Performance of Naïve Bayes in Sentiment Analysis of User Reviews Online

Habeebullah Shah Quadri, R. K. Selvakumar

Abstract: ***Both sellers and buyers heavily depend on the opinions of customers in purchasing and selling products online. When it comes to text-based data, sentiment analysis of user reviews has become a prominent facet of machine learning. Text data is generally unstructured which makes opinion mining very challenging. A wide array of pre-processing and post-processing techniques need to be applied. But the major challenge is selecting the right classifier for the job. Naïve Bayes algorithm is a commonly used machine learning classifier when it comes to opinion mining and sentiment analysis. The focus of this survey is to observe and analyze the performance of Naïve Bayes algorithm in sentiment analysis of user reviews online. Recent research from a wide array of use-cases such as sentiment analysis of movie reviews, product reviews, book reviews, blog posts, microblogs and other sources of data have been taken into account. The results show that Naïve Bayes algorithm performs exceptionally well with accuracies between 75% to 99% across the board.***

Keywords: Naïve Bayes Classifier, Multinomial NB, Bernoulli NB, Gaussian NB, Sentiment Analysis, User Reviews, Ecommerce.

I. INTRODUCTION

User reviews play a crucial part in the decision-making process of a potential buyer on ecommerce portals. Customers are keen on finding the experiences of other people who have previously bought / used the product. Finding the opinion (i.e., negative / positive) of a person regarding their experience with a product by utilizing their review is known as *sentiment analysis* or *opinion mining*.

Sentiment analysis of text is a branch of machine learning and is of great importance to companies and their customers alike. It is a multi-step process that includes, *extraction, pre-processing, classification* and *analysis of text-data*. One of the major goals of sentiment analysis is to find the *polarity* i.e., the ratio of positivity / negativity / neutrality in text. Another important objective is to find the *subjectivity* i.e., the ratio of feelings as compared to facts in a text. The third most important aspect is the mining and classification of the sentiments itself, i.e., the ratio of anger, sadness, happiness, etc.,

Sentiment analysis can be performed at three stages of granularity [1]:

- *Document level:* The text (paragraph, document, etc.,) is analysed as whole.
- *Sentence level:* Text is tokenized into sentence level units and analysed individually.
- *Aspect level:* All features of a sentence are examined individually.

Revised Manuscript Received on November 26, 2020.

Mir Habeebullah Shah Quadri, student, Artificial Intelligence from CVR College of Engineering, Hyderabad, India.

Dr. R. K. Selvakumar, Professor Department of Computer Science and Engineering , CVR College of Engineering, Hyderabad, Telangana.

Sentiment analysis falls under the paradigm of Natural Language Processing (NLP), a field of research in Artificial Intelligence (AI). As such, sentiment analysis derives its principles and procedures from NLP. Text processing procedures such as *tokenization* of sentences or words, *removal ofstop words*, *Part-of-Speech (PoS) tagging*, *Named Entity Recognition (NER)*, etc., are some of the commonly used procedures in sentiment analysis. Machine learning classifiers such as Support Vector Machine (SVM), Naïve Bayes (NB), Logistical Regression (LR), k-Nearest Neighbor (kNN), Random Forest (RF), etc., are used for the sentiment detection and classification of user reviews. In this paper, we focus on the Naïve Bayes algorithm for sentiment detection and classification. We are going to begin by formally defining the Naïve Bayes algorithm and its variants. Next we are going to generalize the procedure of sentiment analysis into five steps and briefly discuss each step. This is followed by a detailed discussion of a wide array of methodologies that Naïve Bayes algorithm is currently being used in. Finally, we are going to analyze the performance of Naïve Bayes algorithm across all the different use-cases discussed.

II. NAÏVE BAYES CLASSIFIER

The Naïve Bayes algorithms is a probabilistic classifier used for predictive analysis. It is simpler as compared to other algorithms and has been known to have a higher success rate [1]. Naïve Bayes makes the assumption that all input attributes are conditionally independent [3]. It is highly scalable and works on the principle of learning by doing [6]. It takes preprocessed data with the extracted features required as input for training. Once trained, it can be used to provide polarity of a given input text, i.e., if the text is positive or negative [7].

