SAVITRIBAI PHULE PUNE UNIVERSITY

A PRELIMINARY PROJECT REPORT ON

Predicting Next Word in the sentence using BERT Algorithm

SUBMITTED TO THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE IN THE PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE

BACHELOR OF ENGINEERING
(Computer Engineering)(SEM-I)

SUBMITTED BY

Group ID : B08

Mr. Shaikh Zaid Mohamed Mateen Exam No:4251

Mr. Pathan MohammadAayan Samirkhan Exam No:4221

Mr. Pathan Muaazkhan Naimkhan Exam No:4222

Mr. Waghmare Avinash Bhagwat Exam No:4275

Under The Guidance of

Prof. A. N. Nawathe

DEPARTMENT OF COMPUTER ENGINEERING
Amrutvahini College of Engineering, Sangamner
Amrutnagar, Ghulewadi - 422608

2022-23

Savitribai Phule Pune University

CERTIFICATE

This is to Certify that

Mr. Shaikh Zaid Mohammed Mateen Exam No: 4251

Mr. Pathan MohammadAayan Samirkhan Exam No: 4221

Mr. Pathan Muaazkhan Naimkhan Exam No: 4222

Mr. Waghmare Avinash Bhagwat Exam No: 4275

of B.E. in Computer Engineering was examined in the

Project Examination entitled

"Predicting Next Word in the sentence using BERT Algorithm"

on . . ./. . . /2022

At

Department of Computer Engineering

Amrutvahini College of Engineering, Sangamner

. . . . . .

Internal Examiner

. . . . . .

External Examiner

Acknowledgment

Achievement is Finding out what you have been doing and what you have to do. The higher is submit,the harder is climb. The goal was fixed and we began with the determined resolved and put in a ceaseless sustained hard work. Greater the challenge, greater was our determination and it guided us to overcome all difficulties. It has been rightly said that we are built on the shoulders of others. For everything we have achieved, the credit goes to who had really help us for project and for the timely guidance and infrastructure. Before we proceed any further, we would like to thank all those who have helped us in all the way through. To start we are thankful to Honorable Principal Dr. M. A. Venkatesh sir for his encouragement and support.I would like to take this opportunity to thank to our respected Head of Department Prof. R. L. Paikrao and Seminar Coordinator Prof. R. G. Tambe and Prof. D. R. Patil .And we also thank our guide Prof. A. N. Nawathe for guidance, care and support, which they offered whenever we needed.

Abstract

Long phrases might be tedious to write, but text prediction technology built into keyboards makes this simple. In addition, Next Word Prediction called ”Language Modeling.” The task at hand is predicting the first word that will be said. It has numerous applications and is one of the main tasks of human language technology. This method uses letter to letter prediction and says that it predicts a letter when letter is used to build a word. Long short time memory formula can sense past text and anticipate the words that can be useful for the user to border sentences. Word prediction software has been developed to help people talk more easily and to help those who write more slowly.

Synopsis

- **Title: Predicting Next Word in the sentence using BERT Algorithm**
- **Domain : Artificial Intelligence**
- **Sub-domain : Neural Network/Deep Learning**
- **Objectives:**

  1. By recommending the next word based on past text, this application seeks to minimise human effort.
  2. Using an BERT Algorithm to predict the appropriate word for the user's convenience.

- **Abstract:**

  Writing long sentences is a little difficult, but text prediction technology built into the keyboard has made this simple. Language Modeling is another name for Next Word Prediction. It involves trying to predict the next word that will be spoken. It is one of the main tasks of human language technology and has a variety of uses. Long short time memory formula may recognise previous material and anticipate words that may be helpful for the user to border phrases. This method makes use of letter-to-letter prediction, which denotes that it anticipates a letter when a word is formed from other letters. Word prediction tools have been developed that could make it easier to talk and also help those who write more slowly.

- **Keywords:**

  Natural Language Processing (NLP), Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), Language Modeling (LM)

- **Problem Definition:**

  Develop a graphical user interface based text prediction system, which suggests text based on the previous typed text by the user through recurrent neural

network and Long Short Term Memory.

- **List of Modules:**

  1. Client Interface
  2. Prediction Program - RNN model
  3. Model Dataset

- **Current Market Survey:**

  1. VIPA : VIPA is a software platform that is designed to swallow, process and display large numbers of disparate streaming data flows, including video, audio, text, etc. It is developed by a company named Oceanit and it was founded in the year 1985. It has its primary application in various fields.
  2. Linguamatics : Linguamatics is the world leader in deploying innovative natural language processing (NLP)-based text mining for high-value knowledge discovery and decision support. It is a private company and was founded in the year 2001. Linguamatics I2E is used by top commercial, academic and government organizations.

- **Scope of the Project:**

  1. Every person using modern technology platforms needs to use the text prediction model based on BERT Algorithm.
  2. This system has been significantly improved with useful features for future upgrades. BERT is the algorithm used to improve text that is predicted automatically.

3. Text prediction is widely used across a variety of software platforms, including webpages, computer programmes, and mobile apps. In order to anticipate the text, deep learning or machine intelligence employs a variety of supervised and unsupervised machine-learning methods.

- **Literature Survey:**

1. Next Word Prediction
   Author - Keerthana N, Harikrishnan S, Konsaha Buji M, Jona J B
   Year - 2021
   Summary - Suggests subsequent immediate word supported this out their word. These systems work victimization machine learning algorithms that has limitation to form correct syntax.
2. Recurrent Neural Network based Models for Word Prediction
   Author - S.Ramya, C.S.Kanimozhi Selvi
   Year - 2019
   Summary - Suggest and presented a comparative study on various models like Recurrent Neural Network, Stacked Recurrent Neural Network, Long Short Term Memory network (LSTM) and Bi-directional LSTM that gives solution for the above said problem.
3. Predicting next Word using RNN and LSTM cells: Stastical Language Modeling
   Author - Aejaz Farooq Ganai, Farida Khursheed
   Year - 2019
   Summary - The paper describes how some common structural next word predicting queries would be satisfactorily described inside model.

4. A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2

   Author - Yuanbin Qu, Peihan Liu, Wei Song, Lizhen Liu*, Miaomiao Cheng

   Year - 2020

   Summary - We train the machine for specific tasks and then use it in natural language processing, which will help solve some sentence generation problems.

