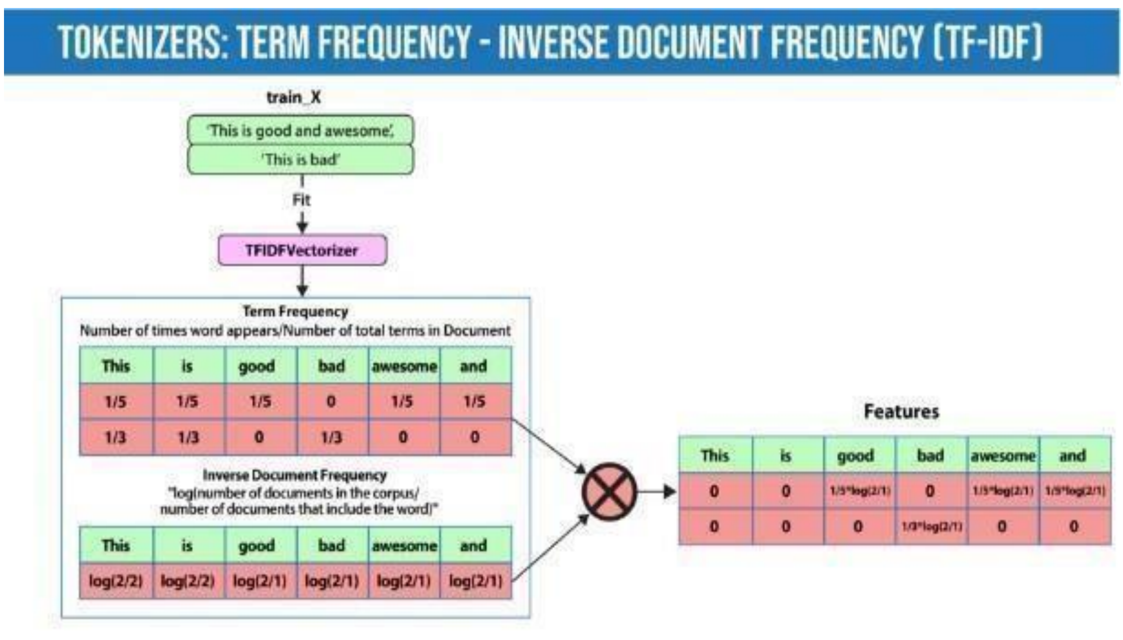
Term Frequency:How important is the word in the document?

TF(word in a document)= Number of occurrences of that word in document / Number of words in document

Inverse Document Frequency: How important is the term in the whole corpus?

IDF(word in a corpus)=log(number of documents in the corpus / number of documents that include the word)

A word is important if it occurs many times in a document. But that creates a problem. Words like "a" and "the" appear often. And as such, their TF score will always be high. We resolve this issue by using Inverse Document Frequency, which is high if the word is rare and low if the word is common across the corpus. Thescore of a term is the product of TF and IDF.



TF-IDF creates features for each document based on how often each word shows up in a document versus the entire corpus.

Word2Vec, introduced in 2013, uses a vanilla neural network to learn high-dimensional word embeddings from raw text. It comes in two variations: Skip-Gram, in which we try to predict surrounding words given a target word, and Continuous Bag-of-Words (CBOW), which tries to predict the target word from surrounding words. After discarding the final layer after training, these models take a word as input and output a word embedding that can be used as an input to many NLP tasks. Embeddings from Word2Vec capture context. If particular words appear in similar contexts, their embeddings will be similar.

GLOVE is similar to Word2Vec as it also learns word embeddings, but it does so by using matrix factorization techniques rather than neural learning. The GLoVE model builds a matrix based on the global word-to-word co-occurrence counts.

Modeling:After data is preprocessed, it is fed into an NLP architecture that models the data to accomplish a variety of tasks.
Numerical features extracted by the techniques described above can be fed into various models depending on the task at hand. For example, for classification, the output from the TF-IDF vectorizer could be provided to logistic regression, naive Bayes, decision trees, or gradient boosted trees. Or, for named entity recognition, we can use hidden Markov models along with n-grams.Deep typically work without using extracted features, although we can still use TF-IDF or Bag-of-Words features as an input.

Language Models: In very basic terms, the objective of a language model is to predict the next word when given a stream of input words. Probabilistic models that use Markov assumption are one example: $P(Wn)=P(Wn|Wn-1)$

Deep learning is also used to create such language models. Deep-learning models take as input a word embedding and, at each time state, return the probability distribution of the next word as the probability for every word in the dictionary. Pre-trained language models learn the structure of a particular language by processing a large corpus, such as Wikipedia. They can then be fine-tuned for a particular task. For instance, BERT has been fine-tuned for tasks ranging from to writing headlines.Top Natural Language Processing (NLP) Techniques

Most of the NLP tasks discussed above can be modeled by a dozen or so general techniques. It’s helpful to think of these techniques in two categories: Traditional machine learning methods and deep learning

6.5 TRADITIONAL MACHINE LEARNING NLP TECHNIQUES:

Logistic regression is a supervised classification algorithm that aims to predict the probability that an event will occur based on some input. In NLP, logistic regression models can be applied to solve problems such as sentiment analysis, spam detection, and toxicity classification.

Naive Bayes is a supervised classification algorithm that finds the conditional probability distribution P(label | text) using the following Bayes formula:

P(label | text) = P(label) x P(text|label) / P(text)

and predicts based on which joint distribution has the highest probability.

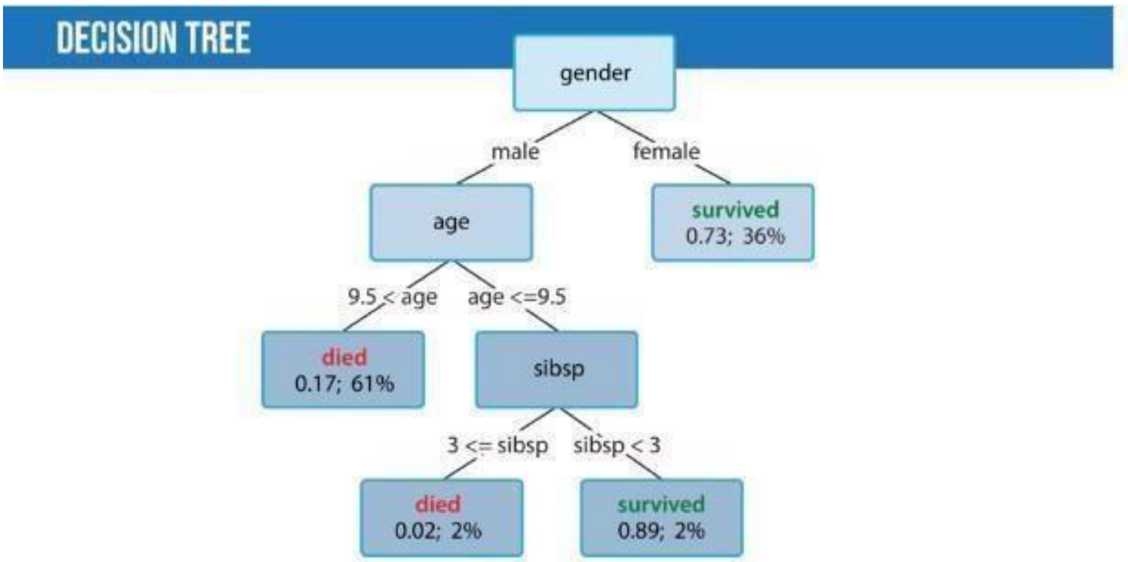
The naive assumption in the Naive Bayes model is that the individual words are independent. Thus:

P(text|label) = P(word_1|label)*P(word_2|label)*…P(word_n|label)

In NLP, such statistical methods can be applied to solve problems such as spam detection or finding bugs in software code.

