A

PROJECT REPORT

ON

"Business Meeting Summarization Using Natural Language Processing (NLP)"

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS OF DEGREE OF

BACHELOR OF ENGINEERING
BY

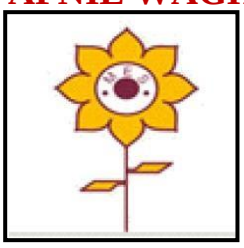
RAJ KSHIRSAGAR

SUYOG MALKAR

CHAITANYA PATHAK

SUPERVISOR

PROF. SWAPNIL WAGHMARE



Department of Computer Engineering
Pillai HOC College of Engineering and Technology, Rasayani,
Pillai's HOC Educational Campus, HOCL Colony,
Rasayani, Tal : Khalapur, Dist- Raigad – 410207
UNIVERSITY OF MUMBAI

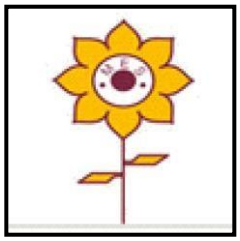
[2020-2021]

MAHATMA EDUCATION SOCIETY'S

PILLAI HOC COLLEGE OF ENGINEERING AND TECHNOLOGY,

RASAYANI-410207

2020-2021



CERTIFICATE

This is to certify that the project entitled **" Business Meeting Summarization Using Natural Language Processing (NLP)"** is a bonafide work of **" Raj kshirsagar, Suyog Malkar and Chaitanya Pathak**" submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Undergraduate"** in **"Computer Engineering".**

Prof. Swapnil Waghmare

(Guide)

Dr. Ashok Kanthe

(Head of Department)

Dr. Mathew Joseph

(Principal)

PROJECT REPORT APPROVAL FOR B.E.

This project report entitled **"Business Meeting Summarization Using Natural Language Processing (NLP)"** by **"Raj kshirsagar, Suyog Malkar and Chaitanya Pathak"** is approved for the degree of Bachelor of Engineering in Computer Engineering.

Examiners

1. _____
2. _____

Date :

Place :

Declaration

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the original sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

_____

Raj kshirsagar

_____

Suyog Malkar

_____

Chaitanya Pathak

Date :

ABSTRACT

Text summarization aims to condense a source text into a shorter version. Automatic data summarization is part of data mining. In order to build a corpus for this task, it is necessary to obtain the transcription of each meeting, and then to segment and align it with the corresponding manual report to produce training examples suitable for training. In this work, an MMS method combining the techniques of natural language processing (NLP), speech processing, computer vision and advanced encryption standard (AES) encryption is used to explore the rich information contained in multi-modal data, to improve the quality and security as well the key idea is to bridge and lessen the semantic gaps between multi-modal data. For audio, speech transcriptions are used. Finally, all the multi-modal aspects are considered to generate a textual summary by maximizing the salience, readability, non- redundancy. The summary so generated by text, audio is encrypted using AES encryption method (to make it secure) and sent to members of meeting on mail, from there the user can retrieve it whenever required by providing the decryption key.