A. Naïve Bayes

Bayes theorem states that the posterior probability, P(A|B) is calculating, from P(A), P(B), and P(B|A) is defined in Equation 1. Naive Bayes classifier assume that the effect of the value of a predictor (y) on a given class (X) is independent of the values of other predictors is defined in Equation 2. What makes Naïve Bayes distinct from the more traditional Bayes Belief Networks (BBN) is that in Naïve Bayes, we assume complete conditional independence of all input attributes [3]. The Bayes theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

where, A and B are two events that are independent upon each other and make an equal contribution to the outcome. If we generalize the Naïve Bayes, we get:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (2)$$

where, y is the class variable and X is the feature vector of size *'n'*, i.e., X = (x_1, x_2, x_3, …, x_n)

B. Multinomial Naïve Bayes

In sentiment analysis, for certain cases, finding the word frequency or discrete count can be beneficial in increasing the accuracy of the machine learning model. In such cases, Multinomial Naïve Bayes, a variant of the standard Naïve Bayes can be used. In MNB, the assumption is that the distribution of each feature, i.e., $P(f_i|C)$, is a multinomial distribution.

Multinomial distribution is defined in Equation 3.

$$P = \frac{n!}{(n_1!)(n_2!)\ldots(n_x!)} P_1^{n_1} P_2^{n_2} \ldots P_x^{n_x} \quad (3)$$

where, n is the total number of events, n_1 … n_x are the number of outcomes for each event and P_1 … P_x are the probabilities of the occurrence of each event.

C. Bernoulli Naïve Bayes

Bernoulli Naïve Bayes (BNB) is often used when the features are binary in nature, i.e., the features lie in the interval [0,1] exclusively. MNB takes the frequency of the occurrence of each token into account whereas BNB does not. BNB is generally used for document classification in machine learning.

Generation of a document using BNB is defined in Equation 4.

$$b_t = 1 \text{ iff } t \text{ occurs in } d \quad (4)$$

where, t is the term to search for and d is the document such that $d = \{e_1, \ldots, e_i, \ldots, e_m\}$ *and* $e_i \in \{0,1\}$.BNB performs best when the document size is small [5].

III. GENERALIZED PROCEDURE FOR SENTIMENT ANALYSIS

The procedure for sentiment analysis of text-based data using machine learning can be generalized to five major sub-processes that work iteratively to increase the overall performance of the machine learning classifier is shown in Figure 1.

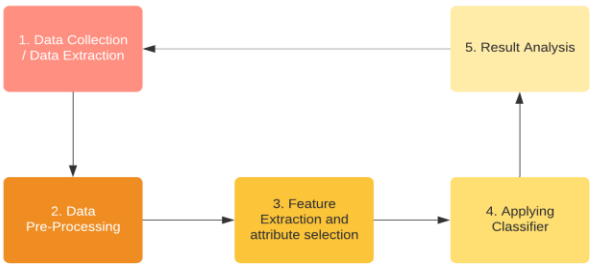

Fig 1.Design flow of sentiment analysis for text data

A. Data Collection / Data Extraction

As shown in Figure 1, data collection / data extraction is the first step in which the required text data for analysis is taken from the source. For the purpose of training/testing the model, the two most commonly used methods of data collection are using a predefined dataset [1] or by extracting the required data from the source to form a dataset [5].

B. Data Pre-processing

Almost all forms of text data require some type of pre-processing such as cleaning, tokenizing, removal of stop words etc., This increases the accuracy of the model used for the classification of sentiments [10].

C. Feature extraction and attribute selection

Not all of the data that is extracted may be of use for the specific problem that we are trying to address. Therefore, extracting the appropriate features and selecting the attributes that produce the most optimum results [11] can be crucial, both in terms of reducing time as well as space complexity.

D. Applying classifier

Applying a trained classifier to the data is where machine learning truly comes into picture for sentiment analysis. Mining the opinion of the user and reaching a conclusion as to whether the opinion is negative, positive or neutral with highest possible accuracy is the major challenge in this step [9]. A lot of the success of this step is predicated on the efficacy of the previous steps.

