

**Faculty of Computer Studies  
Information Technology & Computing**

**SummarizeIt**

**(text summarizer with ML&DL)  
  
Submitted to  
Arab Open University  
  
Supervised by  
Dr. /** **Mustafa Abdul-Salam**Presented by **Awad Safwat**

**ID: 1851710330**

**Cairo  
2021/2022**

|  |  |
| --- | --- |
|  | **Declaration of No Plagiarism**  I hereby declare that this submitted report work is a result of my own efforts and I have not plagiarized any other person's work. I have provided all references of information that I have used and quoted in my work.  Name of Student: **Awad Safwat Rady**  Signature: **Awad Safwat**  Student ID: **1851710330**  **Abstract.**  Text Summarization is the task of extracting salient information from the original text document, and it's one of the most important tasks of NLP (Natural Language Processing) . In this process, the extracted information is generated as a condensed report and presented as a concise summary to the user. It is very difficult for humans to understand and interpret the content of the text. In this project we will work on the problem of text Summarization , using ML & DL algorithms. the data we will use to train the model in this project is "Amazon Fine Food reviews " from kaggle.  Acknowledgements  I needed a lot of help and asked a lot of questions to finish this project, and my professor, Dr. Mustafa Abdul-Salam, always responded, even if the time was not appropriate, so he has my gratitude and respect for everything he did for me in terms of assistance and encouragement to complete this project. I'd want to express my gratitude to Dr. Eid Emary for his assistance. and i will never forget the help from the ENG ( "Mohamed Anwar" and "karem Eldepaisy"). and I'd want to thank the AOU for everything it has done to help me get to the last course, which is graduation  project.    Table of content.  ***Contents***  [1. Chapter One: Introduction. 1](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111065)  [1.1 Overview 1](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111066)  [1.2 Motivations of the project 2](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111067)  [1.3 Aims and Objectives 5](file:///F:\\AOU\\summarizer\\report%20Pa\\Project%2520Sample.docx" \l "_Toc40111069)  [1. 4Scope and Constraints of the project 5](file:///F:\\AOU\\summarizer\\report%20Pa\\Project%2520Sample.docx" \l "_Toc40111070)  [1.5Project Plan 9](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111072)  [Project plan. 9](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111073)  [Gantt Chart. 10](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111074)  [2. Chapter Two: Literature Review 11](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111075)  [2.1 Background Information 11](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111076)  [2.2 Similar applications 14](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111077)  [3. Chapter Three: Requirements and Analysis 18](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111078)  [3.1 Functional Requirements 18](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111079)  [3.2 Non-Functional Requirements 18](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111080)  [3.3 Software Tools & Hardware Specifications 20](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111081)  [3.4 Analysis Diagrams 21](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40111082)  ***List of figures:***  [Fig1. 1 Solution process.. 9](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112437)  [Fig1. 2 Project Plan. 10](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112438)  [Fig2. 1 Text Summarizer 14](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112446)  [Fig2. 2 Text Summarizer\_AI 15](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112447)  [Fig2. 3 Sumlt! Text Summarization 16](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112448)  [Fig2. 4 Summarizer 17](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112449)  [Fig3. 1 Use Cases. 18](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112453)  [Fig3. 2 Use case Diagram. 21](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112454)  [Fig3. 3 Sequence Diagram. 23](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112456)  [Fig3. 4 Block Diagram 27](file:///F:\AOU\summarizer\report%20Pa\Project%2520Sample.docx#_Toc40112460) |

# 

# **Chapter One: Introduction.**

## **Overview**

**“SummarizIt”** (Text Summarizer using DL) is a flutter application that does the task of summarizing a text entered by the user by scanning an image and extracting the text from it, or uploading a file , using one of the ML algorithms and then display the summarized text and can pronounce it .

The automatic text summarizing system creates a summary, condensed version of the text that only includes a few key sentences. Text summarization began in the late 1950s, and the discipline has progressed significantly since then. In this field of study, a large number of techniques and approaches have been developed.