LIST OF FIGURES

LIST OF TABLES

TABLE OF CONTENTS

CHAPTER – I

INTRODUCTION

1.1 BACKGROUND

Today's world is centralized on computers and data. Data are our intangible thoughts and imagination. We are producers and consumers of data at the same time. Every little thing in our mundane lives are either a source or receiver of data. For, example when we drive there's data involved, the speed of the car, mileage, distance traveled, etc. Since 20th century, data has been a significant part of our lives, but these days we infer more from data. We store and access them through electronic and wireless systems. Automatic meeting summarization is valuable, especially if it takes advantage of multi-modal sensing of the meeting environment, such as microphones to capture speech each participant's voice. It is the task of writing the report corresponding to a meeting. We focus on so-called exhaustive reports, which capture, all the information of a meeting, keeping chronological order and speakers' interventions. Such reports are typically written by professionals based on their notes and the recording of the meeting. We use abstractive text summarization method for summary generation, In Abstractive Summarization they do not select sentences from the originally given text passage to create the summary. Instead, they produce a paraphrasing of the main contents of the given text, using a vocabulary set different from the original document. We built such a corpus using Speech Recognition to generate transcription, then aligning segments manually and automatically. When NLP processing has been used as a technique to summarize text documents, we see that python libraries such as speech recognition, spacy, and pydub, Tkinter, pyaudio, AES has been used. Text summarization has grown into a crucial and appropriate engine for supporting and illustrate text content in the latest speedy emergent information age. It's far very complex for humans to physically summarize oversized documents of text. We use AES Encryption and decryption algorithm to provide security to generated summary. The Advanced Encryption Standard (AES) is a symmetric block cipher, is implemented in software and hardware throughout the world to encrypt sensitive data. AES is very fast and secure, and it is the de facto standard for symmetric encryption, and the whole framework depends on the linguistic treatment of the original document. The system provides benefited for both abstractive and extractive approaches and also summarizes multiple documents of the same topic.

CHAPTER – II

LITERATURE REVIEW

2.1 LITERATURE REVIEW

SR. NO	AUTHOR NAME	PUBLLICATION DETAILS	DESCRIPTION
1	Rahul, Surabhi Adhikari, Monika	"NLP based Machine Learning Approaches for Text Summarization", International Conference on Computing Methodologies and Communication 2020.	Rahul, Surabhi Adhikari have used a sentence framing method to generate EXT text summaries. In this method, the documents are examined for their applicability and scored accordingly. Similarly, sentences were clustered collectively to find the most meaningful sentences, and Documents were selected based on sentence scores. The summaries extracted using these approaches may not always accurate. In some conditions, it's also inappropriate to the native documents.
2	Soe Soe Lwin, Khin Thandar Nwet	"Extractive Summarization for Myanmar Language", IEEE 2018	Soe Soe Lwin, Khin Thandar Nwet proposed Extractive Summarization for Myanmar Language system, which was the Myanmar text Summarizer developed using latent semantic analysis model. There are various projects developed with text summarization for English and European languages and also in other languages such as Chinese and

			Japanese. But there are very few projects are implemented in Myanmar text summarization. This project compares the sentence selection methods of latent semantic analysis but mainly focused on a single document.
3	Ravali Boorugu, Dr. G. Ramesh	"A Survey on NLP based Text Summarization for Summarizing Product Reviews", International Conference on Inventive Research in Computing Applications, IEEE 2020	The European Monitoring Centre for Drugs and Drug Addiction is the hub of drug-related information in Europe. Its mission is to provide the European Union and its Member States with 'factual, objective, reliable and comparable information' on drugs and drug addiction and their consequences.
4	Madhuri, Ganesh Kumar.	"Extractive Text Summarization Using Sentence Ranking", IEEE 2019	Ravali Boorugu, Dr. G. Ramesh Presented "A Survey on NLP based Text Summarization for Summarizing Product Reviews". This paper recounted the most important techniques of summarizations. Basically, the project compares all the single document summarization, multi-document summarization, and domain-specific summarization. Different hybrid classifiers such as Naive Bayes and SVM are used to develop summaries.
5	Kavya Kishore, Greeshma N Gopal, and Neethu P H	"Document Summarization in Malayalam With	Madhuri, Ganesh Kumar. R proposed a system for text summarization which

		Sentence Framing", IEEE 2016	is implemented n by using a statistical novel method that is based on sentence ranking. The highly-rated sentences are taken out from the original documents to generate a high- quality test summary. The summaries of documents are extracted and converted into audio form. The summaries of documents are extracted and converted into audio form. This method gives more accuracy as compared to a traditional approach.