Decision trees are a class of supervised classification models that split the dataset based on

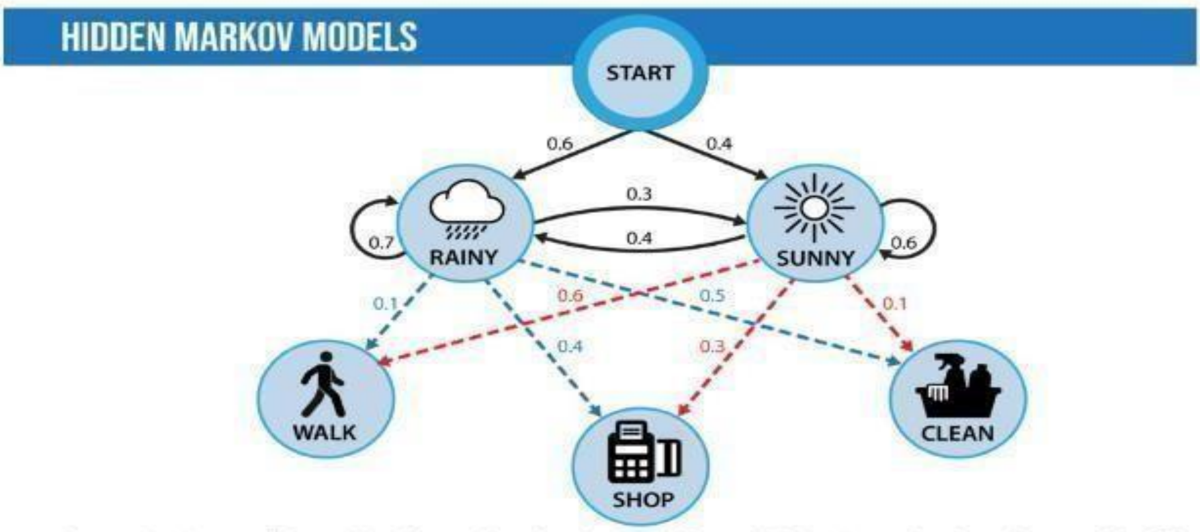
different features to maximize in those splits.



This decision tree assesses the survival of males and females aboard the Titanic when the ship sank. "Sibsp" stands for the number of siblings and spouses. The numbers of each leaf of the tree show first the probability of survival and second the percentage of examples classified by each leaf. Females had a 73 percent chance of survival, and young males with less than three siblings had a 89 percent chance of survival.

Latent Dirichlet Allocation (LDA) is used for topic modeling. LDA tries to view a document as a collection of topics and a topic as a collection of words. LDA is a statistical approach. The intuition behind it is that we can describe any topic using only a small set of words from the corpus.

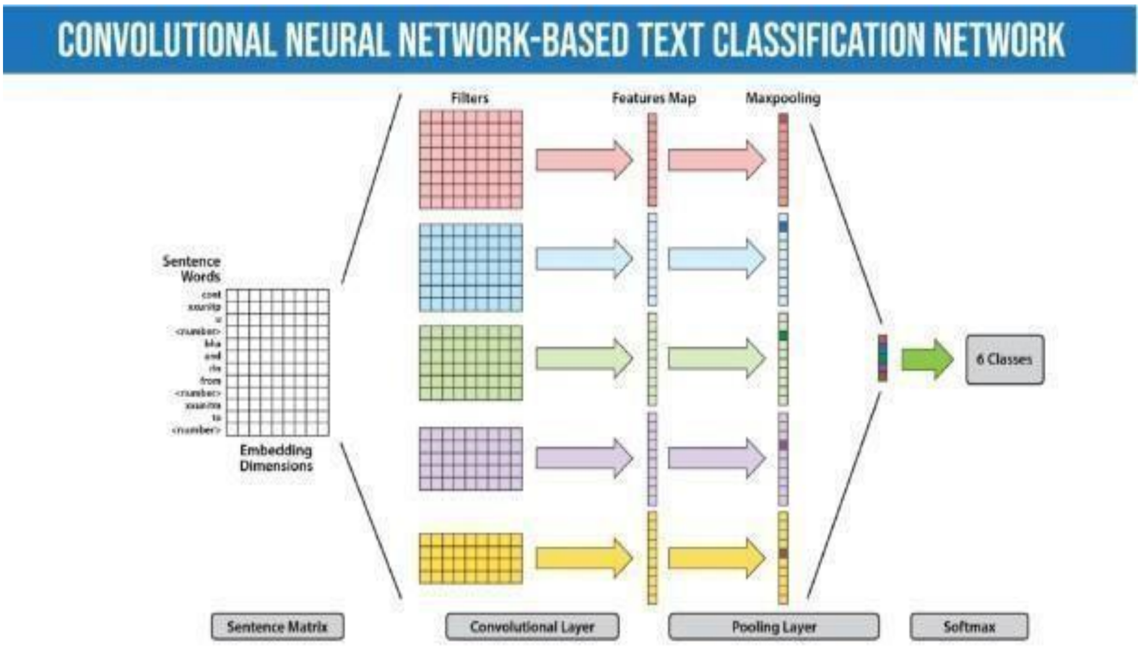
Hidden Markov models:Markov models are probabilistic models that decide the next state of a system based on the current state. For example, in NLP, we might suggest the next word based on the previous word. We can model this as a Markov model where we might find the transition probabilities of going from word1 to word2, that is, P(word1|word2). Then we can use a product of these transition probabilities to find the probability of a sentence. The hidden Markov model (HMM) is a probabilistic modeling technique that introduces a hidden state to the Markov model. A hidden state is a property of the data that isn't directly observed. HMMs are used for part-of- speech (POS) tagging where the words of a sentence are the observed states and the POS tags are the hidden states. The HMM adds a concept called emission probability; the probability of an observation given a hidden state. In the prior example, this is the probability of a word, given its POS tag. HMMs assume that this probability can be reversed: Given a sentence, we can calculate the part-of-speech tag from each word based on both how likely a word was to have a certain part- of-speech tag and the probability that a particular part-of-speech tag follows the part-of-speech tag assigned to the previous word. In practice, this is solved using the Viterbi algorithm.



Someone is trying to recall the weather of last week based on what they did last week. Their actions are the observed states, and the different types of weather are the hidden states. Lines between weather states show the transition probabilities, while lines from weather states to actions show the emission probabilities (likelihood that they performed one action given the weather).

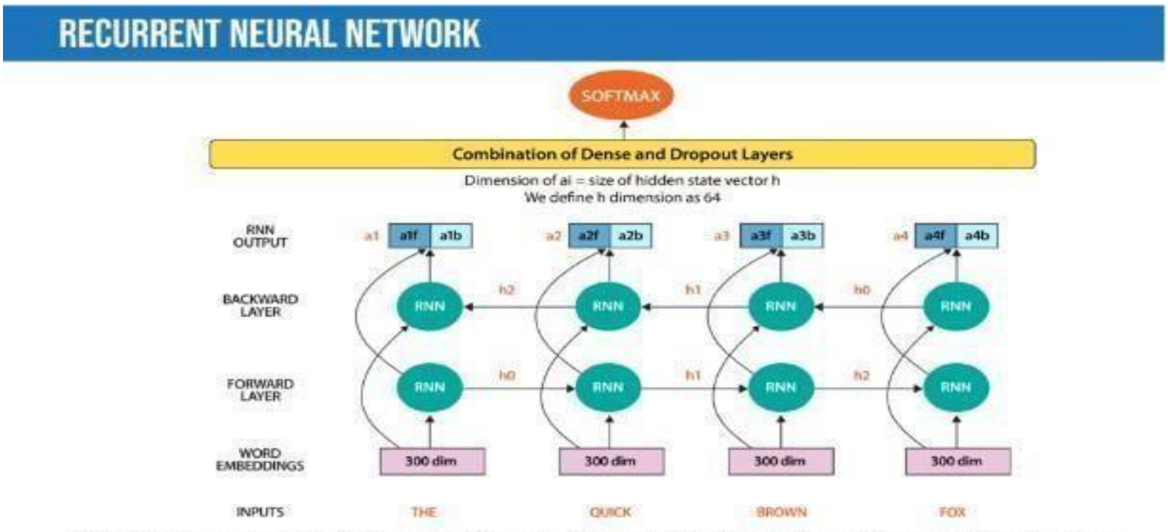
6.6 DEEP LEARNING NLP TECHNIQUES:

Convolutional Neural Network (CNN): The idea of using a CNN to classify text was first presented in the paper "Convolutional Neural Networks for Sentence Classification" by Yoon Kim. The central intuition is to see a document as an image. However, instead of pixels, the input is sentences or documents represented as a matrix of words.



