Mini Project Report on

NLP IN SOCIAL MEDIA

Submitted in partial fulfilment of the requirement for the award of the degree of

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE & ENGINEERING

Submitted by:

Student Name : KM Neha Andola **University Roll No. : 2018888**

Under the Mentorship of

Mr. Kireet Joshi
Professor

Department of Computer Science and Engineering
Graphic Era (Deemed to be University)
Dehradun, Uttarakhand
January 2023

Graphic Era
Deemed to be
University
DEHRADUN

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the project report entitled **"NATURAL LANGUAGE PROCESSING IN SOCIAL MEDIA"** in partial fulfilment of the requirements for the award of the Degree of Bachelor of Technology in Computer Science and Engineering of the Graphic Era (Deemed to be University), Dehradun shall be carried out by the under the mentorship of **Mr. Kireet Joshi, Professor,** Department of Computer Science and Engineering, Graphic Era (Deemed to be University), Dehradun.

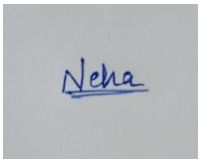
Name	University Roll No.	Signature
KM NEHA ANDOLA	2018888	

Table of Contents

CHAPTER 1

INTRODUCTION

Our society is revolutionized by Information and Communications Technology (ICT), and artificial intelligence is presently leading such a revolution, taking a central part suitable to remarkably impact the near future of humankind. thus, multitudinous researchers devoted to artificial intelligence raised questions Could a machine replace some people's functionalities and come a central axis for the coming generations in certain aspects of their lives? Different advances have been made regarding this question, and in this paper, we specifically review the capability of artificial intelligence to understand mortal language.

INTRODUCTION

Natural Language Processing (NLP) refers to the branch of computer wisdom, more specifically, the branch of Artificial Intelligence or AI concerned with giving computers the capability to understand text and spoken words in important the same way mortal beings can. In other words, Natural Language Processing strives to make machines that understand and respond to text or voice data and respond with text or speech of their own in important the same way humans do.

NLP has been for further than 50 times and has roots in the field of linguistics. It has a variety of real- world operations in several fields, including medical disquisition, quest machines and business intelligence. These technologies enable computers to exercise mortal language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or pen's intent and sentiment.

NLP drives computer programs that translate text from one language to another, respond to spoken commands, and epitomize large volumes of text swiftly in real time. There's a good chance you 've interacted with NLP in the form of voice- operated GPS systems, digital assistants, speech- to- text dictation software, and other consumer conveniences.

1.2 Working of NLP

Working of NLP enables computers to understand natural language as humans do. NLP uses artificial intelligence to take real- world input, process it, and make sense of it in a way a computer can understand. There are two main phases to NLP Data Pre- processing and Algorithm Development.

Data pre- processing involves preparing and" drawing" text data for machines to be suitable to assay it. Pre- processing puts data in workable form and highlights features in the text that an algorithm can work with.

- Tokenization- This is when text is broken down into lower units to work with.
- Stop word junking This is when common words are removed from text so unique words that offer the most information about the text remain.
- Lemmatization and stemming- This is when words are reduced to their root forms to exercise.
- Part- of- speech running - This is when words are pronounced predicated on the part- of speech they are analogous as nouns, verbs and adjectives.

Once the data has been pre- processed, an algorithm is developed to exercise it.

1.3 NLP in social media

Natural language processing (NLP) is continuously being used in the social media age. The reason behind this is because of the ever- growing number of social media channels, where users are posting, blogging, and editorializing on various motifs. By using natural language processing, you can easily prize information from the social media posts and thus, help your business in various ways. With its use, you can easily prize information and also use that information to needleworker your marketing campaigns specifically to them. This helps your business to reach the right target cult, who is more likely to convert. NLP can be used to break a wide range of problems, from helping people search for information to encapsulating text to detecting sentiment. It has witnessed an adding handover in the world of social media. Social media is a boon for the businesses as it connects them directly with the guests.

1.4 Sentiment Analysis

NLP allows various businesses to interact with the guests and avail the feedbacks and suggestions on their products and services. Sentiment analysis, also known as opinion mining, is the use of natural language processing (NLP), text analysis, and computational linguistics to identify and prize private information from source paraphernalia. The thing of sentiment analysis is to determine the stations, opinions, and passions of a speaker or pen with respect to some content or the overall contextual opposition of a document. Sentiment analysis is considerably applied to voice of the customer paraphernalia analogous as reviews and check responses, online and social media.

1.5 Natural Language Tool Kit (NLTK)

Natural language processing (NLP) is a field that focuses on making natural mortal language usable by computer programs. NLTK, or Natural Language Toolkit, is a Python package that you can use for NLP. NLTK Library is a suite that contains libraries and programs for statistical language processing. It's one of the most Important NLP libraries used in python, which contains packages to make machines understand mortal language and reply to it with an applicable response. It's designed for fast trial and allows for easy structure of modules.