2.2 EXISTING SYSTEM

Text summarizations were developed with different methods and implemented in different fields. Different text summarization techniques are sentence compression, template and graph-based methods, sentence framing, etc. The summary was generated using extractive summarization method, it is the traditional method developed first. The main objective is to identify the significant sentences of the text and add them to the summary. You need to note that the summary obtained contains exact sentences from the original text. sentences were clustered collectively to find the most meaningful sentences, and Documents were selected based on sentence scores. The summaries extracted using these approaches may not always accurate. Extractive methods work by selecting a subset of existing words, phrases, or sentences in the original text to form the summary. In Previous Text Summarization the sentences were clustered collectively to find the most meaningful sentences, and Documents were selected based on sentence scores. The summaries extracted using these approaches may not always accurate. The summary was sent to meeting participants manually there was not any automated system for it. In the existing system the summary generates with less security and summary were not send to the members of the meeting automatically. They were generated summary using extractive summarization method, which is less effective as compare to abstractive summarization.

2.3 PROBLEM STATEMENT

In the existing system the summary generates with less security and summary were not send to the members of the meeting automatically. They were generated summary using extractive summarization method, which is less effective as compare to abstractive summarization. Extractive summarization means identifying important sections of the text and generating them verbatim producing a subset of the sentences from the original text.

CHAPTER – III

REQUIREMENT GATHERING

3.1 SOFTWARE AND HARDWARE REQUIREMENT

HARDWARE REQUIREMENT

- Processor : - Pentium core
- RAM : - 4 GB and More
- Hard Disk : - 160 GB

SOFTWARE REQUIREMENT

- Operating System : Windows
- Developing Tools : Python

CHAPTER – IV

PLAN OF THE PROJECT

4.1 METHODOLOGY

Natural language processing (NLP) is a branch of artificial intelligence that helps computers understand, interpret and manipulate human language. NLP draws from many disciplines, including computer science and computational linguistics, in its pursuit to fill the gap between human communication and computer understanding. The approach is to identify the important sections, interpret the context and reproduce in a new way. This ensures that the core information is conveyed through shortest text possible. Note that here, the sentences in summary are generated, not just extracted from original text. Some Practical examples of NLP are speech recognition for eg: google voice search, understanding what the content is about or sentiment analysis etc. Python provides developers with an extensive collection of NLP tools and libraries that enable developers to handle a great number of NLP-related tasks such as document classification, topic modeling, part-of-speech (POS) tagging, word vectors, and sentiment analysis. Abstractive summarizers are so-called because they do not select sentences from the originally given text passage to create the summary. Instead, they produce a paraphrasing of the main contents of the given text, using a vocabulary set different from the original document. This is very similar to what we as humans do, to summarize. We create a semantic representation of the document in our brains. For converting speech to text, we used speech recognition module. Speech must be converted from physical sound to an electrical signal with a microphone, and then to digital data with an analog-to-digital converter. Once digitized, several models can be used to transcribe the audio to text. By using Crypto library provide security to the summary.

4.2 IMPLEMENTED SYSTEM ARCHITECTURE

We will be discussing various roles of the system in overview. Text summarization means the technique of making short the long pieces of text. In the System, the meeting conversation will record by using python libraries such as speech recognition and pyaudio.

