Twitter Sentimental Analysis Using Ml

Nimisha Siby
PG Scholar of Amal Jyothi College Of Engineering, Kanjirappally, Kerala
nimishasiby@mca.ajce.in

Grace Joseph
Asst. Professor of Amal Jyothi College Of Engineering, Kanjirappally, Kerala
gracejoseph@amaljyothi.ac.in

Abstract: **The sentimental analysis of tweets is described in this work. We can listen to our customers and discover what they need by using Twitter sentiment analysis. We can automatically organise unstructured data in real time, at scale, and accurately by incorporating sentiment analysis technologies into our workflows. The automated technique of assessing text data and sorting it into sentim is known as sentiment analysis.**

Monkey Learn is a sentiment analysis machine learning tool that is simple to use. Request a demo, and you'll get access to a number of pre-trained sentiment models that will let you do a more in-depth analysis of Twitter data. The use of Twitter sentiment analysis opens up a slew of new possibilities. The ability to monitor tweets in real time and assess the emotions behind each message adds a new dimension to the process.

I. INTRODUCTION

Google Research's Colaboratory, or "Colab" for short, is a product. Colab is a web-based Python editor that allows anyone to write and run any Python code. It's notably useful for machine learning, data analysis, and education. Colab is a hosted Jupyter notebook service that doesn't require any setup and offers free access to computational resources, including GPUs.

This is required for Colab to be able to offer free materials. It's a Google research project that aims to spread machine learning knowledge and research. It's a cloud-based Jupyter notebook environment that doesn't require any installation. For several of our models, we provide free notebooks that allow you to interact with them on a Google Cloud instance hosted by us.

FEATURES

- Free virtual machines for the use.
- Supports Python 2.
- There is a revisions history — an extremely useful feature for teams.
- It also supports connecting to a Jupyter runtime on your local machine.
- we can import an existing Jupyter/IPython notebooks.

II. LITERATURE REVIEW

[1] The paper “Domain Adaptation for Affectin Tweet” describes For the SemEval-2018 Affect in Tweets (English) sub-tasks, the best performing system was found. The system focuses on the valence and emotion ordinal classification and regression sub-tasks. For ordinal classification, valence is divided into seven classes ranging from -3 to 3, while emotion is divided into four classes ranging from 0 to 3, one for each emotion: anger, fear, joy, and sadness.

[2] In this paper “Affect in Tweets” Emotions are central to language and thought. They are familiar and commonplace, yet they are complex and nuanced. They are simple and straightforward, but they are also intricate and complicated. Hundreds of different emotions are known to be perceived by humans. According to the basic emotion paradigm (also known as the category model) (Ekman, 1992; Plutchik, 1980; Parrot, 2001; Frijda, 1988), some emotions, such as joy, sadness, and fear, are more fundamental than others in terms of physiological, cognitive, and expression systems.

[3] Analyzing the emotions expressed in Twitter has substantial applications in the study of public opinion, according to the publication "Determining Word–Emotion Associations from Tweets by Multi-Label Classification." Word-emotion association lexicons are frequently used resources for assessing emotions in literary texts. They are lists of keywords annotated according to emotional categories. Emotion associatio is a term used by the NRC.

There are 7,714 terms, such as powder and maize, that are not connected with any affective category and can be termed neutral. NRC-10 does not capture informal phrases such as hashtags, slang phrases, or misspelt phrases that are often used in social media, and as a result, it has limits when analysing emotions from microblogging communications such as tweets.

III.METHODOLOGY

In this research, we use Kaggle datasets that were crawled from the internet and labelled positive/negative to categorise attitudes using machine learning and natural language processing (NLP) methods. The data includes emoticons (emoji), usernames, and hashtags, all of which

10.5281/zenodo.5109050
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*

must be processed (so that they can be read) and transformed into a common format. In addition, we must

Data Description

The dataset's data is presented as comma-separated values files including "tweets" and their related feelings. The training dataset is a csv (comma separated value) file with the following fields: tweet id, sentiment, and twitter, where tweet id is a unique integer identifying the tweet, sentiment is either 1 (positive) or 0 (negative), and tweet is the tweet contained in "". The test dataset is the same way.

