International Journal of Computer Trends and Technology
ISSN: 2231-2803 / https://doi.org/10.14445/22312803/IJCTT-V68I4P141

Volume 68 Issue 4, 261-265, April 2020

Original Article

Text Sentiment Analysis using Naïve Baye's Classifier

Afreen Jaha [1], N.Satya Deepthi [2], G.Suryakanth[3] ,G. Surya Kala Eswari[4]

[1,2,3] *Students Computer science Department & Pragati Engineering College,1-378, ADB Road, Surampalem, E.G.Dist.,A.P.*
[4]*Faculty, Computer science Department & Pragati Engineering College, 1-378, ADB Road, Surampalem, E.G.Dist., A.P.*

Received Date: 25 March 2020
Revised Date: 07 May 2020
Accepted Date: 09 May 2020

Abstract *- Sentiment analysis may be a study that analyses people's opinions, sentiments, evaluations, attitudes, and emotions from written communication. It's one of the foremost active research areas in tongue processing and is additionally widely studied in processing, Web mining, and text mining. This research has pervaded outside of computing to the management sciences and social sciences to business and society as a whole. The evolving importance of sentiment analysis coincides with the expansion of social media like reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. For the primary time in human history, we now have an enormous volume of opinionated data recorded in digital form for analysis. Today Sentiment analysis is being applied in most businesses and social domains because opinions are key influencers to most human activities and behaviours. Our beliefs and perceptions of reality, and thus the alternatives we make, are largely conditioned on how others see and evaluate the earth. For this reason, once we need to make a choice, we often seek out the opinions of others. This is often true not only for individuals but also for organizations. In this paper, we use Naïve Bayes classifiers in determining the sentiment embedded within the textual data.*

Keywords *- Sentiment Analysis, Naïve Bayes Algorithm, Natural Language Toolkit*

I. INTRODUCTION

In this paper, we've discussed how sentiment is extracted from a text. It is a region where the users give their views and opinions to support the case. The main objective of our proposed system is to perform analysis on the text having sentiment, which causes great help to business intelligence on predicting the long run. This document addresses the sentiment analysis on the text, that initial classification is performed on text using Naïve Bayes classifier. Each text is represented within the sort of sentiment asserted in terms of positive, negative and neutral. Performing sentiment analysis is important, which is utilized for hunting out the pros and cons of their products within the market by a public that finishes up in improving their business productivity.

II. PROPOSED MECHANISM

A. Introduction To Sentiment Analysis

The history of the web has changed the way people talk about their perceptions. Nowadays, it is all done through different blog posts, online discussion forums, product review websites etc. People usually rely upon user-generated content on any product to a good extent when it involves performing any desired action. When people want to shop for a product online, they will first search its reviews therein particular product website online before making a call. Some analysis is to be done on of these reviews in order that the ultimate outcome says whether the merchandise is nice to shop for or not. There are different sentiment analysis techniques that are available with many applications for various domains, like in business, to induce feedback for products from customers. Cognitive content and Machine learning techniques are two techniques that are mainly used for sentiment analysis. Within the case of the database approach, this needs an oversized database with predefined emotions and an efficient and effective knowledge representation for identifying sentiments. Within the case of Machine learning approach doesn't require any predefined set of emotions. This makes use of a training set so as to develop a sentiment classifier that classifies sentiments from the tweets, and then the machine learning approach is very simpler than the knowledge-based approach. There are different machine learning techniques that are accustomed classify data, i.e., they are naïve Bayes classifier, support vector machine, decision tree, random forest, neural networks etc. Classification may be a technique that is employed to perform classification on different sets of information into different classes. These classification techniques are divided into two categories Supervised and unsupervised. In the supervised learning approach, the pc learns from the labelled input file that's given thereto and so makes the pc use this learning which is employed to classify output data. During this, the dataset could also be within the style of a class, i.e., identifying whether the

statement is positive or negative or it's going to be multi-class too, i.e., the statement could also be positive or negative or neutral. In an unsupervised learning approach, the pc learns with the unlabelled input file and which is employed in grouping the info, for instance, cluster analysis.

B. Phases Involved

The aim of this project is to develop a classification technique using machine learning which provides accurate results and automatic sentiment classification of a text by predicting the longer term. In this document, sentiment analysis is completed on text data. So in order to classify data first, we'd like to perform the subsequent steps.

● Tokenization: it's a technique that divides the range of documents into small parts called tokens. These tokens are also within the sort of words or numbers or punctuation marks.