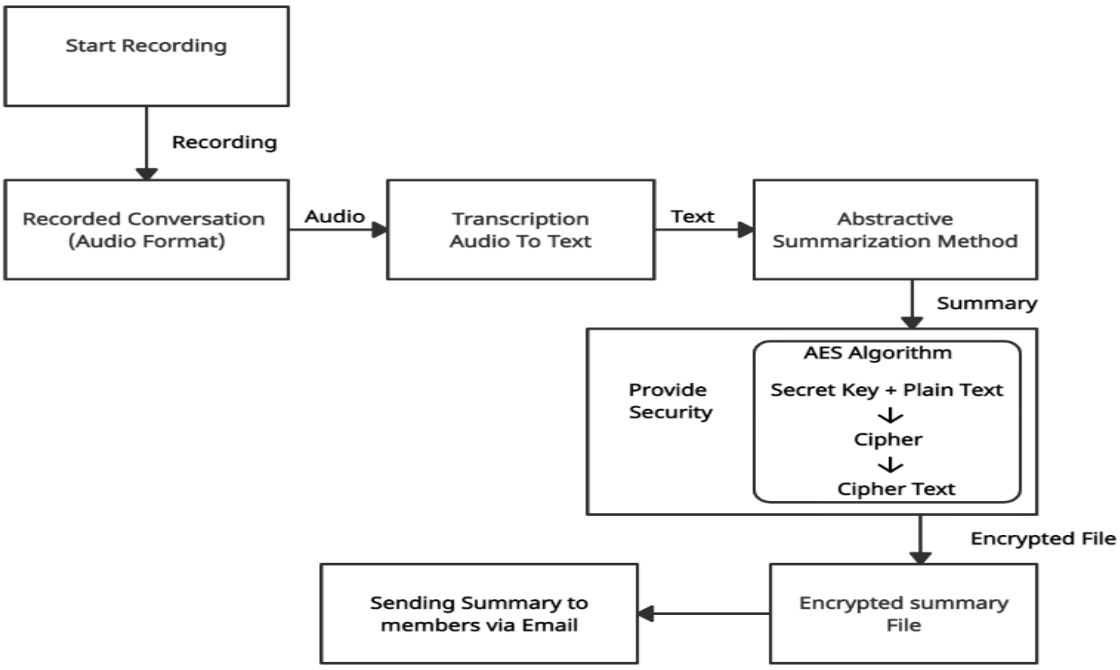


Fig 4.2: - System Architecture

Speech Recognition is a library for performing speech recognition, with support for several engines and APIs, online and offline. If you want to use microphone as input the speech recognition is required. After the recording the audio file is generated. Then we will transcript the audio file (Audio to text conversion) using pyaudio library. The summary will generate from transcription by using abstractive summarization method and for that we use Spacy library. To provide security to the generated summary we used AES algorithm. For encryption and decryption of the summarized file the Crypto library is required. We made one model for decrypt the encrypted file. This model and the encrypted file will send to the meeting members via Email. We have used smtplib library for sending mail. We created GUI of this system using Tkinter library, Tkinter has several strengths. Tkinter is most commonly used. The layered approach used in designing Tkinter gives Tkinter all of the advantages of the TK library. Therefore, at the time of creation, Tkinter inherited from the benefits of a GUI toolkit that had been given time to mature.

After generating the summary, system will take Email id of members of Meeting and send over an email automatically and also create the text file of summary. This system will reduce the human efforts.

4.3 PROJECT PLANNING

Project planning is the process of planning the development of software application completely defining the time and cost of software. The decision to make Business meeting summary generation System is to provide effective summary of meeting to members. The Audio should record and the transcription should create from recorded audio for summary. After generating summary, the security should be provided and send to meeting participants on their mail so they can get summary easily.

CHAPTER V

PROJECT ANALYSIS

5.1 USECASE DAIGRAM

The purpose of use case diagram is to capture the dynamic aspect of a system. However, this definition is too generic to describe the purpose, as other four diagrams (activity, sequence, collaboration, and State chart) also have the same purpose. We will look into some specific purpose, which will distinguish it from other four diagrams. Use case diagrams are used to gather the requirements of a system including internal and external influences. These requirements are mostly design requirements. Hence, when a system is analyzed to gather its functionalities, use cases are prepared and actors are identified.

When the initial task is complete, use case diagrams are modeled to present the outside view.

In brief, the purposes of use case diagrams can be said to be as follows −

- Used to gather the requirements of a system.
- Used to get an outside view of a system.
- Identify the external and internal factors influencing the system.
- Show the interaction among the requirements are actors.

