Sentiment Analysis : A Review and Comparative Analysis on Colleges

Toshita Chandurkar[1], Dr. Pritish Tijare[2]

[1]Computer Science and Engineering Department, Sipna College of Engineering and Technology, Amravati, Maharashtra, India

[2]Associate Professor, Sipna College of Engineering and Technology, Amravati, Maharashtra, India

ABSTRACT

Sentiment analysis is the process of detecting positive and negative sentiment in text. It's often used by businesses to detect sentiment in social data, gauge brand reputation, to understand it and make better. Sentiment analysis models focus on polarity (positive, negative and neutral) and even intentions (interested v. not interested). Depending on how you want to interpret feedback and queries, you can define and tailor your categories to meet your sentiment analysis needs. This paper focuses the reviews of college which are an important form of opinion contents. The basic objective of this work is to classify every sentence's semantic orientation (e.g. positive, negative and neutral) of the reviews. It is a really useful analysis since we could possibly determine the overall opinion about the colleges.

Keywords : Sentiment Analysis, Machine Learning, Opinion Mining

I. INTRODUCTION

Nowadays the undeniable truth is that social media users are more likely to share how they feel about a current "hot topic" on social media platforms. It is current affairs that may trend regionally or globally. Users may post positive, negative or neutral opinions about that topic or a particular product they are using. The advancement of Machine Learning has opened doors in detect and categorize online opinions [1]. This study aims at proposing a Machine Learning model to detect opinion in unstructured texts for different college regarding different factors. We analyze the sentiments of user views about the recently controversial issues and compare them with the related different major issues like campus placement, infrastructure facilities, LAB facilities, Teaching methodology, etc. Analysis of sentiments is the method of deciding whether the sentiment in the text is positive, negative or neutral. It is also known as material polarity or mining of opinions.

Existing research has produced numerous techniques for various tasks of sentiment analysis, which include both supervised and unsupervised methods [2]. In the supervised setting some papers used all types of supervised machine learning methods (such as Support Vector Machines (SVM), Maximum Entropy, Naïve Bayes, etc.) and feature combinations. About a decade ago, deep learning has emerged as a very

powerful machine learning technique and produced state-of-the-art results in specific domains, ranging from computer vision and speech recognition to NLP. Applying deep learning method to sentiment analysis has also become very popular recently. But in practice, the long-range dependencies are still problematic to handle. So this paper presents maximum entropy for college reviews sentiment analysis.

Machine learning approach is proposed for the efficient classification of college data. To increase the efficiency of classification process, the technique parallel processing is implemented at each node. After training the dataset the machine is analyses with various set of test data [3]. Every day massive amount of data is generated by users which can be used to analyze their opinion about any college. Conventional tools like Machine learning analyze and process data in real time making analyzing and processing of real time data possible.

II. Related Work

The research in the college field, identify and predicts the views besides offering the right guidance with machine learning techniques. The machine learning techniques have been utilized by a wide verity of works in the analyzing method including: twitter handling, emotions detection and the like. Information associated with the reviews prevailing in the form of comment i.e. teaching, course content, examination, lab work, library facilities, extracurricular; were employed in all these works [4]. The suggested model included pre-processing comments as first step on given data and with the feature extraction model, an effective feature was generated. The outcomes have demonstrated the opinion about given college. Other researchers proposed semi supervised approaches for sentiment classification with a small amount of label data. In

the recent time, the machine learning approach was utilized by many researchers to analyses twitter data and many more for diverse result.

Sentiment Analysis

Sentiment analysis of college review technique is one of the prediction detection systems. The benefit of this is to know public opinions and extract their emotions by considering them and explained how database gives advantage during admission process. This system suggested that how to choose good one for bright future. It opted two stage approach for their framework- first preparing training dataset from given database using machine learning relevant features, after collecting and preprocessing the dataset training data set was created [5]. Next by using Sentimental Analyzer output gets the polarity in percentage. This approach reduced the number of training set and further they applied Machine learning approach and classification algorithm to determine the polarity of database.
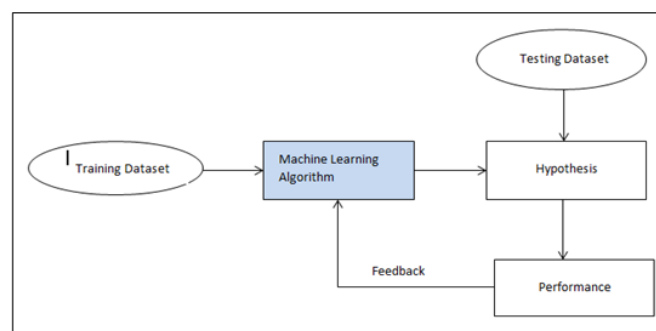

Fig. 1. Sentiment Analysis Using Machine Learning

The common approach found in almost all relevant research works constitutes data collection, preprocessing of data, filtering of data then approaches in feature extraction, classification and pattern analysis makes the distinction. This kind of information forms a basis for people to evaluate, rate about the performance of not only any college but about other products and to know about whether it will be a success or not. Therefore, sentiment analysis

has wide applications and includes emotion mining, polarity and influence analysis.