E. Result Analysis

The final step is the analysis of the results. A less optimal result may indicate an issue with the selection of attributes, or the classifier used. A lot of the research on sentiment analysis is focused on comparing the performance of different machine learning algorithms [3][6][12], which helps in gaining critical insight as to which algorithm fits a model best.

IV. METHODOLOGIES

Noor et al. [1] performed sentiment analysis on user reviews for women's products on the ecommerce website amazon.com using Weka. The main objective of the research was to perform opinion mining on user reviews and gain critical insights necessary for boosting sales of products that are specifically target towards women. Naïve Bayes, JRip, J48 and Sequential Minimal Optimization (SMO) algorithms were used for the training the model. While all the four algorithms used performed well, the highest accuracies were achieved using SMO and Naïve Bayes. Ahmed et al. [2] performed sentiment analysis on movie reviews using the aclimdb movie dataset. The research focused on creating a new sentence level sentiment analysis system for user reviews. It also focused on the effects that stop words and other attributes of a sentence have on the overall accuracy of the model. SVM, Naïve Bayes and Multi-layer Perceptron (MLP) were used for training of the model. Two variations of Naïve Bayes, namely the Bayes Net (BN) and Multinomial Naïve Bayes (MNB) algorithms were used. MNB performed with a higher accuracy. Alrehili et al. [3] applied sentiment analysis on a user reviews dataset from amazon.com. The main objective of this research was to use an ensemble method of machine learning, titled 'voting', which used classifiers, namely Naïve Bayes, SVM, Random Forest (RF), Bagging and Boosting to provide the sellers with a comprehensive picture of the user-reviews on their products.

The highest accuracy achieved was by Random Forest, followed closely by SVM and Naïve Bayes. Athuraliya et al. [4] used sentiment analysis along with other techniques to devise an application targeted towards the hotel management industry. The primary objective was to identify and extract maintenance related issues from the reviews posted by the customers of the hotel. A dataset from bookings.com was used for the analysis. SVM, MNB, Logistic Regression (LR) and RF algorithms were used to train the model. SVM performed with highest accuracy followed by MNB. Chakraborty et al. [5] proposed a new model of sentiment analysis for game reviews. The objective was to perform opinion mining of a given game using the popular microblogging platform twitter. A dataset of game related user reviews from amazon.com was used to train the model. Naïve Bayes, SVM, LR and Stochastic Gradient Descent (SGD) were used for training purposes. Naïve Bayes algorithm performed classification with the highest accuracy. Dholpuria et al. [6] used movie reviews to study the performance of sentiment analysis using different types of classifiers with a focus on studying the efficacy of Convolutional Neural Networks (CNN) in sentiment analysis w.r.t to classical methods such as NB, MNB, SVM, KNN, etc., The study showed that CNN, followed closely by Naïve Bayes and Logistical Regression (LR) performed with the highest accuracies. Gautam et al. [7] performed sentiment analysis to analyze customer reviews on various products from the internet. Datasets from major online retail companies such as amazon.com were used. Naïve Bayes algorithm, along with SVM, Maximum Entropy (MaxEnt) and Sematic Orientation (SO) based WordNet were used for the training of the model. The research showed that Naïve Bayes performed with the highest accuracy across datasets for sentiment analysis. Iqbal et al. [8] used the Stanford Twitter Sentiment140 and IMDb datasets to perform sentiment analysis. The focus of the research was to try different variations of feature set selection to improve the overall accuracy of the machine learning algorithms. Naïve Bayes, SVM and Maximum Entropy (MaxEnt), were used for training the model. Results showed the Naïve Bayes performed on par with MaxEnt with high accuracy for twitter dataset whereas MaxEnt performed slightly better than Naïve Bayes for the IMDb dataset. Kalra et al. [9] used sentiment analysis on news articles from the Indian news outlets. The focus of this research was to predict the fluctuations in stock market prices in India using historical news data to train the model. Naïve Bayes algorithm was selected for this task and performed with high accuracies with negligible amount of variations across different feature sets. Moh et al. [10] designed a multi-tier sentiment analysis approach with specific predefined stages such as data cleaning, data preprocessing, classification, etc., The analysis was performed on a dataset containing over 150,000 movie reviews. Four machine learning algorithms were used, namely, Naïve Bayes, SVM, Random Forest and Stochastic Gradient Descent (SGD). Results showed that Naïve Bayes outperformed all of the other algorithms in both single tier as well as multi-tier classification.