5. Natural Language Word Prediction Model Based on Multi-Window Convolution and Residual Network

   Author - Jingyun Yang , Hengjun Wang, Kexiang Guo

   Year - 2020

   Summary - Proposed MCNN-ReMGU model based on multi-window convolution and residual-connected minimal gated unit (MGU) network for the natural language word prediction.

- **Software and Hardware Requirement of the Project:**

  You can write software and hardware requirement of the project here

  Software:

  1. VS code
  2. Keras
  3. Tensorflow
  4. NLTK(Natural Language Toolkit)

  Hardware:

  1. Processor : 2.6 GHz
  2. RAM : 4GB
  3. Hard Drive : 40GB

- **Contribution to Society:**

  Our Next Text Prediction Model provides a better platform for all text typers by automating text suggestion based on the user's identified pre-typed text. These assist users in fast typing and getting relevant text suggestions based on their pre-typed text, putting them into lesser efforts. A text prediction system improves people's user experience of typing and searching text across the internet.

- **Probable Date of Project Completion:** March 2023

- **Outcome of the Project:**

  1. The proposed model will suggest the text by predicting the next text with the help of : Deep Learning/Machine Learning (Recurrent Neural Network).
  2. Previously typed text will be analysed for further text prediction.

Abbreviation

RNN	Recurrent Neural Network
LSTM	Long Short Term Memory
BERT	Bidirectional Encoder Representations from Transformers
NLP	Natural Language Processing
NLTK	Natural Language Toolkit

List of Figures

List of Tables

INDEX

CHAPTER 1

INTRODUCTION

1.1 PROJECT IDEA

- As we type in what is the weather we already receive some predictions. We can see that certain next words are predicted for the weather. When a user is texting or typing, the next word prediction can be fantastic. Understanding the user's messaging tendencies would save a tonne of time. Additionally, our virtual assistant might use this to finish some statements. Overall, we will be implementing the next word prediction feature of the predictive search system.

1.2 MOTIVATION OF THE PROJECT

- When autocomplete successfully guesses the term a user intends to write after only a few characters have been entered into a text input box, it speeds up human-computer interactions. Its influence permeates every part of our life, from messaging to looking up information or files online. For some people, this is what enables them to perform even the most fundamental duties with ease, including self-expression, interpersonal communication, searching, and simple job-related tasks. The user would find the entire process to be much more convenient thanks to this project, and they may also save time and effort. By enhancing the effectiveness and speed of word prediction through our project, we sincerely hope to accelerate this procedure and provide the public with a solution to this specific problem.

CHAPTER 2

LITERATURE SURVEY

2.1 LITERATURE SURVEY

1. Recurrent Neural Network based Models for Word Prediction

    - Author - S.Ramya, C.S.Kanimozhi Selvi
    - Suggest and presented a comparative study on various models like Recurrent Neural Network, Stacked Recurrent Neural Network, Long Short Term Memory network (LSTM) and Bi-directional LSTM that gives solution for the above said problem.

2. Predicting next Word using RNN and LSTM cells: Stastical Language Modeling

    - Author - Aejaz Farooq Ganai, Farida Khursheed
    - The paper describes how some common structural next word predicting queries would be satisfactorily described inside model.

3. A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2

    - Author - Yuanbin Qu, Peihan Liu, Wei Song, Lizhen, Miaomiao Cheng
    - We train the machine for specific tasks and then use it in natural language processing, which will help solve some sentence generation problems.

4. Natural Language Word Prediction Model Based on Multi-Window Convolution and Residual Network

    - Author - Jingyun Yang , Hengjun Wang, Kexiang Guo
    - Proposed MCNN-ReMGU model based on multi-window convolution and residual-connected minimal gated unit (MGU) network for the natural language word prediction.

Sr. No.	Paper Title	Year of Publication	Method Algorithm Used
1	A text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2	2020	BERT and GPT-2
2	Predicting next Word using RNN System	2020	RNN
3	Natural Language Word Prediction and LSTM cells: Stastical Language Mod eling and Residual Network	2019	LSTM
4	Predicting next Word using RNN and LSTM cells: Stastical Language Modeling and Residual Network	2019	RNN and LSTM

Table 2.1: Comparative Analysis

CHAPTER 3

PROBLEM DEFINITION AND SCOPE

3.1 PROBLEM STATEMENT

Develop a graphical user interface based text prediction system, which suggests text based on the previous typed text by the user through recurrent neural network and Long Short Term Memory.

3.1.1 Goals and objectives

Goal and Objectives:

- This application is focused on reducing human efforts by suggesting the next word based on previous text.
- Utilizing RNN neural network which will predict the relevant word for convenience of user.

3.1.2 Statement of scope

- In technology platforms, the text prediction model based on the Recurrent Neural Network system is extremely important for every person.
- This system has been significantly improved with useful features for future upgrades. Recurrent neural network is the technology used to improve text that is predicted automatically.
- Text prediction is widely used across a variety of software platforms, including websites, computer programs, and mobile apps. In order to predict the text, deep learning or machine intelligence employs a variety of supervised and unsupervised machine-learning algorithms.

3.2 SOFTWARE CONTEXT

The entirety of the numerous features a developer requires to finish a task can be referred to as a programming context. Programmers interpret the same information differently based on their context, which includes information from multiple sources goals for programming. A context model outlines the management and structure of

context data. It aims to formalise or semi-formalize descriptions of the contextual data found in context-aware systems. In other words, the model offers the mathematical interface and behavioural description of the environment, while the context is the components of the system that surround it. utilised to display reusable contextual data for a component.

3.3 MAJOR CONSTRAINTS

- Information confidentiality
- Cost of the project
- Errors in the Final Output
- Modifying software specifications

3.4 METHODOLOGIES OF PROBLEM SOLVING AND EFFICIENCY ISSUES

Using theory and research to identify solutions to issue domains, idea testing, and the use of best practises are all part of the problem solving process in software development. Utilizing logic and imagination to find problems and find software solutions is another aspect of problem-solving.

3.5 OUTCOME

- Provide option to select how many number of words to be predicted.
- Predict the next word in the sentences.
- Provide relevant output to the user.