Problem Analysis

Today Sentiment Analysis is widely used in many sectors, but the existing system is getting old and new generation faced many problem and the main problems that exist in the current techniques are: inability to perform well in different domains, inadequate accuracy. Performance in sentiment analysis based on insufficient labeled data, incapability to deal with complex sentences that require more than sentiment words and simple analyzing [6]. So there is scope to implement a new system in the direction of reducing computational complexity and accurate output.

III. Proposed Work

In this paper for the analysis and classification of the colleges' data with various attributes is considered. This dataset contain 6 attributes we refer all of them to generate result. This dataset is freely available to the internet. A comment contains a lot of opinions about the data which are expressed in different ways by different users. The dataset used in this survey work is already labeled into classes' viz. negative, positive and neutral polarity and thus the sentiment analysis of the data becomes easy to observe the effect of various features.

The workflow of project by using Sentiment Analysis given in figure 2, which provide the brief overview of fundamental steps that should be followed to apply machine learning algorithm on dataset:
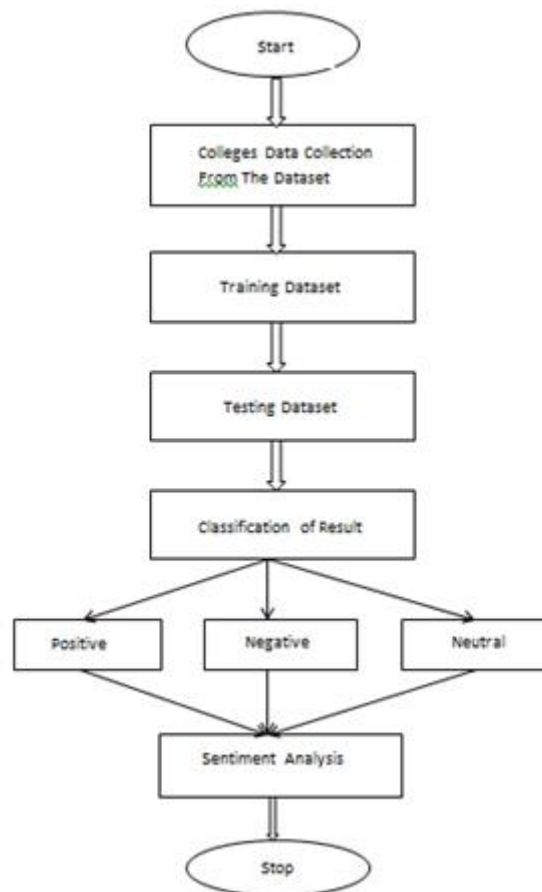

Fig. 2. Fundamental Steps

IV. PROCESS OF SENTIMENT ANALYSIS

4.1 Data Collection:

Dataset is in the form of raw comments which is retrieved. This dataset is nothing but the reviews on colleges stored. It contains sufficient number of colleges and opinion about the same, to be given as input to the training network.

4.2 Training Dataset:

Data in the training dataset must be pre-processed before the evaluation by the Sentiment analysis. It involves Tokenization which is the process of splitting the comments into individual words called tokens. Basically, in machine learning, training data is the data use to train a machine learning algorithm or model.

4.3 Testing Database:

Test data is used to measure the accuracy and efficiency of the algorithm used to train the machine - to see how well it can predict new answers based on its training. The accuracy increased due to the data testing, dataset should be verified by means of testing dataset different from the training dataset. The test data is given as the input to the trained dataset and actual output is obtained by classifying data, in the form of Positive, Negative and Neutral.

V. Comparison and Analysis

In this paper we are going to compare various machine learning algorithms and they are,

5.1 Naïve Bayes:

Naive Bayes is a classification algorithm suitable for any type of classification. But, Naive Bayes classifiers are highly scalable; it required a number of parameters linear in the number of variable in a learning problem. It does not satisfy for probabilistic prediction [7]. First we compute all the similar possible probability condition by training the dataset. Second, and then decide the actual probability by testing the dataset.

5.2 Support Vector Machine:

It is a supervised machine learning algorithm which can be widely used for both classification and regression challenges. And, SVM is better at computation speed and memory but worse at performance compared to deep learning.

5.3 Random Forest:

In a random forest algorithm, the set of many classification trees is generated by many decision trees taken together and proven effective for many sentiment tasks. This can be used mainly for regression and classification. The Decision Tree algorithm has rules for the given training data set with targets the features.

5.4 Decision Tree:

For classification the decision tree may be used with a structure like a tree. The best attribute of a dataset is put in a decision tree building algorithm, and then the training data set is divided into subsets. Decision trees are designed to generate a training model to predict the value of the destination variable.

5.5 Maximum Entropy:

It is a learning method based on empirical data and provides the probabilities for which sentence belongs to a particular class. While comparing to other it shows good result. Its principle is based on the most uniform significance method.

