ISSN No:-2456-2165

Sentimental Analysis using NLP

Terisri Paladugula[1]
B.Tech III Year
Department of AIM
Sri Vasavi Engg. College
Tadepalligudem

Hiranmayee Nandyala[2]
Sr. Asst Prof
Department of CSE,
Sri Vasavi Engg. College,
Tadepalligudem

S V V S S C Ekantha[3]
B. Tech III Year,
Department of AIM,
Sri Vasavi Engg. College,
Tadepalligudem

Puthin Dungala[4]
B. Tech III Year,
Department of AIM
Sri Vasavi Engg. College,
Tadepalligudem

Karteek Kishor Ambati[5]
B. Tech III Year,
Department of AIM
Sri Vasavi Engg. College,
Tadepalligudem.

Jyothi Tanmai Ramisetti[6]
B. Tech III Year,
Department of AIM,
Sri Vasavi Engg. College,
Tadepalligudem.

Abstract:- Sentiment analysis is a subset of text analysis techniques that uses automatic text polarity detection. One of the main responsibilities of NLP (Natural Language Processing) is sentiment analysis, often known as opinion mining. In recent years, sentiment analysis has gained a lot of popularity. It is meant for people to build a system that can recognize and categorize sentiment or opinion as it is expressed in an electronic text. Nowadays, people who wish to purchase consumer goods prefer to read user reviews and participate in public online forums where others discuss the product. This is because consumers frequently have to make trade-offs when making purchases. Before making a purchase, a lot of customers read other people's reviews. Individuals frequently voice their opinions about several things. Opinion mining has grown in significance as a result. Sentiment analysis is the process of determining if the expressed opinion about the subject is favorable or negative. Customers must choose which portion of the available data to utilize. Sentiment analysis is the technique of locating and removing subjective information from unprocessed data. If we could accurately forecast sentiments, we could be able to gather online opinions and anticipate the preferences of online customers. This information could be useful for study in marketing or economics. As of right now, sentiment classification, feature-based classification, and handling negations are the three main issues facing this research community.

Keywords:- *Numpy, Pandas, TF-IDF, Tfidf Vectorizer, Linear SVC, Train-Test Split, Accuracy Score, Classification Report, Confusion Matrix, user Input, Vectorization, Prediction, Preprocessing, Text Classification, Supervised Learning, Machine Learning Model, Scikit-Learn.*

I. INTRODUCTION

Sentiment analysis, sometimes referred to as opinion mining, is the process of locating and extracting subjective information from source materials using computational linguistics, natural language processing, and text analysis. Put another way, sentiment analysis is a kind of natural language processing that monitors public opinion toward a specific good or subject. Its main responsibility is to recognize and extract sentiment from a given string. It accepts a string as input and returns a sentiment rating between very negative and very positive ([-1 t0 1]). It entails developing a system to gather and analyze product-related feedback expressed in tweets, blog posts, reviews, and comments. Sentiment analysis has various applications. For instance, in marketing, it assists in determining whether an advertisement campaign or traditional text mining focuses on the examination of facts, while sentiment analysis focuses on attitudes. Sentiment analysis research primarily focuses on three primary areas: opinion summarization, feature-based sentiment classification, and sentiment classification. Sentiment classification is the process of categorizing entire writings based on the thoughts of the authors regarding particular things. On the other hand, feature-based sentiment classification takes into account users' judgments about certain object features. The task of summarizing opinions differs from standard text summarization in that it focuses solely on the product attributes that have prompted consumer evaluations. Unlike the standard text summary, opinion summarization does not select a selection of the reviews or rework some of the original lines from the reviews to capture the essential points. Sentiment analysis determines the tone of a given text by applying a variety of classification algorithms. It makes clear if the text is neutral, negative, or favorable. Large data sets can be used with this technique to produce information that is useful in a variety of situations. For instance, the judgment expressed in the line "The life of the battery of this mobile is too compressed" is negative and focuses on the "life of the battery" of the mobile object (target). This level of in-depth analysis is necessary for many everyday applications because it is necessary to know which aspects of a product are liked and disliked by customers to upgrade it. Subjectivity and sentiment have not been used to categorize such information. Linguistics is the study of human (natural) language interaction and computer science, artificial intelligence, and natural language processing (NLP). For example, it identifies the product features that the reviewer has commented on and indicates whether the comments are neutral, positive, or negative.

