A PCL Report

on

"NLP Automation For Text Processing and Model Building"

Submitted in partial fulfilment for the award of the degree of

BACHELOR OF TECHNOLOGY (HONOURS)

IN

COMPUTER SCIENCE (DATA SCIENCE)

Submitted by

M R NAVEEN KUMAR
19BTRCR005

A RISHAB VANIGOTHA
19BTRCR018

ABHIRAM K V
19BTRCR023

KEERTHI U S
19BTRCR024

MILAN HUNDIA JAIN
19BTRCR026

Under the guidance of
Guide name

Dr S Vijayakumar
Designation
Associate Professor

Faculty of Engineering & Technology

Jain (Deemed-To-Be University)

B. Tech (Honours) in Computer Science (Data Science)

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

2022-2023.

B. Tech (Honours) in Computer Science (Data Science)

Jain Global Campus, Kanakapura Taluk - 562112

Ramanagara District, Karnataka, India

CERTIFICATE

This is to certify that the PCL work titled **"NLP Automation For Text Processing and Model Building"** is carried out by **M R Naveen Kumar (19BTRCR005), A Rishab Vanigotha (19BTRCR018), Abhiram K V (19BTRCR023)**, **Keerthi U S (19BTRCR024), Milan Hundia Jain (19BTRCR026),** a bonafide students of Bachelor of Technology at the Faculty of Engineering & Technology, Jain (Deemed-to-be University), Bangalore in partial fulfilment for the award of degree, Bachelor of Technology (Honours) in Computer Science (Data Science), during the Academic year **2022-2023**.

Guide Name **Dr. S VIJAYAKUMAR** **Designation** **Associate Professor** Faculty of Engineering & Technology, Jain (Deemed-to-be University) Date: Signature:	**Dr. S RAMESH** **Associate Professor and** **Programme Head,** **Dept. of CSE** Faculty of Engineering & Technology, Jain (Deemed-to-be University) Date: Signature:

Name of the Examiner Signature of Examiner

1.

2.

DECLARATION

We, **M R NAVEEN KUMAR (19BTRCR005)**, **A RISHAB VANIGOTHA (19BTRCR018), K V ABHIRAM (19BTRCR023), KEERTHI U S (19BTRCR024), MILAN HUNDIA JAIN (19BTRCR026)** are students of seventh semester B. Tech (Honours) in **Computer Science (Data Science)**, at Faculty of Engineering & Technology, **Jain (Deemed-To-Be University)**, hereby declare that the project work titled **"NLP Automation For Text Processing and Model Building"** has been carried out by us and submitted in partial fulfilment for the award of degree in **Bachelor of Technology (Honours) in Computer Science (Data Science)** during the academic year **2022-2023**. Further, the matter presented in the project has not been submitted previously by anybody for the award of any degree or any diploma to any other University, to the best of our knowledge and faith.

Name: Signature

M R Naveen Kumar
USN: 19BTRCR005

Name: Signature

A Rishab Vanigotha
USN: 19BTRCR018

Name: Signature

Keerthi U S

USN:19BTRCR024

Name: Signature

K V Abhiram

USN:19BTRCR023

Name: Signature

Milan Hundia Jain

USN: 19BTRCR026

Place: Bengaluru

Date:

ACKNOWLEDGEMENT

It is a great pleasure for us to acknowledge the assistance and support of a large number of individuals who have been responsible for the successful completion of this project work.

First, we take this opportunity to express our sincere gratitude to the Faculty ***of Engineering & Technology, Jain (Deemed-to-be University),*** *for providing us with a great opportunity to pursue our Bachelor's Degree (Honours) in this institution.*

In particular, we would like to thank ***Dr. Geetha G, Director, School of Computer Science and Engineering, Jain (Deemed-to-be University)****, for her constant encouragement and expert advice.*

It is a matter of immense pleasure to express our sincere thanks to ***Dr. S. Ramesh, Professor and Program Head, Department of Computer Science (Data Science), School of Computer Science & Engineering, Jain (Deemed-to-be University)****, for providing the right academic guidance that made our task possible.*

We would like to thank our guide ***Dr. S Vijay Kumar, Associate Professor , Dept. of Computer Science & Engineering****,* ***Jain (Deemed-to-be University),*** *for sparing his valuable time to extend help in every step of our project work, which paved the way for smooth progress and fruitful culmination of the project.*

We would like to thank our Project Coordinator ***Dr. S. Vijayakumar*** *and all the staff members of Computer Science & Engineering for their support.*

We are also grateful to our family and friends who provided us with every requirement throughout the course.