An automatic text summarizer should produce a summary that contains the most important information in a document while also taking up less space than the original document. Automatic summary production, on the other hand, is a difficult undertaking. When summarizing a large number of documents, various concerns such as duplication, temporal dimension, sentence sequencing, and so on require special attention, making the process more difficult. (Halifax, 2017)

For example:

The original text is : “Product arrived labeled as Jumbo Salted Peanuts...the peanuts were actually small sized unsalted. Not sure if this was an error or if the vendor intended to represent the product as "Jumbo".”

The summary is : “Not as Advertised”

## **Motivations of the project**

These days we feel that time is passing quickly and there is not enough time to spend our daily need completely, because of the many daily requirements of each of us, so each person should try to create enough time for himself and not waste his time on things that are supposed to take less time than usual .

Such as reading the daily news, or reading an entire article to reach specific information and not needing the rest of the article,

Or that the person is supposed to see important texts and there is not enough time to do so.

All these reasons and more made my project something that has become a constant presence in the life of each of us, and these reasons are considered a great incentive to complete this project.

## **Problems of the project**

Nowadays anyone can be a publisher, and this is with the help of the Internet and social media. It has also become very easy to publish different books and articles daily in large numbers and in different languages, as well as the large number of different sites in very similar specialties and topics, all of this has led to an increase in the volume of data There are very large daily on the Internet, these data are difficult to read all in order to extract some of the information from it, also sometimes we need to analyze this data to know whether it meets the ethical standards or not.

Therefore, we must try to find automatic solutions to this problem.

## **Aims and Objectives**

In this project, we will seek to create an application that works using deep learning algorithms, to summarize texts in a way similar to the way humans summarize texts.

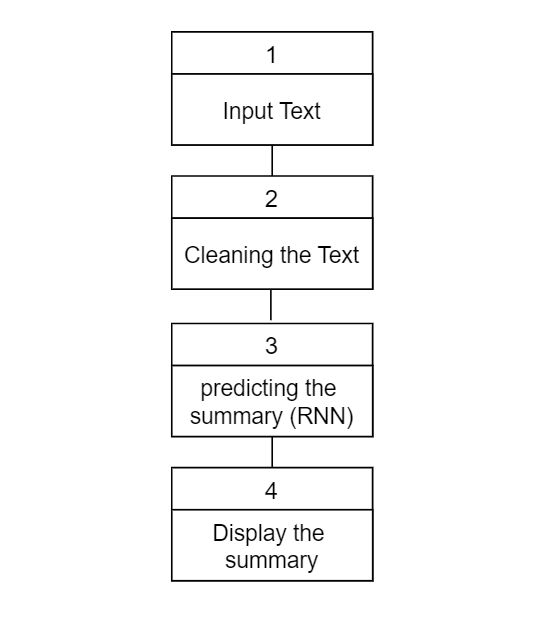
We also strive to make this application work accurately and quickly.

## **Scope and Constraints of the project**

In this project, we target readers in general and academics in particular, and we will seek to develop the algorithm so well that we can rely on the program to summarize scientific and research articles.