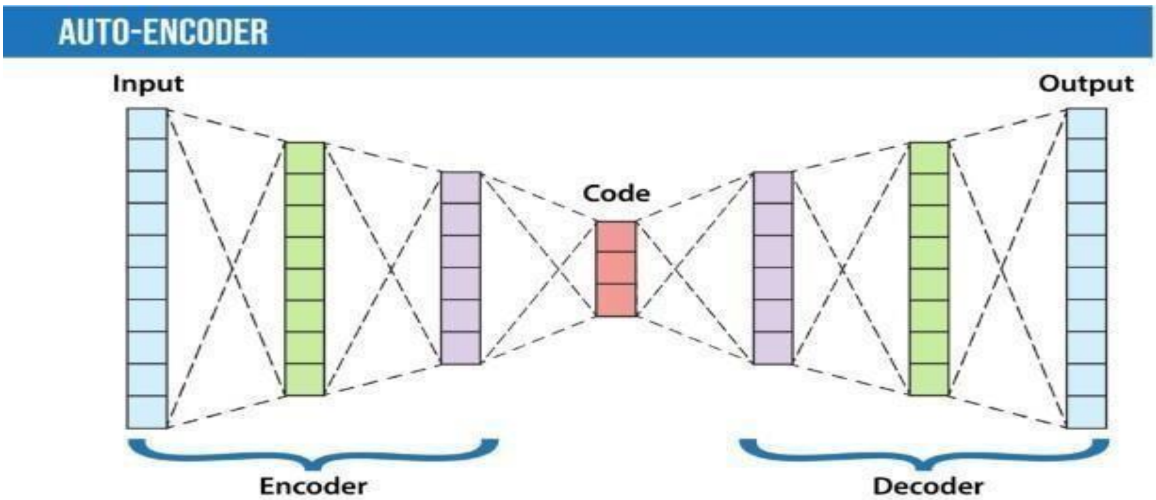
Given a sentence, a convolutional neural network uses convolutional layers to refine representations of input words, before combining them to render a classification.

Recurrent Neural Network (RNN): Many techniques for text classification that use deep learning process words in close proximity using n-grams or a window (CNNs). They can see "New York" as a single instance. However, they can't capture the context provided by a particular text sequence. They don't learn the sequential structure of the data, where every word is dependent on the previous word or a word in the previous sentence. RNNs remember previous information using hidden states and connect it to the current task. The architectures known as Gated Recurrent Unit (GRU) and long short-term memory (LSTM) are types of RNNs designed to remember information for an extended period. Moreover, the bidirectional LSTM/GRU keeps contextual information in both directions, which is helpful in text classification. RNNs have also been used to into words.



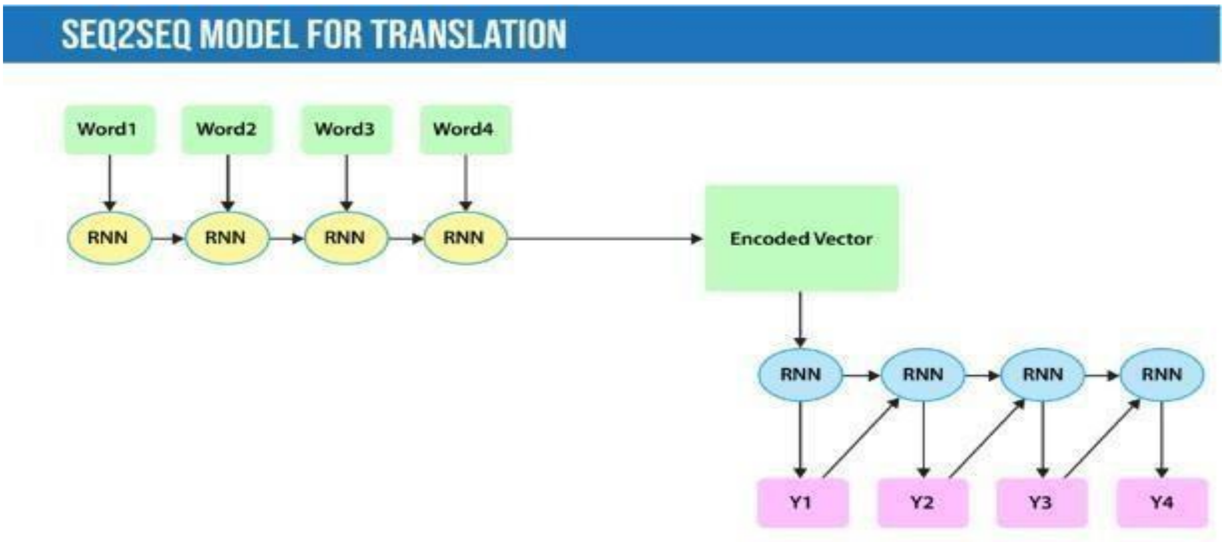
A bidirectional recurrent neural network processes the input both forward and backward to improve the representations it produces.

Autoencoders are deep learning encoder-decoders that approximate a mapping from X to X, i.e., input=output. They first compress the input features into a lower-dimensional representation (sometimes called a latent code, latent vector, or latent representation) and learn to reconstruct the input. The representation vector can be used as input to a separate model, so this technique can be used for dimensionality reduction. Among specialists in many other fields, geneticists have applied autoencoders to in amino acid sequences.



An autoencoder uses an encoder to compress an input into a representation and a decoder to reconstruct the input from the representation.

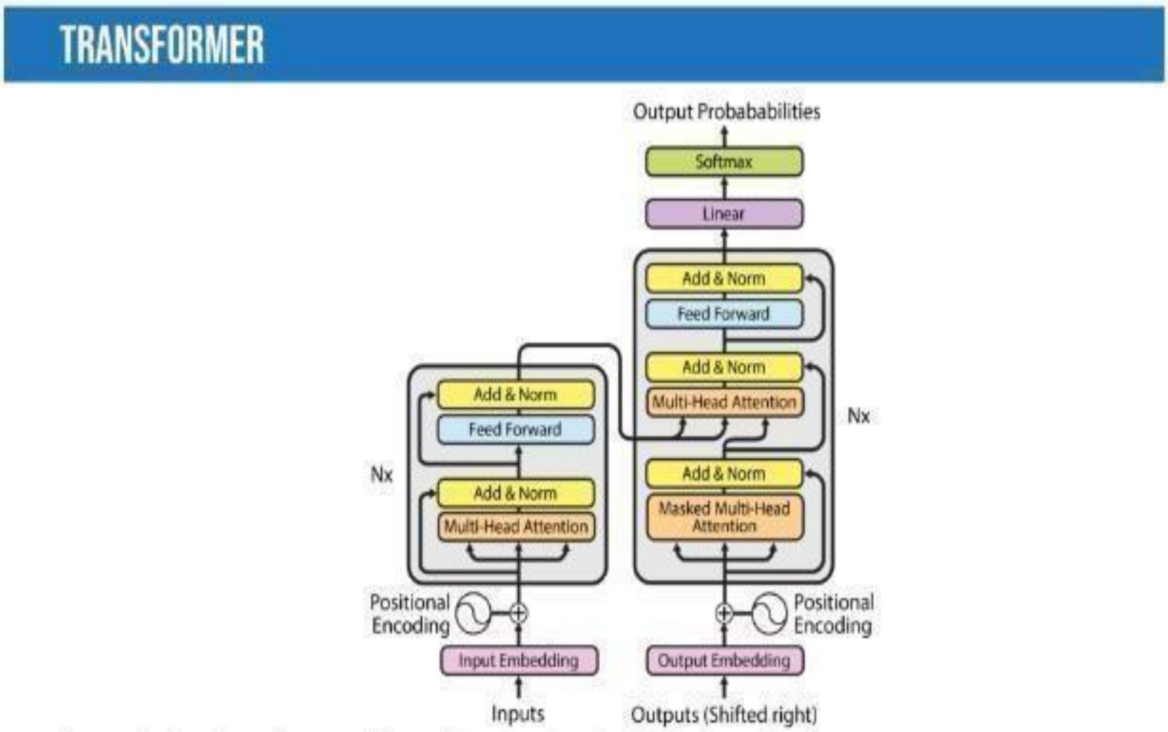
Encoder-decoder sequence-to-sequence:The encoder-decoder seq2seq architecture is an adaptation to autoencoders specialized for translation, summarization, and similar tasks. The encoder encapsulates the information in a text into an encoded vector. Unlike an autoencoder, instead of reconstructing the input from the encoded vector, the decoder's task is to generate a different desired output, like a translation or summary.



