

Sentiment Analysis of Comments in Social Media



Anik Barua, Mahabuba Monju, MD. Mohuuddin Munna
Course Title: Machine Learning and Data Mining Lab

Abstract

The Web is a vast virtual environment where people may express and share their personal thoughts. This has ramifications for both marketing and communication. The field has matured to the point where it may be used in a variety of industries, including marketing, health, banking, and politics. Currently, Facebook dominates the digital marketing space, followed closely by Twitter. This paper describes a Sentiment. This study included several enhancement methods, including feature extraction, classification algorithms, NLP techniques, and more, including the ability to recognize polarity. This paper evaluates a sentiment analysis study that was conducted on 999 Facebook posts and 162980 tweets, which were then combined and further processed. Here, I apply the ML algorithms SVM, DT, XGBoost, and deep learning-based algorithm LSTM.

Keywords: NLP, Facebook, Twitter, ML Algorithm, Deep-Learning Algorithm.

Introduction

The Web is a vast virtual environment where people may express and share their personal thoughts. This has ramifications for both marketing and communication. More than 30% of Internet users have reviewed items or services online, according to a recent Forrester study. The term "opinion mining" also applies to sentiment analysis. It is a technique for determining the tone of speech. Sentiment analysis comes under Natural Language Processing. The sentiment is an attitude or an emotion or a feeling. . Over 1.5 million tweets are produced daily by users, and every website has its benefits and drawbacks.

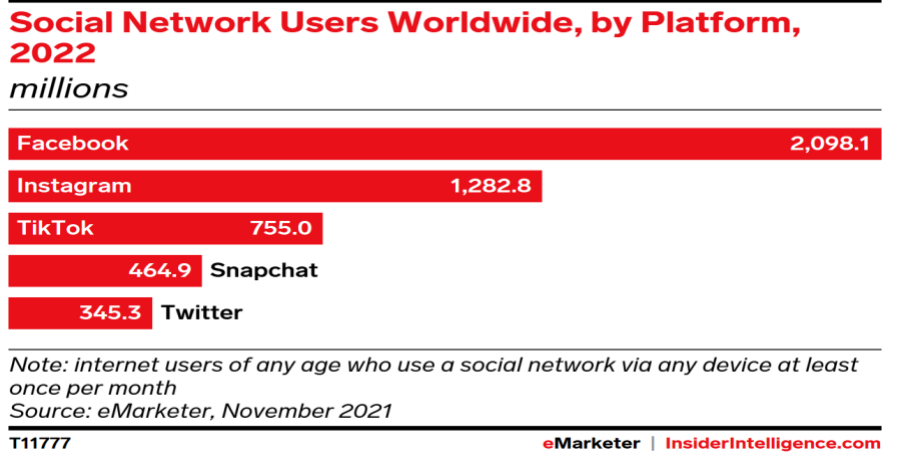


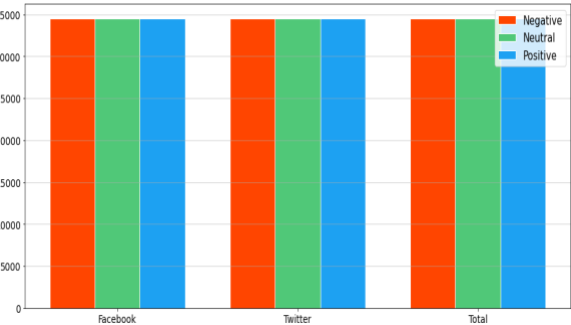
Figure-1
To determine the emotional state of tweets, Twitter data was retrieved, transformed, and analyzed before being divided into the positive, negative, and neutral categories.



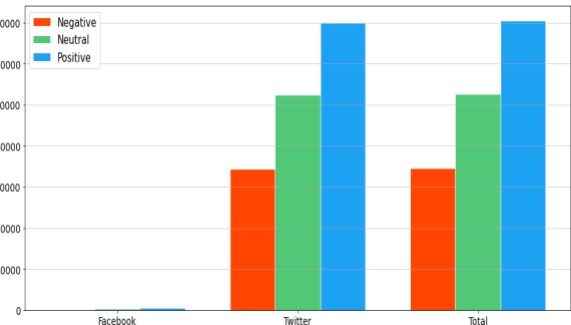
Figure-2

Methodology

Data Collection:
For sentiment analysis, we gathered and manually categorized a sizable dataset from twitter and Facebook comments. This dataset I collected from Kaggle and preprocess by own way. This dataset includes thoughts from viewers about various English comments and postings. As we are interested in Social media sentiment data we collected data based on negative, positive and another is neutral.



