

A MINI PROJECT REPORT

On

TEXT PREDICTOR USING N-GRAMS

Submitted in partial fulfillment of the requirement of University of Mumbai for the Course

**In**

# Computer Engineering (IV SEM)

Submitted By

**Janhavi Anap**

**Riddhi Narkar**

Subject Incharge

**Prof. Merlin Priya Jacob**



**CERTIFICATE**

This is to certify that the requirements for the project report entitled ‘**Text Predictor using n-grams ’** have been successfully completed by the following students:

|  |  |
| --- | --- |
| **Name** | **Moodle Id** |
| Janhavi Anap | 19102043 |
| Riddhi Narkar | 19102003 |
|  |  |

In partial fulfilment of the course Python Programming (CSL405) in Sem: IV of Mumbai University in the Department of Computer Engineering during academic year 2020-2021.

Sub-in-Charge

# 

# PROJECT APPROVAL

The project entitled ‘**Text Predictor using n-grams**’ by **Janhavi Anap**, and **Riddhi Narkar** are approved for the course of Python Programming (CSL405 ) in Sem: IV of Mumbai University in the Department of Computer Engineering.

Subject-in-Charge

Date:

Place: Thane

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr**. **No.** | **Topic** | **Page No.** |
| 1. | Abstract (150-200 words) | 6 |
| 2. | List of Figures |  |
| 3. | List of Tables |  |
| 4. | Problem Definition |  |
| 5. | Introduction |  |
| 6. | Description of the modules used |  |
| 7. | Implementation details with screen-shots (stepwise) |  |
| 8. | Conclusion and Future Scope |  |
| 9. | References |  |
| 10. | Acknowledgement |  |

**FOLLOW THE GUIDELINES GIVEN BELOW FOR PREPARING DOCUMENT IN THE GIVEN FORMAT**

1. The project report should be neatly typed.
2. Avoid using Abbreviations.
3. The text should be justified and typed in the Font style ‘Times New Roman’ and Font size ‘12’.
4. Heading and subheading should be bold with font 14.
5. As of now only prepare soft copy of the document.
6. Report should be of 15-20 pages

**Abstract**

One of the major technological marvels of this tech-infused 21st century is the ability of us to connect digitally. With the advent of sophisticated communication tools, being able to express oneself through a digital medium effectively and efficiently needed a reliable mechanism. A very popular method of expressing thoughts, ideas or sharing information, is texting a message or posting a typed text piece on the internet. Often, typing can be prone to errors and not much efficient and quick when it comes to long typing sessions.

A word predictor is a piece of software which suggests words after processing your input. The input here could be a string of any number of words, and it normally depends on the algorithm used. An algorithm could be processing in a static way, which doesn’t learn, no matter how much it is trained and tested, or it can be dynamic, and thus, adapting the changes in diction, grammar and style of the user. A dynamic word predictor may use a corpus or a database that changes every time a user uses the service and thus updates itself with the changes and advancements of the user. Also, such a system can very easily survive the changing language style, new words, new phrases, or new slangs and hence is versatile and needs less maintenance. In addition to texting and using this software as just a medium for faster and more efficient typing, a slightly differently built model can also be used to enhance the communication experience of persons with disabilities. Such a model would use the Augmentative and Alternative Communication devices as an API.

In this Python project, we tried to build a text predictor using the concept of n-grams which could be deployed for use where one would want to type something using a hardware or software keyboard. It would process the user input using bi-grams and suggest 3 words which would make the sentence grammatically correct in the order of the decreasing frequency which in turn, is based on the typing habits of the user.

**Problem Definition**

One important aspect of effective and fast communication is proper grammar, and choice of words, spellings and more meaningful sentences. In this fast age of electronics and internet, racing our fingers with the speed of our thoughts is error prone.

A text predictor not only focuses on suggesting next words which suit your style, but also to the fact that how much the suggested word would make a complete sense in the context of the information being conveyed through the sentence being typed. In addition to its listed expectations, if a software would be trained to your style of grammar and your diction, and which would adapt the changes of the language you try to adapt in yourself would be an extremely powerful tool and would cater to a wide range of users. Such a tool if provided with the communication service itself would be an added benefit.

**Introduction**

**Conclusion**

Hence, word prediction can have a major impact when it comes to

**References**

## [Coursera - Java Programming: Principles of Software Design (Duke University / Chapter 3: N-grams: predictive text)](https://www.coursera.org/programs/a-p-shah-institute-of-technology-on-coursera-05d2i/browse?collectionId=&currentTab=CATALOG&productId=3-NBikXzEeWwQw6izX918Q&productType=s12n&query=java+duke+university&showMiniModal=true)

1. [Geeks for Geeks - word prediction using n-grams](https://www.geeksforgeeks.org/word-prediction-using-concepts-of-n-grams-and-cdf/)
2. [Kdnuggets - word prediction corpora](https://www.kdnuggets.com/2017/11/building-wikipedia-text-corpus-nlp.html)
3. [StackExchange – predicting the best length of n-grams](https://stats.stackexchange.com/questions/155483/estimating-the-best-length-of-n-gram)
4. [Towards data science – good grams: how to find predictive n-grams for your problem](https://towardsdatascience.com/good-grams-how-to-find-predictive-n-grams-for-your-problem-c04a5f320b39)
5. [StackAbuse – Python for NLP development](https://stackabuse.com/python-for-nlp-developing-an-automatic-text-filler-using-n-grams/)
6. [RealPython – Regex tools](https://realpython.com/lessons/regex-query-tool/)
7. [Tutorials Point – Python regular expressions](https://www.tutorialspoint.com/python/python_reg_expressions.htm)