We would like to thank one and all who directly or indirectly helped us in completing the Project work successfully.

Signature of Students

TABLE OF CONTENTS

ABSTRACT

We have created a state-of-the-art NLP application that automates numerous jobs in the discipline. The app's goal is to speed up the NLP procedure while increasing accuracy. Text pre-processing, automatic annotation, sentiment analysis, named entity identification, machine translation, chatbot capability, and text summarizing are just a few of the essential features offered by the programme. The software opens up NLP to a wider audience by utilizing cutting-edge algorithms and a user-friendly interface, and it offers insightful analysis of text data. Our NLP software is the ideal solution for automating NLP processes and maximizing the potential of your text data, regardless of whether you are a research scientist, data analyst, or business professional.

LIST OF FIGURES

NOMENCLATURE USED

NLP	Natural Language Processing
EDA	Exploratory Data Analysis
NER	Named Entity Recognition
CSV	Comma-Separated Values
GUI	Graphical User Interface

Chapter 1

INTRODUCTION

1.1. Overview

The field of computer science known as "natural language processing" (NLP) is more particularly the field of "artificial intelligence" (AI) that is concerned with providing computers the capacity to comprehend written and spoken words in a manner similar to that of humans.

NLP blends statistical, machine learning, and deep learning models with computational linguistics—rule-based modelling of human language. With the use of these technologies, computers are now able to process human language in the form of text or audio data and fully "understand" what is being said or written, including the speaker's or writer's intentions and sentiment.

A computer program's capacity to comprehend natural language, or human language as it is spoken and written, is known as natural language processing (NLP). It is a part of machine intelligence (AI). Natural Language Processing (NLP) is a field of study focused on allowing computers to understand, interpret, and generate human language. The automation of NLP tasks has become increasingly important as the amount of unstructured text data continues to grow.

The development of web interfaces and tools that simplify and automate NLP tasks is crucial in allowing users to analyze and understand their data without the need for extensive technical knowledge.NLP is widely used for applications like NER(Named Entity Recognition), Text Classification, Text Generation, Text Mask Prediction etc.NLP requires a lot of tedious tasks to be done. It requires a lot of effort and time.

A. What is Automation?

Automation is the process of using technology to carry out tasks and processes without requiring human involvement. This might comprise a variety of tasks, from straightforward ones like flipping on a switch to more intricate ones like managing a manufacturing facility. Several technologies, including robots, artificial intelligence, and machine learning algorithms, can be used to automate processes.

Automation aims to increase productivity, decrease expenses and errors, and free up human workers to concentrate on more imaginative and strategic duties. Several industries, including manufacturing, transportation, finance, and healthcare, can benefit from automation.

It's crucial to remember that automation can also have negative effects, including worsening social and economic inequality, job loss, and diminished job security for some individuals. The secret to effective automation is to strike a balance between the advantages and any potential disadvantages, and to make sure the procedure is just, moral, and sustainable.

B. What is Natural Language Processing?

Natural Language Processing, or NLP for short, is a subfield of artificial intelligence (AI) that deals with the use of natural language in communication between machines and people. With the use of NLP, computers can analyze, comprehend, and produce human language in order to communicate with people more naturally and humanely.

Language translation, sentiment analysis, text categorization, speech recognition, and chatbots are just a few examples of the many industries that use NLP technologies. NLP can be used, for instance, to translate text between languages in real-time or to assess customer evaluations and categorize them as good, negative, or neutral.

To process and analyze human language, NLP algorithms and models apply methods from linguistics, computer science, and mathematics. Additionally, they use a lot of annotated linguistic data to train machine learning algorithms and boost their efficiency.

NLP has the ability to change how people engage with technology and make it easier for computers to comprehend and process human language, which could open the door to new applications and use cases in the future.

C. How Automation can help in NLP?

Natural language processing (NLP) is a field where automation can be very helpful in a number of ways.

Text Pre-processing: Automated text pre-processing activities like tokenization, stemming, and stop-word elimination can speed up the process and increase the precision of NLP models.