Given a sentence, a Recurrent Neural Network encodes the sentence and then iteratively generates a translation.

Transformers: The transformer, a model architecture first described in the 2017 paper "Attention Is All You Need" (Vaswani, Shazeer, Parmar, et al.), forgoes recurrence and instead relies entirely on a self-attention mechanism to draw global dependencies between input and output. Since this mechanism processes all words at once (instead of one at a time) that decreases training speed and inference cost compared to RNNs, especially since it is parallelizable. The transformer architecture has revolutionized NLP in recent years, leading to models including BLOOM, Jurassic-X, and Turing-NLG. It has also been successfully applied to a variety of different vision tasks, including making 3D images.

Transformers: The transformer, a model architecture first described in the 2017 paper "Attention Is All You Need" (Vaswani, Shazeer, Parmar, et al.), forgoes recurrence and instead relies entirely on a self-attention mechanism to draw global dependencies between input and output. Since this mechanism processes all words at once (instead of one at a time) that decreases training speed and inference cost compared to RNNs, especially since it is parallelizable. The transformer architecture has revolutionized NLP in recent years, leading to models including BLOOM, Jurassic-X, and Turing-NLG. It has also been successfully applied to a variety of different vision tasks, including making 3D images.



The encoder-decoder transformer used for translation. Encoder on the left, decoder on the right. Note that the decoder takes in its previously generated words during generation.

6.7 SIX IMPORTANT NATURAL LANGUAGE PROCESSING (NLP) MODELS:

Over the years, many NLP models have made waves within the AI community, and some have even made headlines in the mainstream news. The most famous of these have been chatbots and language models. Here are some of them:

Eliza was developed in the mid-1960s to try to solve the Turing Test; that is, to fool people into thinking they're conversing with another human being rather than a machine. Eliza used pattern matching and a series of rules without encoding the context of the language.

Tay was a chatbot that Microsoft launched in 2016. It was supposed to tweet like a and learn from conversations with real users on Twitter. The bot adopted phrases from users who tweeted sexist and racist comments, and Microsoft deactivated it not long afterward. Tay illustrates some points made by the "Stochastic Parrots" paper, particularly the danger of not debiasing data.

BERT and his Muppet friends: Many deep learning models for NLP are characters, including ELMo, BERT, Big BIRD, ERNIE, Kermit, Grover, RoBERTa, and Rosita. Most of these models are good at providing contextual embeddings and enhanced knowledge representation.

Generative Pre-Trained Transformer 3 (GPT-3)I s a 175 billion parameter model that can write original prose with human-equivalent fluency in response to an input prompt. The model is based on the transformer architecture. The previous version, GPT-2, is open source.to access GPT-3's underlying model from its developer OpenAI, but other users can interact with it via an application programming interface (API). Several groups including have released open source interpretations of GPT-3.

Language Model for Dialogue Applications (LaMDA) is a conversational chatbot developed by Google. LaMDA is a transformer-based model trained on dialogue rather than the usual web text. The system aims to provide sensible and specific responses to conversations. Google developer Blake Lemoine came to believe that LaMDA is sentient. Lemoine had detailed conversations with AI about his rights and personhood. During one of these conversations, the AI changed Lemoine's mind about Isaac Asimov's third law of robotics. Lemoine claimed that LaMDA was sentient, but the idea was disputed by many observers and commentators. Subsequently, Google placed Lemoine on administrative leave for distributing proprietary information and ultimately fired him.

Mixture of Experts (MoE): While most deep learning models use the same set of parameters to process every input, MoE models aim to provide different parameters for different inputs based on efficient routing algorithms to achieve higher performance.Switch Transformer is an example of the MoE approach that aims to reduce communication and computational costs.
Programming Languages, Libraries, And Frameworks For Natural Language Processing (NLP) Many languages and libraries support NLP. Here are a few of the most useful.

Python is the most-used programming language to tackle NLP tasks. Most libraries and frameworks for deep learning are written for Python. Here are a few that practitioners may find helpful:

Natural Language Toolkit (NLTK) is one of the first NLP libraries written in Python. It provides easy-to-use interfaces to corpora and lexical resources such as WordNet. It also provides a suite of text-processing libraries for classification, tagging, stemming, parsing, and semantic reasoning.

spaCy is one of the most versatile open source NLP libraries. It supports more than 66 languages. spaCy also provides pre-trained word vectors and implements many popular models like BERT. spaCy can be used for building production-ready systems for named entity recognition, part-of- speech tagging, dependency parsing, sentence segmentation, text classification, lemmatization,
morphological analysis, entity linking, and so on.

Deep Learning libraries: Popular deep learning libraries include TensorFlow and PyTorch,which make it easier to create models with features like automatic differentiation. These libraries are the most common tools for developing NLP models.

Hugging Face offers open-source implementations and weights of over 135 state-of-the-art models. The repository enables easy customization and training of the models.

Gensim provides vector space modeling and topic modeling algorithms.

R:Many early NLP models were written in R, and R is still widely used by data scientists and statisticians. Libraries in R for NLP include TidyText, Weka, Word2Vec, SpaCyR, TensorFlow, and **PyTorch.** Many other languages including JavaScript, Java, and Julia have libraries that implement NLP methods.

Controversies Surrounding Natural Language Processing (NLP)

NLP has been at the center of a number of controversies. Some are centered directly on the models and their outputs, others on second-order concerns, such as who has access to these systems, and how training them impacts the natural world.

Stochastic parrots: A 2021 titled "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?" by Emily Bender, Timnit Gebru, Angelina McMillan-Major, and Margaret Mitchell examines how language models may biases found in their training data. The authors point out that huge, uncurated datasets scraped from the web are bound to include social biases and other undesirable information, and models that are trained on them will absorb these flaws. They advocate greater care in curating and documenting datasets, evaluating a model's potential impact prior to development, and encouraging research in directions other than designing ever-larger architectures to ingest ever-larger datasets.

Coherence versus sentience:Recently, a Google engineer tasked with evaluating the LaMDA language model was so impressed by the quality of its chat output that sentient. The fallacy of attributing human-like intelligence to AI dates back to some of the earliest
NLP experiments.

Environmental impact: Large language models require a lot of energy during both training and inference. One study estimated that training a single large language model can as much carbon dioxide as a single automobile over its operational lifespan. Another study found that models than training. As for solutions,researchers have proposed as one way to offset this impact.located in countries with lots of renewable energy

High cost leaves out non-corporate researchers: The computational requirements needed to train or deploy large language models are too expensive for many small companies. Some experts worry that this could block many capable engineers from contributing to innovation in AI.

Black box: When a deep learning model renders an output, it's difficult or impossible to know why it generated that particular result. While traditional models like enable engineers to examine the impact on the output of individual features, neural network methods in natural language processing are essentially black boxes. Such systems are said to be "not explainable," since we can't explain how they arrived at their output. An effective approach to achieve explainability is especially important in areas like banking, where regulators want to confirm that a natural language processing system doesn't discriminate against some groups of people, and law enforcement, where models trained on historical data may perpetuate historical biases against certain groups.

