INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.165

A Survey on Twitter Sentimental Analysis Using Machine Learning

Yashraj Ghodke, Abhishek Katte, Shruti Kokane, Prof.Mayuri Vengurlekar

Student, Dept. of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Student, Dept. of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Student, Dept. of Computer Engineering, K J College of Engineering and Management Research, Pune, India

Assistant Professor, Dept. of Computer Engineering, K J College of Engineering and Management Research, Pune, India

ABSTRACT: In this age of exponentially evolving technology, the emergence of social media has captured all life forms on Earth. Twitter is a social media platform where millions of users tweet their opinions and feelings on various topics every day. Sentiment analysis is a technique for analysing source material, determining its sentiment, and classifying it as favourable, negative, or fair. Twitter Sentiment Analysis is the use of multiple libraries to collect data from the Twitter API about a topic, evaluate it, and derive public and private opinion about it. Using opinion mining to validate your online reputation and maximize profits is critical for businesses. For this study, we took a dataset named Sentiment140 and pre-processed the data to transform it from unstructured to structured format, perform feature selection on the dataset, and train and score various machine learning models. Ring, compare, and aim to use the classifier that achieves the best accuracy results. Vectorization by a pipeline that identifies the sentiment of new tweets given as input.
.

KEYWORDS: Sentiment Analysis, Twitter.

I. INTRODUCTION

Python is an open source, interpreted, high-level language that offers an excellent approach to object-oriented programming. It is one of the best languages used by data scientists in various data science projects/applications. Python offers excellent facilities for processing mathematics, statistics, and scientific functions. It provides an excellent library for handling data science applications. One of the main reasons Python is so widely used in the scientific and research community is its ease of use and simple syntax. Also suitable for rapid prototyping.

According to academic and industrial engineers, the deep learning frameworks available in the Python API, in addition to the scientific packages, have made Python incredibly productive and versatile. The deep learning Python framework has evolved significantly and will be updated soon.

When it comes to application areas, ML scientists also prefer Python. When it comes to areas like building fraud detection algorithms and network security, developers gravitate to Java, while developers of applications like natural language processing (NLP) and sentiment analysis chose Python. Problem Solving helps you solve business problems simply by building powerful systems and data applications.

II. RELATED WORK

1. **User Interface: -**Web Interface based on 'Sentimental Analysis of Twitter data Using Machine Learning Algorithms.

2. **Hardware Interfaces**: -
   RAM: 8 GB RAM minimum required is 8 GB.
   Processor: Intel i5 Processor
   Operating System: Windows 10 /11 Latest Operating System that supports all type of installation and development Environment.

3. **Software Interfaces Operating System**: - Windows 10/11.

4. **IDE**: Jupiter Notebook Programming Language: Python Modules – Numpy, Pandas, Matplotlib, Sklearn, Wordnet, Stop words, punkt, word2vec, genism, keras, tensorflow, tokenizer.

5. **Performance Requirements**: - The performance of the functions and every module must be well. The overall performance of the software will enable the users to work efficiently. All the python libraries and modules need to run error free.

6. **Safety Requirement**: - The dataset which we are using in this project does not include any private data or any user's personal identity.

III. PROJECT WORK

A. Web Based View:-

There Will be website for replication of sentimental analysis of Twitter Data. Website is going to be developed in python with Flask web framework with all the required data analysis python libraries.

B. Sentimental Analysis: -

Basically, with Twitter will be the host Data Source for our Project. On website we will have section to add text to search on twitter. With this searched data using sentimental data analysis technique the graphs with result as Positive, negative, neutral, etc. The Sentimental analysis itself say's something related to sentiments so, sentiment analysis is to classify the polarity of a given text at the document, sentence, or functional/aspect level. Whether the opinion expressed in the document, sentence or feature/aspect is positive, negative or neutral.

C. Website Deployment: -

We are going to use Streamlit for website live deployment. Streamlit is a free and open-source framework to rapidly build and share beautiful machine learning and data science web apps.

IV. FLOW DIAGRAM

Pre-processing

↓

Create word2vec

↓

Tokenize and Padding

↓

Model Creation

↓

Testing
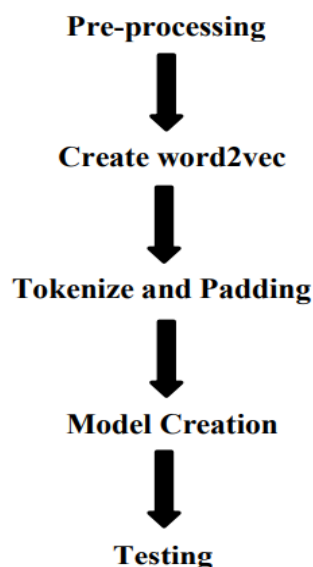
1. **Pre-processing: -** Whenever you have text data, you need to apply some pre-processing steps to the data to convert words into numerical features that machine learning algorithms can work with. Not all steps need to be applied to every problem, as the pre-treatment steps for the problem depend largely on the domain and the problem itself.

2. **Create Word2Vec: -** The purpose and utility of word2vec is to group vectors of similar words in vector space. That is, it recognizes similarity mathematically. Word2vec creates a vector that is a distributed numerical representation of word features such as individual word context.

3. **Tokenize and Padding: -** Applying tokenization is the first step in converting words into numbers that can be processed by machine learning models. Once the set is tokenized, TensorFlow returns the number associated with each token.

V. CONCLUSION AND FUTURE WORK

Twitter sentiment analysis falls into the category of text and opinion mining. It focuses on analyzing the sentiment of tweets, feeding the data to train a machine learning model, checking its accuracy, and depending on the results, making this model available for future use. This includes steps such as data collection, text preprocessing, sentiment detection, sentiment classification, training, and model testing. This research topic has evolved over the past decade, with models reaching efficiencies of nearly 85% to 90%. However, the dimension of data diversity is still missing. Also, the slang and contracted words used have many usage problems. Many analyzers do not perform well when the number of classes increases. We have also not yet tested how accurate the model is for topics other than those considered. Therefore, sentiment analysis has great potential for future development

REFERENCES

1. Abdullah Ikram, Mohit Kumar, GeetikaMunjal."Twitter Sentiment Analysis using Machine Learning ", IEEE 2022.
2. Sai Madhu K, Chakradhar Reddy B. "Real Time Sentimental Analysis on Twitter", 2021.
3. Nithyashree T, Nirmala M.B "Analysis of the Data from the Twitter account using Machine Learning", 2020.
4. Hayder A. Alatabi, Ayad R. Abbas "Sentiment Analysis in Social Media using Machine Learning Techniques", Iraqi Journal of Science. 2019.
5. Megha Rathi, Aditya Malik, Daksh Varshney, Rachita Sharma, Sarthak Mendiratta."Sentiment Analysis of Tweets using Machine Learning Approach", 2018.
6. Jaspreet Kaur, Brahmaleen Kaur Sidhu. "Sentiment Analysis Based on Deep Learning Approaches", 2018.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

📱 9940 572 462 💬 6381 907 438 ✉️ ijircce@gmail.com

Scan to save the contact details