Annotations: Automated annotation techniques can be used to annotate significant amounts of text material that will be utilized to train NLP models.

Sentiment analysis: Automated systems can assess the sentiment of a lot of text data, giving us useful information about what customers think and say.

Named Entity Recognition: Automated named entity recognition programmes can recognise and classify entities from text data, such as persons, businesses, and locations.

Machine Translation: Automated machine translation software may convert text from one language to another, opening up NLP models to a larger audience.

Chatbots: Automated chatbots can interpret and reply to user requests using natural language processing (NLP), which makes them helpful in customer care and other applications.

Summarization: Automated text summarizing systems can reduce long texts to shorter ones, which makes them simpler to comprehend and evaluate.

Automation generally helps to streamline NLP processes, allowing practitioners to concentrate on creating and enhancing NLP models rather than wasting time on monotonous manual operations.

1.2. Problem Definition

The goal of this project is to develop a web interface using Streamlit that simplifies and automates various Natural Language Processing (NLP) tasks, including Exploratory Data Analysis (EDA), named entity recognition (NER), text classification, masking and other related tasks.

The interface should provide the user with a convenient and user-friendly environment for performing NLP tasks with ease, without requiring extensive knowledge of programming and NLP algorithms.

The aim is to provide a tool that streamlines the NLP workflow, allowing users to focus on the analysis of their data, rather than on technical details.

1.3. Existing Methodologies

The existing system is essentially a Python library that does automated exploratory data analysis, data cleaning and preprocessing for machine learning and natural language processing.It also offers widget-based data analysis, which gives a graphical user interface and allows users to do any function by just ticking a checkbox. It also performs text preprocessing using regular expressions to remove noise.Word Analysis and Basic EDA Analysis are two forms of exploratory data analysis that are available

Chapter 2

LITERATURE SURVEY

2.1. Related Work

1. An Examination System Automation Using Natural Language Processing

Authors: Manjusha Pandey Indrashis Das Siddharth S. Rautaray bharat sharma

Year Of Publication: 2020

Summary: In this study, they attempt to automate the process of scoring answers. Essentially, a descriptive online examination system is where the data comes from. The data is analyzed and the model assigns marks to the answers provided. The back-end is written in Python, and NLTK and the NLTK library is used for natural language processing and database purposes.

2. A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing

Authors: Ayisha Tabassum, Dr. Rajendra R. Patil

Year Of Publication: 2020

Summary: This study discovered that text preprocessing methods are a significant factor in raising the accuracy of any method for text-based machine learning. The sequence of The result is influenced by the NLP pipeline, which is made. It is found that StopWords removal, punctuation, and tokenization are the popular and effective text formatting techniques.

3. Automation in Systematic, Scoping and Rapid Reviews by an NLP Toolkit: A Case Study in Enhanced Living Environments

Authors: Eftim Zdravevski, Petre Lameski, Vladimir Trajkovik, Ivan Chorbev, Rossitza Goleva, Nuno Pombo & Nuno M. Garcia

Year Of Publication: 2019

Summary: In this paper, they present an NLP toolkit for surveying scientific articles and trend analysis meta-studies. By leveraging NLP, it facilitates a robust and comprehensive eligibility and relevance analysis of articles. The framework is able to analyze the abstracts of over 70000 articles automatically.

4. Natural Language Processing approach to NLP Meta model automation

Authors: Mohammad Hossein ,Hassan B. Kazemian, Karim Ouazzane, Chris Chandler

Year Of Publication: 2018

Summary: An intelligent software has been developed which is able to perform as a competent NLP practitioner or psychologist. Results by the software were compared to the obtained results by the practitioner. A more efficient performance of the software, with a high level of accuracy and reliability, was observed.

2.2. Existing System:

The existing system is essentially a Python library that does automated exploratory data analysis, data cleaning and preprocessing for machine learning and natural language processing.

It also offers widget-based data analysis, which gives a graphical user interface and allows users to do any function by just ticking a checkbox.

It also performs text preprocessing using regular expressions to remove noise.

Word Analysis and Basic EDA Analysis are two forms of exploratory data analysis that are available.

2.3. Limitations Of Existing System:

The existing system is considered to be a long tedious process. The existing system doesn't support automation. Since NLP is a process, there are many time-consuming and exhausting tasks involved. It takes a lot of time and work.