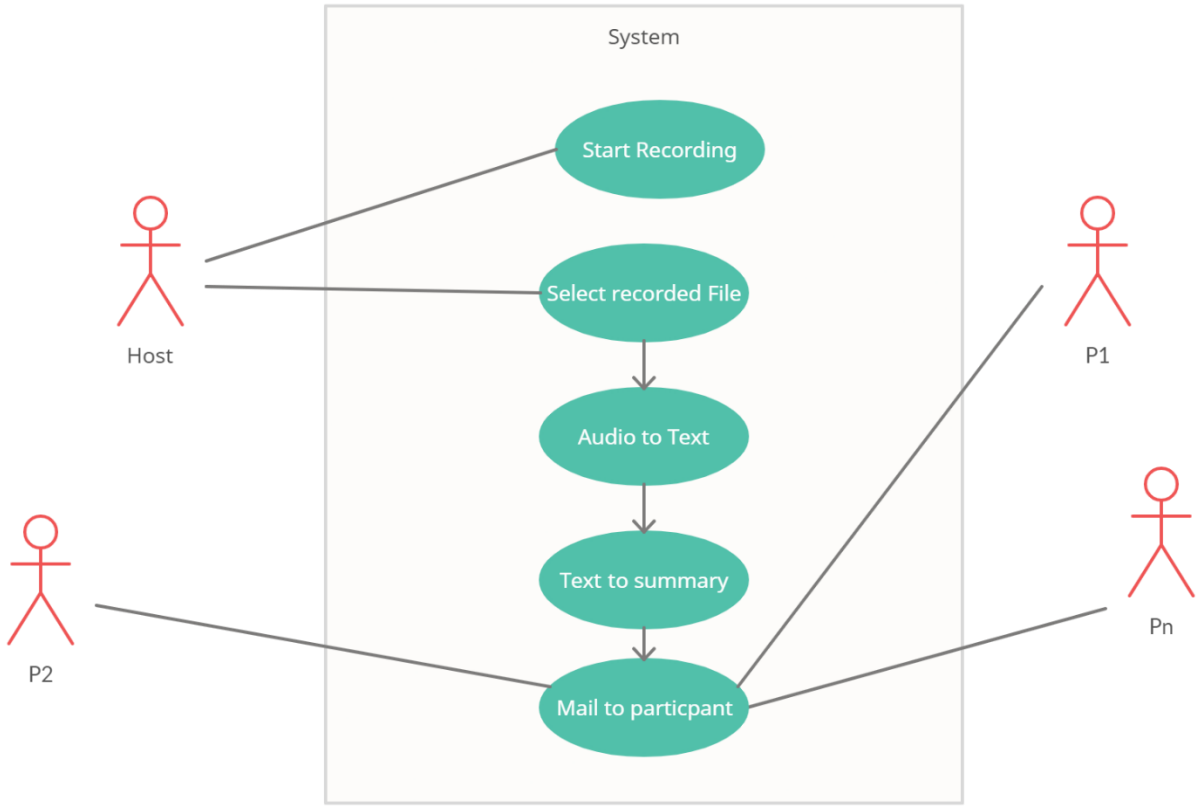


Fig 5.1: - Use Case Diagram

6. RESULT

Fig6.1: - Main Window

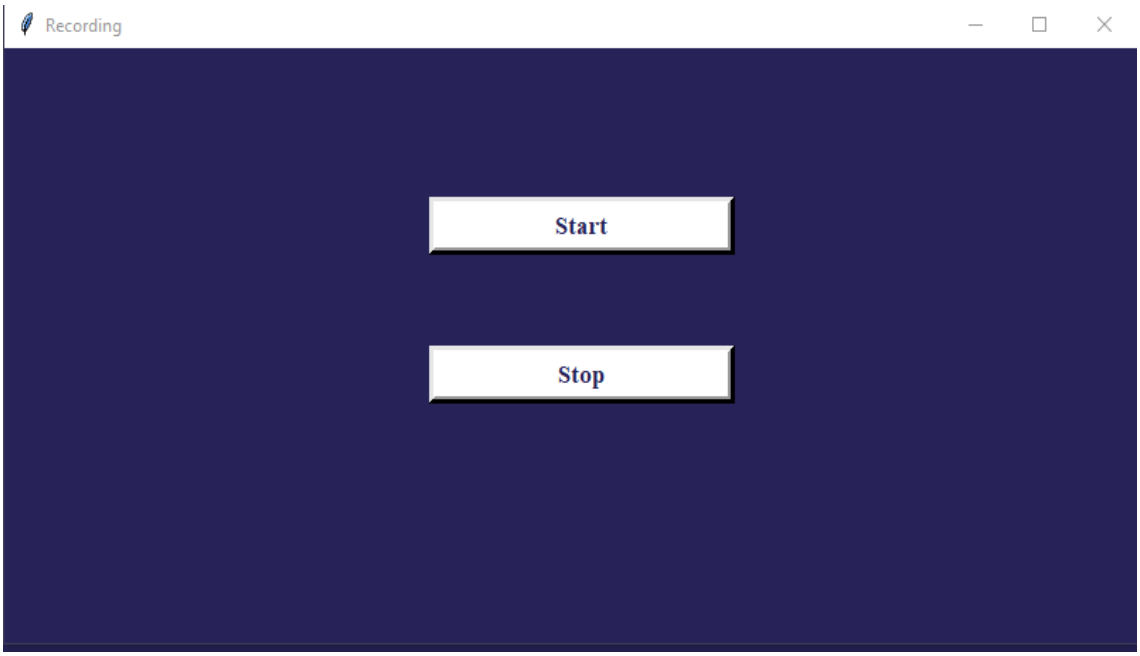


Fig 6.2: -Recording Window

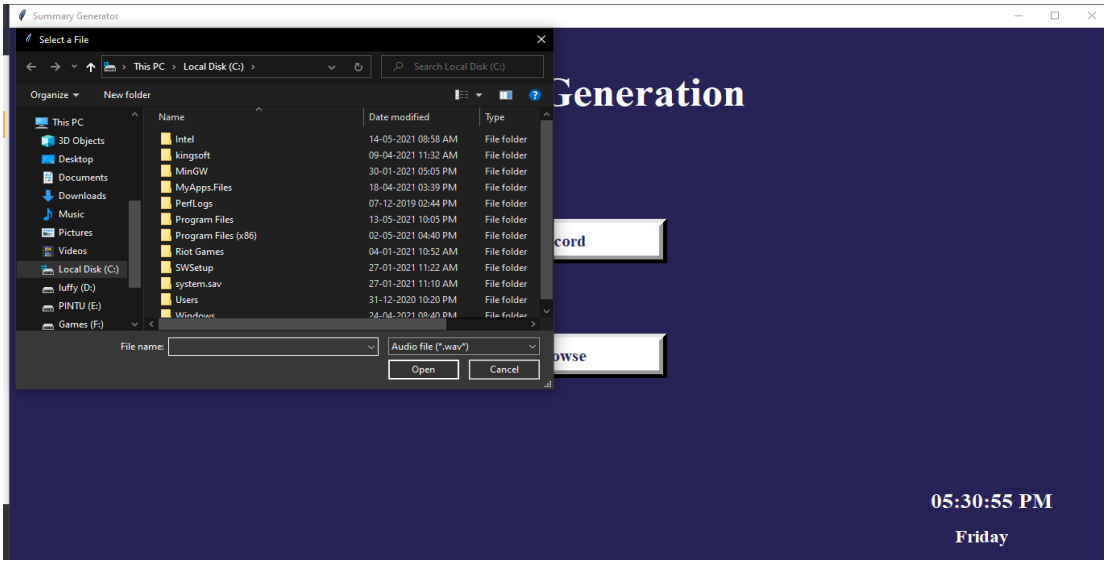


Fig6.3: - Browse File

Fig6.4: - Message Box

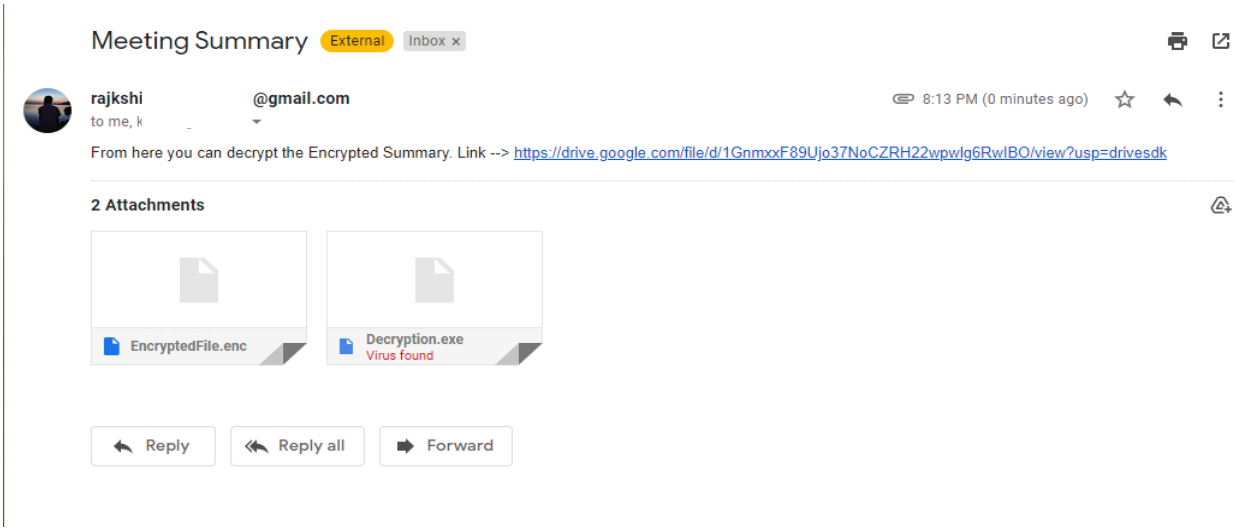


Fig6.5: - Email

Fig:6.6: - Decrypt File

7. RESULT ANALYSIS

- In Business meeting summarization model, record audio and then that recorded audio convert to the text file.
- Text file is converted to the summary.
- Provided security to the generated summary.
- Encrypted summary along with decryption software successfully send on email.

8. CONCLUSION

- The business meeting summary generation system has made easy summary generation.
- It records business meeting, create summary with security and sends the encrypted file to meeting participants on their mail.
- It reduces the human efforts by creating summary automatically.