3.6 APPLICATIONS

- Word processors, search engines, messaging services like WhatsApp, command-line translators, and more can all use Word Predictor.

- Word prediction software is developed to assist people with physical limitations boost their typing speed and reduce the number of keystrokes required to complete a word or a sentence.
- The next word prediction model is useful for the people who are victim of the disease known as Attention deficit hyperactivity disorder (ADHD) , it will help them to type efficiently.

3.7 HARDWARE RESOURCES REQUIRED

Sr. No.	Parameter	Minimum Requirement	Justification
1	CPU Speed	2 GHz	Minimum
2	RAM	4 GB	Minimum
3	Hard Disk	50 GB	Minimum

Table 3.1: Hardware Requirements

3.8 SOFTWARE RESOURCES REQUIRED

Platform :

1. Operating System: Windows OR Linux OR MAC
2. IDE: Microsoft Visual Studio
3. Programming Language : Python

CHAPTER 4

SOFTWARE REQUIREMENT SPECIFICATION

4.1 INTRODUCTION

4.1.1 Purpose and Scope of Document

A series of use cases that define how users will interact with the code are sometimes included in a software requirements specification, which is a detailed description of the behaviour of a system that needs to be constructed. Additionally, it has non-functional criteria as well. Nonfunctional requirements, such as performance engineering requirements, quality standards, or design limitations, place restrictions on the design or implementation. All criteria needed for the project's development are listed in the software requirements specification document. We must have a complete and accurate grasp of the items that will be developed in order to determine the needs. This was created following thorough discussions with the project team and the client. A software requirements specification is a detailed explanation of the intended use and setting for software that is currently being developed.

There are many excellent definitions of System and Software Requirements Specifications that can be used as a solid foundation for defining a fantastic specification and pointing out flaws in earlier efforts. Additionally, there is a tonne of excellent information about developing strong specifications online. Lack of information about how to prepare specifications appropriately or even what should be included in specifications is not the issue.

4.1.2 Overview of responsibilities of Developer

- Supervising the project plan's testing, documentation, and training initiatives.
- Analysis of risk management is done.
- ensuring that the project is finished on schedule and within the given budget.
- Regularly monitor project progress and report to the guide.
- Work on creating the project timeline, budget, and plan.
- Give team members different responsibilities to do.
- Coordinate with team members.

4.2 FUNCTIONAL REQUIREMENTS

The primary functional need is our programme, for which we will use a programming language, that is Python. The dataset file must always be accessed by the system. The user will enter data as a word (partial or full). The words that match the letters you've typed so far will be shown on the screen. Every time a word prediction cycle is finished, the system prompts the user to input further words. If so, a new cycle begins; otherwise, the system halts.

4.3 NONFUNCTIONAL REQUIREMENTS

4.3.1 Performance Requirements

- The system should be interactive to users.
- The interface is simple and easy to use.
- System is user friendly, self-explanatory
- This system can be used by everyone.
- Speed: The system should be made as fast as possible to reduce response time.

- Throughput: The throughput should be as high as possible. We should be able to attain maximum output in minimum time.
- Resource Utilization: Resources are modified according to user requirements.

4.3.2 Safety Requirements

- Operation of regular updation for the dataset should take place.
- The system should be tough and not prone to breakdowns.

4.3.3 Security Requirements

- The administrators maintain the system as per the maintenance contract.
- The system has to be secure from attacks.
- In case of breakdown should be stabilized soon.

4.3.4 Software Quality Attributes

- Try to attain maximum reliability.
- Reliability will also be higher since we try to attain maximum accuracy.
- Maintain proper and updated dictionary files to improve reliability.
- The information provided in the dataset files should be correct.
- Minimize the errors.
- All operations will be done correctly to increase the level of accuracy.

4.4 SYSTEM REQUIREMENTS

4.4.1 Database Requirements

4.4.1.1 Hardware Requirements

- Minimum CPU – P3/AMD Athlon 1.0 GHz+

- Minimum Disk Space – 512MB
- Minimum Memory – 500MB
- Touch Screens/Keyboard

4.4.1.2 Software Requirements

- Natural Language Toolkit (NLTK)
- Keras
- Tensorflow
- Numpy
- Streamlit
- Pytorch

4.5 ANALYSIS MODELS: SDLC MODEL TO BE APPLIED

Waterfall Model

The waterfall SDLC model moves step by step through the phases of analysis, projecting, realisation, testing, implementation, and support to show the development process as a flow. The SDLC model involves stage-by-stage execution. With waterfall, strict documentation is implied. Each phase of this SDLC model includes predefined features that are required of it.

This strategy enables preventing a number of errors that could occur due to inadequate project control. It leads to widespread documentation development.

4.6 SYSTEM IMPLEMENTATION PLAN:

Month	Week	Date	Project Activity
August	1st Week	05/08/2022	Project topic searching
August	2nd Week	12/08/2022	Project topic selection
August	3rd Week	18/08/2022	Finalization of base paper
August	4th Week	26/08/2022	Submition of synopsis
September	1st Week	10/09/2022	Preparing presentation
September	2nd Week	16/09/2022	Make presentation slides
September	3rd Week	23/09/2022	Prepared for review
September	4th Week	30/09/2022	Do changes suggested by guide
October	1st Week	07/10/2022	discussed about dataset
October	2nd Week	14/10/2022	finalized the dataset
October	3rd Week	21/10/2022	Discussed software requirement
October	3rd Week	04/11/2022	Dicuss abount content of next presentation
November	1st Week	04/11/2022	Draw and discussed UML diagram
November	2nd Week	11/11/2022	Prepared for next presentation
November	3rd Week	11/11/2022	Prepared sem1 report

