

Ex.No.13
Date: 19.04.23

Mini Project
Twitter Sentiment Analysis using VADER module from NLTK toolkit

Problem Statement:

Social media platforms like Twitter generate vast amounts of user-generated content that can provide valuable insights into public sentiment and opinion on various topics. However, the unstructured nature of this data can make it challenging to extract useful information. Can sentiment analysis of tweets using natural language processing techniques provide insights into people's attitudes and opinions on a particular topic, and help businesses, organizations, and policymakers make better decisions?

Aim:

To perform sentiment analysis of tweets using the VADER module from NLTK toolkit.

Introduction

A business or a brand's success depends solely on customer satisfaction. Suppose, if the customer does not like the product, you may have to work on the product to make it more efficient. So, for you to identify this, you will be required to analyze the sentiment of their opinions. Therefore, Sentiment analysis is basically defined as the process of identifying and categorizing opinions from a piece of text, thereby determining whether the opinion is positive or negative.

In this article, we will work on how to perform sentiment analysis using VADER. Sentiment analysis gives meaning to the text. Semantics helps us to interpret symbols, their types, and their relation with each other. Let us briefly understand what NLP actually is and also about the NLTK Library.

What is NLP?

NLP is an automatic way of manipulating or processing human language. We use NLP to extract meaningful data from textual data. There are various applications of NLP such as Sentiment Analysis, Chatbot, Speech Recognition, Machine Translation, spell checking, Information Extraction, Keyword search, Advertisement matching, etc. Some real-world examples are Google Assistant and Google translate.

Software & Hardware Requirements: Python IDE(Google Colab),nlk toolkit, VADER module.

NLTK

Natural Language Toolkit (NLTK) is one of the most powerful NLP libraries which contains packages to make machines understand human language and reply to them in an appropriate desired response. NLTK has many built-in packages to process our textual data at every stage. There are various stages in nltk processing such as data cleaning, visualization, vectorization, etc.

Sentiment Analysis

Sentiment analysis is used to find out the polarity of the text, which is positive, negative, or neutral. It is one of the advanced research areas in natural language processing. This is widely used in data mining, text mining, etc. It helps collect and analyze opinions about a brand or a product by processing blog posts, comments, reviews, tweets, etc.

Applications of Sentiment Analysis

- **Social media monitoring:** As we all know, social media is taking over the world. More than 55% of customers share their reviews about purchases socially on many social networking sites. It's almost difficult to analyze the reviews manually. Sentiment analysis lets us analyze and derive meaning from them.
- **Brand monitoring:** Brand owners use sentiment analysis tools to keep track of the bad reviews about their brand. They can also use machine learning algorithms to predict outcomes based on the results derived using semantic analysis.
- **Voice of customer:** Various sentiment analysis algorithms let us analyze the voice of the customers, such as the product that are most needed by the customers and also the products that are highly rated, etc. The brand owners can create a personalized customer experience based on these evaluations.
- **Customer service:** Chatbots are a widespread way of delivering good customer service. Using sentiment analysis, you can transfer the chat to a customer service associate whenever needed. Also, you can automate the tasks such as booking a ticket, an appointment for a salon, etc.
- **Market research:** Using sentiment analysis, you can research how well your competitors are growing and what are their positive feedbacks from the customers. You can also analyze the way they deal with their customers. You can, in turn, work on the issues related to your product's failure.
- **Product Analysis:** You can do keyword research to identify the products in demand and the highly rated products. You can also determine what features of a particular product are highly appreciated by the customers or the end users.

NLTK's VADER module

VADER(Valence Aware Dictionary for Sentiment Reasoning) is an NLTK module that provides sentiment scores based on the words used. It is a rule-based sentiment analyzer in which the terms are generally labeled as per their semantic orientation as either positive or negative.

SOURCE CODE:

```
import pandas as pd
import numpy as np
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
import re
import nltk

data = pd.read_csv('newyears.csv',on_bad_lines='skip',encoding = 'unicode_escape')
print(data.head())
nltk.download('stopwords')
stemmer = nltk.SnowballStemmer("english")
from nltk.corpus import stopwords
import string
stopword=set(stopwords.words('english'))

def clean(text):
    text = str(text).lower()
    text = re.sub('[.*?]', "", text)
    text = re.sub('https?://\S+|www\.\S+', "", text)
    text = re.sub('<.*?>+', "", text)
    text = re.sub('[%s]' % re.escape(string.punctuation), "", text)
    text = re.sub('\n', "", text)
    text = re.sub('\w*\d\w*', "", text)
    text = [word for word in text.split(' ') if word not in stopword]
    text="" ".join(text)
    text = [stemmer.stem(word) for word in text.split(' ')]
    text="" ".join(text)
    return text
data["tweet"] = data["tweet"].apply(clean)
from nltk.sentiment.vader import SentimentIntensityAnalyzer
nltk.download('vader_lexicon')
sentiments = SentimentIntensityAnalyzer()
```

```

data["Positive"] = [sentiments.polarity_scores(i)["pos"] for i in data["tweet"]]
data["Negative"] = [sentiments.polarity_scores(i)["neg"] for i in data["tweet"]]
data["Neutral"] = [sentiments.polarity_scores(i)["neu"] for i in data["tweet"]]
data = data[["tweet", "Positive",
             "Negative", "Neutral"]]
print(data.head())
x = sum(data["Positive"])
y = sum(data["Negative"])
z = sum(data["Neutral"])

def sentiment_score(a, b, c):
    if (a>b) and (a>c):
        print("Positive 😊 ")
    elif (b>a) and (b>c):
        print("Negative 😞 ")
    else:
        print("Neutral 😐 ")
sentiment_score(x, y, z)
print("Positive: ", x)
print("Negative: ", y)
print("Neutral: ", z)

