

International Journal of Computer Science and Mobile Computing



A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X

IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 3, March 2022, pg.37 – 41

Review Sentimental Analysis

Samiksha Rahate; Vaishnavi Dehanka; Tanvi Teppalwar; Prof. Vaishali R Surjuse

samiksharahate39356@gmail.com; vaishnavidehankar17@gmail.com; teppalwartanvi@gmail.com;
surjuse.vaishu@gmail.com

KDK College of Engineering, Nagpur, India

DOI: <https://doi.org/10.47760/ijcsmc.2022.v11i03.005>

Abstract:- Sentiment analysis or opinion mining is the computational study of people's opinions, sentiments, attitudes, and emotions expressed in written language. It is one of the most active research areas in natural language processing and text mining in recent years. Its popularity is mainly due to two reasons. First, it has a wide range of applications because opinions are central to almost all human activities and are key influencers of our behaviors. Whenever we need to make a decision, we want to hear others' opinions

Keywords— Sentimental analysis, Reviews, SVM, RF, KNN

I. INTRODUCTION

People's ways of expressing themselves have changed dramatically as a result of the internet era. This is more easily accomplished through blog articles, forums, product testing blogs, social media, and other means. Today millions of people are using social networking sites like Facebook, Twitter, Google and others to express their feelings. We get immersive media from online groups that use forums to inform and influence others. Tweets, updates, blog posts, notes, feedback, and other forms of social media generate a large amount of sentimentally rich data. Additionally, social media allows businesses to engage with their ad clients. The majority of people rely on user-generated content on the internet. For instance, if someone wants to buy or utilise a product, they first look up the product's review on the internet, Then, before making a decision, address it on social media. The volume of user data is too large for an ordinary user to analyse. As a result, this must be automated, and various methods for processing feelings are commonly used.

Natural language processing (NLP), review, and study, is often referred to as "opinion mining." The computational processing and subjectivity of the document are dealt with by SA, also known as sentiment classification, sentiment mining, examination mining, subjectivity analysis, opinion mining, or evaluation extraction and polarity classification. The textual analysis aims to define a speaker's decision on a trustworthy issue.

In this paper higher performance rate has been produced when using hybrid SVM, KNN and Radom forest and deep learningbased sentiment classification techniques.

II. RELATED WORK

M. Butler and S. Robila(2016) The creation and deployment of the IMDb dataset file extraction and database import method are described in this study. The previous study used relational databases, which differs from the majority of published methods and analyses. This solution makes use of document-driven data structures and allows anyone to add and alter structures to suit their needs. The project entailed the use of current technology for site developers and software engineers, as well as assisting other developers in forking variants of the work and using it for their own study.

A. Yenter & A. Verma (2017) For five high-performance models, researchers are experimenting with a wide range of regularisation methods, network architectures, and kernel sizes. With a precision of over

89 percent, these models can predict the polarity of sentiments in IMDb dataset feedback. To begin with, the best-performing model's precision exceeds that of prior models, and it also greatly improves the baseline CNN+LSTM model Sandesh Tripathi et al. (2020) They used SA on the Kaggle Bag of Words dataset of IMDb film reviews in the study schemes to demonstrate how helpful insights may be generated from a variety of textual data acquired across the Internet. These concepts are created from the use of standard machine learning algorithms: NB, LR, RF, and DT. Furthermore, the results of all four algorithms were assessed using six test measures:

III. PROBLEM STATEMENT AND OBJECTIVE

A. Problem Statement Various machine learning and lexicon-based algorithms have been developed in the literature to automate the sentiment analysis task. Despite the fact that these strategies have been widely employed for sentiment categorization, they have failed to produce the best results in terms of accuracy and resolution of all difficulties. As a result, new automated procedures must be developed to address all of the obstacles and provide the best results. Positive and negative reviews will be separated into two groups.

B. Objectives

- To classify the review
- Use of NLP to analyse the data
- Give one word analysis

IV. FLOWCHART

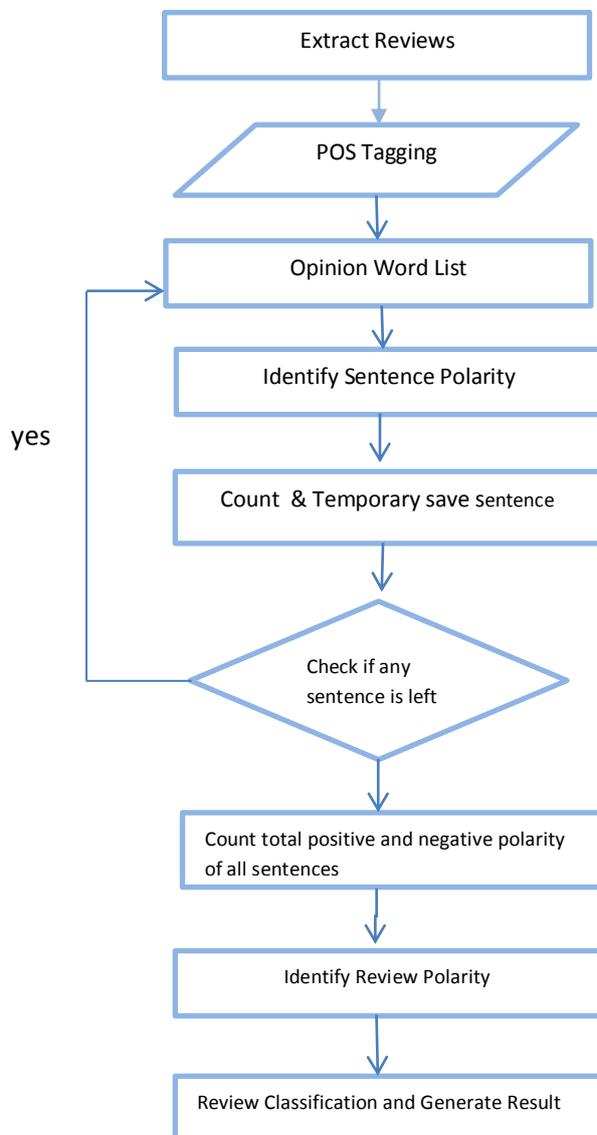


Figure 1 Flow chart

V. PROPOSED WORK

A. Data Input

There are two ways to provide feedback to the emotion analyzer for movie reviews. After retrieving reviews, the system analyses the sentiment of the first 10 reviews for each Title.

B. Part of Speech Tagging

To extract features from a sentence, POS is utilised to disambiguate it [2]. Every word is labelled in POS tagging. It's utilised to figure out where a word belongs in a sentence. In a sentence, POS tagging assists in the identification of nouns, noun phrases, verbs, and adjectives. For feature selection and opinion words, there is a small chance that a chosen word will be discarded after POS Tagging.

C. Features and Opinion Words Extraction

The sentence's entire set of opinion words is chosen. The algorithm collects all nouns, noun phrases, verbs, and adjectives from the movie review and compares them to a word list already in existence. The polarity of these words is used to classify them. The word "good" has a positive polarity. On the other hand, features are chosen based on the frequency with which opinion words appear. If an incidence of the word "opinion" in a review is greater than the threshold value, it is added to the features list. This system's API has only been trained for movie reviews, using a keyword and phrase vocabulary that includes terms such as "excellent acting," "solid storey," and "great action."

D. Identify Sentence Polarity

It is very simple to determine the polarity of the sentence after extracting all features and Opinion words. The same rules apply to sentence polarity as they do to arithmetic expressions. A negative sentiment is made up entirely of negative words, while a good sentiment is made up entirely of positive opinion words. A positive opinion word could be included in a negative mood. In a movie review, for example, "This movie's storey is not good." The word "good" has a positive polarity in this sentence, whereas "not" has a negative polarity. As a result, this sentence's overall polarity will be negative.

E. Identify Review Polarity

The total number of positive or negative sentences discovered in a review determines the overall polarity of the review. When the overall number of positive sentences exceeds the total number of negative sentences, the review polarity is positive. A review's polarity will also be negative if the total number of negative sentences exceeds the total number of positive sentences.

F. Classification of Review

Once that is done, the review polarity is computed. Review the polarity percentage and polarity classification (Positive or Negative) [28] and store it for further investigation.

VI. ALGORITHM

Machine learning techniques concentrate on algorithms that evaluate sentiments, which is a common difficulty in the classification of syntaxes and language functions. These algorithms employ a number of methods in which the data package offers the structure for identifying and detecting different Patterns. The trained models are then used to identify and classify previously unknown patterns based on their emotions.

Support Vector Machine

The goal of SVM classification is to distinguish between two groups by preparing relevant data with a feature and producing a classifier that works well on hidden data. The maximum range classification is the most basic sort of SVM. The main classification problem is frequently solved by binary classification of linear separable training data.

K- Nearest Neighbor (KNN)

The NN approach is used to find the unknown data point by focusing on the nearest neighbour whose value has already been determined. Search for the closest point. The NN mechanism can be divided into two ways. Structure and function are less used NN classification approaches. The scheme classifies K-NN as a less method. The KNN method makes use of the NN for the value of k, which specifies how many NN to a sample data point must be checked in the class description. There are two types of NN strategies: KNN dependent structure and KNN less structure

Random Forest (RF)

It's a classification algorithm for detecting sentiment analysis and emotion analysis in numerically significant data. Random forest is a classification method that involves merging tree architectures with training on the available data. The creation of Extreme Gradient Boosting (XGBoost) is a treebased classification technique. This algorithm creates trees in the same way that Random forest does, and it also uses gradient descent and boosting.

A. Details of hardware and software

HARDWARE REQUIREMENTS:

- HARD DISK – 500 GB
- SYSTEM – I5 PROCESSOR
- RAM-4 GB

Software Requirements:

- LANGUAGE – Python
- FRONT END: HTML, CSS
- APP- Java
- Web – python

VII. EXPERIMENTAL EVALUATION

The method was put to the test for more than 50 distinct movie titles, each with a maximum of 10 reviews, with the ultimate results of 500 reviews presented. It displays the number of available features, the system's accuracy (Accurate Result Percentage), Error Percentage, and False Negative and False Positive Percentages. False Negative refers to a positive polarity review that is mistakenly seen as negative. False Positive, on the other hand, refers to a negative polarity review is misinterpreted as positive.

VIII. CONCLUSION

The polarity of reviews is characterised in this paper as positive or negative. The author's proposed technique can be used to categorise a large database of reviews. The system's best feature is that it's a web-based sentiment analysis API for reviews with JSON output that can be shown on any operating system. This will allow users to see how far their review has progressed. This API can be trained for various reviews in the future, such as cellphones, computers, or clothing.

REFERENCES

- [1] S. M. Metev and V. P. Veiko, *Laser Assisted Microtechnology*, 2nd ed., R. M. Osgood, Jr., Ed. Berlin, Germany: Springer-Verlag, 1998.
- [2] J. Breckling, Ed., *The Analysis of Directional Time Series: Applications to Wind Speed and Direction*, ser. Lecture Notes in Statistics. Berlin, Germany: Springer, 1989, vol. 61.
- [3] (2002) The IEEE website. [Online]. Available: <http://www.ieee.org/>
- [4] M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: <http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/>
- [5] *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
- [6] "PDCA12-70 data sheet", Opto Speed SA, Mezzovico, Switzerland.
- [7] A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
- [8] J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.