Marathwada Shikshan Prasarak Mandal's
Deogiri Institute of Engineering and Management Studies, Aurangabad

Project Report

on

Next Word Prediction

Submitted By

Sumit Dobhal(AI3146)
Pranav Vyavahare (AI3152)
Sharayu Iralepatil (AI3158)

Dr. Babasaheb Ambedkar Technological University
Lonere (M.S.)

Department of Computer Science and Engineering
(Artificial Intelligence and Machine Learning)
Deogiri Institute of Engineering and Management Studies, Aurangabad
(2023- 2024)

Project Report

Next Word Prediction

Submitted By

Sumit Dobhal (AI3146)

Pranav Vyavahare (AI3152)

Sharayu Iralepatil (AI3158)

In partial fulfillment of

Bachelor of Technology CSE (AI & ML)

Guided By

Dr. Sugandha Nandedkar

Department of Computer Science and Engineering
(Artificial Intelligence and Machine Learning)
Deogiri Institute of Engineering and Management Studies,

Aurangabad

(2023- 2024)

CERTIFICATE

This is to certify that, the Mini Application entitled "**Next Word Prediction**" submitted by **Sumit Dobhal (AI3146), Pranav Vyavahare (AI3152) Sharayu Iralepatil (AI3158),** is a bonafide work completed under my supervision and guidance in partial fulfillment for award of Bachelor of Technology CSE (AI&ML) Degree of Dr. Babasaheb Ambedkar Technological University, Lonere.

Place: Aurangabad
Date:

Dr. Sugandha Nandedkar
Guide

Dr. S. A. Shaikh
H. O. D.

Director,
Deogiri Institute of Engineering and Management Studies,
Aurangabad

DECLARATION

This is to certify that, the partial project report entitled, "**Next Word Prediction**" Submitted by **Sumit Dobhal (AI3146), Pranav Vyavahare (AI3152), Sharayu Iralepatil (AI3158)** is a bonafide work completed under my supervision and guidance in partial fulfillment for award of Bachelor degree in CSE (AI&ML) of Deogiri Institute of Engineering and Management Studies, Aurangabad under Dr. Babasaheb Ambedkar Technological University, Lonere.

Place: Aurangabad

Date:

Guide: Dr **Sugandha Nandedkar**

External Examiner: _____

Contents

List of Abbreviations i
List of Figures ii

ABSTRACT

1. INTRODUCTION 1 - 4
- 1.1 Introduction to Existing System
- 1.2 Problems in Existing System
- 1.3 Need for Improvement
- 1.4 Objectives
- 1.5 Problem Statement

2. LITERATURE SURVEY 5 - 6
- 2.1 Introduction
- 2.2 Research Paper- I
- 2.3 Research Paper- II
- 2.4 Research Paper- III

3. PROPOSED SYSTEM 7 - 10
- 3.1 Concept Diagram
- 3.2 Sequence Diagram
- 3.3 Block Diagram
- 3.4 Dataset
- 3.5 Concept Involved
- 3.6 Steps of Implementation

4. SYSTEM DEVELOPMENT 11 - 14
- 4.1 Requirement Specification
  - 4.1.1 Software Requirement Specification (SRS)
  - 4.1.2 Data Flow Diagram

4.2 Methodology
4.3 User Interface Design
4.3 Dataset Description

5. Performance Analysis **15 – 17**

5.1 Identify the effective technique for training optimization and hyper-parameter selection on deep learning model
5.2 User Interface Design
5.3 Performance Analysis Metrics

6. Conclusion **18**

6.1 Conclusion
6.2 Limitations
6.3 Future Scope

REFERENCES

ACKNOWLEDGEMENT

List of Abbreviation

Sr. No	Acronym	Abbreviation
1.	BERT	Bidirectional Encoder Representations from Transformers
2.	DL	Deep Learning
3.	ML	Machine Learning
4.	LSTM	Long Short-Term Memory
5.	GRU	Gated Recurrent Unit
6.	NLP	Natural Language Processing
7.	RNN	Recurrent Neural Network
8.	ReLU	Rectified Linear Unit
9.	RMSE	Root Mean Squared Error
10.	Seq2Seq	Sequence-to-Sequence

List of Figures

Figure No.	Illustration	Page
3.1.1	Concept Diagram	vii
3.1.2	Concept Diagram	vii
3.1.3	Concept Diagram	vii
3.2.1	Sequence Diagram	viii
4.1.2	Data Flow Diagram	xii
4.3.1	Interface Diagram (Sketch)	xii
5.2.1	Interface Design	xvi

Abstract

The Next Word Prediction project aims to develop a sophisticated model that can accurately predict the subsequent word in a given text sequence. Utilizing advanced Natural Language Processing (NLP) techniques and deep learning architectures such as Long Short-Term Memory (LSTM) networks, the project seeks to enhance text entry efficiency and user experience in digital communication platforms.

The project encompasses the collection and preprocessing of large text corpora to train the predictive model. By implementing language modeling and machine learning algorithms, the model learns patterns and dependencies within the language data. The core of the project lies in the ability of the LSTM network to remember long-term dependencies, making it particularly suited for the sequential nature of language.

The predictive model operates by analysing the context provided by the preceding words and generating probabilities for the next word. This process not only speeds up text entry but also aids in error correction and assists users with limited typing abilities.

The anticipated outcome of the project is a robust and efficient next word prediction system that can be integrated into various applications, providing a seamless and intuitive typing experience. The project's success will be measured by the accuracy of the predictions and the system's adaptability to different languages and contexts.

This abstract provides a concise overview of the project's objectives, methodology, and expected outcomes. It highlights the use of NLP and LSTM networks in creating a model that improves text entry and assists users in digital communication.

References

[1] **Dataset Link:** raw.githubusercontent.com/Jaynwabueze/Simple_Chat_bot/main/topical_chat.csv

[2] https://medium.com/@evertongomede/next-word-prediction-enhancing-language-understanding-andcommunication-1322f3b57632

[3] https://www.geeksforgeeks.org/next-word-prediction-with-deep-learning-in-nlp/

[4] https://www.analyticsvidhya.com/blog/2023/07/next-word-prediction-with-bidirectional-lstm/

[5] Next Word Prediction Model | AmaKharwal (thecleverprogrammer.com)

[6] https://github.com/topics/next-word-prediction

[7] https://www.youtube.com/watch?v=Zn22qt7j2dM

[8] https://www.youtube.com/watch?v=fiqo6uPCJVI

[9] https://www.kaggle.com/code/ysthehurricane/next-word-prediction-bi-lstm-tutorial-easy-way

ACKNOWLEDGEMENT

We would like to place on record my deep sense of gratitude to Dr. S. A. Shaikh, HOD-Dept. of CSE(AIML), Deogiri Institute of Engineering and Management Studies Aurangabad, for his generous guidance, help and useful suggestions.

I express my sincere gratitude to Dr. Sugandha Nandedkar, Dept. of CSE(AIML), Deogiri Institute of Engineering and Management Studies Aurangabad, for her stimulating guidance, continuous encouragement and supervision throughout the course of present work.

I am extremely thankful to Dr. Ulhas Shiurkar, Director and Dr. S V. Lahane, Dean Academics, Deogiri Institute of Engineering, and Management Studies Auranagabad, for providing me infrastructural facilities to work in, without which this work would not have been possible.

Signature of Students

Sumit Dobhal (AI3146)

Pranav Vyavahare (AI3152)

Sharayu Iralepatil (AI3158)