2.4. Proposed System:

This system performs the basic exploratory data analysis and text preprocessing required for NLP. It has the ability to create models through a web-based graphical user interface.

It just requires a dataset as input, and our web GUI outputs a dataset based on the user's option of word or phrase analysis.

The user does not need any prior coding knowledge.

Chapter 3

METHODOLOGY

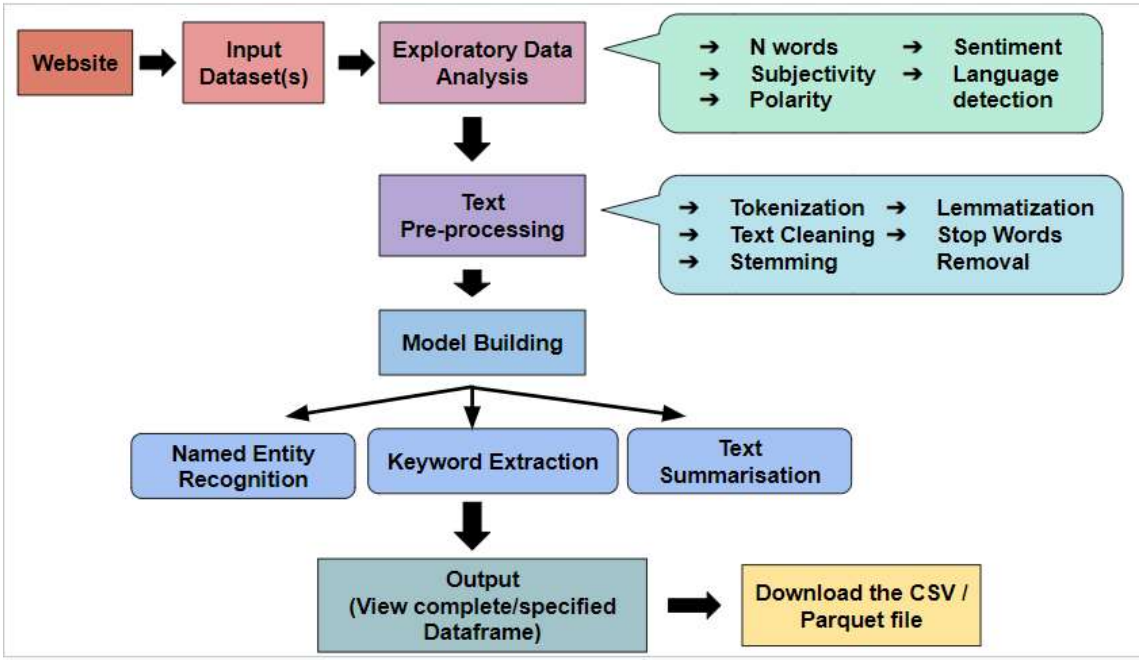
3.1. Architecture

Our model consists of a user application based model,where a user interface connects a deep rooted backend consisting of NLP techniques.

This system performs the basic exploratory data analysis and text preprocessing required for NLP. It has the ability to create models through a web-based graphical user interface. It just requires a dataset as input, and our web GUI outputs a dataset based on the user's option of word or phrase analysis. The user does not need any prior coding knowledge.

3.2. Sequence Diagram

Figure 1:



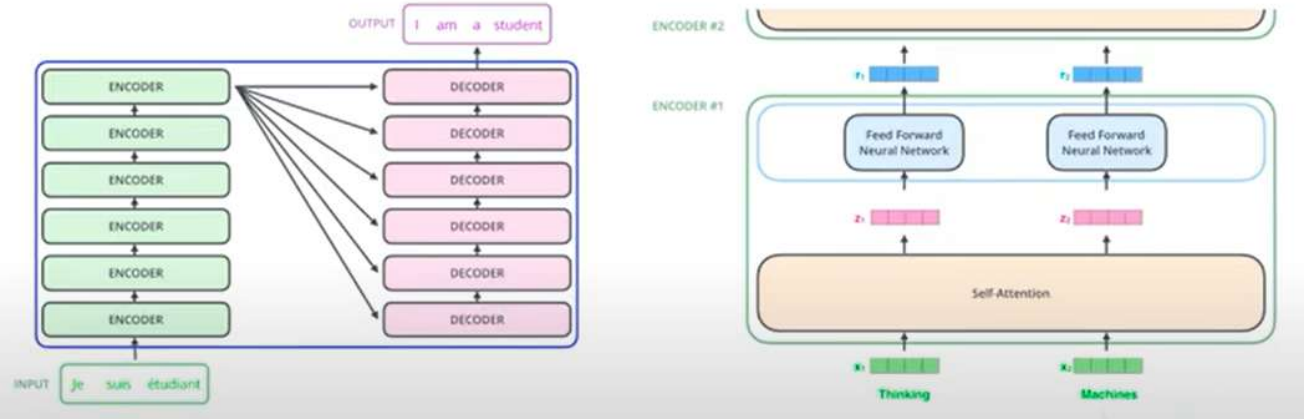
3.3. Understanding of Exploratory Data Analysis

Exploratory data analysis (EDA) for text data is the process of exploring, analyzing and understanding textual data in order to gain insights, detect patterns and identify relationships between different pieces of information. EDA helps to uncover hidden trends and gain a better understanding of the data. It is an important step in the data science process and can be used to inform the subsequent stages of data analysis.

The main techniques used in EDA for text data include data visualization, text mining, and natural language processing (NLP). Text mining can be used to uncover relationships between different words and phrases in the text, as well as sentiment analysis. Natural language processing (NLP) can be used to identify topics, identify sentiment, and extract important information from the text.

In addition to these techniques, EDA for text data can also involve data cleaning and preprocessing, such as removing stop words and punctuation, lemmatization, and normalization. These steps are important in order to ensure the data is in a suitable format for the analysis.

Transformers



Transformers are a type of machine learning model used for natural language processing (NLP). They are based on the concept of self-attention, where the model attends to different parts of the input sequence to compute a representation of the entire sequence. Transformers have become popular due to their ability to capture long-term dependencies in text and to handle large amounts of data more efficiently than traditional recurrent neural networks. They are typically used for tasks such as text classification, question answering, and machine translation.

3.4. Understanding of Named Entity Recognition

The most common data preprocessing activity is named entity recognition (NER). It entails locating important information in the text and classifying it into a number of predetermined categories. A constant subject of discussion or reference in a book is referred to as an entity. NER is a variation of NLP. The two steps that make up the fundamental process of NLP are as follows:

- identifying items in the text
- putting them into various categories

3.5. Understanding of Text Summarization

Text summary is the process of condensing lengthy articles into concise words or paragraphs. The process preserves the meaning of the paragraph while also extracting pertinent information. This expedites the process of understanding lengthy resources like research articles without leaving out important details.

Written summarising is the process of condensing a longer text content into a coherent, fluid, and brief summary while also emphasising its key ideas.

Text summarization involves a number of challenges, such as text detection, interpretation, creation of the summary, and examination of the finished summary. In extraction-based summarising, it is crucial to locate key terms in the text and use them to unearth pertinent information to include in the summary.

3.6. Understanding of Text Classification

The practice of classifying text into ordered groupings is called text classification, commonly referred to as text tagging or text categorization. Text classifiers can automatically assess text using Natural Language Processing (NLP), and then based on its content assign a set of predefined tags or categories.

Because it makes it simple to extract insights from data and automate business procedures, text classification is becoming an increasingly significant component of enterprises. The following are a few of the most typical uses and examples of automatic text classification:

Sentiment Analysis: Sentiment analysis is the method of determining if a text is speaking favourably or unfavourably about a particular issue (e.g. for brand monitoring purposes).

Topic Detection: Finding the theme or topic of a text is known as "topic detection" (e.g. know if a product review is about Ease of Use, Customer Support, or Pricing when analysing customer feedback).

Language Detection: Determine whether an incoming support ticket is written in English or Spanish to automatically route it using the language detection process.