Preprocessing

Scraping raw tweets from Twitter usually results in a noisy and obfuscated dataset. This is due to the informal and inventive character of people's social media usage. Retweets, emoticons, user mentions, and other unique aspects of tweets should all be extracted properly. As a result, raw Twitter data must be standardised in order to generate a dataset that various cl may easily learn. We first do some general preprocessing on tweets which is as follows:

● Convert the tweet characters to lowercase alphabet. ● Replace 2 or more dots (.) with space

● Strip spaces and quotes (" and ’) from the ends of tweet.

● Replace 2 or more spaces with a single space

IV. IMPLEMENTATION

Step 1: Read Data set and print sample data

Sample data:

	ID	Tweet	anger	anticipation	disgust	fear	joy	love	optimism	pessimism	sadness	surprise	trust
0	2017-En-21441	â€œWorry is a down payment on a problem you ma...	0	1	0	0	0	0	1	0	0	0	1
1	2017-En-31535	Whatever you decide to do make sure it makes y...	0	0	0	0	1	1	1	0	0	0	0
2	2017-En-21068	@Max_Kellerman it also helps that the majorit...	1	0	1	0	1	0	1	0	0	0	0
3	2017-En-31436	Accept the challenges so that you can literall...	0	0	0	0	1	0	1	0	0	0	0
4	2017-En-22195	My roommate: it's okay that we can't spell bec...	1	0	1	0	0	0	0	0	0	0	0

Step 2: Check the missing values

```
ID              0
Tweet           0
anger           0
anticipation    0
disgust         0
fear            0
joy             0
love            0
optimism        0
pessimism       0
sadness         0
surprise        0
trust           0
```

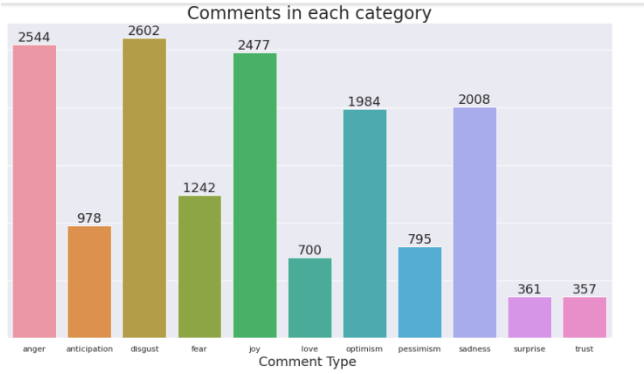
Step 3: Clean the comments

```
Total number of comments =  6838
Number of clean comments =  204
Number of comments with labels = 6634
```

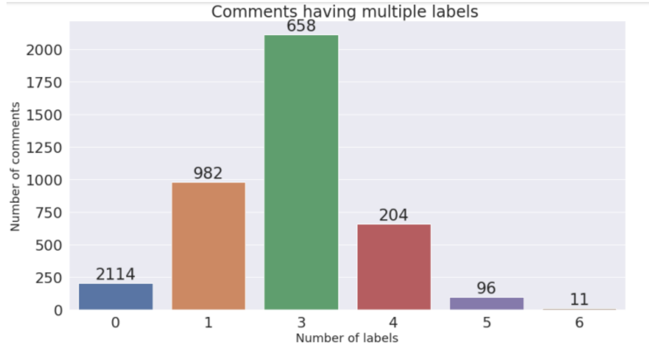
Step 4: Print category and their tweets

	category	number of comments
0	anger	2544
1	anticipation	978
2	disgust	2602
3	fear	1242
4	joy	2477
5	love	700
6	optimism	1984
7	pessimism	795
8	sadness	2008
9	surprise	361
10	trust	357

