**Logo

Description automatically generated**

**AMITY UNIVERSITY MUMBAI**

**NATURAL LANGUAGE PROCESSING LAB**

**FILE**

**SUBMITTED BY – PRATHMESH PATIL**

***Certificate***

**This is to certify that Mr. Prathmesh Patil is a Bonafede student of Amity Institute of Information Technology, at Amity University Maharashtra and he has done the LAB WORK on “Natural Language Processing” at Amity University Mumbai as prescribed by AIIT, AUM in partial fulfillment of the requirement of BCA Program for the academic year 2020-21.**

**Teacher Signature:**

**Natural Language Processing Lab**

**Index**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr.No** | **Name of the Practical** | **Date** | **Remarks** |
| 1 | Implementation of regular expression function | 22/07/2021 |  |
| 2 | Text Processing using nltk-Stemming(PorterStemmer, LancasterStemmer, SnowballStemmer) | 22/07/2021 |  |
| 3 | Tokenization using nltk(word, sentence tokenization) | 28/07/2021 |  |
| 4 | Stop words removal from text file and tokenize into word tokenizer | 28/07/2021 |  |
| 5 | Creating and displaying hindi stop words using nltk | 29/07/2021 |  |
| 6 | Text Nomalization-   1. WordNetLemmatizer, 2. synsets-similarity between two words | 29/07/2021 |  |
| 7 | Generating n-grams from text. | 29/07/2021 |  |
| 8 | Use of spacy library   1. Checking if the tokens of a text are alphabets with.is\_alpha(). 2. Accessing information about tokens through the attributes. 3. Printing all the tokens of a text with spacy library | 04/08/2021 |  |
| 9 | Displaying POS tags from text using nltk | 04/08/2021 |  |
| 10 | Displaying chunks from text using nltk | 04/08/2021 |  |
| 11 | Displaying frequency distributions of text and plotting it using matplotlib | 05/08/2021 |  |
| 12 | Generating wordcloud | 05/08/2021 |  |
| 13 | To build a machine learning model to classify whether a particular tweet is  hate speech or not,Text Classification-final\_dataset\_basicmlmodel.csv | 4/8/21 |  |
| 14 | Natural Language processing for Indian Languages-inltk, stanford,indicnlp library   1. Sentence and word tokenization 2. Finding similar sentence 3. Finding embedding vectors 4. Prediction of next word 5. Generating pos tags for hindi text | 4/8/21 |  |
| 15 | Implementation of term frequency and inverse term frequency(TF-IDF) and  Text “Features” | 4/8/21 |  |
| 16 | Sentiment analysis using textblob | 27/10/21 |  |
| 17 | Creating data frame from corpus | 27/10/21 |  |
| 18 | Web Scrapping with beautiful soup python library | 27/10/21 |  |
| 19 | Program to measure the similarity between two sentences using cosine similarity. |  |  |
| 20 | Creating model using genesim, Word2Vec and generating embedding vector | 10/11/21 |  |
| 21 | Using Displacy function to generate dependency parsing | 17/11/21 |  |
| 22 | Use of displacy function to highlight entities in the sentence. |  |  |
| 23 | 1. Text Summarization using Gensim with TextRank 2. Text summarization algorithm—sumy(LexRank summarizer, LsaSummarizer, LuhnSummarizer) | 24/11/21 |  |
| 24 | Transformers (DEEP Learning)   1. Text summarization using   Transformers   1. Text summarization using   Transformers T5   1. Tokenize a text using the `transformers` package 2. Create a Question-Answering system from given context 3. Text generation starting from a given piece of text 4. Classify a text as positive or negative sentiment with transformers. | 1/12/21 |  |