Ex: it's going to rain today

After performing tokenization, the sentence is split into tokens as follows

It, is, going, to, rain, today.

● Stop words: These are the common words that are to be ignored, which reduce the natural language of the dataset also the number of words (tokens). In our programming language, python, we use a tool called Natural language toolkit (NLTK), in which there is a list of stop words in 16 different languages.

Ex: i prefer dancing, so I dance.

After removing stop words, the sentence is as follows

Like, dancing, dance.

● The bag of words concept is applied to these tokens.

● Finally, our classification technique Naïve Bayesian classifier, is applied which calculates the probability of all the words in the document and offers the result, i.e., probability of each text in both and negative.

● Results show the probability of each text saying whether the text is either positive or negative.

C. Naive Bayes Algorithm

1. Consider a training data set X consisting of documents that belong to different classes, say class A and B.

2. Foregoing, probability of both classes A and B is calculated as shown

Class A=Number of objects of class A/total number of objects.

Class B= Number of objects of class B/total number of objects.

3. Now calculate the entire number of word frequencies of both classes A and B, i.e., n_i

n_a=the total number of word frequency of cla

ss A

n_b=the total number of word frequency of class B

4. Calculate the conditional property of keyword occurrence for given class P(word1 / class A) = wordcount / n_i(A) P(word1 / class B) = wordcount / n_i(B) P(word2 / class A) = wordcount / n_i(A) P(word2 / class B) = wordcount / n_i (B) …… …… P (wordn / class B) = wordcount / n_i (B)

5. Uniform distributions are to be performed so on avoid zero frequency problems.

6. Now a replacement document M is assessed supported calculating the probability for both classes A and B P (M/W). a) Find P (A / W) = P (A) * P (word1/class A)* P (word2/ class A)……* P (wordn / class A). b) Find P (B / W) = P (B) * P (word1/class B)*P (word2/ class B)……* P (wordn / class B).

7. After calculating the probability for both classes A and B, the category with higher probability is the one the new document M assigned.

III. FIGURES AND TABLES

The following are the phases undergone in the process of sentiment analysis.

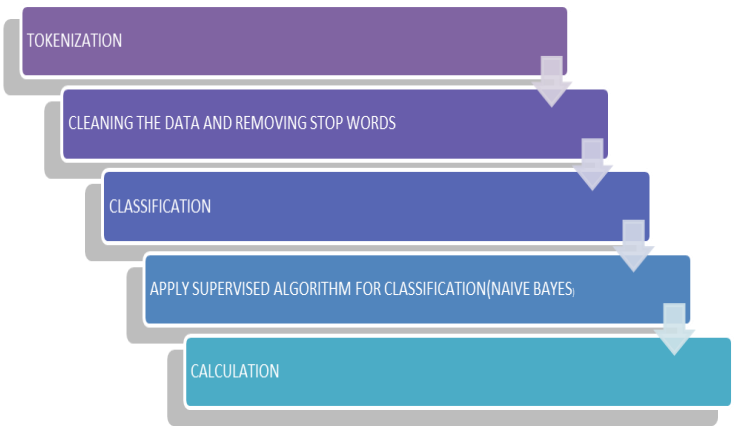

Fig.1 Phases of Sentiment Analysis

Table 1. frequency of each word represente as count

Word	Count
It	1
Is	1
so	1
hot	2
Today	1
I	2
Am	2
Not	1
outside	1
Watch	1
The	1
Season	1
Premiere	1