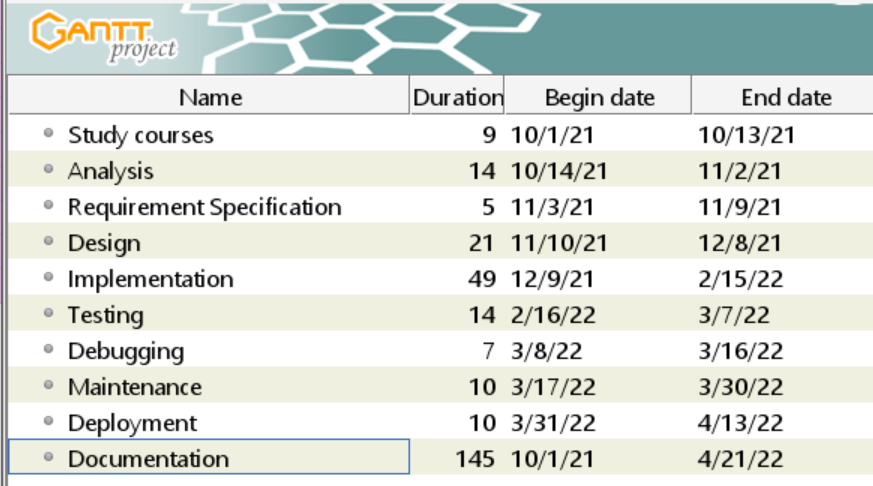
## **Suggested Solution**

The targeted solution to solve this problem, is that the target text is received from the user first, then that text is cleaned to be ready to work on, after that the process of tokenization and transformation into matrices for that text, after which the matrices are inserted into a set of recurrent neural networks ( RNN) that has previously been trained on data similar to this text, in turn predicts the summary text of that text



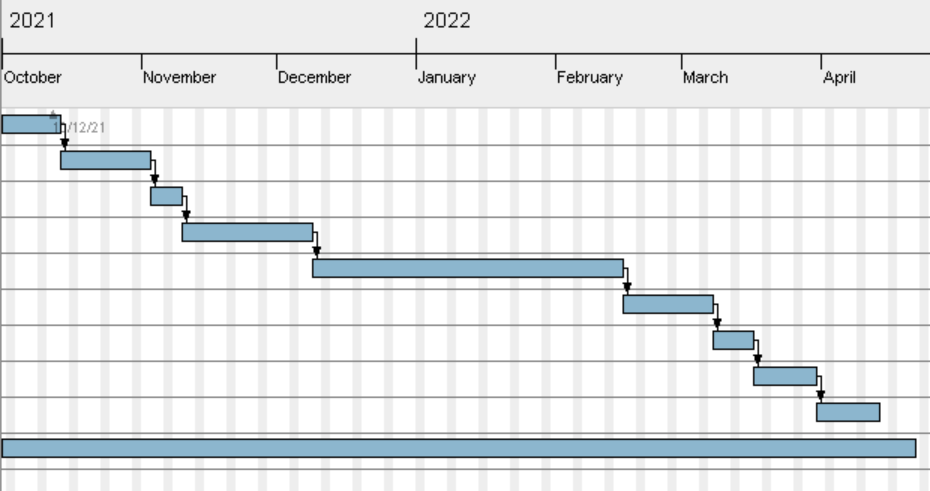
## Fig1. 1 Solution process.

## **Project Plan**



## Fig1. 2 Project plan.

The process sequence of this system is depicted in the diagram above. Starting with the analysis phase, when all components of the project are investigated in order to define the requirements for the design, implementation, and testing stages to be implemented. The user interface and navigation for the application are introduced during the design phase. The test types are applied according to software testing principles in the code for the back-end application and the AI model. All faults and errors detected during the test phase will be fixed after this debug and maintenance phase.



## Fig1. 3 Gantt Chart.

# **Chapter Two: Literature Review**

## **Background Information**

People are getting overwhelmed by data as more and more digitalized text becomes available, especially as the Internet grows. It becomes critical to figure out how to assist individuals in successfully and efficiently extracting information from data. Text summarization is one of several strategies that have been developed to achieve this aim.

Since the 1950s, text summary has existed in some form. two major factors have dominated study in this field. A emphasis on methods for creating extracts from scientific articles has emerged as a result of work in library science, office automation, and information retrieval, including the use of "shallow" language analysis and term statistics.

The other factor has been artificial intelligence research, which has looked at "shallow" knowledge-based ways for condensing data. While there are still some issues to be resolved, extraction-based approaches have made significant development in the recent decade. (devopedia, 2020)