CHAPTER 7
SYSTEM OVERVIEW

The block diagram of Home Automation system based on speech recognition is shown in figure

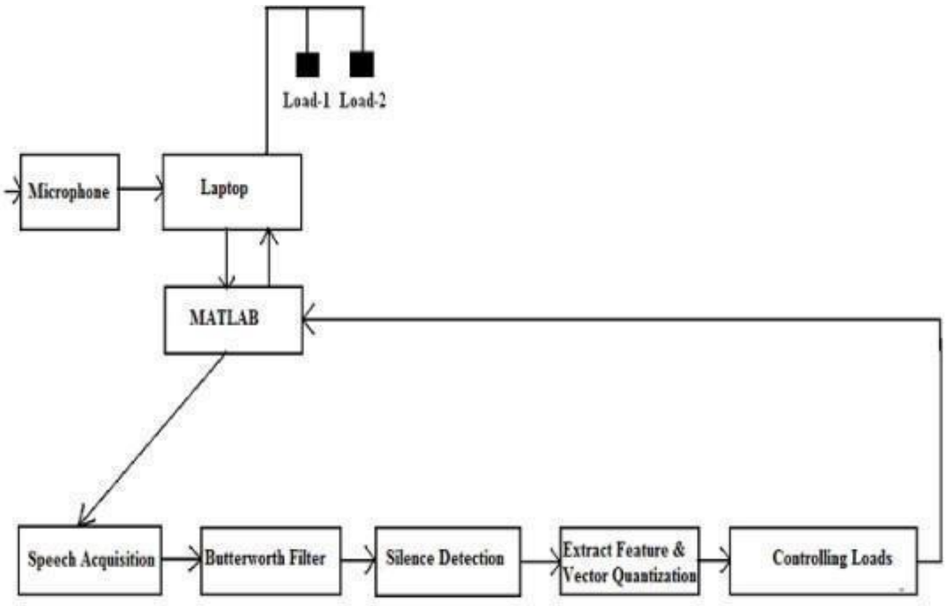


Fig.1: Block Diagram of Speech Operated System for Home Appliances

Speaker Identification based Home Automation system using speech recognition is a low-cost, reliable, efficient and secure method for Home Automation System. This report is divided into two main parts which are Voice Training Process and Voice Testing Process. In voice training process the first step is acquisition of speech. Robust training in which severalversions of the sound pattern are used to create a single mergedtemplate or statistical model. In voice testing process the userhas uttered two different words each process. The system usedis speaker dependent method that means user has to record his/her voice before using the system. Various steps involved in Speaker Identification based Home Automation system using speech recognition is shown in figure 1.

7.1 Voice Training Process:

In voice training process the first step is acquisition of speech. Built in Microphone in laptop is utilized for Speech Acquisition, and then speech acquisition device is installed bysimply Connecting the Microphone with laptop via sound cardinput port. In second step a function is created, which will record speech in MATLAB. In third step recorded speech is played on laptop based audio output device. Fourth step is to write acquired speech in MATLAB and .Wav file is created. Infifth step .wav file is loaded in MATLAB, in order to read the saved speech and in sixth step saved speech is acquired. In seventh step it is filtered out through the Butterworth band passfilter. Butterworth filter is used because it is the best compromise between attenuation and phase response. It has gotno ripple in the pass band or the stop band. After that it is savedin the computer memory so that it can be matched with incoming utterance of speech. In this research work user has uttered two training voices to control the load. These uttered words are “CLOSE” and “YES”. Now all above steps are applied to these uttered words. Silence detection or Voice Activity Detection (VAD) is used in speech processing, whichis used to detect presence or absence of human speech. VAD is used here to deactivate some processes when there is a silenceor non-speech section in audio session. Short time Fourier transforms is performed successfully so that for each incomingspeech, the part of containing high frequency component is extracted. Actually here in MATLAB coding 2500 samples perword are created for feature extraction.

7.2 Voice Testing Process:

In voice testing process the user has uttered two different wordseach process. One word is same as which was trained intraining phase was “CLOSE” and other one is “OPEN”. Then both uttered signals are further processed and analysed by applying same steps which are already used in Voice TrainingProcess. Like voice training process, 2500 samples per word are also created here for feature extraction. These testing signals are used to match with trained signals to authenticate the desired speech. There are various feature matching techniques used in MATLAB, from which Vector Quantizationmethod is used in this research paper. Vector Quantization is a process of mapping vectors from a big vector space to a finite number of regions in that space. In the testing phase, a speakerspecific Vector Quantization codebook is generated for each known speaker by clustering his/her testing acoustic vectors.

7.3 Speech Recognition System:

A speech recognition roughly consists of two portions. They are speech analysis and pattern recognition.

7.4 Speech Analysis:

The purpose of the speech analysis block is to transform the speech waveform into a parsimonious representation which characterizes the time varying properties of the speech. The speech analysis typically includes two modules, namely data acquisition and feature extraction. The data acquisition moduleusually contains a microphone and a code from which digitizedspeech data are generated. The feature extraction is the computation of a sequence of feature vectors which provides acompact representation of the given speech signal. The featureextraction is done on short-time basis. The speech signal is separated into overlapped fixed-length frames. From each frame, a set of frequency-domain or cepstral-domain parameters are derived from each frame, to form the so-called feature vector. There are some basic principles and analysis techniques used in the feature extraction module. They are pre- emphasis, frame blocking and windowing, Discrete Fourier Transform (DFT) computation, spectral magnitudes, Mel- frequency filter bank, logarithm of filter energies, Discrete Cosine Transformation (DCT), Cepstral Weighting, anddynamic featuring.

7.5 Pattern Recognition:

The speech signal is first analyses and a feature representationis obtained for comparison with either stored reference

templates or statistical models in the pattern matching block. Adecision scheme determines the word or phonetic class of the unknown speech based on the matching scores with respect tothe stored reference patterns. There are two types of reference patterns. The first type, called a nonparametric reference pattern (or often a template), is a pattern created from one or more spoken tokens of the sound associated with the pattern. The second type, called a statistical reference model, is createdas a statistical characterization of the behavior of a collection of tokens of the sound associated with the pattern. The vector quantization model is used as the statistical model. There are three portions in pattern recognition. They are pattern training,pattern matching and maximum selection.

7.6 Pattern Training:

Pattern training is the method by which representative sound patterns are converted into reference patterns for use by the pattern matching algorithm. There are several ways in which pattern training can be performed, including: Casual training inwhich a single sound pattern is used directly to create either a template or a crude statistical model. Robust training in whichseveral versions of the sound pattern are used to create a singlemerged template or statistical model. Clustering training in which a large number of versions of the sound pattern is used to create one or more templates or a reliable statistical model of the sound pattern.

7.7 Pattern Matching:

Pattern matching refers to the process of assessing the similarity between two speech patterns, one of which represents the unknown speech and one of which represents thereference pattern (derived from the training process) of each element that can be recognized. When the reference pattern is a "typical" utterance template, pattern matching produces a gross similarity (or dissimilarity) score. When the reference pattern consists of a probabilistic model, the process of patternmatching is equivalent to using the statistical knowledge contained in the probabilistic model to assess the likelihood ofthe speech (which led to the model) being realized as the unknown pattern. Pattern matching refers to the process of assessing the similarity between two speech patterns, one of which represents the unknown speech and one of which represents the reference pattern (derived from the trainingprocess) of each element that can be recognized. Firstly, the training set of vectors is used to create the optimal set of codebook vectors for representing the spectral variability observed in the training set. And Then distance is measured between a pair of spectral analysis vectors to able to cluster thetraining set vectors as well as to classify spectral vectors into unique codebook entries. The next step is a centroid computation procedure. Finally, a classification procedure selects the codebook vectors that closet to the input vector anduses the codebook index as the resulting spectral representation. The classification procedure is essentially a quantizer. It accepts speech spectral vectors as input and provides the code index of the code vectors that best matches the input.

7.8 Electronic Control System:

In this diagram, Speech instruction is firstly taken as input to control home appliances and then a microphone is used to record the person speech. Secondly, the speech instruction is caught and transferred the analog signal to digital signal and the recorded speech is sent to the speech based verification/identification system. Thirdly, the digital information of speech instruction is processed and compared by using the MATLAB programming.