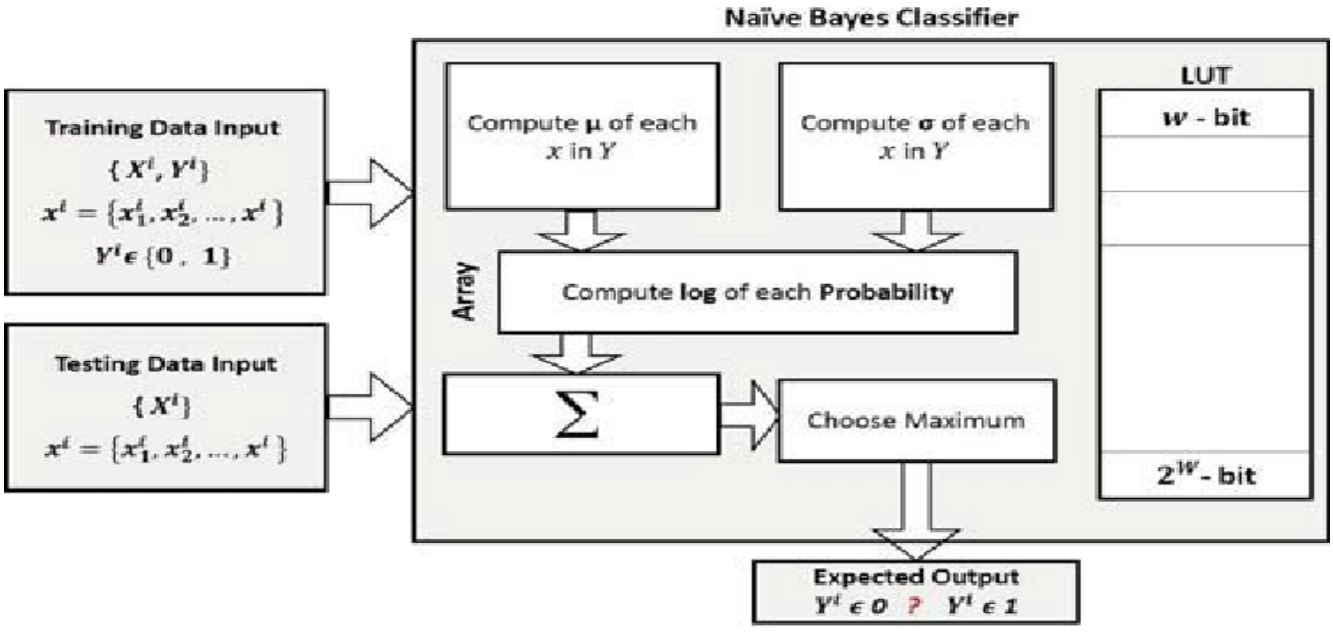

Fig. 2 Naïve bayes classification

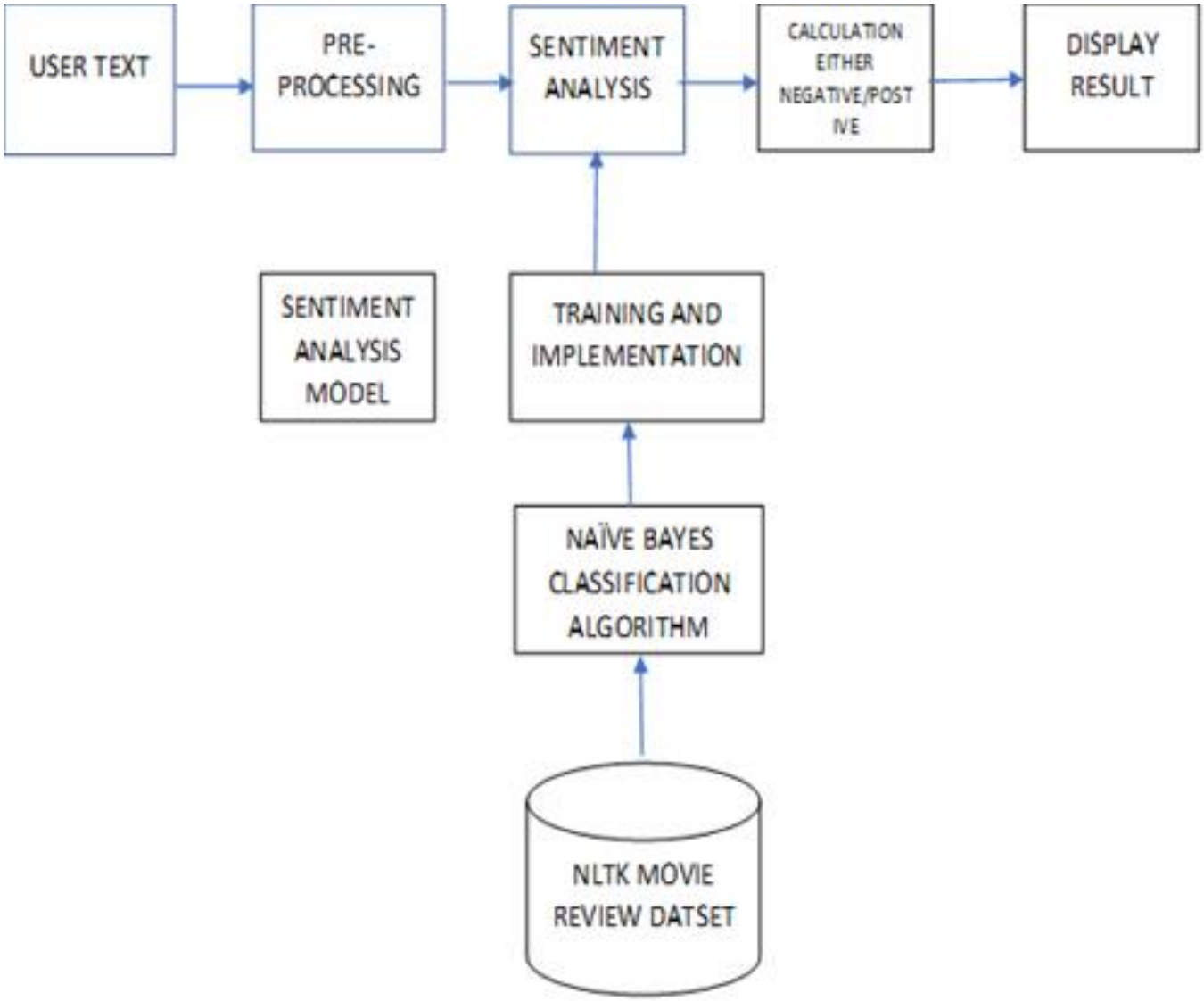

Fig. 3 Architecture of sentiment analysis

IV.CONCLUSION

As we can see, even a really basic implementation of the Naive Bayes algorithm can cause surprisingly good results for the task of sentiment analysis. Notice that this model is actually a binary classifier, meaning that it is often applied to any dataset during which we've two categories. There are all types of applications for it, starting from spam detection to bitcoin trading supported sentiment. With an accuracy of 82%, there is really a lot that you could do, all you need is a labelled dataset, and of course, the larger it is, the better!

V.ACKNOWLEDGMENTS

We express our deep sense of gratitude to our respected and learned guides, Prof. G.Surya Kala Eswari for their valuable help and guidance. We are thankful to them for the encouragement they have given us in completing the project.

We are also grateful to respected Prof. Dr M.Radhika Mani HOD (Department of CSE), for permitting us to utilize all the necessary facilities of the institution.

We are also thankful to all the other faculty & staff members of our Department for their co-operation and help.

Lastly, we would like to express our deep appreciation towards our classmates and our indebt ness to our parents for providing us with moral support and encouragement.

REFERENCES

[1] M. Vadivukarassi, N. Puviarasan and P. Aruna, Sentimental Analysis of Tweets Using Naive Bayes Algorithm, [Publication Details: World Applied Sciences Journal, 35(1) (2017) 54-59. DOI: 10.5829/idosi.wasj.2017.54.59]

