iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 7 **Issue:** XII **Month of publication:** December 2019

DOI: http://doi.org/10.22214/ijraset.2019.12146

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: 7.177
Volume 7 Issue XII, Dec 2019- Available at www.ijraset.com

Sentiment Analysis using Three Different Algorithms

Shivani Pathak[1], Piyusha Mahajan[2], Ankita Patil[3], Rutuja Patil[4], Mrs. M. M. Swami[5]
[1, 2, 3, 4, 5]*Department of Computer Engineering, All India Shri Shivaji Memorial Society's College of Engineering, Pune*

Abstract: Sentiment analysis is an application of NLP (Natural Language Processing). It is also known as emotion extraction or opinion mining. It is a very popular field of research in text mining. It helps us to understand the human decision making or to categorize or it can be used as a task for analyzing and extracting opinions from review documents on web sites, blogs, social media, and others in order to understand the opinions of consumers. The mild sentiments of users when analyzed can help improving the brand's reputation. We have proposed an approach which combines feature engineering with DAN2 to identify mild sentiments.

Keywords: Natural language Processing, Sentiment analysis, Machine learning, Feature engineering, DAN2 classifier, SVM, Naïve Bayes

I. INTRODUCTION

In today's world social media is in trend such as Facebook, Twitter with which anyone can comment regarding their opinion on any topic. These comments can be analyzed to find the hidden motive behind them. This can be done with the help of Sentiment Analysis. Sentiment analysis is an application of NLP which can help us in opinion mining by analyzing polarity of sentence. Polarity of sentence is nothing but the subjective expression categorized as positive, negative, neutral.

Sentiment analysis is widely used as an application of review system, survey response system, marketing, for analysis of product recommendation, etc. These systems can categorize sentences into polarity, can identify emotion or sentiment (happy, sad, angry) or can mark product (interested or not interested). The data of various companies related to their services or product is stored in unstructured format so to analyse such data manually is time-consuming and exhaustive. Sentiment Analysis helps to scale the data efficiently and is cost-effective. It also helps in analysing situations by identifying critical information from the reviews and then performing action and spreading awareness of the situations in Real-time. Sentiment Analysis can be done in different type in which it can perform analysis of both text and voice.

There are various algorithms used for sentiment analysis such as SVM (Support Vector Machine), Naïve Bayes classifier, neural network classifier, etc. SVM and Naive Bayes algorithm can be efficient for small dataset but as the data increases the efficiency of these algorithms may decrease. Hence to overcome this problem neural network classifier such as DAN2 (Dynamic architectural artificial neural networks) algorithm can be used. It can help in the analysis of huge amount of data and it also provides scalability. Classifiers can identify only strong opinion but mild opinion also play vital role in brand management or for improvising reputation of brand. Feature engineering used in sentiment analysis with which mild opinion can be identified. DAN2 classifier with Feature engineering can be used for training and testing of data by which polarity of both strong and mild opinions and emotions can be identified.

II. LITERATURE SURVEY

Sentiment analysis depends on the nature of the text, which is either subjective or objective. The objective statements are factual in nature and cannot be used for extracting any emotions. For example, a sentence like "McDonalds has many outlets in India", demonstrates a fact and does not contain any emotions. The subjective statements can be used for identifying the emotions at document, sentence or phrase level. While understanding the sentiments in a text (subjective), first the text is pre-processed using different techniques and tagged into different POS (Parts of Speech). Sentiments are classified using two main approaches- Subjective lexicons and Machine Learning [1].

Lexicon-based approaches extract opinions on the basis of dictionaries. Hassan Saif, Yulan He, Miriam Fernandez, Harith Alani suggested a lexicon-based approach- SentiCircle, which combines the context of the words used in the text for analysing the sentiments. The outcomes of SentiCircle are better than the other commonly used lexicon-based approaches. However sentiment lexicons are a costly and time consuming strategy [2].

Supervised machine learning techniques provide a better accuracy than subjective lexicons. Some of the most common classifiers used are SVM (Support Vector Machine) and Naïve Bayes. Naive Bayes classifier is a probabilistic approach which is derived from Bayes theorem. The implementation of Naïve Bayes is easy but the assumption of strong independence between features is not accurate [3].The SVM classifier is used in a dual prediction technique [4] where the emotions are extracted by finding the average of evaluation of sentences in both forward and reverse directions. Another approach uses optimized SVM by adding Radial Basic Kernel function (RBF) to traditional SVM [5]. The optimized SVM performs better than the traditional SVM and Naïve Bayes.