```

OUTPUT:

```

other_topic \
0    Read moore books, read less facebook.
1                                NaN
2                                NaN
3  Help More\nspread pet therapy|helping other
4                                NaN

resolution_topics  gender      name \
0      Eat healthier  female  Dena_Marina
1  Humor about Personal Growth and Interests Reso...  female  ninjagirl325
2      Be More Confident  male  RickyDelReyy
3      Other  male  CalmareNJ
4      Be more positive  female  welovatoyoudemi

Resolution_Category  retweet_count \
0  Health & Fitness      0.0

```

1	Humor	1.0
2	Personal Growth	0.0
3	Philanthropic	0.0
4	Personal Growth	0.0

	tweet	tweet_coord \
0	#NewYearsResolution :: Read more books, No scr...	NaN
1	#NewYearsResolution Finally master @ZJ10 's pa...	NaN
2	#NewYearsResolution to stop being so damn perf...	NaN
3	My #NewYearsResolution is to help my disabled ...	NaN
4	#NewYearsResolution #2015Goals #2015bucketlist...	NaN

	tweet_created	tweet_date	tweet_id	tweet_location \
0	12/31/14 10:48	12/31/14	5.500000e+17	Southern California
1	12/31/14 10:47	12/31/14	5.500000e+17	New Jersey
2	12/31/14 10:46	12/31/14	5.500000e+17	Hollywood
3	12/31/14 10:45	12/31/14	5.500000e+17	Metro NYC
4	12/31/14 10:44	12/31/14	5.500000e+17	Pittsburgh, Pennsylvania

	tweet_state	user_timezone	tweet_region
0	CA	Pacific Time (US & Canada)	West
1	NJ	Central Time (US & Canada)	Northeast
2	CA	Eastern Time (US & Canada)	West
3	NY	NaN	Northeast
4	PA	Eastern Time (US & Canada)	Northeast

[nltk_data] Downloading package stopwords to /root/nltk_data...

[nltk_data] Unzipping corpora/stopwords.zip.

[nltk_data] Downloading package vader_lexicon to /root/nltk_data...

	tweet	Positive	Negative \
0	newyearsresolut read book scroll fbcheck emai...	0.000	0.00
1	newyearsresolut final master part kitchen sink	0.000	0.00
2	newyearsresolut stop damn perf ùððùðé	0.000	0.62
3	newyearsresolut help disabl patient discov emo...	0.434	0.00
4	newyearsresolut continued%öü	0.000	0.00

	Neutral
0	1.000
1	1.000
2	0.380
3	0.566

4 1.000

Neutral 😊

Positive: 641.87200000000008

Negative: 369.00700000000004

Neutral: 4000.1309999999994

```
➡ Neutral 😊  
Positive: 641.87200000000008  
Negative: 369.00700000000004  
Neutral: 4000.1309999999994
```

INFERENCE:

- The given output is the result of sentiment analysis performed on a dataset of tweets related to New Year's resolutions using the VADER (Valence Aware Dictionary and sEntiment Reasoner) tool from the NLTK library.
- The first column shows the actual tweet text. The second, third and fourth columns represent the positive, negative, and neutral sentiment scores respectively, assigned by the VADER tool to each tweet.
- The overall sentiment of a tweet is determined by the combination of these three scores, with positive and negative scores indicating the degree of positive and negative sentiment in the tweet, and the neutral score indicating the degree of sentiment that is neither positive nor negative.
- The last three lines of the output show the overall sentiment of the dataset, calculated by summing up the positive, negative and neutral scores for all tweets in the dataset. In this case, the dataset is found to be mostly neutral, with a slightly higher positive sentiment score compared to the negative sentiment score.
- Overall, the VADER sentiment analysis tool has been applied to the given dataset of tweets related to New Year's resolutions to provide a quantitative understanding of the general sentiment of the tweets in the dataset.

CONCLUSION:

In conclusion, sentiment analysis of tweets can be a powerful tool for understanding public opinion and attitudes towards a particular topic. Through the use of natural language processing techniques and sentiment analysis tools like Vader, we were able to extract sentiment scores for tweets related to New Year's resolutions. Our analysis revealed that the majority of tweets were neutral, but there were also many positive and negative tweets expressing a range of emotions and attitudes towards resolutions. By examining the common themes and factors that contribute to successful resolutions, we can gain insights into how people can better achieve their goals. The findings of this study can be useful for businesses, organizations, and policymakers looking to understand public sentiment and make informed decisions.