[2] Ankur Goel, Jyoti Gautam; Sitesh Kumar, Sentimental Analysis of Tweets Using Naive Bayes Algorithm, [Publication Details., 2nd International Conference on Next Generation Computing Technologies (NGCT), DOI:10.1109/NGCT.2016.7877424], (2016).

[3] Vik Paruchuri, Predicting Movie Review Sentiment with Naive Bayes, [blog:https://www.dataquest.io/blog/naive-bayes-tutorial/]

[4] Filip Knyszewski Machine teacher @ University of Amsterdam Linkedin:https://www.linkedin.com/in/filip-knyszewski-11141fk/

[5] HumaParveen and ShikhaPandey Sentiment analysis on Twitter Dataset using Naive Bayes algorithm, 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), (2016), 416-419 @article{Parveen2016SentimentAO}.

[6] http://www.international journalssrg.org/ssrg-journals.html

[7] Le B., Nguyen H., Twitter Sentiment Analysis Using Machine Learning Techniques. In: Le Thi H., Nguyen N., Do T. (Eds) Advanced Computational Methods for Knowledge Engineering, Advances in Intelligent Systems and Computing, Springer, Cham, 358 (2015).

[8] Sayali P. Nazare, Prasad S. Nar, Akshay S. Phate, Prof.Dr. D. R. Ingle Sentiment Analysis in Twitter International Research Journal of Engineering and Technology (IRJET), 5 (2018).

[9] Anuja Prakash Jain and Padma Dandannavar Application of machine learning techniques to sentiment analysis, 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), (2018) 628-632. article {Jain2016ApplicationOM}

[10] Mejova, Yelena., Sentiment Analysis, An Overview, (2019).