Chapter 4

TOOL DESCRIPTION

4.1. Software Requirements:

1. Operating System : Windows / Linux / Mac
2. IDE : Jupyter Notebook
3. Streamlit for Website

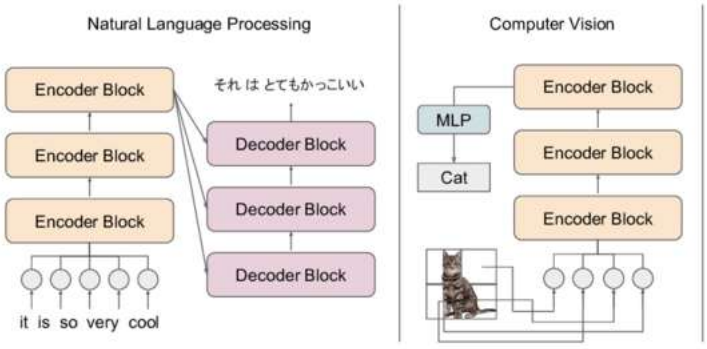
4.2. Hardware Requirements:

1. Ram 2GB+
2. Rom 2GB+
3. Intel i3 or better processor

Chapter 5

IMPLEMENTATION

- In this part, we begin with loading the dataset file in csv format in the webpage.
- We explore the selected text column using various exploratory data analysis techniques.
- Then, we perform text processing like data cleaning , lexical analysis (tokenization, stop word removal, stemming, etc.,), followed by sentence level analysis (semantic, syntactic, pragmatic, and disclosure analysis).
- After finishing the textual analysis, we perform Feature extraction where we extract useful features from the processed data.
- Then, you can select a model for building out of given 3 models:
    - Named – Entity Recognition
    - Text Summarization
    - Text Classification
- For model building we use Transformers , a type of neural network architecture that uses encoders and decoders with positional embeddings to process the language.
- They have been shown to outperform recurrent neural networks (RNNs) on a variety of natural language processing (NLP) tasks, such as text classification, machine translation, question answering, and text generation
- They have ability to capture long term dependencies in text and process the text parallelly
- This section concludes with the creation of a user-friendly web-based interface that generates the desired output.



Chapter 6

RESULTS AND ANALYSIS

6.1. Result Discussion:

The streamlit based web interface developed for automating various NLP tasks was tested on several datasets, including text classification, sentiment analysis, and text summarization, keyword extraction, Named Entity Recognition etc.

Our Web interface allows the user to choose multiple text files to analyze and perform various EDA tasks on them with the help of different NLP models.

A. Exploratory Data Analysis

This entails studying text data to learn about its properties and structure. EDA can assist NLP practitioners in identifying patterns in data, comprehending its distribution, and determining how to prepare it for further analysis.

N-words: N-words or N-grams are collections of N words that appear together in a text corpus and are used in exploratory data analysis (EDA) for natural language processing (NLP). Any positive integer N can be used, and N-grams can provide important information about the distribution and frequency of words in a corpus in addition to spotting recurring themes and grammatical structures.

Subjectivity: Subjectivity is a measure of the amount of personal opinion and factual information in a writing. Higher the value of subjectivity indicates that the text contains high personal opinion when compared to factual information.

Polarity: The overall sentiment expressed by a specific sentence, phrase, or word is referred to as polarity. This polarity can be expressed numerically as a "sentiment score."

Sentiment: Sentiment indicates if a given phrase, sentence or a word is negative or positive.

Language Detection: As the name suggests this feature detects the language of the text.

	subjectivity text1	sentiment text1	polarity text1	languages text1	n_words text1
0	0.4017	Positive	0.0867	en	403
1	0.2074	Negative	-0.1352	en	259
2	0.3334	Positive	0.0136	en	241
3	0.4491	Positive	0.0391	en	288
4	0.3774	Positive	0.0734	en	409
5	0.2678	Negative	-0.0667	en	249
6	0.5141	Positive	0.2873	en	266