Before applying classifier models, first the features of the data need to be extracted efficiently. One of the approaches first identifies relationships and constructs concepts based on them using ConceptNet source. These concepts are condensed significantly using the Minimum Redundancy and Maximum Relevance (mRMR) technique [6]. The automatic keyword selection (AKS) is another feature selection technique which outperforms the mRMR technique. The AKS technique reduces the training time required for classifiers like RBF, MLP (Multi-Layer Perception), Naïve Bayes and Decision Tree. This technique is very efficient for huge training datasets [7].

The common supervised learning classifiers used for sentiment analysis like SVM and Naïve Bayes are very old and require excessive computations apart from classification like selection of a kernel function for SVM. A more recent classifier used for sentiment analysis is the Dynamic Architecture for Artificial Neural Networks (DAN2). Ghiassi, M. and Saidane, H developed the DAN2 model which is a modification of the traditional ANN. It is a purely feed-forward model. DAN2 provides more scalability as compared to previous classifiers. Also it uses all of the samples for training which reduces the training SSE (Sum of squared errors) or MSE (mean squared error). DAN2 accumulates results at every stage and the user does not have to decide the number of hidden nodes for every hidden layer [8]. The DAN2 classifier can also be used effectively for automated text classification and gives better accuracy than SVM and kNN (k-Nearest Neighbour). This is especially because of the dimensionality reduction property of DAN2 which is necessary due to the complexity of the texts used for classification [9].

Most of the approaches used for sentiment analysis usually display only the strong sentiments of the users. The users having mild sentiments can be targeted by the brand for change in opinion. These mild sentiments of the users can be identified by using feature engineering along with DAN2. David Zimbra, M. Ghiassi, and Sean Lee [10] presented an approach which combines feature engineering and DAN2 classifier for sentiment analysis. The outcome of this approach outdoes the accuracy of SVM and Naïve Bayes [10].

III.PROPOSED SYSTEM

The proposed system will take input of twitter dataset as excel/csv sheet. The data is then pre-processed using NLP (Natural Language Processing) techniques like tokenization, stemming, removal of stop word, etc. This preprocessed data is passed on for transformation of it into desired output. The system is going to find emotion on the basis of polarity. SVM and Naïve Bayes algorithm categorizes polarity into positive, negative and neutral.

The diagrammatic representation of our system is given as below in Fig. no. 1. The system will perform sentiment analysis on the input dataset using SVM classifier, Naïve Bayes classifier and feature engineering combined with DAN2. The outputs of these classifiers will be compared on the basis of their accuracy.

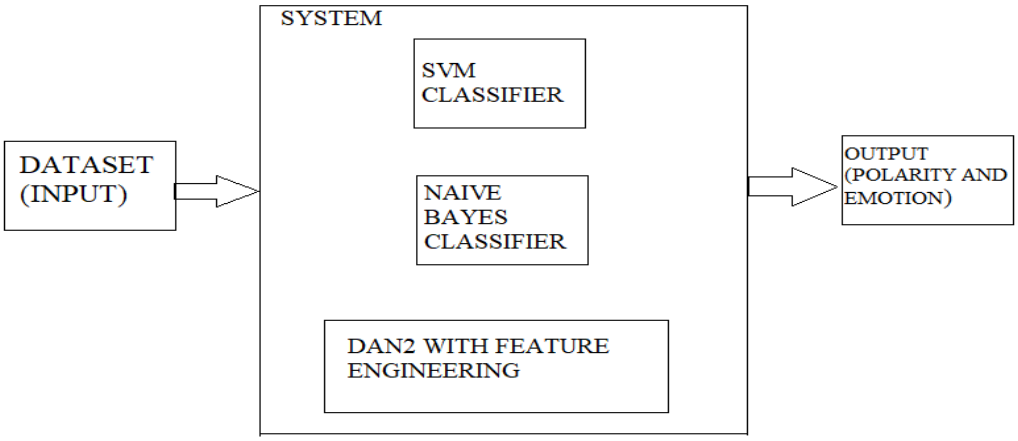


Fig no. 1: The Proposed System

For the purpose of identifying mild sentiments in the dataset, DAN2 classifier can be combined with feature engineering.

A. Working of DAN2 with Feature Engineering

The opinions are extracted in the form of scores which represents polarity of given sentence using feature engineering functions. These scores can be found using feature engineering technique like Senti-Word-Net dictionary which help in the scoring of sentence and finding polarity of sentence.

The output of data of feature engineering can be used as an input for the DAN2 algorithm. DAN2 algorithm is a purely feed-forward model. DAN2 categories the sentence into 5 type of polarity as mildly positive, strongly positive, neutral, mildly negative, strongly negative, etc. With the help of this polarity and polarity score we can further classify sentence to find type of it like angry, happy, sad or maybe even sarcastic.