- This results in a confidential and more comfortable session for the individual.

9. FUTURE SCOPE

- We can improve security.
- We will assign unique id to meeting members so they can access the summary.

REFERENCES

- Rahul, Surabhi Adhikari, Monika" NLP based Machine Learning Approaches for Text Summarization", International Conference on Computing Methodologies and Communication 2020.
- Soe Soe Lwin, Khin Thandar Nwet" Extractive Summarization for Myanmar Language", IEEE 2018
- Ravali Boorugu, Dr. G. Ramesh" A Survey on NLP based Text Summarization for Summarizing Product Reviews", International Conference on Inventive Research in Computing Applications, IEEE 2020
- Madhuri, Ganesh Kumar. R" Extractive Text Summarization Using Sentence Ranking", IEEE 2019
- Kavya Kishore, Greeshma N Gopal, and Neethu P H" Document Summarization in Malayalam With Sentence Framing", IEEE 2016

ACKNOWLEGMENT

It is privilege for us to have been associated with **Prof. Swapnil Waghmare** our guide and mentor during this project work. We have been greatly benefited by her valuable suggestions and ideas. It is with great pleasure that we express our deep sense of gratitude to them for her valuable guidance, constant encourage and patience throughout this work. We are also indebted to **Dr. Ashok Kanthe (Head of Department of Computer Engineering),** for extending the help to academic literature. We express our gratitude to **Dr. Mathew T Joseph (Principal)** for her constant encouragement, co- operation and support. We take this opportunity to thank all our classmates for their company during the course work and for useful discussion we had with them. We would be failing in our duties if we do not make a mention of our family members including our parents for providing moral support, without which this work not have been completed.

Thank you.

Raj Kshirsagar

Suyoag Malkar

Chaitanya Pathak