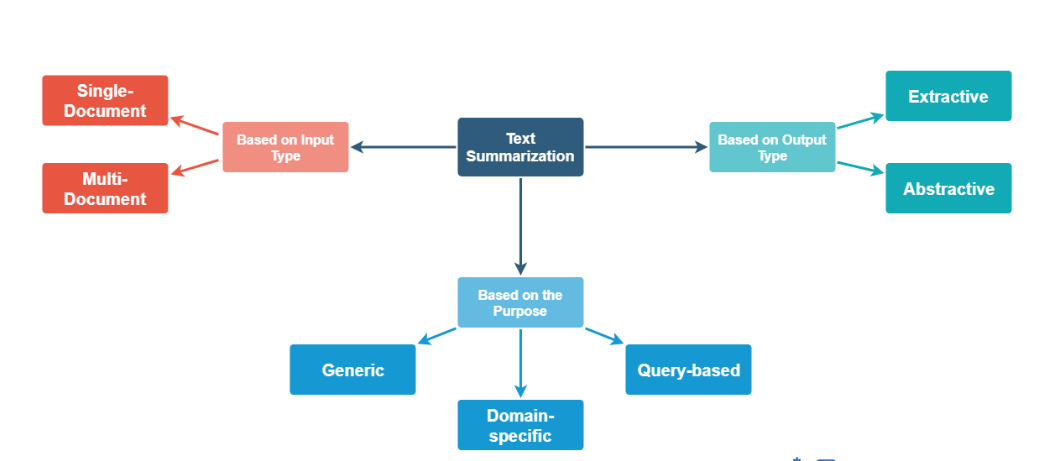


Fig2. 1 Types of text summarization.

Since extraction is frequently utilized, such as the snippets provided by search engines, and it is considerably easier than abstracting, most contemporary text summarizers are extractive. As a result, this post will concentrate on extraction. Extractive summarizers (whether for single-document or multi-document summarizing) often have to overcome three problems:(i) Content selection, or what should be selected from the text to form the summaries, which are typically in the form of sentences or phrases; (ii) Information ordering, or how to order the extracted sentences or phrases; (iii) Sentence realization, or how to clean up the extracted sentences or phrases so that they form a coherent summary. The first issue is clearly the most important for extractive summarizers. (Shen, 2009)

**Abstractive Summarizers:**

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.In our heads, we generate a semantic representation of the document. Then we choose terms from our general vocabulary (words we frequently use) that meet the semantics to construct a brief summary that summarizes the entire content. As you can see, creating this type of summarizer may be tough because it would require Natural Language Generation.

Let's have a look at the most common solution to the problem.

Application of sequence-to-sequence RNNs

The approach was proposed in a paper by Ramesh Nallapati, Bowen Zhou, Cicero dos Santos, Caglar Gulcehre, Bing Xiang from IBM.

The term “sequence to sequence models” is used because the models are designed to create an output sequence of words from an input sequence of words. The input sequence in the considered case is the actual text document and the output sequence is the shortened summary. (Vashisht, 2019)

In this project the model used will be attentional Recurrent Neural Network encoder-decoder model which was first proposed for machine translation by Dzmitry Bahdanau, Jacob’s University, Germany.

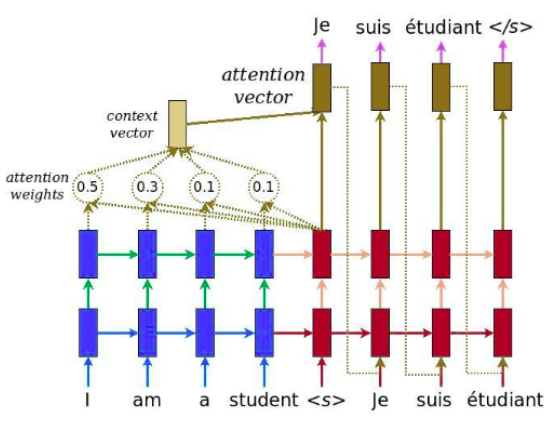


Fig2. 2 Attentional-encoder-decoder

## **Similar applications**

1. (Text Summarizer)

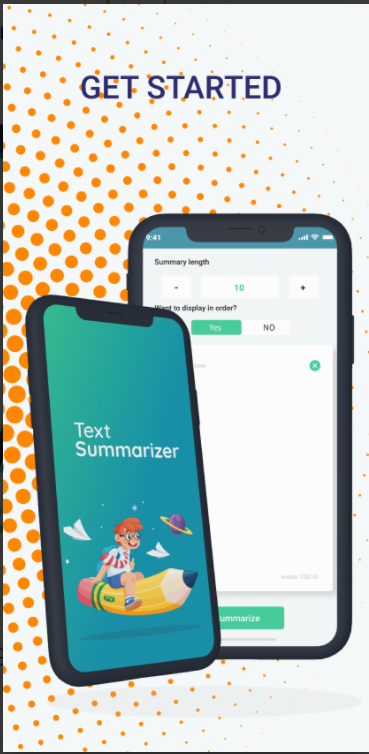


Fig2. 3 Text Summarizer

Text summarizer is an android app that helps you to summarize the educational books ,