B. Architecture Description of DAN2 Algorithm [8]

The number of hidden nodes in each hidden layer is fixed a priori. The DAN2 model architecture contains four types of nodes:

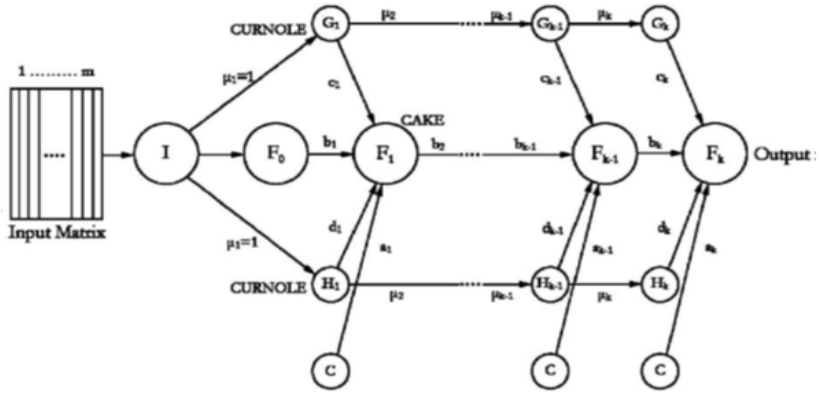


Fig no 2: DAN2 architecture diagram

1) *Constant Input Node(C):* contains input
2) *Current Accumulated Knowledge Element (CAKE or F_k):* contains linear data computed from previous training steps, receives its input from all nodes in the previous layer
3) *Current Residual Nonlinear Component (CURNOLE or G_k):* represent the remaining non-linear component which was not processed earlier
4) *Current Residual Nonlinear Component (CURNOLE or H_k):* represent the remaining non-linear component which was not processed earlier

The number of layers in DAN2 architecture depends on the input dataset and they are generated dynamically. DAN2 (Dynamic Architecture for Artificial Neural Network) algorithm is a purely feed-forward model. It's architecture is similar to traditional ANN but no. of hidden nodes in hidden layers are fixed in DAN2 due to which it's accuracy is better than other algorithms [9].