Table 4.1: Implementation Plan

CHAPTER 5

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

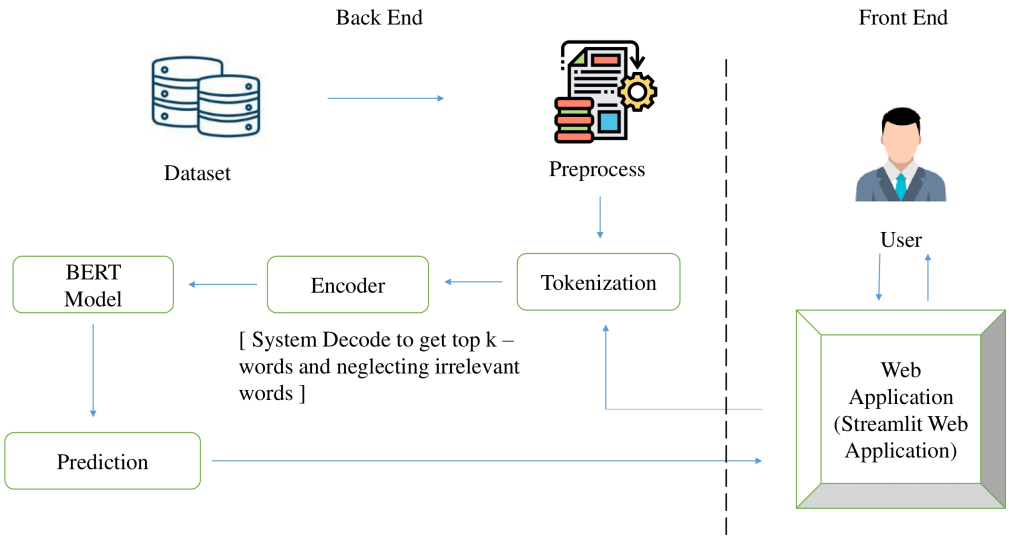

Figure 5.1: System Architecture

5.2 DATA FLOW DIAGRAM

DFD 0:
Level 0 Data flow diagram consist of basic overview of the flow diagram in which external entity as input that is sentences and output as the relevant words predicted.

DFD 1:
First level data flow diagram consist of small deep overview of the system working flow that is in process we include Bidirectional Encoder Representation from Transformer. Here user can enter the sentences or the paragraphs to predict the further relevant text. Here the previous data will be taken as input and processed to generate the required output in the model.

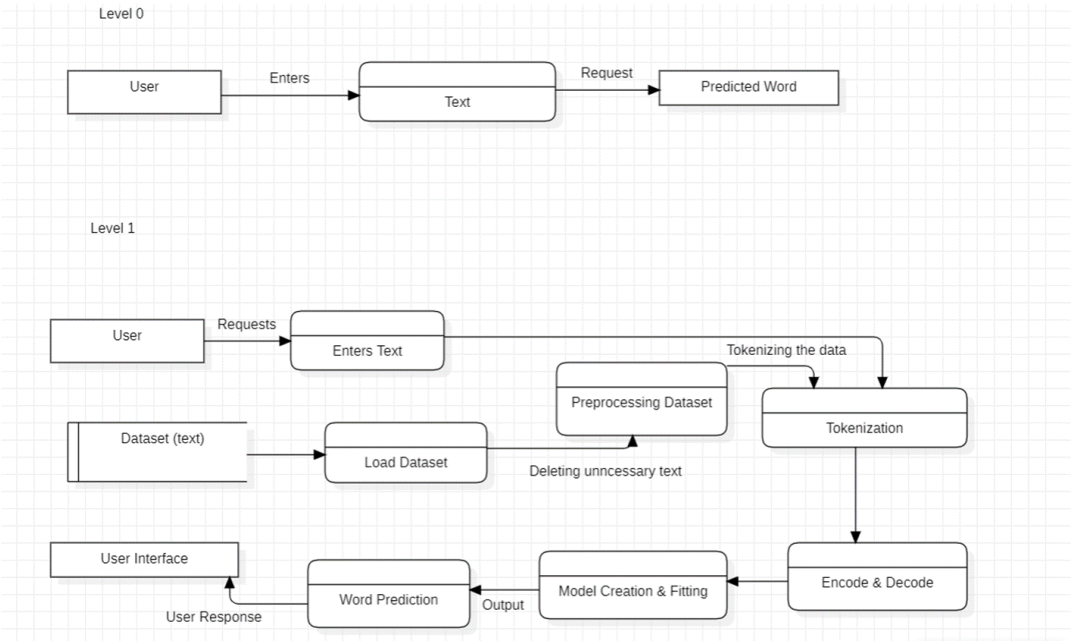

Figure 5.2: Data Flow Diagram

5.3 ENTITY RELATIONSHIP DIAGRAM

ER Diagram: Entity-Relationship model is referred to as an ER model. This data model is on a high level. The data items and relationships for a given system are defined using this model. An Entity Relationship Diagram is a diagram that shows the relationships between various entities in a system. It creates the database's conceptual design. Additionally, it creates a very straightforward and straightforward data view. The database structure is represented by an entity-relationship diagram in ER modeling.