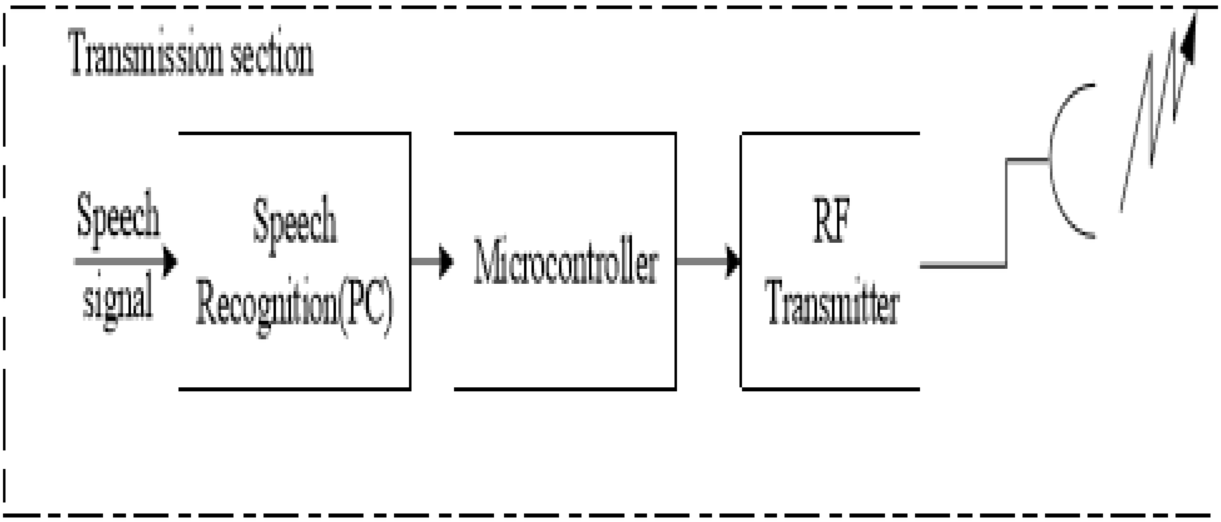


Fig. 3: Block Diagram of Transmission Section of HomeAppliances Control System for Speech Recognition

The speech signal is first analyzed and a feature representationis obtained for comparison with either stored reference templates or statistical models in the pattern matching block. Speech Recognition is a technology allowing the computer to identify and understand words spoken by a person using a microphone. Speech Recognition is a technology allowing thecomputer to identify and understand words spoken by a personusing a microphone. Then signal goes to microcontroller unit then the signal is transmitted.

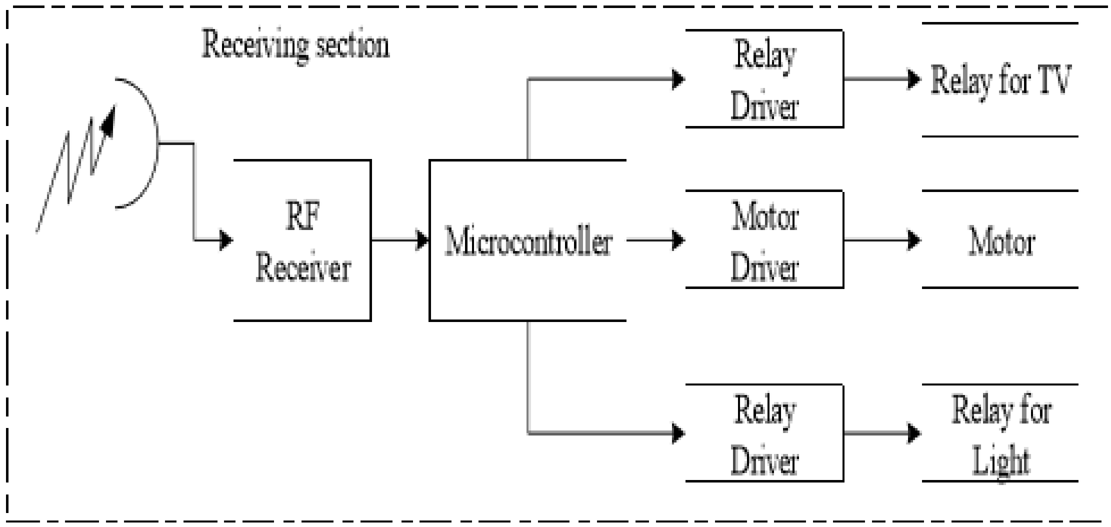


Fig. 4: Block Diagram of Receiving Section Home Appliances Control System for Speech Recognition

In the receiver section receiver accept radio signal and then microcontroller read the signal and then send to drives relay and motor driver.Pulse width modulation (PWM) is a method for binary signals generation, which has two signal periods (high and low). The width (W) of each pulse varies between 0and the period (T). The main principle is control of power by varying the duty cycle. Here the conduction time to the load iscontrolled. The duty cycle can be varied from 0 to 1 by varyingt_{on} or T. Therefore, the average output voltage V_{avr}can be changed between 0 and V_{in} by controlling the duty cycle, thus, the power flow can be controlled. The on-off switching is performed by power MOSFETs.

7.9 Control:

The transmitted control characters are received by the microcontroller and compared with some predefinedcharacters. If there is a match, the microcontroller will switch the corresponding relay and turn on/off the appliance connected to it. On the control side, the microcontroller has tobe programmed to be able to receive control characters from the receiver and activate/control the required relays accordingly.

7.10 Advantages:

7.10.1 Speech is a very natural way to interact & it is not necessary to sit at a keyboard or work with aremote control.

7.10.2 No training required for users.

7.10.3 Beneficial for aging population.

7.11 Disadvantages:

7.11.1 Even the best speech recognition system must make errors.If there is noise of some other soundin the room (e.g. Television), the no. of errors will increase.

7.11.2 Speech recognition work best if the microphone is close tothe user (e.g.in a phone or if the useris wearing a microphone).More distance microphones (e.g. on a table or wall) will tend to increase no. of errors.

7.12 Applications:

7.12.1 Speech to text processing (word processors or emails)

7.12.2 Optimizing use of low cost electricity.

7.12.3 Can be used in all electrical appliances.

CHAPTER 8

8.1 CIRCUIT DETAILS:

Transmission Section: In the transmission section, there are KS232 module, PIC 16F887 and KST-TX01 (Radio Frequencytransmitter module). The KS232 module is used to carry the signal from PC to Microcontroller unit. The signal is retransmitted with baud rate 1200 for RF transmission by KST-TX01 module. This module has four pins: supply pin, data pin, GRN pin, and ANT pin. KST-TX01 technical specific data forwireless transmitter module are:

(1) Transmit power: 1W,

(2) Operating frequency: 315MHZ~433.92MHZ,

(3) Operating temperature:-40℃~80℃,

(4) Operating voltage: 3V~5V

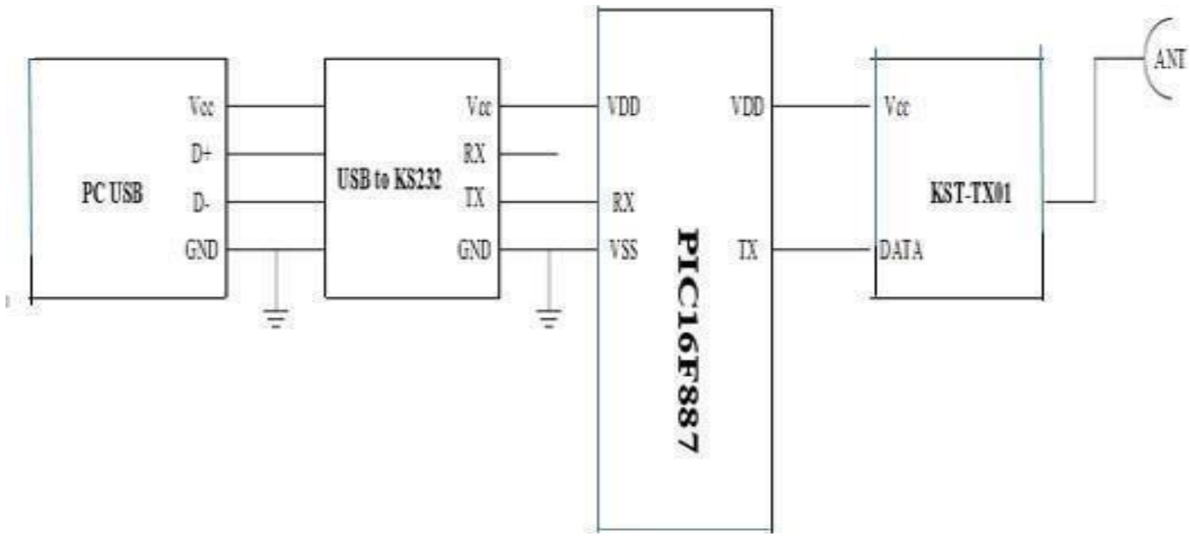


Fig. 5: Circuit Diagram of Transmission Section

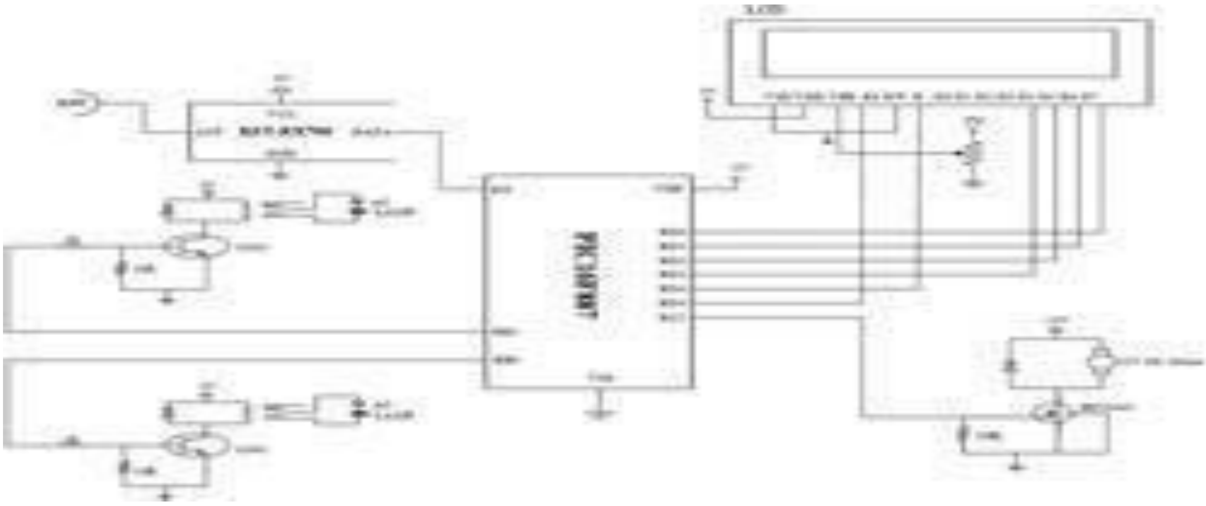


Fig. 6: Circuit Diagram of Receiving Section

Receiving sections**:** The receiver section consists of KST- RX706 (RF receiver module), PIC microcontroller, relays,relay drivers and motor driver. In this section, KST-RX706 firstly accept radio signal and then microcontroller read radio signal with baud rate 1200. Microcontroller drives relay and motor driver. The speed of motor is controlled by using Pulse Wide Modulation (PWM) module. KST-RX706 firstly accept radio signal and then microcontroller read radio signal with baud rate 1200. Microcontroller drives relay and motor driver.The speed of motor is controlled by using Pulse Wide Modulation.

8.2 SOFTWARE DETAILS

MATLAB Millions of engineers and scientist worldwide useMATLAB® to analyse and design the systems and products transforming our world. MATLAB is in automobile active safety systems, interplanetary spacecraft, health monitoring devices, smart power grids, and LTE cellular networks. It is used for machine learning, signal processing, image processing, computer vision, communications, computational finance, control design, robotics, and much more. MATLAB isthe easiest and most productive software for engineers and scientists. MATLAB window, whether you're analysing data, developing algorithms, or creating models, MATLAB provides an environment that invites exploration and discovery. It combines a high-level language with a desktop environment tuned for iterative engineering and scientific workflows. The desktop environment invites experimentation, exploration, and discovery. These MATLAB tools and capabilities are all rigorously tested and designed to work together.

8.3 Key Features of MATLAB:

High-level language for scientific and engineering computing.

- Desktop environment tuned for iterative exploration,design, and problem solving.
- Graphics for visualizing data and tools for creatingcustom plots.
- Apps for curve fitting, data classification, signalanalysis, and many other domain-specific tasks.
- Add-on toolboxes for a wide range of engineering andscientific applications.
- Tools for building applications with custom userinterfaces.
- Interfaces to C/C++, Java®, .NET, Python®, SQL,Hadoop®, and Microsoft® Excel®.
- Royalty-freedeployment options for sharingMATLAB programs with end users.

8.4 SOFTWARE IMPLEMENTATION

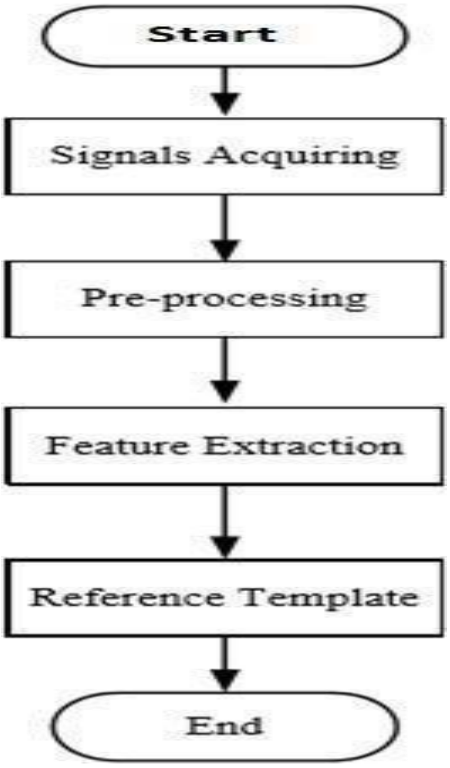


Fig.7: Flow Chart of Training phase

The software implementation part of voice recognition based home automation system implemented using the Arduino controller. It consists of training of voice recognition module. The voice recognition module needs to be trained first with thevoice commands before it can be put to recognizing function. This section explains the methods used for speech recognition.These methods are training phase and testing phase.

Initially, the user must prepare the training files. Figure.4.4 shows the flow chart of the step of training phase. In that signalacquiring, pre-processing, features extraction, reference template, these steps are include. The speech files are recorded from the microphone and MFCC features are extracted from the input file. Then these features are stored. In this case, the collection of training files is called database. Then, the user must train the system using the files in the database. This is called training phase or pre-processing.

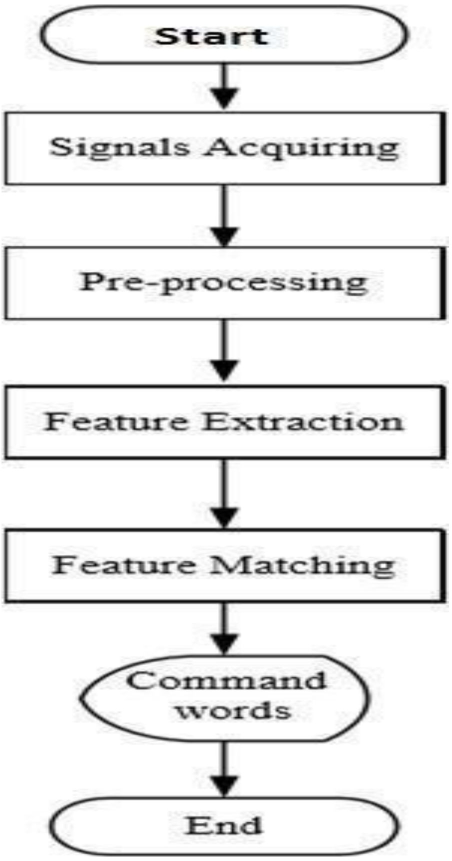
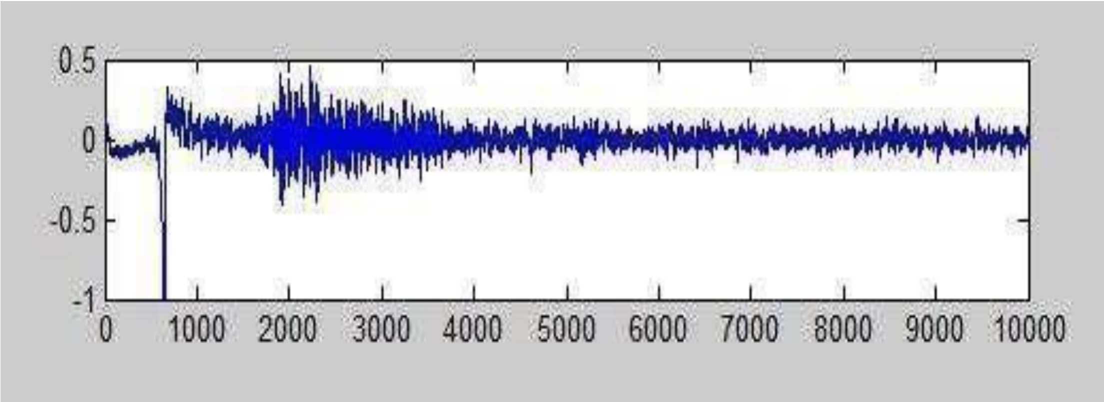


Fig. 8: Flow Chart of Testing Phase

In the testing phase, users have to provide the command wordsas input. In this case, user may use two ways of testing. If userchooses to use the pre-recorded sound file, one of the samplesare loaded from test files and read. Then, the modified MFCC features are extracted from the input file. In the next step, the distances between the modified MFCC features and the storedreference models are calculated using Euclidean Distance. Finally, the minimum distance is selected among the distances between the input vectors and codebook vectors. If this minimum distance falls below the local threshold, the system outputs the command word as result. Otherwise, the system determines it is wrong command word. If the user wants to testthe system with spoken commands in real time, the sound file to be recognized is recorded from the microphone. To do so, the user must choose time length typical time length is 2 seconds. In this system, sound files are recorded within this time length. Then the subsequence processes, as above, are carried out and recognition decision.