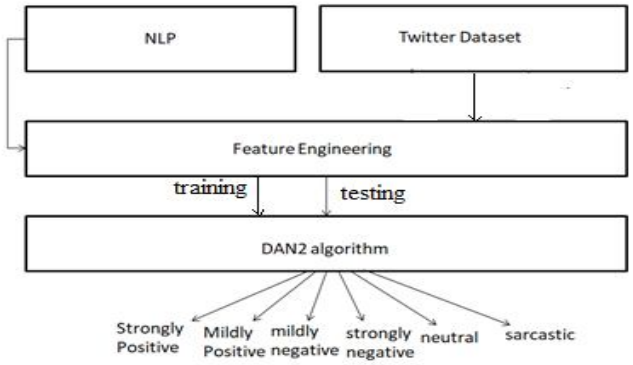


Fig no 3: Working of DAN2 with feature engineering

IV. CONCLUSIONS

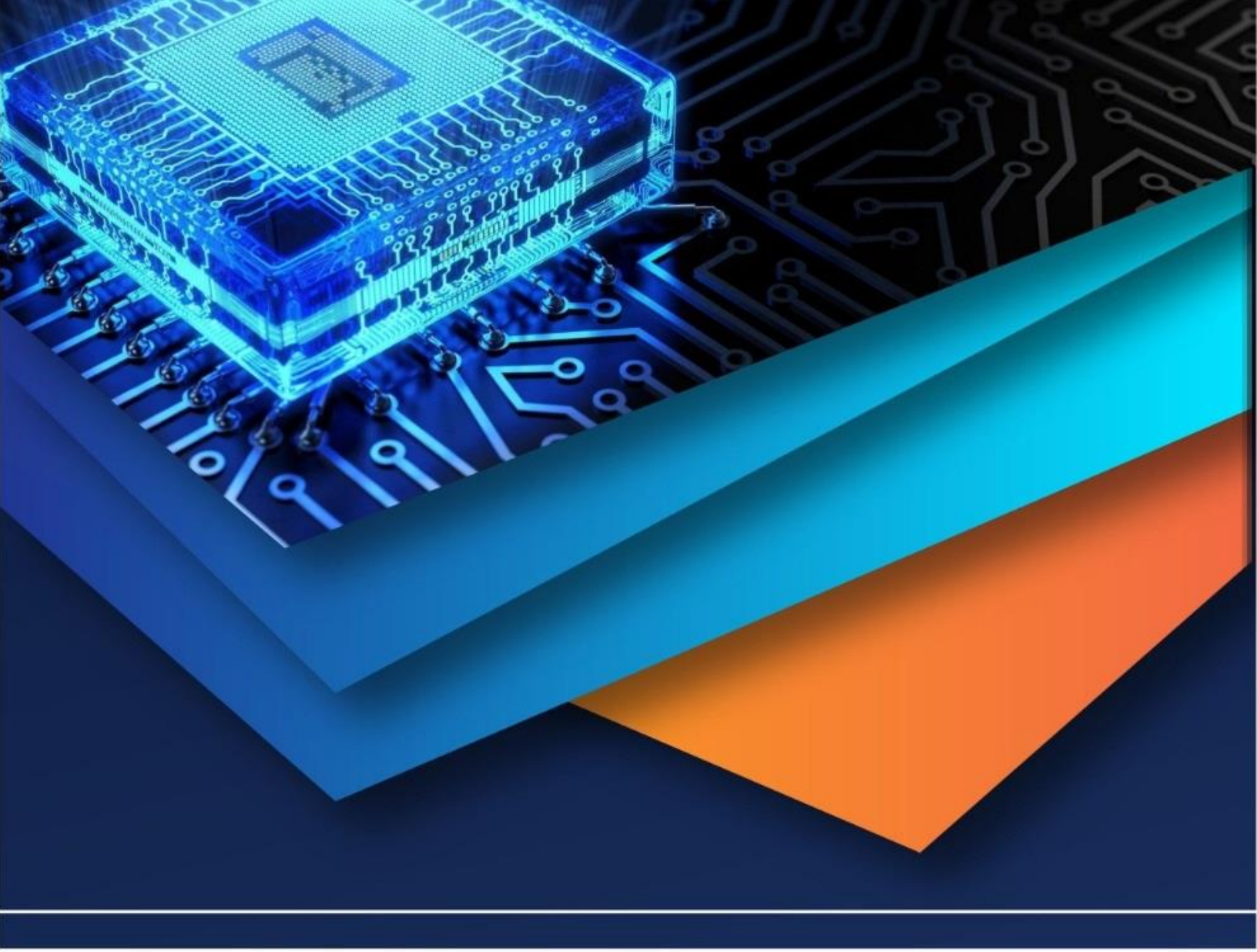
This paper studies the merits and demerits of the classifiers Naïve Bayes, SVM and DAN2. Among all the three classifiers, DAN2 is expected to give better accuracy when combined with feature engineering. This approach can be further enhanced by identifying sarcastic comments as well.

V. ACKNOWLEDGMENT

We would like to extend our sincere gratitude to our guide Mrs. M.M.Swami for her invaluable guidance and for giving us useful inputs and encouragement time and again, which inspired us to work harder.
We are extremely grateful to Mr. D.P.Gaikwad, Head of Department of Computer Engineering, All India Shri Shivaji Memorial Society's College of Engineering, Pune for his encouragement during the course of the project work.

REFERENCES

[1] Harpreet Kaur, Veenu Mangat, Nidhi, "A Survey of Sentiment Analysis techniques", International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC 2017).
[2] Hassan Saif, Yulan He, Miriam Fernandez, Harith Alani, "Contextual semantics for semantic analysis of Twitter", Information Processing Management, Volume 52, Issue 1, Jan 2016, ScienceDirect.
[3] Mykhailo Granik and Volodymyr Mesyura, "Fake news detection using naïve Bayes classifier", First Ukraine Conference on Electrical and Computer Engineering(UKRCON), IEEE, 2017.
[4] Minara P Anto, Nivya Johny, Mejo Antony, Vinay James, Muhsina K M, Aswathy Wilson, "Product Rating Using Sentiment Analysis" ,International Conference on Electrical, Electronics and Optimization Techniques(ICEEOT) – 2016.
[5] Bhumika M. Jadav, Vimalkumar B. Vaghela, "Sentiment Analysis using Support Vector Machine based on Feature Selection and Semantic Analysis", International Journal of Computer Applications, Volume 146 - No. 13,2016.
[6] Basant Agarwal, Soujanya Poria, Namita Mittal, Alexander Gelbukh, Amir Hussain, "Concept-Level Sentiment Analysis with Dependency-Based Semantic Parsing: A Novel Approach, Volume 7, Issue 4, August 2015, Springer.
[7] Niphat Claypo Department of Computer Science, Saichon Jaiyen Department of Computer Science, "Automatic Keyword Selection for Sentiment Analysis using Class Dependency and Dissimilarity", 2016.
[8] Ghiassi, M. and Saidane, H., "A Dynamic Architecture for Artificial Neural Network", Neuro computing, (63), 2005.
[9] M. Ghiassi, M. Olschimke, B. Moon, P. Arnaudo, "Automated Text Classification using a Dynamic Artificial Neural Network Model", Expert Systems with Applications, (39), 2012.
[10] David Zimbra, M. Ghiassi, and Sean Lee, "Brand-Related Twitter Sentiment Analysis Using Feature Engineering and the Dynamic Architecture for Artificial Neural Networks." 49th Hawaii International Conference on System Sciences (HICSS), IEEE, 2016.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY