iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** I **Month of publication:** January 2023

DOI: https://doi.org/10.22214/ijraset.2023.48706

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue I Jan 2023- Available at www.ijraset.com

Sentiment Analysis Using Machine Learning

A. Samyuktha[1], M. Pallavi[2], L. Jagan[3], Dr .Y. Srinivasulu[4], Mr. M. Rakesh[5]
[1, 2, 3]*Student Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India*
[4, 5]*Professor Department of Electronics and Communication Engineering, Sreenidhi Institute of Science and Technology, Hyderabad, Telangana, India*

Abstract: Sentiment analysis falls within the category of analytics research. This can make sense by reading raw data using computational methods. This is what analysis is. Written expressions that are neutral, unfavourable, or indifferent can be assessed using sentiment analysis. People use a variety of social media platforms, including Facebook and Twitter, which is a useful tool for gauging public sentiment. This uses a variety of machine learning techniques. We have considered a variety of sentiment analysis techniques in this study. Using machine learning classifiers, sentiment analysis has been carried out. Users' tweets are categorised as having "positive" or "negative" sentiment using polarity-based sentiment analysis and deep learning models. Sentiment Analysis, one of the branches of computer science that is now gaining the most ground.

I. INTRODUCTION

A machine learning tool called sentiment analysis looks for positive or negative polarity in texts. Using textual examples of emotions as training material, machine learning tools learn to automatically recognize emotion without human intervention. Simply put, machine learning enables computers to acquire new skills without being explicitly programmed to do so. It is possible to train sentiment analysis models to read beyond just definitions to comprehend things like context, sarcasm, and misused words. We are prompted to consider the phrase "You're so smart!" by sentiment compliment? It is evident that the speaker is heaping praise on a person of the highest intelligence. Sentiment Analysis, a subfield of Natural Language Processing (NLP), uses the sentiment of the words to classify the reviews as positive or negative. Opinions about any entity can be categorized as positive or negative based on the sentiment that is expressed in the words. The phrase "I am not excited by this product though it is quite cheap," for instance, conveys a negative opinion of the product. The level of the feeling utilized is additionally thought about. For instance, the phrase "I love this product" conveys a more enthusiastic attitude than the phrase "I like this product." Aside from ordinary descriptors like 'great', 'terrible' and 'excellent', conjunctions like 'yet', 'despite the fact that", 'while' additionally have something to do with the general extremity of the sentence. There is a lot of information on the Internet that can help people and organizations make decisions, but it also makes it hard for people and organizations to understand what other people think and how they feel about things. Unfortunately, finding, monitoring, and analyzing opinion sources is a monumental task.
Online opinion sources cannot be manually retrieved, sentiments extracted, and then expressed in a standard format.

A. Scope

Initially the extent of feeling examination was restricted to understanding public discernment, over the long haul it has extended to incorporate input and client perspectives on items and administrations. An explosion of online opinion channels, tech-savvy customers, and a generation that lives online to provide and absorb opinions have all contributed to an exponential increase in the complexity of sentiment understanding.

B. Overview

Software and Hardware requirements

1) Software Requirements

- Operating System: windows
- Tool: Anaconda with Jupyter Notebook

2) Hardware Requirements

- Processor: core i3/i5
- Hard disk: min 300GB
- RAM: min 4GB

©IJRASET: All Rights are Reserved

II. LITERATURE SURVEY

Taking into consideration a dataset consisting of more than 5.1 million product reviews from Amazon.com, the sentiment polarity categorization is the most fundamental issue in sentiment analysis. The products in this dataset fall into four categories. To classify the sentence's words, a second Python program is used, max-entropy POS tagger, to speed up the process. Adverbs include negation words like "no" and "not," among others, but Negation of Adjective and Negation of Verb are specifically used to identify phrases. The various classification models chosen for classification are as follows: Logistic Regression, Support Vector Machine, and Naive Bayesian Pang and Lee suggested extracting subjective sentences from objective sentences when performing feature selection. They came up with a text-categorization method that uses minimum cut to find subjective content. Gann and co. selected 6,799 tokens using data from Twitter. Each token has a sentiment score, called the TSI (Total Sentiment Index), which determines whether it is a positive or negative token. A TSI for a particular token is specifically calculated as follow tp/tn is the ratio of the total number of positive tweets to the total number of negative tweets, where p is the number of times a token appears in positive tweets and n is the number of times a token appears in negative tweets.

III. GENERAL PURPOSE OF SENTIMENT ANALYSIS

The following operations use sentiment analysis : Locate and extract the opinionated data—also known as sentiment data—pertaining to a particular platform (such as customer support, reviews, etc.).

Identify the opinion holder (both on its own and in relation to the existing audience segments) and define the subject matter (what is being discussed specifically and in general).

The following scenarios can be used with the sentiment analysis algorithm, depending on the purpose:

Record level - for the whole text.

Obtains the meaning of a single sentence at the sentence-level.

Obtains the meaning of sub-expressions within a sentence at the sub-sentence level .It is difficult to extract an opinion due to its subjective nature. Opinions vary. Compared to others, some are more valuable. An opinion is further characterized by four subcategories:

The opinion that makes a clear statement is the direct opinion. "The responsiveness of the buttons in application X is poor," for instance. Here you have a genuine point.

A comparative opinion is one in which X and Y are compared using specific criteria. For instance, "the responsiveness of the button in application X is worse than in application Y" serves as micro competitive research in addition to providing insight into your product.

Everything is clearly defined in the explicit opinion. Take, for instance, "this chair is rocking."

Implicit opinions are those that are implied but not explicitly stated. For instance, "the application began slacking in two days." It is essential to keep in mind that implicit opinions may also contain metaphors and idioms , making sentiment analysis more difficult.

IV. SENTIMENT ANALYSIS DATASETS

Obtaining a suitable source of training data is the first step in the development of any model, and sentiment analysis is no different. There are a couple of standard datasets in the field that are much of the time used to benchmark models and look at exactnesses, yet new datasets are being fostered consistently as marked information keeps on opening up.

Market research, brand monitoring, social media monitoring, customer service monitoring, and the voice of the customer (VoC) monitoring all make extensive use of sentiment analysis. To collect data from datasets, R's sentiment analysis makes use of hybrid, rule-based, or machine learning-based NLP algorithms and methods.

The sentiment analysis needs a lot of specialized data in very large quantities. Finding large amounts of data is the hardest part of the sentiment analysis training process; instead, it is locating the relevant datasets. These datasets ought to cover a wide range of applications and use cases for sentiment analysis. The Stanford Sentiment Treebank is the first of these datasets. It stands out because it contains over 11,000 sentences that were precisely parsed into labelled parse trees from movie reviews. Recursive models can thus train at each level of the tree, enabling them to predict the sentiment first for the sentence as a whole and then for its subphrases. Using product ratings as a proxy for the sentiment label, machine learning practitioners can train sentiment models with the help of the over 142 million Amazon product reviews that are included in the Amazon Product Reviews Dataset.

There are 50,000 highly polarized movie reviews in the IMDB Movie Reviews Dataset, split 50-50 train/test.

Training sentiment models to work with social media posts and other informal text is made easier with the help of the Sentiment140 Dataset. It gives you 1.6 million training points, all of which can be either positive, negative, or neutral.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

V. SENTIMENT SENTENCE EXTRACTION AND POS TAGGING

The fundamental requirement for POS tagging is the tokenization of reviews following the removal of STOP words that have no relation to sentiment. After legitimate expulsion of
STOP words like "am, is, are, the, however, etc the leftover sentences are changed over in tokens. These tokens partake in POS labeling
In normal language handling, grammatical form (POS) taggers have been created to group words in light of their grammatical forms. A POS tagger is very useful for sentiment analysis for the following two reasons: 1) Words like "noun" and "pronoun" rarely convey any emotion. It can sift through such words with the help of a POS tagger; 2) Words that can be used in various parts of speech can also be distinguished with the help of a POS tagger.

A. Multinomial Naïve Bayes

In machine learning, one of the variations of the Naive Bayes algorithm is the Multinomial Naive Bayes. When applied to a multinomially distributed dataset, it is extremely beneficial. Natural language processing-based classification tasks particularly benefit from this algorithm. One of the applications for this algorithm is spam detection. This article is for you if you have never used this algorithm before to solve classification-based machine learning problems. In this article, you will learn about the Multinomial Naive Bayes algorithm, which is used in machine learning, and how Python is used to implement it.
The probability determined by the Bayes theorem is P(c|x), where c denotes the class of possible outcomes and x denotes the given instance that needs to be classified and represents particular characteristics.
$P(c|x)=P(x|c)*P(c)/P(x)$

B. Logistic Regression

The probability of an outcome that can only have two values is predicted using logistic regression. The use of one or more predictors (numerical and categorical) is the foundation of the prediction. Maximum likelihood estimation (MLE) is used in logistic regression to discover the model coefficients that link predictors to the target. The procedure is repeated until Log Likelihood(LL) does not change significantly after this initial function is estimated.