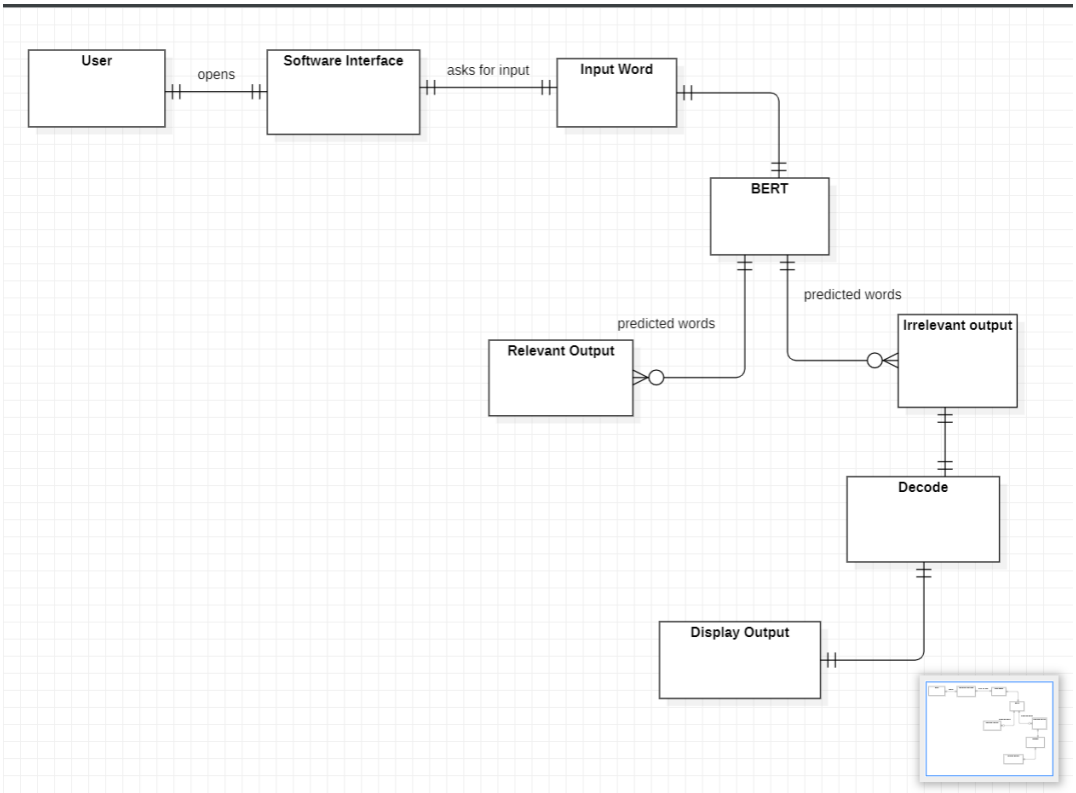

Figure 5.3: Entity Relationship

5.4 UML DIAGRAMS

5.4.1 Class Diagram

Class Diagram:
Class diagrams describe the static structure of a system, or how it is structured rather than how it behaves. These diagrams contain the following elements: 1. Classes: which represent entities with common characteristics or features. These features include attributes, operations, and associations. 2. Relationships: which represent relationships that relate two or more other classes where the relationships have common characteristics or features. These features include attributes and operations.

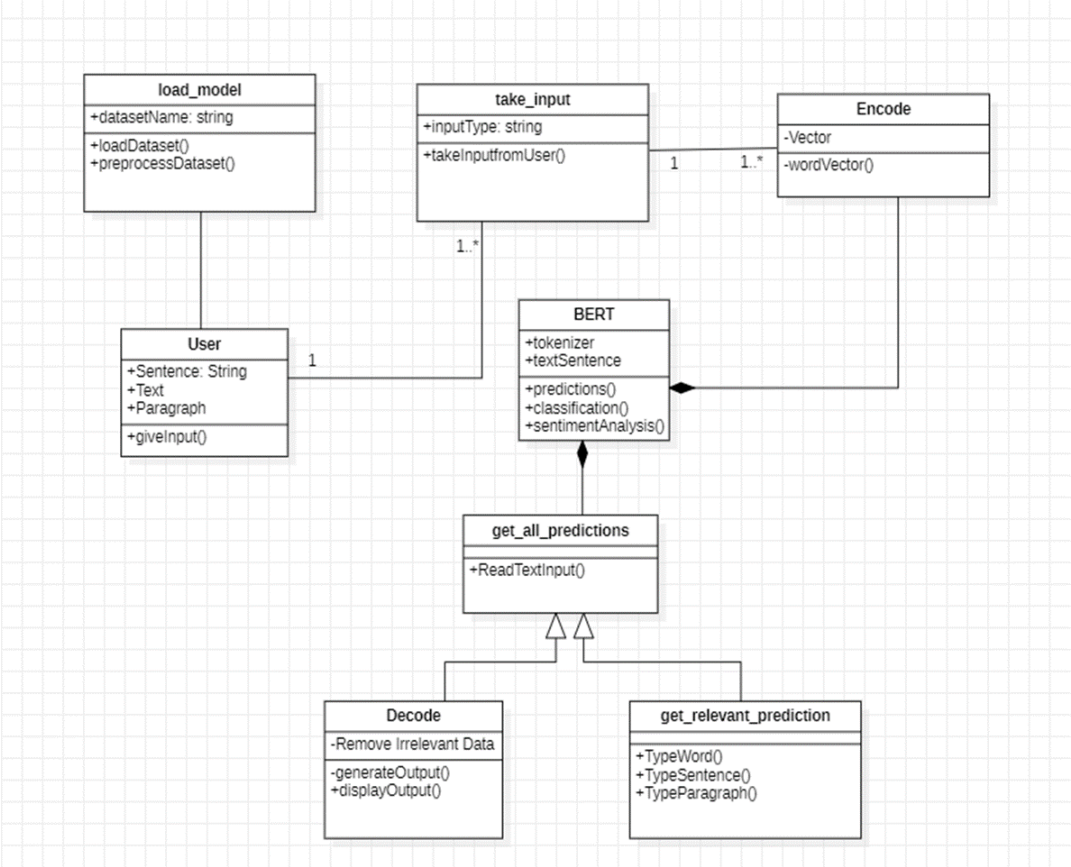

Figure 5.4: Class Diagram

5.4.2 State Diagram

State Diagram: State transition diagrams provide a way to model the various states in which an ob ject can exist. While the class diagram shows a static picture of the classes and their relationships, state transition diagrams model the dynamic behaviour of a system in response to external events (stimuli).