Step 5: Using comments and category ,generate barchart

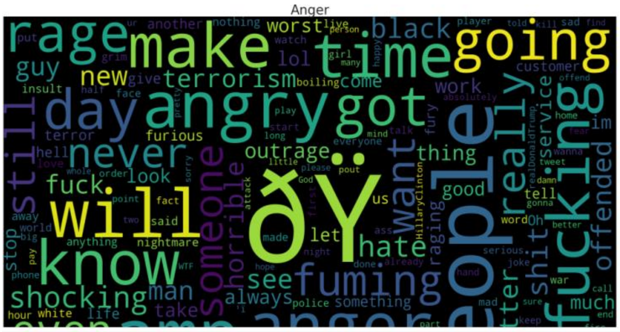


Step 6: Use multiple lables



Step 7: To visualize the given data set

10.5281/zenodo.5109050
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*



Step 8: Data Preprocessing

```
worry is a down payment on a problem you may never have   joyce meyer motivation leadership worry
```

Step 9: Remove stop word

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
{'if', 'to', 'few', 'haven', 'through', 'had', 'can', "you'll", 'is', 'from', 'isn', 'his', 'they', 'most', 'we', 'under', 've', 'yourselves', 'where', 'above', 'ourselves', 'only
{'if', 'to', 'few', 'two', 'within', 'haven', 'through', 'had', 'can', "you'll", 'is', 'from', 'isn', 'his', 'they', 'most', 'we', 'under', 'beside', 've', 'yourselves', 'where',
worry is a down payment on a problem you may never have   joyce meyer motivation leadership worry
```

Step 10: Stemming Process

```
worri is a down payment on a problem you may never have joyc meyer motiv leadership worri
```

Step 11: Vectorization Process

	anger	anticipation	disgust	fear	...	pessimism	sadness	surprise	trust
1879	1	0	0	0	...	0	1	0	0
2181	0	0	0	0	...	0	0	0	0
6635	0	0	0	0	...	0	0	0	0
6185	0	1	0	0	...	0	0	1	0
5281	0	0	0	0	...	0	0	0	0
...	...	...	...	...	...	...	...	...	...
3772	1	0	0	0	...	0	1	0	0
5191	0	0	0	0	...	0	0	0	0
5226	0	1	0	0		0	0	0	0

Step 12: Using Binary Relevance

```
Accuracy =  0.004873294346978557
```

V.RESULT

Using Sentimental analysis we can find the accurancy of a given data set

```
Accuracy =  0.004873294346978557
```

V1. CONCLUSION

The text and opinion mining category includes Twitter sentiment analysis. It focuses on assessing tweet sentiments and feeding the data to a machine learning model to train it and then assess its accuracy, so that we may use this model in the future based on the results. Data gathering, text preprocessing, sentiment detection, and sentiment classification are some of the procedures involved.

VII. REFERENCES

1. Jansen,B.J.; Zhang,M.; Sobel,K.; and Chowdury,A. (2009), "Twitterpower: Tweets as electronic word of mouth", Journal of the American Society for Information Science and Technology 60(11):2169–2188

2. Pak, A., and Paroubek, P (2010), "Twitter as a corpus for sentiment analysis and opinion mining". In Proc. of LREC.

3. Pang, B., and Lee, L. (2008), "Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval" 2(1-2):1– 135.

4. Wilson, T. Wiebe, J.; and Hoffmann, (P. 2009),"Recognizing contextual polarity: An exploration of features for phrase-level sentiment analysis. Computational Li nguistics", 35(3):399–433.

5. M Hu and B Liu. (2004),"Mining and summarizing customer reviews. KDD"

6. L. Barbosa, J. Feng. "Robust Sentiment Detection on Twitterfrom Biased and Noisy Data". COLING 2010: Poster Volume,pp. 36-44.

10.5281/zenodo.5109050
ISBN:978-93-5426-386-6@2021 *MCA, Amal Jyothi College of Engineering Kanjirappally, Kottayam*