

Final Year Project

Twitter Sentiment Analysis using NLP

Synopsis

Bachelor of Technology
Information Technology



Institute of Engineering and Technology, Ayodhya
2022

Submitted To:

Er. Praveen Mishra
Assistant Professor
Department of IT

Submitted By:

Anurag Gupta (19317)
Anshita Shrivastav (19316)
Aaysha Shukla (19301)
IT 4th Yr. (VIIth Sem.)

Dr. Ram Manohar Lohia Avadh University
Ayodhya, Uttar Pradesh - 224001

INDEX

S.N.	Topics	Page No.
1.	Introduction	3
2.	Objective	3
3.	Model	3-4
4.	Techniques and Methods	5
5.	Applications of Sentiment Analysis	5-6
6.	Future Scope	6

SENTIMENT ANALYSIS

Sentiment analysis (also known as opinion mining or emotion AI) is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information. Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media, and healthcare materials for applications that range from marketing to customer service to clinical medicine.

With the rise of deep language models, such as RoBERTa, also more difficult data domains can be analyzed, e.g., news texts where authors typically express their opinion/sentiment less explicitly.

Objective

To accurately extract people's opinions and feelings expressed in positive or negative comments, questions and requests, by analyzing a large number of unstructured texts and classifying them into sentiment classes.

Model

It is a Natural Language Processing Problem where Sentiment Analysis is done by Classifying the Positive tweets from negative tweets by machine learning models for classification, text mining, text analysis, data analysis and data visualization.

Natural Language Processing (NLP) is a hotbed of research in data science these days and one of the most common applications of NLP is sentiment analysis. From opinion polls to creating entire marketing strategies, this domain has completely reshaped the way businesses work, which is why this is an area every data scientist must be familiar with.

Thousands of text documents can be processed for sentiment (and other features including named entities, topics, themes, etc.) in seconds, compared to the hours it would take a team of people to manually complete the same task.

We will do so by following a sequence of steps needed to solve a general sentiment analysis problem. We will start with preprocessing and cleaning of the raw text of the tweets. Then we will explore the cleaned text and try to get some intuition about the context of the tweets. After that, we will extract numerical features from the data and finally use these feature sets to train models and identify the sentiments of the tweets.

Here, we will use machine learning to see whether the emotion of a tweet is positive, neutral or negative.

And it is quite a challenging task as you can see tweets are different than other type of text data that we have because:

- a) they are normally in conversational language, and
- b) they are of short length

So, we will do sentiment analysis on some tweets data.

Now when we are talking about the text data of a tweet, We will consider it in into four categories:

- i) one there is text regarding the tweet itself,
- ii) then there is emoji,
- iii) then there is normally mentions of a user, and
- iv) there could also be some link to some website inside a tweet.

In order to do the analysis we will do preprocessing on the text to make it appropriate for the way the model is trained.

The preprocessing of the text data is an essential step as it makes the raw text ready for mining, i.e., it becomes easier to extract information from the text and apply machine learning algorithms to it. If we skip this step then there is a higher chance that you are working with noisy and inconsistent data. The objective of this step is to clean noise those are less relevant to find the sentiment of tweets such as punctuation, special characters, numbers, and terms which don't carry much weightage in context to the text.

Techniques and Methods

There are several techniques and methods used in sentiment analysis, including:

Lexicon-based methods - This method involves using dictionaries or lexicons to identify the sentiment in a piece of text. Lexicons are lists of words that are categorized as positive, negative, or neutral based on their semantic meaning.

Machine learning-based methods - This method involves training machine learning algorithms on large datasets to identify patterns and predict the sentiment of a piece of text. This method is more complex than lexicon-based methods and requires large amounts of labeled data for training.

Deep learning-based methods - This method involves using neural networks to analyze text and extract sentiment information. This method is the most advanced form of sentiment analysis and has shown promising results in recent years.

Applications of Sentiment Analysis

Sentiment analysis has a wide range of applications, including:

Politics: In the political field, it is used to keep track of political views, to detect consistency and inconsistency between statements and actions at the government level. It can be used to predict election results as well!

Public Actions: Sentiment analysis also is used to monitor and analyze social phenomena, for the spotting of potentially dangerous situations and determining the general mood of the blogosphere.

Marketing - Sentiment analysis can be used to understand consumer opinions on products and services, which can help companies improve their offerings and increase customer satisfaction.

Customer feedback - Sentiment analysis can be used to analyze customer feedback on products and services, which can help companies understand customer satisfaction and identify areas for improvement.

Stock market analysis - Sentiment analysis can be used to analyze news and social media posts to predict stock market trends, which can help investors make informed decisions.

Future Scope

The future scope of sentiment analysis is very promising, and here are a few reasons why:

Increased demand: With the growing volume of online customer reviews, social media posts, and other forms of user-generated content, there is a growing demand for sentiment analysis technology to help organizations understand the opinions and sentiments of their customers.

Advancements in AI and NLP: The advancements in artificial intelligence and natural language processing technologies are making sentiment analysis more accurate and efficient. This is leading to the development of new applications and use cases for sentiment analysis.

Integration with other technologies: Sentiment analysis is likely to be integrated with other technologies such as virtual assistants, chatbots, and voice-activated devices, making it more accessible and useful for a wider range of applications.

Improved data privacy and security: As sentiment analysis relies on large amounts of personal data, there is a growing concern about data privacy and security. In the future, sentiment analysis technology will likely include better privacy and security features to protect sensitive data.

Overall, sentiment analysis has a bright future ahead and is expected to continue to grow and evolve as more and more organizations adopt it to better understand their customers and improve their products and services.