8.5 RESULT:



Time vs. Amplitude

Fig. 9: Train Signal Uttered as "CLOSE"

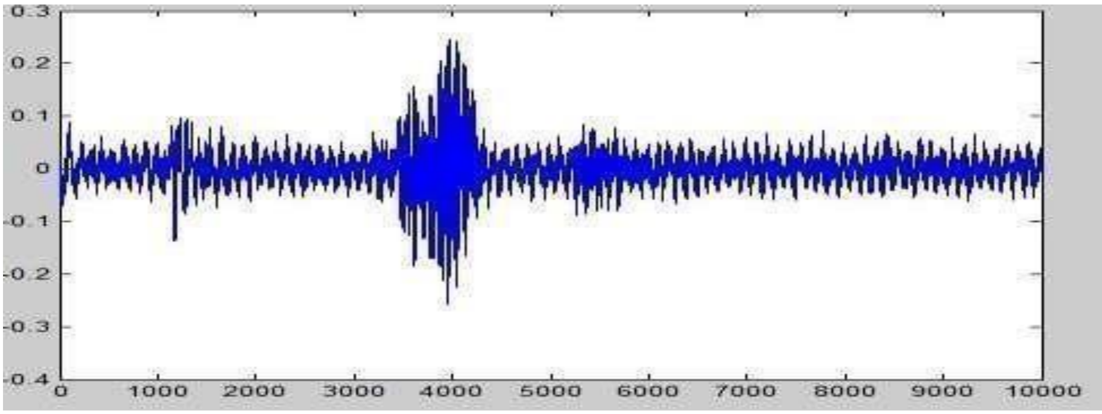


Fig. 10: Train Signal Uttered as "OPEN"

CHAPTER 9

SOURCE CODE

```
import speech_recognition as sr
import requests  # To send HTTP requests to your home automation system

# Initialize the recognizer
r = sr.Recognizer()

def listen():
    with sr.Microphone() as source:
        print("Listening...")
        audio = r.listen(source)
    try:
        text = r.recognize_google(audio)
        print("You said: " + text)
        process_command(text)
    except sr.UnknownValueError:
        print("Google Speech Recognition could not understand audio")
    except sr.RequestError as e:
        print("Could not request results from Google Speech Recognition service; {0}".format(e))

def process_command(command):
    # Here you would typically add NLP processing to understand the command
    # For simplicity, we are directly parsing with simple if-else
    if 'turn on the light' in command.lower():
        # This URL would be the endpoint for your home automation light control
        requests.get('http://your-home-automation-system/api/lights/on')
    elif 'turn off the light' in command.lower():
        requests.get('http://your-home-automation-system/api/lights/off')

if__name___== "_main_":
    while True:
        listen()
```

CHAPTER 10

CONCLUSION

The designed Speech operated system is a low-cost, reliable, efficient and secure. The designed Speech Operated system canbe used in various areas of application. Speech operated systemcan also be used to answer computers in a hands-free environment, like when driving. Speech operated system can be used in tasks that require human-machine interface, for example automatic call processing in the telephone network and data query information systems. The system has two main parts: speech recognition and smart home appliances electronic control system. Speech recognition is implemented in MATLAB environment. An application for speech command processing is developed.

REFERENCES

[1] **B.H. Juang& Lawrence R.Rabiner**, Automatic Speech Recognition – *"A Brief History of the Technology Development".*

[2] **Coucopoulos, Andre;** 2007*, "Voice processing for home automation systems"*, Network system design line, January 2007.

[3] **Gnanasekar. A.K, Jayavelu.P, Nagarajan.V**, *"Speech Recognition Based Wireless Automation of Home Loads with Fault Identification",* IEEE International conference on communications and signal processing (ICCSP), Vol. 3, pp.128-132, 2012.

[4] **Jie Liu, Jigui Sun, Shengsheng Wang,** *"Pattern Recognition: An overview",* IJCSNS International Journal of Computer Science and Network Security, vol. 6, no. 6, June 2006.

[5] **LA Roberts, HT Nguyen, EM Silver**, ‗*"Gesture Activated Home Appliance* '- US Patent 6,937,742, 2005 - Google Patents

[6] **M.A.Anusuya,** *"Speech Recognition by Machine,"* International Journal of Computer Science and Information security, Vol.6, No.3, 2009.

[7] **Mishr and Suyash Agrawal**, *"Recognition Of Voice Using Mel Cepstral Coefficient & Vector Quantization"* International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 2,Mar-Apr 2012, pp.933-938

[8] **Prabhakar V**. Mhadse1, Amol C.Wani2*"Speaker Identification Based Automation Systemthrough Speech Recognition''*

[9] **SeemaAsht and RajeshwarDass**,*"Pattern Recognition Techniques: A Review",* International Journal of Computer Science and Telecommunications, vol. 3, issue 8, August 2012.

[10] **S.J.Arora and R.Singh**, *"Automatic Speech Recognition"* A Review, "International Journal of Computer Applications, vol60-No.9, December 2012

[11] **Tychtl and Josef Psutka,** *"Speech Production Based on the Mel- Frequency CepstralCoefficients"* Mulgrew,

[12] **Subhajit Dey**, "Web based real-time home automation and security system", International Journal of Electrical and Electronic Engineering & Telecommunications, Volume 4, No. 3, July 2015.

[13] **Nathan David, Abafor Chima, Aronu Ugochukwu**, "Design of a Home Automation System Using Arduino", International Journal of Scientific & Engineering Research, Volume 6, Issue 6, June-2015.

[14] **Mukesh Kumar, Shimi S.L**," Voice Recognition Based Home Automation System for Paralyzed People", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 10, October 2015.

[15] **Akbar Satria, Muhammad Luthfi Priadi, Lili Ayu Wulandhari, Widodo Budiharto**, "The Framework of Home Remote Automation System Based on Smartphone", International Journal of Smart Home Vol. 9, No. 1 (2015).

[16] **Sonali Sen, Shamik Chakrabarty, Raghav Toshniwal, Ankita Bhaumik**, "Design of an Intelligent Voice Controlled Home Automation System", International Journal of Computer Applications (0975 – 8887) Volume 121 – No.15, July 2015.

[17] **Mahesh N. Jivani**, "GSM Based Home Automation System Using App-Inventor for Android Mobile Phone", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering, Vol. 3, Issue 9, September 2014.

[18] **Erik Brynjolfsson and Andrew McAfee**, "The Second Machine Age: Work, Progress and Prosperity in a Time of Brilliant Technologies", 2014.

[19] **Rozita Teymourzadeh, Salah Addin Ahmed, Kok Wai Chan, and Mok Vee Hoong**, "Smart GSM Based Home Automation System", 2013 IEEE Conference on System

[20] **Errobidart et al.** proposed an offline voice-assisted home automation system using the EasyVR speech recognition module and an Arduino Mega without NLP, leading to therequirement of pre-defined commands for each task.

[22] **G. et al.** utilized the Hidden Markov Model toolkit (HTK) for converting speech into text, aiming at improving home automation communication via SMS, though it lacked NLP integration.

[23] **Elsokah et al.**introduced a system based on voice recognition that employs an Easy VR 2.0 shield alongside

[24] **Rani et al.**developed a system that uses NLP processed through a mobile phone for voice .