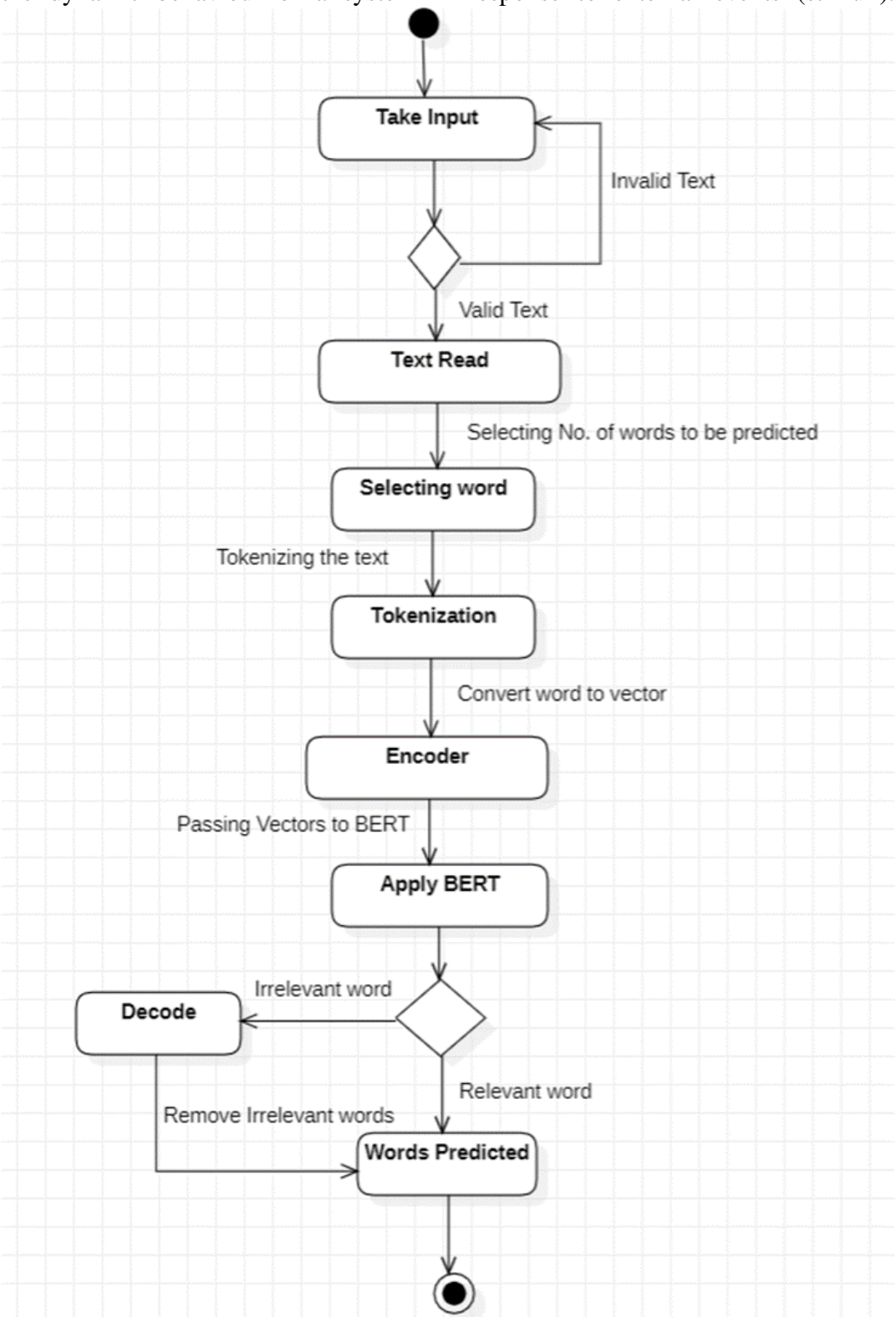

Figure 5.5: State Diagram

5.4.3 Use Case Diagram

Use case diagram: Use case diagrams describe the functionality of a system and users of the system. They contain the following elements:
1. Actors: which represent users of a system, including human users and other systems.
2. Use cases: which represent functionality or services provided by a system to users.

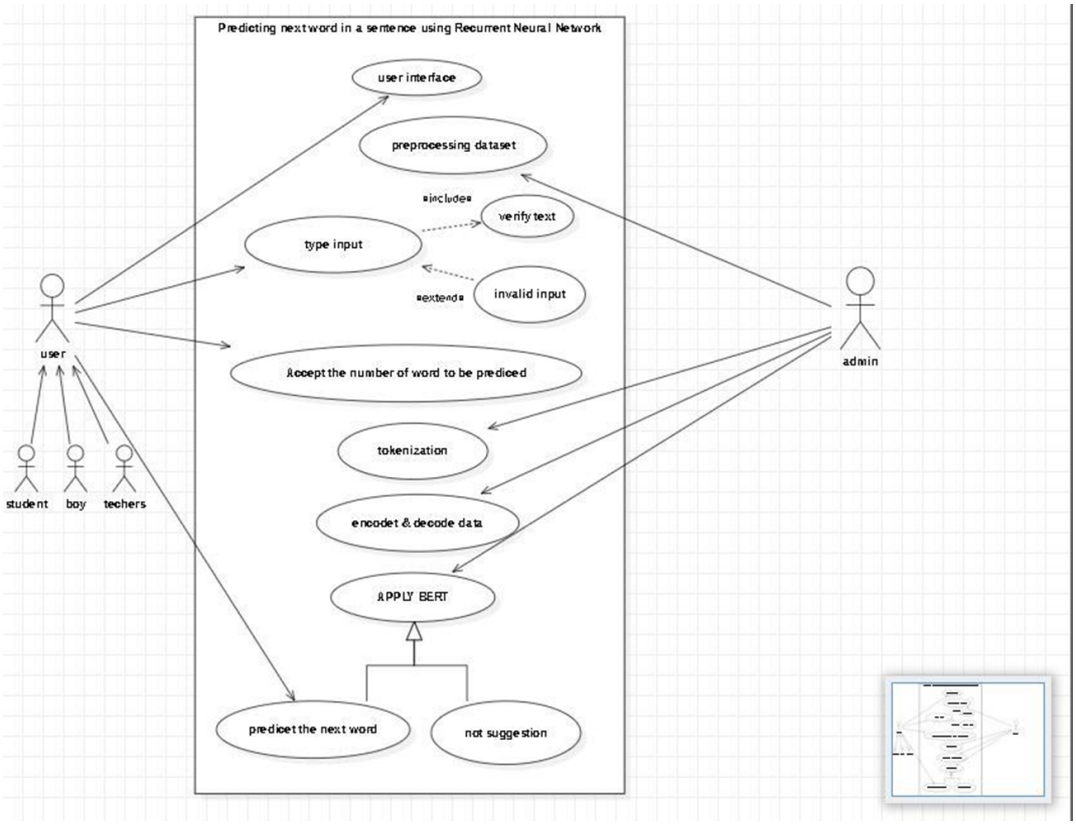

Figure 5.6: Use Case Diagram

5.4.4 Activity Diagram

Activity Diagram:
Activity diagrams describe the activities of a class. They are similar to state transition diagrams and use similar conventions, but activity diagrams describe the behaviour/states of a class in response to internal processing rather than external events. They contain the following elements:
1. Action States: which represent uninterruptible actions of entities, or steps in the execution of an algorithm.
2. Action Flows: which represent relationships between the different action states on an entity.
3. Object Flows: which represent utilization of objects by action states, or influence of action states on objects.