1.6 Problem Statement

The main end of this design is to assay and determine the text (ex- customer reviews) in positive and negative terms. The problem is faced in getting to understand the various modules and the library functions of python and when to use those libraries.

CHAPTER 2

LITERATURE SURVEY

Check exertion that take place or are told as a result of opinions being made are told by opinions at the root position. Analysis of opinions or sentiment analysis plays a vital part in trying to make as close approximation as possible. The implicit to gather analogous information and assay the data can yield vital perceptivity into how products, services, political personalities, companies, governments, etc. are perceived and viewed. While there have been various styles that have been supposed and developed for sentiment analysis, there yet remains to be an effective approach in lodging and producing accurate sentiment analysis on a harmonious base.

Modern mortal societies are largely told by opinions in nearly all spheres and sphere of mortal civilization. Sentiment Analysis of opinions can, therefore, be viewed as an integral tool to assay the mood and current disposition of any sample group of individualities concerning any product, service, event or content expressed in text form and published on social media platforms, blog posts, commentary, web reviews, etc. The idea of Sentiment Analysis has been, therefore, to table a summary of opinions isolated into positive, neutral and negative reviews predicated on analyses of handbooks posted by users.

To prize sentiment, SA uses three terms. These are object and point, opinion holder, and opinion and exposure. The task of bearing SA is not easy with several technical hurdles that can be thrown in the way. These hurdles can range from relating objects, opinion exposure type through to point birth. SA is also known as opinion birth, opinion mining, effect analysis, sentiment mining, emotion analysis, review mining, etc. It has set up global acceptance as an important tool in analysing data to help predict election results, cast stock request positions of businesses and help estimate deals of products across different request conditions. In conducting SA, it's constantly assumed that the data analysed would have opinions expressed explicitly. In a disquisition, Tripathyet.al. proposed four ML algorithms for sentiment type. These were Naïve Bayes, Maximum Entropy, Stochastic Gradient Descent and Support Vector Machine. Their disquisition demonstrated that delicacy can be

achieved through progressive type. Planting an n- gram approach, the authors were suitable to produce harmonious high delicacy using a combination of TF- IDF (Term frequency – Inverse document frequency) and count vectorize fashion.

To fight the problem of words and punctuation symbols that although were accessible to humans, had no formal description in the English dictionary, the authors developed a new list of analogous words to help in SA. Lemmatizers, stemmers, n- grams, negations etc were explored and presented in a detailed study. The researchers go on to compare these models with alternate type ways and grade them on the lines of perfection, recall, f- measure, performance and delicacy. The researchers used Naïve Bayes, Support Vector Machine and K- Nearest Neighbour to conduct comparison tests. Their trials revealed that while SVM was most useful in turning in topmost perfection, it was KNN that achieved topmost recall constantly.

- Naïve Bayes- Naive Bayes predict the marker of a text. They calculate the probability of each marker for a given text and also affair the marker with the topmost one.
- Support Vector Machine- Support vector machine (SVM) is a supervised machine knowledge model that uses type algorithms for two- group type problems. After giving an SVM model sets of labelled training data for each order, they're suitable to classify new text.
- K- Nearest Neighbour- The k- nearest neighbours' algorithm, also known as KNN or k- NN, is a non- parametric, supervised knowledge classifier, which uses contiguity to make groups or prognostications about the grouping of an individual data point.

CHAPTER 3

METHODOLOGY

Verbal analysis is an important part of semantic analysis. verbal semantics is the study of the meaning of any word. In semantic analysis, the relation between verbal particulars is linked. Some of the relations are hyponyms, antonyms, Synonyms, Homonyms etc. Through relating these relations and considering different symbols and punctuations, the machine is suitable to identify the environment of any judgment or paragraph. The methodology used to train a language model using a textbook train and assay the true sentiment behind the given input expression. The model is trained using the NLTK library in Python, which is one of the most important NLP libraries used in python, which contains packages to make machines understand mortal language and reply to it with an applicable response. The model is created using the Sentiment Intensity Analyser from the NLTK library's models' module. It helps us to assay the emotion of the textbook. It helps us find the opposition of textbook, which is positive, negative or neutral.

The first step is to read the input given by the user. Sentiment Analyser is also initialized. It's also reused using Sentiment Intensity Analyser. It's done through Two Processes. The following are two processes of Semantic Analysis

1. Word Sense disambiguation-

   It's an automatic process of relating the environment of any word, in which it's used in the judgment. In natural language, one word can have numerous meanings.

   For partner- The word ‘light’ could be meant as not veritably dark or not veritably heavy. The computer must understand the entire judgment and pick up the meaning that fits the stylish. This is done by word sense disambiguation.

2. Relationship birth –

   In a judgment, there are a many realities that are co-related to each other. Relationship birth is the process of rooting the semantic relationship between these realities.

   For partner- In a judgment, “I’m learning mathematics”, there are two realities, ‘I ’ and ‘ mathematics ’ and the relation between them is understood by the word ‘ learn ’. After recycling the data, the score of the analysis is stored in a variable which is published using print statement.