Fig. 2. EDA tasks

B. Text Preprocessing

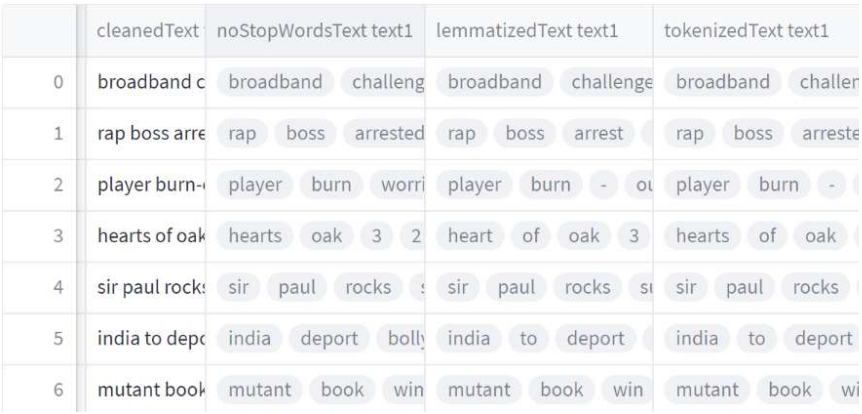
It involves cleaning and modifying raw text data so that it can be analyzed further. Preprocessing removes noise and extraneous information from text input while maintaining the most useful information. This entails studying text data to learn about its properties and structure. EDA can assist NLP practitioners in identifying patterns in data, comprehending its distribution, and determining how to prepare it for further analysis.

Text Cleaning: It includes removing irrelevant and distracting information from raw text data. Text cleaning is required to boost the accuracy of following NLP tasks including sentiment analysis, text categorization, and machine translation.

Stop Words Removal: Stop words are often used terms in a language that have little to no meaning and are frequently eliminated from text data during NLP preparation. These terms include "the," "a," "an," "in," "of," "to," and "and," among others. Eliminating stop words from text input can help reduce noise and enhance NLP model accuracy.

Tokenization: The practice of breaking down a piece of text into smaller pieces called tokens, which could be words, phrases, or even individual letters, is known as Tokenization. Tokenization is an important preprocessing step in NLP that allows targeting multiple tasks such as sentiment analysis, text classification, and machine translation.

Lemmatization and Stemming: They are Natural Language Processing (NLP) strategies for reducing words to their base or root forms. These strategies are used to reduce the complexity of text data and make it easier to examine.

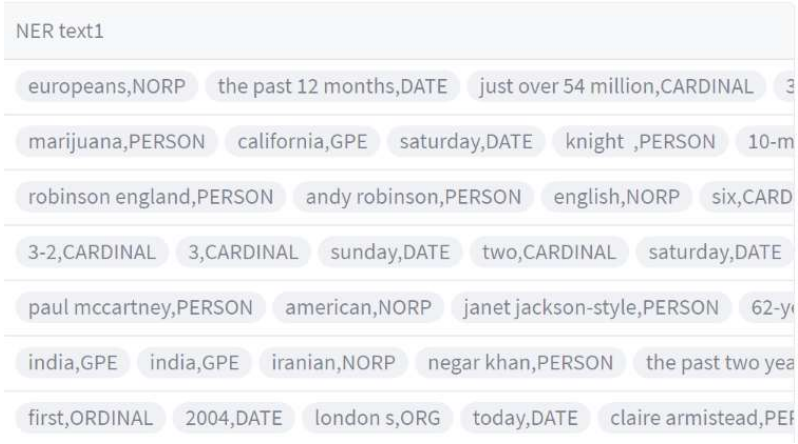


	cleanedText	noStopWordsText text1	lemmatizedText text1	tokenizedText text1
0	broadband c	broadband challeng	broadband challenge	broadband challer
1	rap boss arre	rap boss arrested	rap boss arrest	rap boss arreste
2	player burn-	player burn worri	player burn - ou	player burn -
3	hearts of oak	hearts oak 3 2	heart of oak 3	hearts of oak
4	sir paul rocks	sir paul rocks s	sir paul rocks su	sir paul rocks
5	india to depo	india deport bolly	india to deport	india to deport
6	mutant book	mutant book win	mutant book win	mutant book wi

Fig. 3. Text Preprocessing Tasks

C. NLP Models

Named Entity Recognition (NER): Named Entity Recognition (NER) is an NLP task that requires recognizing and retrieving entities (people, organizations, locations, dates, and other named entities) from textual information that is unorganized. Several NLP implementations, such as knowledge discovery, machine translation, and question-answering systems, rely on NER.