Mtetwa et al. [11] focused on understanding the effects of different feature selection techniques on supervised machine learning algorithms for sentiment analysis. A dataset of movie reviews was used for the analysis. Three supervised machine algorithms, Multinomial Naïve Bayes (MNB), SVM and Random Forest (RF) were used. Results showed MNB outperformed other classifiers when using Bi-grams and SVM and RF performed slightly better than MNB for Counter Vector and TF-IDF based selection.

Singh et al. [12] performed sentiment classification of movie reviews and blog posts using the SentiWordNet approach. The performance of SentiWordNet was then compared against SVM and Naïve Bayes. While the study showed that the performance of SentiWordNet was better than the classical methodologies, Naïve Bayes performed better than SVM. Yasen et al. [13] used tokenization, stemming and feature selection to aid the sentiment analysis of movie reviews. A comparative analysis of the classification was done using eight different classifiers, four of which performed significantly well, namely, Naïve Bayes, Decision Tree (DT), SVM and Bayes Network (BN).

V. PERFORMANCE ANALYSIS

From the performance metrics obtained by thirteen different research initiatives that we surveyed in this paper, we observe that Naïve Bayes algorithm and its variants (MNB, BN, BNB), perform exceptionally well with high accuracies ranging between 76-99%. Table 1 shows the methodologies used in each research initiative and the results obtained. When compared with its variants, Naïve Bayes performs on equal footing with Multinomial Naïve Bayes (MNB) and Bernoulli Naïve Bayes. For instance, in experiment [2], NB obtained 88.73% accuracy whereas MNB performed with 91.16% accuracy. In another instance, NB performed with 90.22% accuracy as compared to 89.32% accuracy of MNB

Table - I: Performance Analysis of Naïve Bayes Algorithm in different use-cases

No.	Methodology	Classifiers	Accuracy %
1	Sentiment Analysis for Women's E-commerce Reviews using Machine Learning Algorithms	SMO	80.87
		NB	**76.62**
		JRip	72.25
		J48	71.25
2	Challenges, Comparative Analysis and a Proposed Methodology to Predict Sentiment from Movie Reviews Using Machine Learning	SVM	97.2
		NB	**88.73**
		MNB	**91.16**
3	Sentiment Analysis of Customer Reviews Using Ensemble Method	SVM	87.87
		NB	**84.4**

		RF	89.87
4	"Revyew" Hotel Maintenance Issue Classifier and Analyzer using Machine Learning and Natural Language Processing	SVM	92.17
		MNB	**89.18**
		LR	90.88
		RF	92.09
5	Rating Generation of Video Games using Sentiment Analysis and Contextual Polarity from Microblog	**NB**	**90.22**
		MNB	**89.32**
		BNB	**90.01**
		LinearSVC	92.02
6	A Sentiment analysis approach through deep learning for a movie review	**NB**	**98.15**
		LR	98.82
		SVM	99.24
		KNN	98.41
7	Sentiment Analysis of Twitter Data Using Machine Learning Approaches and Semantic Analysis	**NB**	**88.2**
		SVM	85.5
8	Enhancing the Performance of Sentiment Analysis by Using Different Feature Combinations	**NB**	**89.0**
		SVM	89.0
		MaxEnt	90.0
9	Efficacy of News Sentiment for Stock Market Prediction	KNN	91.2
		SVM	83.8
		NB	**85.0**
10	On Multi-Tier Sentiment Analysis using Supervised Machine Learning	**NB**	**80.53**
		RF	83.71
		SGD	82.19
11	Feature Extraction and Classification of Movie Reviews	SVM	87.0
		MNB	**88.0**
		RF	84.0
12	Sentiment Analysis of Movie Reviews and Blog Posts	**NB**	**82.9**
		SVM	77.95
13	Movies Reviews Sentiment Analysis and Classification	**NB**	**81.83**
		DT	91.28
		SVM	87.45
		BN	**81.47**