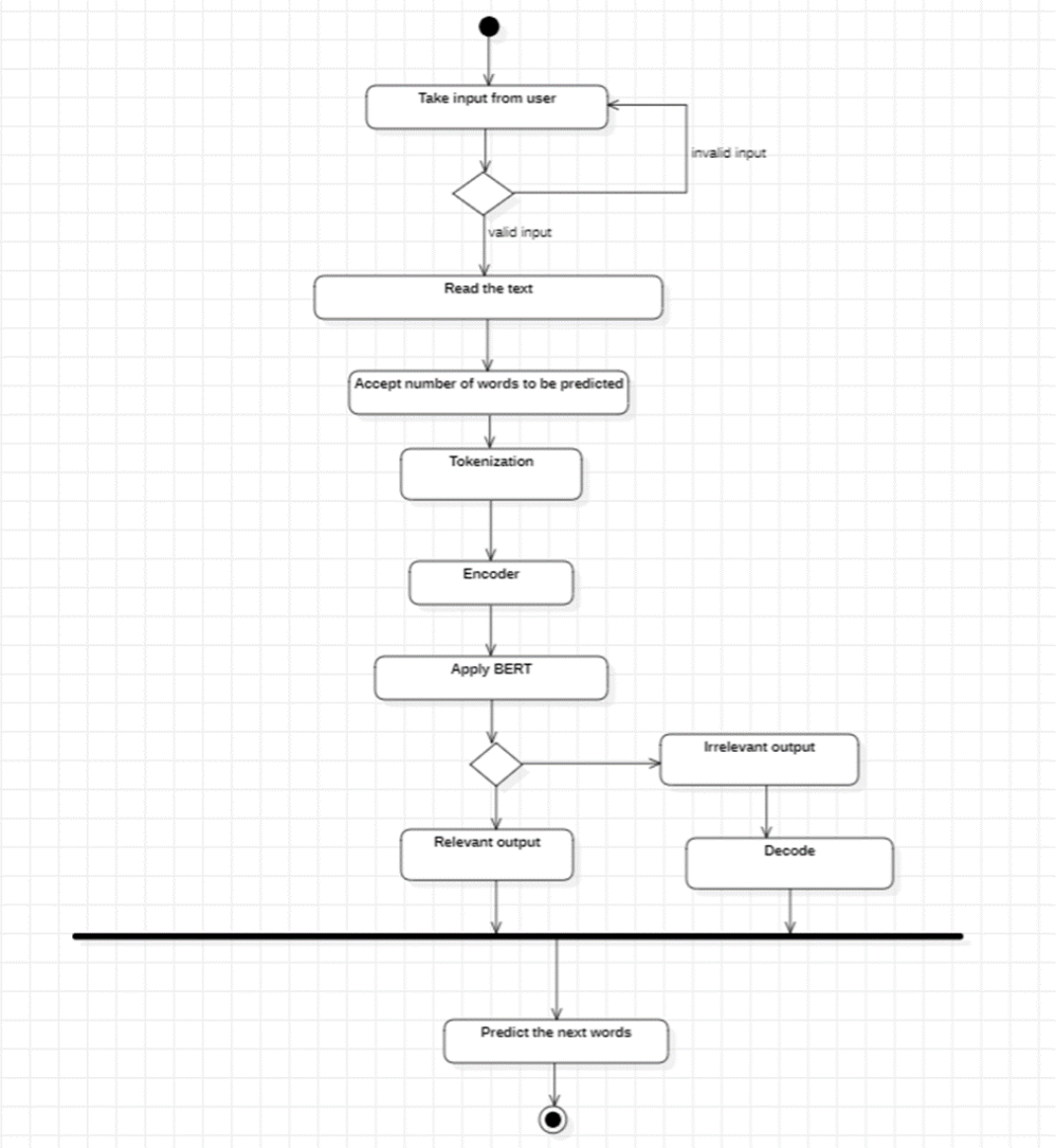

Figure 5.7: Activity Diagram

5.4.5 Sequence Diagram

Sequence Diagram:
Sequence diagrams typically show the flow of functionality through a use case, and consist of the following components:
1. Actors: involved in the functionality.
2. Objects: that a system needs to provide the functionality 3. Messages: which represent communication between objects.