C. Support Vector Machine (SVM)

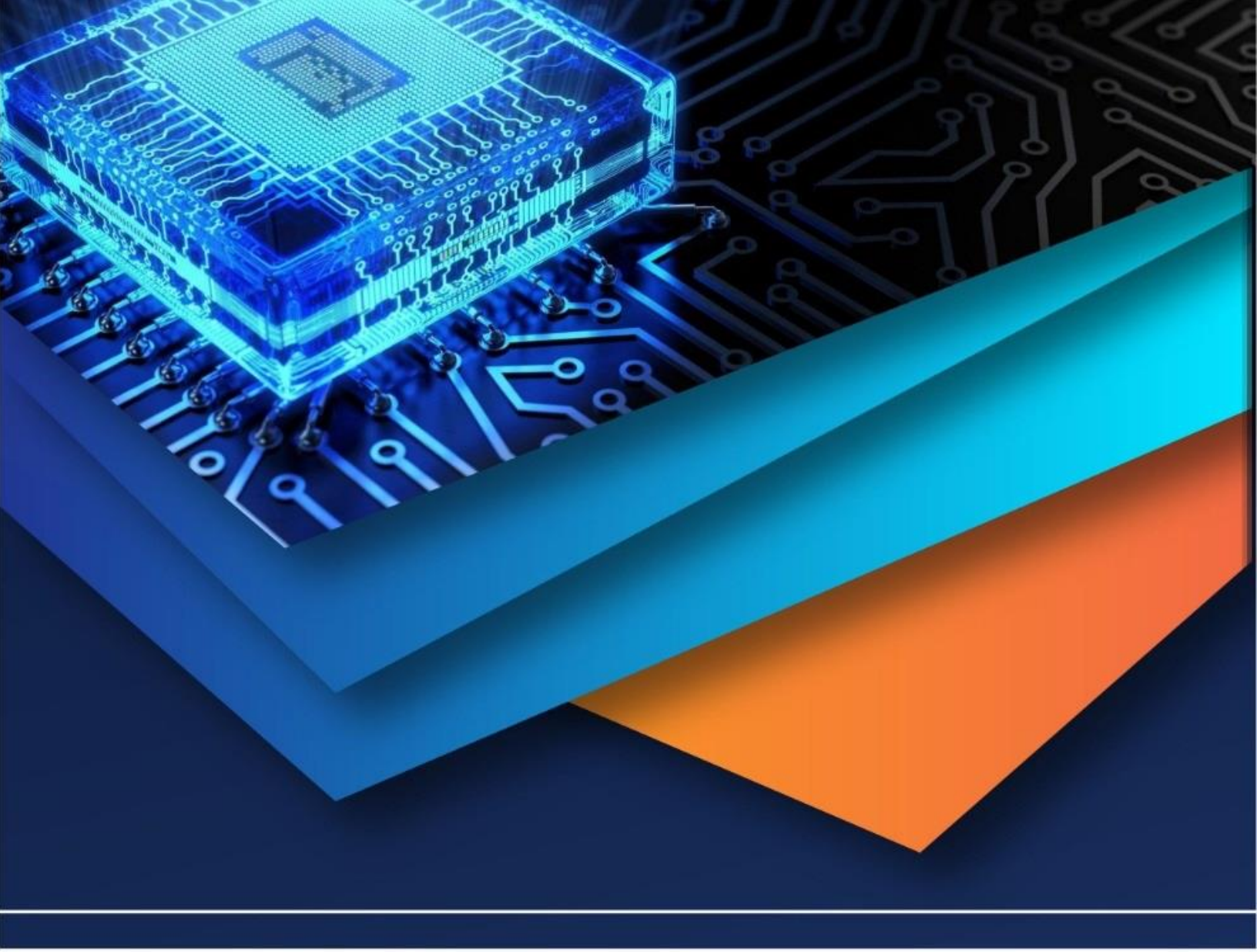
SVM (support vector machine) is a supervised machine learning algorithm that can be applied to regression or classification problems. Regression is the prediction of a continuous value, whereas classification is the prediction of a label or group. The hyper-planes that distinguish the classes we plotted in n-dimensional space are found by SVM for classification.

VI. CONCLUSION

The classification of texts according to the emotions they convey is the subject of sentiment analysis. Data preparation, review analysis, and sentiment classification are the three core steps of a typical sentiment analysis model that are the focus of this project, and representative techniques for each step are discussed. On the Dataset product reviews, a variety of machine learning algorithms, including Linear Regression, SVM, Randomforest classifier, Decision Tree, and Naive Bayes, were utilized. According to the study's findings, the accuracy of the SVM approach on the data set is superior to that of the other approaches. Individuals, large organizations, and governments all benefit from the use of sentiment analysis. Because it provides a comprehensive overview of the public opinion behind a variety of topics, such as product reviews, politics, movie reviews, and other facets of everyday life, sentiment analysis is crucial. Sentiment analysis is used in education to predict students' performance and learning curves, as well as to understand students' needs so that teachers can teach effectively. Sentiment analysis aids in the monitoring of trends in customers' overall opinions of a product or brand in the business sector. An accurate depiction of the feelings being expressed can be found in movie review sentiment analysis. Policy and politics are examples of topics that the government uses sentiment analysis to examine.

REFERENCES

[1] https://www.datarobot.com/blog/using-machine-learning-for-sentiment-analysis-a-deep-dive/
[2] https://www.geeksforgeeks.org/what-is-sentiment-analysis/
[3] https://monkeylearn.com/blog/sentiment-analysis-machine-learning/
[4] https://www.bing.com/search?q=https%3A%2F%2Ftowardsdatascience.com%2Fsentiment-analysis-using-logistic&form=IPRV10
[5] https://thecleverprogrammer.com/2021/08/06/multinomial-naive-bayes-in-machine-learning/
[6] https://www.geeksforgeeks.org/support-vector-machine-algorithm/
[7] https://www.academia.edu/41359036/Sentiment_Analysis_Using_Machine_Learning_Technique
[8] https://www.tutorialspoint.com/machine_learning_with_python/classification_algorithms_decision_tree.htm



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)