You can use such app to summarize a text for four times free then you need to pay for it to use ,

1. (Text Summarizer\_AI)

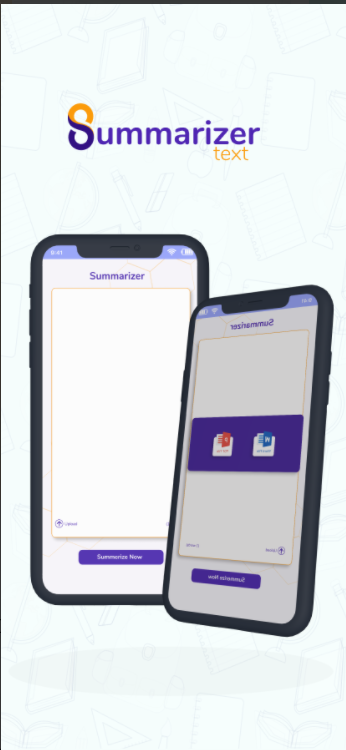


Fig2. 4 Text Summarizer\_AI

Text Summarizer\_AI is an IOS app that helps you to summarize the wiki articles , you can copy past the summary of the text and you can upload a file from your devise , the app is based on extracting model .

## **Disadvantages of similar apps**

1. The Accuracy: the accuracy of such apps is not good at all

You will fined a lot of problems on the summary

Like : repeated words , wrong summary

1. Privacy : such apps asks you about sum of privet information

Put the mechanism of the app not need it

1. Mechanism : there a lot of errors that happen suddenly while using such apps like , disconnect to the model.
2. Need a lot of money to use that apps many times

Our app (SummarizeIt ) will solve these problems with the future versions

# Chapter Three: Requirements and Analysis

## Functional Requirements / use case

* User must be able to open the camera and scan a text from image
* User must be able to Upload file from the memory
* User must be able to listen to the summarized text
* The application must be connected to the internet
* The application must be able to connect to the Summarizer model
* The application must be able to send a text and receive the summarized text from the Summarizer model
* The application must be able to pronounce (text to speech) the summarized text
* The application must be able to run on multiple OS

Fig3. *1* Use Cases

## Non-Functional Requirements

## Usability:

the system should be easy to use and simple to understand:

* Ease of use: Ease of Entering a text with any way of the allowed.
* Easy to work with the result text.

1. Security :

* Using different kind of firewall systems.
* Encrypting the network in high level of encryption.
* The system should be secure and saving people privacy.

1. Maintainability :

* Continuous enhancement, improvement, and development of the system
* There must be periodic maintenance on the servers and devices from time to time (ranging from one to 3 months)
* Ease and speed of handling errors.

## Software Tools & Hardware Specifications

* Mobile device specification:

Supporting Android Platform and IOS

(Android 4.0.4 (Ice Cream Sandwich), 4.3 (Jelly Bean)).

Ram: 1 GB

CPU Quad-core 1.4 GHz Cortex-A9

* Software Requirements:

Operating System : Windows 10 Home Premium x64

• IDE : Visual Studio Code

• Browser : Google Chrome

* Technologies and Programing Languages:

• Dart

• Flutter

• Google Colab

• TensorFlow

• Keras

• NLP

• Recurrent Neural Network (RNN)

• Machine Learning

• Deep Learning

## Analysis Diagrams

* + 1. Use case Diagram**:**

A use case diagram is a graphic representation of the interactions between users and the various aspects of a system. A use case is a strategy for identifying, clarifying, and organizing system needs in system analysis. The use cases are the roles that the actors (User, Admin, and mobile application) play within and around the system.