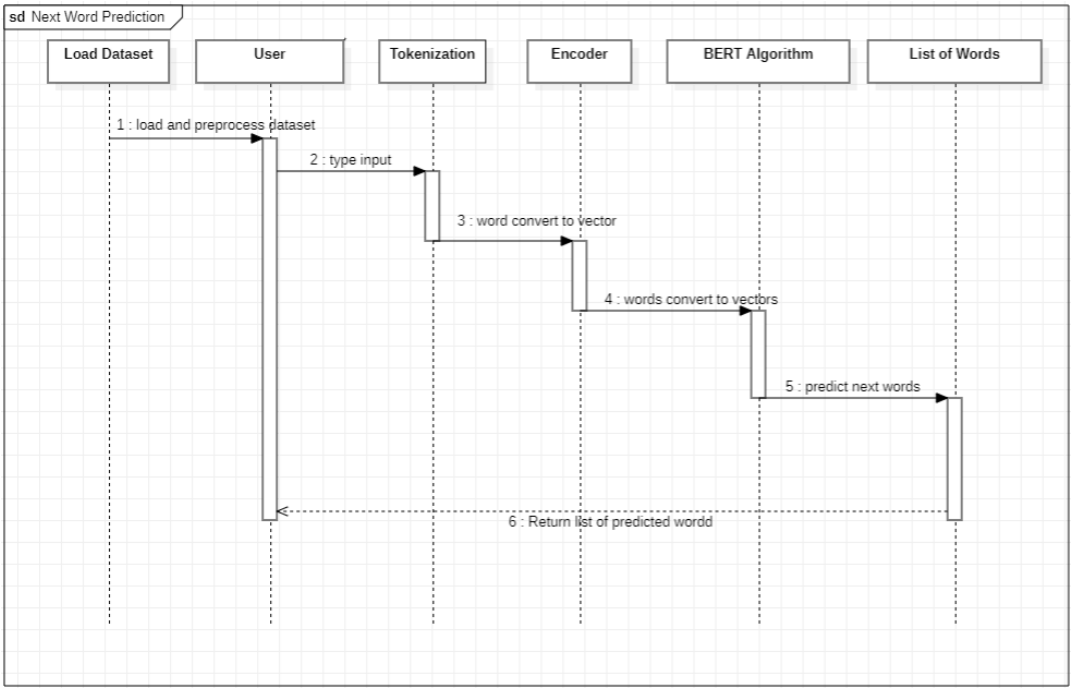

Figure 5.8: Sequence Diagram

CHAPTER 6

OTHER SPECIFICATION

6.1 ADVANTAGES

- Word prediction is a piece of assistive technology that offers word suggestions.
- This technology can help kids who have writing difficulties.
- Significantly improved performance compared to legacy methods.
- A simple method for utilising pre-trained models (transfer learning)
- The ability to tailor your data to the particular language environment and issue you are facing.

6.2 LIMITATIONS

- It is slow to train(algorithm) because it is big and there are a lot of weights to update.
- It is expensive. It requires more computation because of its size, which comes at a cost.
- It is designed to be input into other systems (not a standalone program), and because of that, it has to be fine-tuned for downstream tasks, which can be fussy.

6.3 APPLICATIONS

- Word processors, search engines, messaging services like WhatsApp, command-line translators, and more can all use Word Predictor.
- Word prediction software is developed to assist people with physical limitations boost their typing speed and reduce the number of keystrokes required to complete a word or a sentence.

CHAPTER 7

SUMMARY AND CONCLUSION

7.1 CONCLUSION

Word processors, web search engines, messaging services like WhatsApp, command-line translators, and more can all use Word Predictor. Word prediction software was initially developed to assist people with physical limitations boost their typing speed and to lessen the number of keystrokes required to complete a word or a sentence. Thus, utilising the Bert algorithm, we created our own word prediction tool that unquestionably improves user productivity.

7.2 SUMMARY

Communication is one of the most important characteristics of humanity. We communicate with one another mostly through language. How can our brain digest language so efficiently? How do words work in communication, and how are they read and understood? These are the only immediate, essential concerns that might concern us. I won't go into great detail on the neuroscience and neurolinguistics aspects, but I would want to encourage all readers who are interested to read the reference materials given in the reference section. Remaining on topic, NLP focuses on how computers can process, analyse, and interpret vast volumes of natural human language data, or more specifically, how computers interact with human language.

Modern, cutting-edge networks enable machines to replicate and learn from human-like tasks. In order to demonstrate how these language models may predict the following collection of words for the given input text, we shall cover various key network architectures.

An important area of natural language processing is the creation of sentences from provided starting words or the completion of incomplete phrases. It shows, in one way, whether a machine is capable of human thought and creativity. In order to tackle some problems with sentence production, we train the machine for specific tasks before using it in natural language processing. This is especially useful for application situations like summary generation, machine translation, and automatic question answering. Currently, the BERT models are a popular language model. for the prediction and creation of text. The impressive performance of this model in the

area of text production has been supported by numerous experiments.

To improve its capacity for linguistic comprehension, BERT is trained on a variety of different tasks. Future Sentence Prediction With BERT BERT has developed three methods to anticipate the next sentence: In the first, sentences are used as input, while the output is a single class label, like in the case of the job below: The MNLI is an important classification task (Multi-Genre Natural Language Inference). The goal is to assess if the second assertion supports, refutes, or is neutral with respect to the first. Natural Language Inference Question (QNLI): The following actions by the model are required for this task:

Identify whether the second statement provides an answer to the query posed in the first. It is necessary to assess whether the second statement. whether the second builds on the first or not. The second type just requires one sentence as input, but the result is the same as the label for the next class. The task and data sets utilised for it are as follows: The Stanford Sentiment Treebank, or SST-2: It is a binary sentence classification job that uses sentences that were taken from movie reviews and annotated to show how they felt. This algorithm's objective is to anticipate words in a phrase based on context.

CHAPTER 8

REFERENCES

ANNEXURE A

PROBLEM STATEMENT FEASIBILITY

- The proposed model is reliable and can be used in our day today life.
- Our model is trained on paragraph but in real life we can implement it by training it on large scale.
- For large scale implementation we will use user's search history.

ANNEXURE B

DETAILS OF THE PAPERS REFERRED

1. Next Word Prediction

   Author - Keerthana N, Harikrishnan S, Konsaha Buji M, Jona J B

   Year - 2021

   Summary - Suggests subsequent immediate word supported this out their word. These systems work victimization machine learning algorithms that has limitation to form correct syntax.

2. Recurrent Neural Network based Models for Word Prediction

   Author - S.Ramya, C.S.Kanimozhi Selvi

   Year - 2019

   Summary - Suggest and presented a comparative study on various models like Recurrent Neural Network, Stacked Recurrent Neural Network, Long Short Term Memory network (LSTM) and Bi-directional LSTM that gives solution for the above said problem.

3. Predicting next Word using RNN and LSTM cells: Stastical Language Modeling

   Author - Aejaz Farooq Ganai, Farida Khursheed

   Year - 2019

   Summary - The paper describes how some common structural next word predicting queries would be satisfactorily described inside model.

4. A Text Generation and Prediction System: Pre-training on New Corpora Using BERT and GPT-2

   Author - Yuanbin Qu, Peihan Liu, Wei Song, Lizhen Liu*, Miaomiao Cheng

   Year - 2020

   Summary - We train the machine for specific tasks and then use it in natural language processing, which will help solve some sentence generation problems.

5. Natural Language Word Prediction Model Based on Multi-Window Convolution and Residual Network

   Author - Jingyun Yang , Hengjun Wang, Kexiang Guo

   Year - 2020

Summary - Proposed MCNN-ReMGU model based on multi-window convolution and residual-connected minimal gated unit (MGU) network for the natural language word prediction.

ANNEXURE C

PLAGIARISM REPORT FOR THIS REPORT

All must attach certificate/report of Plagiarism issued by Urkund Software. Percentage of Similarity should not be more than 30%