The key observations are shown in table 1, when Support vector machine (SVM) compared to Max entropy the SVM performance was poor. When no opinion word is detected, we can only make random predictions. We compared commonly used methods that use bag of words features with Naïve Bayes, SVM's, Random Forest and Decision Tree. We analyze performance on positive, negative and neutral prediction [8].

Table 1. Comparison Table

Algorithm	Accuracy	Implementation	Efficiency	Use
Naïve Byes	80%	Easy	Less	Limited
SVM	85%	Hard	Maximum	Versatile
Random Forest	90%	Hard	Maximum	Versatile
Decision Tree	70%	Easy	Maximum	Wide
Maximum Entropy	95%	Easy	Maximum	Wide

VI. Application

Sentiment analysis is applicable to many real world business problems and marketing such as social media monitoring, twitter comment, customer feedback etc. In the background of sentiment analysis, advanced AI algorithms apply language deconstruction techniques, like tokenization, part-of-speech tagging, parsing, and lemmatization to break down and make sense of text [9]. Only then can machine learning software classify unstructured text by emotion and opinion. Sentiment Analysis is a term that includes many tasks such as sentiment extraction, sentiment classification, and summarization of opinions or opinion spam detection, among others. It aims to analyze people's sentiments, attitudes, opinions emotions, etc. towards elements such as, products, individuals, topics, organizations, and services.

Sentiment Analysis of College Reviews

In the current era reviews and opinions play crucial role in planning sees from claiming clients and also influence businessman to the success of a brand and service of products? The same things also apply when deciding the best among various alternatives we have [10]. The work in this report is to perform sentiment analysis on the colleges. Sentimental analysis ascribes to utilization of ML and is the formalization for studying and constructing sentiments.

VII. CONCLUSION

Sentiment Analysis has led to development of better products, public behavior and good business management reviews. In this paper, we tried to show the basic way of classifying comments into positive, negative, neutral category using Machine Learning as baseline. We have performed conclusion examination for utilizing colleges' information. Therefore, look at the details, make a selection of the features, apply transformations and filter the less relevant data making machine learning methods generalize and effective since the computers these days have limits and can't handle them all the data without prior review of any kind [11]. It can find unfair positive reviews and unfair negative reviews, reputation issues, and collusion and control through this work.

VIII. REFERENCES

[1]. Nageswara Rao Moparthi, Dr. N.Geethanjali, "Design and implementation of hybrid phase based ensemble technique for defect discovery using SDLC software metrics", An International Conference by IEEE, AEEICB16 (978-1-4673-9745-2) PP. 269-276, ©2016 IEEE.

[2]. Wei Zhao, et.al., "Weakly-Supervised Deep Embedding for Product Review Sentiment Analysis", IEEE Transactions on Knowledge and Data Engineering, Vol. 30, No. 1, 185, 2018.

[3]. Rui Xia, Feng X, ChengqingZong, Qianmu Li, Yong Qi, Tao Li,Dual, "Sentiment Analysis: Considering Two Sides of One Review", IEEE Transactions On Knowledge And Data Engineering, VOL. 27, NO. 8, AUGUST 2015.

[4]. Ch. Nanda Krishna, Dr. P. Vidya Sagar, Dr. Nageswara Rao Moparthi, "Sentiment Analysis of Top Colleges", 4th International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB-18).

[5]. Chaturvedi, S., Mishra, V., & Mishra, N. (2017), "Sentiment analysis using machine learning for business intelligence", International Conference on Power, Control, Signals and Instrumentation Engineering IEEE (ICPCSI), 1(1), 2162–2166.

[6]. Dr.S.Gomathi, Rohini, Punitha. R, "Sentiment Analysis of Google Reviews of a College",

International Conference on Computing Intelligence and Data Science (ICCIDS 2018).

[7]. Suci Laila Ramdhani, Rachmadita Andreswari, Muhammad Azani Hasibuan, "Sentiment Analysis of Product Reviews using Naïve Bayes Algorithm: A Case Study", The 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT), 123-127, IEEE 2018.

[8]. Vivek Narayanan, Ishan Arora, Arjun Bhatia, (2013), "Fast and Accurate Sentiment Classification Using an Enhanced Naïve Bayes Model", Intelligent Data Engineering and Automated Learning – IDEAL, Springer, Vol. 8206.

[9]. Nontobeko Hlongwane, Yo-Ping Huang, Li-Jen Kao, "Using Sentiment Analysis to Determine Users' Likes on Twitter", 16th Int. Conf. on Dependable, Autonomic & Secure Comp., ©2018.

[10]. Dr.K.A.Waghmare, Miss. Sheetal K. Bhala, "Survey Paper on Sentiment Analysis for Tourist Reviews", 2020 International Conference on Computer Communication and Informatics (ICCCI -2020), Jan. 22 – 24,IEEE 2020.

[11]. Mais Yasen, Sara Tedmori, "Movies Reviews Sentiment Analysis and Classification", Jordan International joint conference on Electrical and Information Technology (JEEIT), IEEE 2019.

Cite this article as :