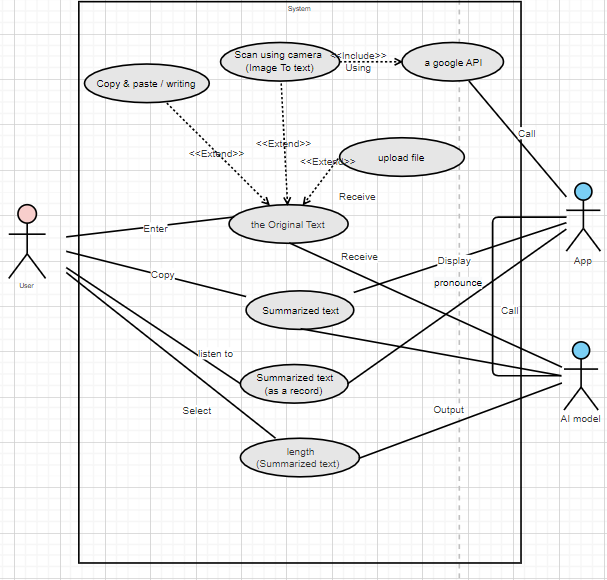


Fig3. 2 Use case Diagram

* + 1. Sequence diagram:

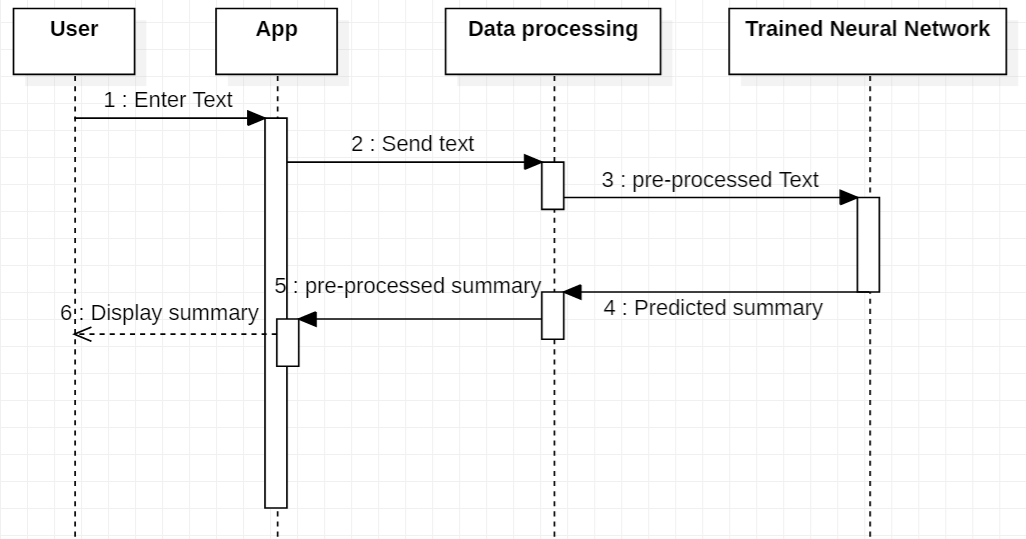
****

Fig3. *3* Sequence Diagram

* + 1. Block Diagram

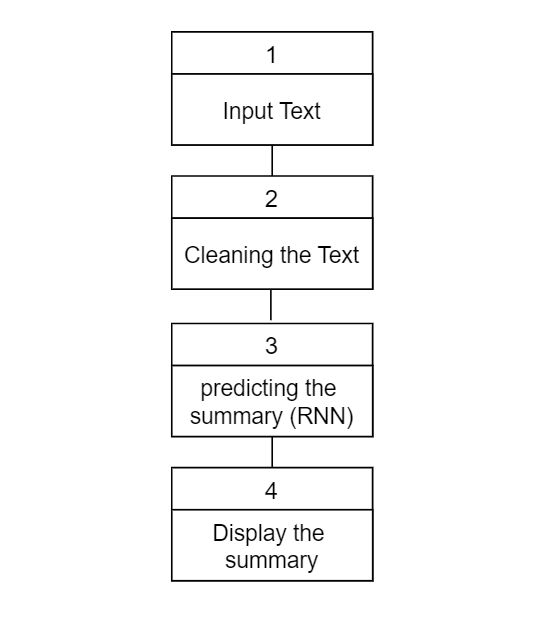


Fig3.4 Block Diagram

* + 1. Flowchart Diagram

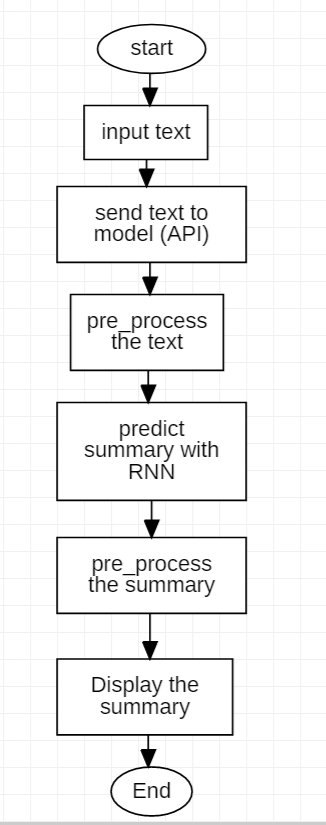


Fig3.5 Flowchart Diagram

* + 1. State Diagram

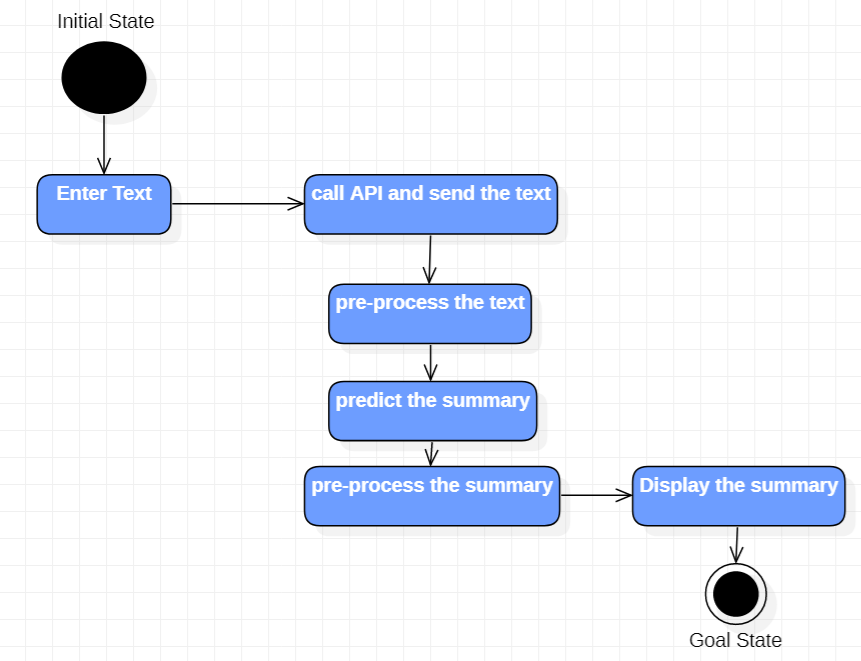


Fig3.5 State Diagram

# References

devopedia, 2020. *Text Summarization.* [Online]   
Available at: https://devopedia.org/text-summarization  
[Accessed 5 6 2022].

Gambhir, M., 2016. Recent automatic text summarization techniques:. *Recent automatic text summarization techniques*, 29 3, p. 66.

Halifax, K. K., 2017. A Brief Survey of Text Mining: Classification, Clustering and. *Classification, Clustering and*, 28 9, p. 13.

Shen, D., 2009. *Springer link.* [Online]   
Available at: https://link.springer.com/referenceworkentry/10.1007/978-0-387-39940-9\_424  
[Accessed 5 6 2022].

Vashisht, A., 2019. *OpenGenus IQ \_Text Summarization using RNN.* [Online]   
Available at: https://iq.opengenus.org/text-summarization-using-rnn/  
[Accessed 5 6 2022].