ISSN No:-2456-2165

Phrase-based SA involves classifying the phrases in a sentence based on their polarity. Determining whether a text expresses positive or negative weather views is part of the emotion analysis task.

➢ *Objective:*

The goal of our task is to create a model that predicts the polarity of the given text i.e. whether the given text is positive, negative, or neutral.

II. RELATED WORK

➢ *TF-IDF Vectorizer:* It is a numeric that reflects the common words that are used in Natural Language Processing.

- *Term Frequency (TF):* It tells the no. of times a particular term is repeated in the given document.
- *It is Calculated as below:*

$$TF(t, d) = \frac{\text{Number of times term t apperars in document d}}{\text{Total number of terms in document d}}$$

- *Inverse Document Frequency (IDF):* It measures the importance of a word in the document.
- *It is Calculated as below:*
- IDF(t, D) = log (total no. of document in corpus D/ no. of documents containing terms t+1)

➢ *TF-IDF Score:* It is the multiplication of TF and IDF. The higher the value higher the importance of the word.

- *It is Calculated as below:*

$$TF\text{-}IDF(t,d,D)=TF(t,d)\times IDF(t,D)$$

➢ *Advantages of TF-IDF:*

- Term Importance
- Common Word Penalization
- Dimensionality Reduction
- Document Similarity
- Language Independence
- Information Retrieval

III. CONFUSION MATRIX

Confusion matrix is a totally popular measure used while fixing type problems. It could be applied to binary classification as well as to multiclass-type problems.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Fig 1 Confusion Matrix

Confusion matrices represent counts from anticipated and actual values. The output "TN" stands for proper poor which indicates the variety of bad examples categorized appropriately. Further, "TP" stands for actual effectiveness which indicates the number of high-quality examples labeled accurately. The period "FP" suggests a fake fantastic price, i.e., the wide variety of actual bad examples classified as tremendous; and "FN" approaches a false bad fee which is the variety of actual effective examples classified as terrible. Performance metrics of a set of rules are accuracy, precision, recollect, and F1 rating, which are calculated based on the above-stated TP, TN, FP, and FN.

A. Accuracy:

The accuracy of an algorithm is represented as the ratio of correctly classified patients to thetotal number of patients.

➢ *Accuracy=TN+TP / TN+FP+FN+TP*

- *Precision:*

The precision of an algorithm is represented as the ratio of correctly classified patients with the disease to the total patients predicted to have the disease.

➢ *Precision=TP / TP+FP*

- *Recall:*

Recall metric is defined as the ratio of correctly classified diseased patients (*TP*) divided by the total number of patients who have the disease. The perception behind recalls how many patients have been classified as having the disease. The recall is also called sensitivity.

➢ *Recall=TP / TP+FN*

- *F1 Score:*

The F1 score is also known as the F-measure. The F1 score states the equilibrium between the precision and the recall.

➢ *F1 score=2*precision*recall / precision + recall*

IV. PROPOSED SYSTEM

Sentiment analysis is a natural language processing technique that helps identify the neutral, positive, and negative aspects of data. It is frequently used on textual data to assist companies in tracking customer sentiment toward their brands and products in feedback from customers and in figuring out what those needs are.

There are various ways to determine a text's sentiment score. Using positive and negative word counts with normalization is one such technique. Utilizing Support Vector Machines (SVM) is an additional technique. The SVM model predicts text emotions and ascertains if the text is neutral, positive, or negative.

ISSN No:-2456-2165

➢ *The following Steps could be Included in a Suggested Sentiment Analysis System:*

- *Data Collection:*

Gathering information from a range of sources, including customer feedback forms and social media platforms.

- *Data Preprocessing:*

The removal of stop words, punctuation, special characters, etc. from the collected data.

- *Feature Extraction:*

Taking characteristics like word frequency and n-grams out of the preprocessed data.

- *Model Training:*

Using the features that were extracted, train a machine learning model to categorize the text into positive, negative, or neutral categories.

- *Model Evaluation:*

Assessing the trained model's effectiveness with metrics like recall, accuracy, and precision, among others.

- *Model Deployment:*

Introducing the learned model into a real-world setting so it can be applied to the classification of fresh data.

V. DESIGN

The process of classifying sentiment using machine learning algorithms is illustrated in the flowchart. Six steps make up the process: gathering data, preprocessing the data, machine learning algorithm, results presentation, sentiment classification, and algorithm.

To explain the procedure in more detail, first data is gathered from multiple sources. The gathered data is then preprocessed to eliminate any unnecessary information and transform it into a format that machine learning algorithms can understand. Subsequently, a machine learning algorithm is employed on the preprocessed data to train a model capable of categorizing text data sentiment. Next, the sentiment of fresh text data is classified using the trained model. The format in which the results are presented makes them simple to read and analyze. Using the trained model, the SVM algorithm is then used to categorize the sentiment of fresh text data.

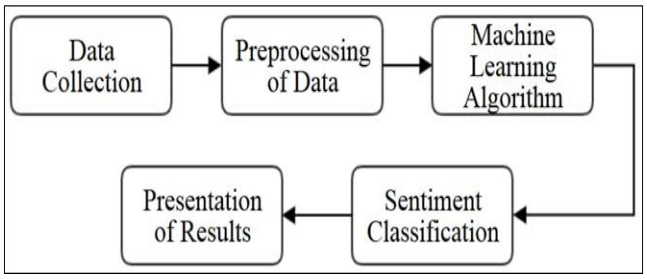

Fig 2 System Archiecture

VI. RESULTS AND DISCUSSIONS

➢ *Get the Dataset:*

The dataset we took contains two columns sentence and sentiment. The sentence contains the text and the sentiment contains the polarity of the sentence.

➢ *Importing Libraries:*

To perform data preprocessing using Python, we need to import some predefined Python libraries. These libraries are used to perform some specific jobs. There are three specific libraries that we will use for data preprocessing, which are: They are two main important libraries which are used, they are:

- *Pandas:*

The last library is the Pandas library, which is one of the most famous Python libraries and is used for importing and managing the datasets. It is an open-source data manipulation andanalysis library. It will be imported as below:

- *Matplotlib:*

The second library is matplotlib, which is a Python 2D plotting library, and with this library, we need to import a sub-library pyplot. This library is used to plot any type of charts in Python for the code. It will be imported as below

✓ Import Pandas as pd
✓ Import Matplotlib.Pyplot as plt
✓ From Sklearn Import *

➢ *Importing Data:*

The downloaded data is imported into the Python code file as a Data frame using pandas module.

```
data = pd.read_csv("data.csv")
```

Fig 3 Importing File

➢ *Feature Extraction:*

In this step, we convert text data to numerical features using TF-IDF vectorization.

```
vectorizer = TfidfVectorizer()
X = vectorizer.fit_transform(data['Sentence'])
y = data['Sentiment']
```

Fig 4 Converting Text to Numericals

- *Split Data:*
  In this step, we are going to split the data into x_train, x_test, y_train, and y_test.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=1)
```

Fig 5 Spliting Data

- *Train the Linear SVC Model:*
  In this step, we train the model.

```
classifier = LinearSVC()
classifier.fit(X_train, y_train)
```

Fig 6 Training

- *Make Predictions:*

```
predictions = classifier.predict(X_test)
```

Fig 7 Predictions

- *Evaluate the Model:*

```
accuracy = accuracy_score(y_test, predictions)
report = classification_report(y_test, predictions)
confusion = confusion_matrix(y_test, predictions)
```

Fig 8 Evaluation

- *Print the Confusion Matrix and Accuracy:*

```
print(f"Accuracy: {accuracy}")
print(report)
print(confusion)
```

Fig 9 Accuracy

The result of the sentimental analysis tells the polarity of the text given. It ultimately takes text as an input and tells the polarity of it. It helps the business people to know how the customers are feeling about their products so that they can make changes accordingly as per the customer's interest. It helps in making changes easily and helps in improving the quality of the product. It not only products but also helps movies understand the pulse of the audience. It makes the directors understand the willingness and type of genre the audience wants and make films accordingly. Not only movies and products sentimental analysis helps in many sectors.

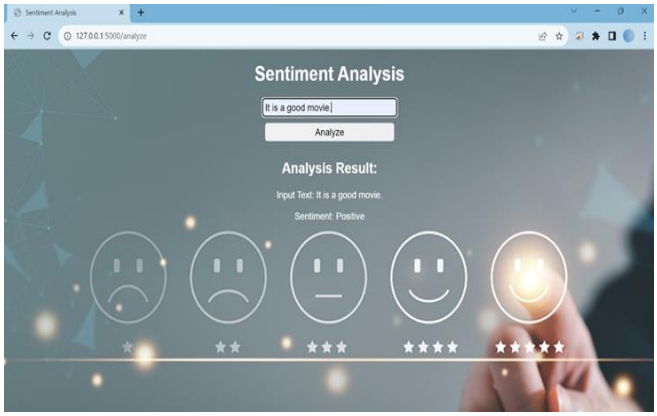

Fig 10 Positive Response

ISSN No:-2456-2165

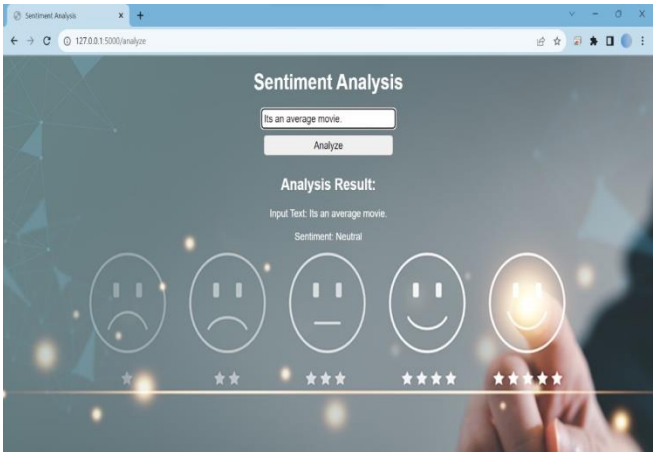

Fig 11 Average Response

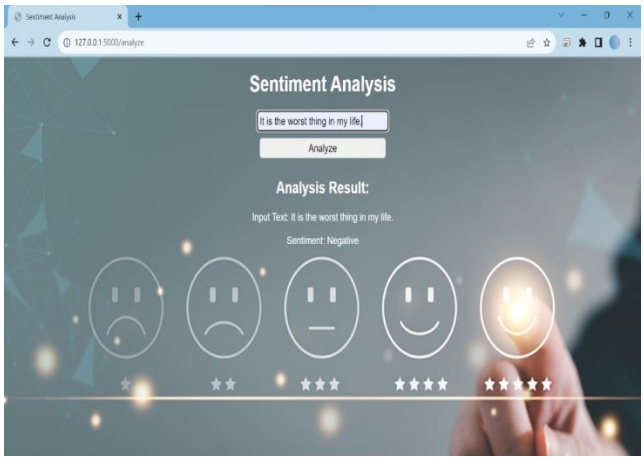

Fig 12 Negative Response

VII. FUTURE SCOPE

- Multimodal Sentiment Analysis Identifying sentiments across various media formats by analyzing emotions in text, photos, audio, and video.
- The Aspect-Based Sentiment Analysis method identifies sentiments associated with discrete elements in text data by concentrating on particular aspects or features.
- Sentiment expressions within textual content can be used to identify a broad spectrum of emotions and minute differences in sentiment.
- Sentiment analysis in healthcare refers to the process of examining patient feedback and healthcare-related content to ascertain attitudes regarding mental health and healthcare services.
- Sentiment Analysis in Conversational AI: Adding sentiment analysis skills to chatbots and virtual assistants will help them comprehend user emotions and react to them sympathetically.

REFERENCES

[1]. In 2019, IEEE Access published a paper titled "Twitter Sentiment Analysis Based on Ordinal Regression," written by S. E. Saad and J. Yang.
[2]. Multi-Strategy Sentiment Analysis of Consumer Reviews Based on Semantic Fuzziness, Y. Fang, H. Tan, and J. Zhang, IEEE Access, vol. 6, pp. 20625-20631, 2018.
[3]. The article "Tourism Mobile App With Aspect-Based Sentiment Classification Framework for Tourist Reviews" was published in May 2019 in IEEE Transactions on Consumer Electronics, volume 65, issue 2, pages 233-242, by M. Afzaal, M. Usman, and A. Fong.
[4]. "Halal Products on Twitter: Data Extraction and Sentiment Analysis Using Stack of Deep Learning Algorithms," A. Feizollah, S. Ainin, N. B. Anuar, N. A. B. Abdullah, and M. Hazim, IEEE Access, vol. 7, pp. 83354–83362, 2019.
[5]. A study titled "An efficient hybrid filter and evolutionary wrapper approach for sentiment analysis of various topics on Twitter" was published in Knowledge-Based Systems on March 15, 2020. It was completed by Mohammad A. Hassonah, Ibrahim Aljarah, Hossam Faris, Rizik Al-Sayyed, Ali Radan, Ala' M. Al-Zoubi, and Almaraz.