CHAPTER 4

RESULTS AND DISCUSSIONS

The model is created using the programming language which is Python. It's a popular, high-level programming language known for its clear syntax, dynamic semantics, and support for object- acquainted programming. It’s erected in data structures and support for dynamic typing and binding make it a seductive choice for rapid-fire operation development and connecting being factors. This design makes use of NLTK (Natural Language Toolkit) Library of python which is substantially used in any operation of NLP (Natural Language Processing) to prop in its functionality. Sentiment Intensity Analyser is used to break the textbook in commemoratives and pre-process it using two ways. They're Word Sense Disambiguation and Relationship birth.

Word Sense Disambiguation is used to remove ambiguation that can arise from a word having two or further meaning to it. For partner- “Light” can have two meanings – not veritably heavy and the other as not veritably dark.

Relationship birth is also used to prize or find relation between two realities.

The Sentiment Analysis is initialized. The score (tells us about negative, positive and neutral points and status) is also calculated by passing the input textbook from user in polarity_scores which is called with the help of sentiment analysis analyser.

A simple demonstration of the design is shown in the following runner. Then we can see that how the introductory perpetration of the design is done.

INPUT 1: This is an amazing product.

OUTPUT 1:

```
PS C:\Users\dell\Desktop\mini-project> & C:/Users/dell/AppData/Local/Programs/Python/Python311/python.exe c:/Users/dell/Desktop/mini-p
/sentimentanalysis.py
Input a text (string) from user for sentiment analysis : This is an amazing product.
Polarity Scores are:

{'neg': 0.0, 'neu': 0.513, 'pos': 0.487, 'compound': 0.5859}
```

INPUT 2: The new law is not a good option.

OUTPUT 2:

```
PS C:\Users\dell\Desktop\mini-project> & C:/Users/dell/AppData/Local/Programs/Python/Python311/python.exe c:/Users/dell/Desktop/mini-p
/sentimentanalysis.py
Input a text (string) from user for sentiment analysis : The new law is not a good option.
Polarity Scores are:

{'neg': 0.286, 'neu': 0.714, 'pos': 0.0, 'compound': -0.3412}
```

CHAPTER 5

CONCLUSION AND FUTURE WORK

We've created a model in which take the input from the user and assay the emotion or true sentiment behind it after precessing the input which we've taken. The limitations of the proposed model have been linked as how important emulsion textbook is taken into account.

The ensuing areas of enhancement will be concentrated on in future in order to enhance the capabilities of the coming word vaticination model

- Increase in delicacy through the disquisition of new ways and styles.
- Expansion of the training dataset for the model to learn further patterns and ameliorate performance.
- Testing on different types of textbooks.

Operations for SA are wide- ranging. SA can be used in cases involving politics, product reviews, cinema reviews, social media posts and much further. In a broader and further global environment, SA has been thrust into spotlight because of the eventuality to gauge more directly and considerably data that can be gathered from social media platforms., ex- Twitter, Facebook, LinkedIn, etc.

In any client centric business, it's veritably important for the companies to learn about their guests and gather perceptivity of the client feedback, for enhancement and furnishing better user experience.

With the help of machine literacy models and semantic analysis, machines can fluently prize meaning from unshaped data gathered from their client base in real time. It helps the company get accurate feedback that drives better decision- timber and as a result improves the client base.

Social media monitoring and analysing of the nonstop inflow of user- generated content can be used as a fresh dimension which contains precious information that would not have been available from traditional media and journals. In addition, we mentioned the challenges with social media data, which are due to their large size, and to their noisy, dynamic, and unshaped nature.

SUMMARY

Sentiment Analysis (SA) is a fashion of Natural Language Processing (NLP) which uses the Natural Language Toolkit(NLTK) Library to assay the sentiment behind a given textbook or judgment . This technology has a wide range of implicit uses as mentioned over.

Sentiment Analysis is now by and large considered to be critical in terms of socio-profitable viewpoint. Understanding SA and examining the styles that can help achieve delicacy in a wide variety of input formats is critical for businesses, institutions a, institutions and individuals to survive and succeed.

REFERENCES

[1] GOOGLE

[2] YOUTUBE

[3] UDEMY

[4] NATURAL LANGUAGE PROCESSING FOR SOCIAL MEDIA, SECOND EDITION

GRAEME HIRST

[5] RECENT TRENDS IN DEEP LEARNING BASED NATURAL LANGUAGE PROCESSING REVIEW ARTICLE

[6] https://www.winklix.com/blog/what-is-natural-language-processing-nlp/

[7] https://realpython.com/nltk-nlp-python/

[8] https://realpython.com/nltk-nlp-python/

[9] https://medium.com/geekculture/natural-language-processing-in-the-social-media-age-23e1e10235bd

[10] https://www.analyticssteps.com/blogs/semantic-analysis-working-and-techniques