NER text1
europeans,NORP the past 12 months,DATE just over 54 million,CARDINAL 3
marijuana,PERSON california,GPE saturday,DATE knight ,PERSON 10-m
robinson england,PERSON andy robinson,PERSON english,NORP six,CARD
3-2,CARDINAL 3,CARDINAL sunday,DATE two,CARDINAL saturday,DATE
paul mccartney,PERSON american,NORP janet jackson-style,PERSON 62-y
india,GPE india,GPE iranian,NORP negar khan,PERSON the past two yea
first,ORDINAL 2004,DATE london s,ORG today,DATE claire armistead,PE

Fig. 4. Named Entity Recognition

Keyword Extraction: The process of selecting as well as extracting the most significant phrases or words from a piece of text is known as keyword extraction. Keyword extraction is widely utilised in natural language processing applications also including search engines, text categorization, and content recommender systems.

Text Summarization: It includes producing a shortened version of a longer piece of text while maintaining the important information and meaning. The purpose of text summary is to minimize the amount of information a user must read while also still providing the important points.

The interface was able to generate summaries of the provided text with high fluency and accuracy.

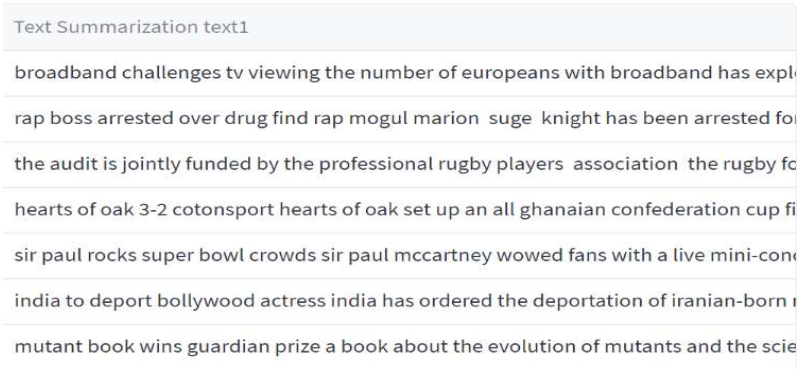


Fig. 5. Text Summarization

Overall, the results show that the streamlit-based web interface is an excellent tool for automating various NLP activities with high accuracy and efficiency. The interface can be enhanced and expanded to support additional NLP tasks and datasets.

Chapter 7

CONCLUSIONS AND FUTURE SCOPE

Conclusion

1. Natural language processing automation is an ongoing process that is becoming increasingly important as businesses strive to become more efficient and productive.By automating natural language processing tasks, individuals and organizations save time and resources while delivering accurate results.
2. Automation helps reduce the cost of human labor while improving the quality of results.
3. With a web-based interface, you can effectively use automation to streamline your natural language processing tasks and improve the accuracy of your results.

Future Scope

In the future, NLP automation will become increasingly important and ubiquitous. NLP automation will be used to generate and analyze large amounts of data quickly and accurately. It will also be used to build virtual assistants and chatbots to help people with their daily tasks. NLP automation will be used to create contextual understanding and natural language processing of text and speech, allowing for more powerful and accurate communication between people and machines. Furthermore, NLP automation will be used to create automated language translation services and to process large volumes of data in real-time. In the future, NLP automation will be used to create more intelligent and personalized customer service experiences and to better understand the needs and preferences of customers.

REFERENCES

[1] Manjusha Pandey Indrashis Das Siddharth S. Rautaray bharat sharma - "An Examination System Automation Using Natural Language Processing" - 2020

➔ https://www.researchgate.net/publication/339406698_An_Examination_System_Automation_Using_Natural_Language_Processing

[2] Ayisha Tabassum, Dr. Rajendra R. Patil - "A Survey on Text Pre-Processing & Feature Extraction Techniques in Natural Language Processing" - 2020

➔ https://www.irjet.net/archives/V7/i6/IRJET-V7I6913.pdf

[3] Eftim Zdravevski, Petre Lameski, Vladimir Trajkovik, Ivan Chorbev, Rossitza Goleva, Nuno Pombo & Nuno M. Garcia - "Automation in Systematic, Scoping and Rapid Reviews by an NLP Toolkit: A Case Study in Enhanced Living Environments" - 2019

➔ https://link.springer.com/chapter/10.1007/978-3-030-10752-9_1#Sec9

[4] Mohammad Hossein ,Hassan B. Kazemian, Karim Ouazzane, Chris Chandler - "Natural Language Processing approach to NLP Meta model automation" - 2018

➔ http://repository.londonmet.ac.uk/3447/1/Paper.pdf