Data Cleaning and Preprocessing:
At first we selected important and common features from both dataset then changed column name on both dataset. Then we replaced all punctuation mark with null and deleted all duplicated comments. Then we merged both cleaned dataset into one and splits into train and test data .At last we vectored data to convert text data to numerical data to understand machine.



Deep Learning Classifier:

- LSTM:** A long short-term memory network is a type of recurrent neural network (RNN).

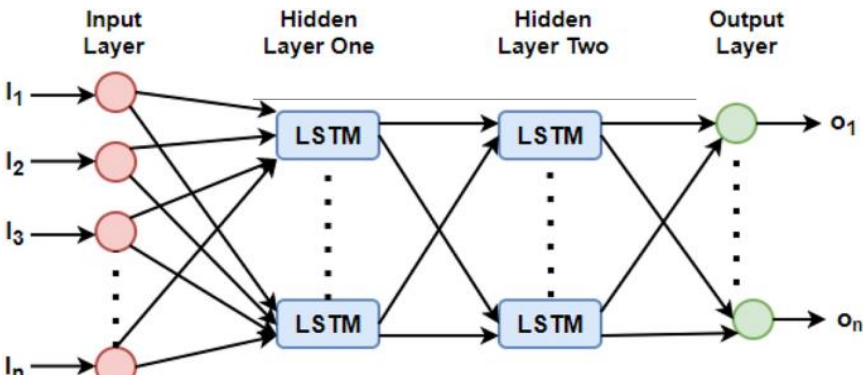


Figure-3

- Traditional Classifiers:**
- Support vector machine:** Support Vector Machine is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems.
 - Decision Tree:** A decision tree is a graph that uses a branching method to illustrate every possible output for a specific input.
 - XGBoost:** It is popular and efficient open-source implementation of the gradient boosted trees algorithm.

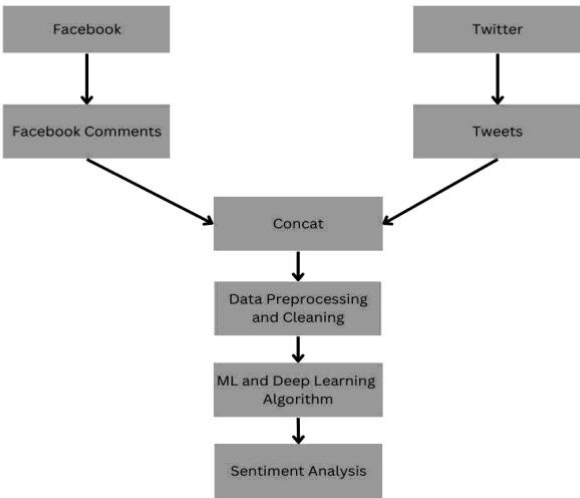


Figure-4
Results and Discussion

NLP or Natural Language Processing is basically an approach to find information out of a text to make it understandable to a machine as the same as humans do. As we all know, the whole idea of machine learning is to provide the human-brain-like capabilities to a machine. This is done to provide the same ability to machines as our human brain, which is capable of understanding text and speech. In this section we discuss about results. Here we uses traditional learning classifiers such as support vector machine, Decision tree, XGBoost and Deep learning classifier such as LSTM. Here SVM given the highest accuracy about 93%.At last we predicted with random test data and predicted correctly.

Table-1

Classifiers	Accuracy
SVM	93%
Decision Tree	82%
XGBoost	71%
LSTM	90%

Future Work

Developers can make use of NLP to perform tasks like speech recognition, sentiment analysis, and translation, auto-correct of grammar while typing, and automated answer generation. NLP is a challenging field since it deals with human language, which is extremely diverse and can be spoken in a lot of ways. The system gathers the necessary data from customers/people by using social media sites like Twitter, YouTube, and Online-News. Real-time analysis is a powerful tool for business, marketing, politics, which has a tendency to extract the result of current trends through feedback and helps in decision making.

Conclusion

Natural Language Processing is the practice of teaching machines to understand and interpret conversational inputs from humans. NLP based on Machine Learning can be used to establish communication channels between humans and machines. Natural language processing can be a huge help to any business, to save time and money, streamline and automate processes, and make real-time, data-driven decisions.

References

[1] A. Alrumaih, "Sentimental Analysis of Arabic Comments in Computer Mediated Communication," MSc Thesis, Sheffield Hallam University, 2018.

[2] H.K.Chan, E.Lacka , R.W.Y. Yee and M.K.Lim "A Case Study on Mining Social Media Data".

[3] MuhammetBaykara, UğurGürtürk "Classification of social media shares using sentiment analysis".

[4] Jabreel, M.; Moreno, A. A Deep Learning-Based Approach for Multi-Label Emotion Classification in Tweets. Appl. Sci.2019, 9, 1123.