and 90% accuracy obtained by BNB [5]. These differences, though miniscule, point to the fact that selecting the right variation of NB is important to obtaining optimum performance of the model. Another interesting observation that can be made is that SVM and Naïve Bayes are the most commonly used combination of classifiers in sentiment analysis. In terms of performance, as can be seen in Figure 2, SVM and NB perform on par with each other for sentiment analysis [2][3][6][7][8][9][12]. The highest variation in performance can be noticed in experiment [2] with SVM obtaining 97.2% accuracy and Naïve Bayes obtaining 88.73%. In the same experiment, MNB outperforms Naïve Bayes by obtaining the accuracy of 91.16%. The lowest variation in performance can be noticed in [8], where both SVM and Naïve Bayes performed with 89% accuracy. The highest accuracy measure achieved by SVM was 99.24% and the highest accuracy measure achieved by Naïve Bayes was 98.15% [6].

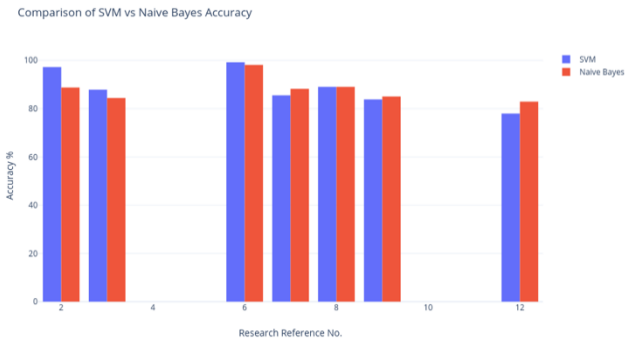

Fig 2.Comparison of performance of SVM vs Naïve Bayes

Another interesting comparison can be made between the performance of Naïve Bayes and Random Forest (RF) algorithm. As shown in Figure 3, Random Forest performed slightly better than Naïve Bayes in three experiments [3][4][10]. Naïve Bayes performed better than RF in experiment [11], obtaining an accuracy of 88% against the 84% obtained by RF.

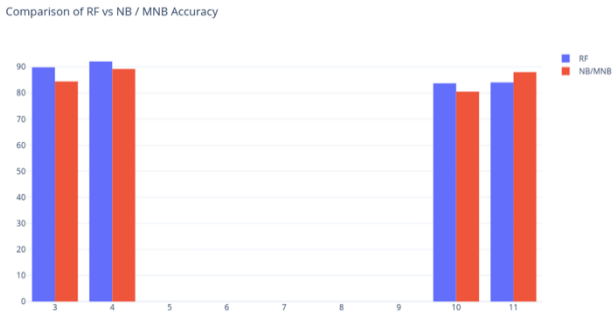

Fig 3. Comparison of performance of Random Forest vs Naïve Bayes

VI. CONCLUSION AND FUTURE WORK

In this paper, we have discussed the Naïve Bayes algorithm, its variants and its different use-cases in sentiment analysis and machine learning. We have observed use-cases of the algorithm (NB) 84.61% on a wide variety of datasets, including movie reviews, book reviews and other product reviews. We also observed the opinion mining of user reviews on microblogging platforms and the social media. In analyzing the results, it becomes abundantly clear that Naïve Bayes algorithm is a sophisticated choice for sentiment analysis of text-based big data. It is also observed that SVM tends to perform slightly better than Naïve Bayes in certain scenarios [2][3][6].

As per our review, we found that the algorithm (NB) is average of 85.92% accuracy got by the various researcher in their works. Choosing the right variant of the Naïve Bayes algorithm for a specific problem has been proven to be a more effective strategy in increasing the overall performance of the machine learning model [2][4][11]. Deep learning has shown a lot of promise in sentiment analysis with better performance than the classical machine learning methods [2][6]. In the future, more focus on using Naïve Bayes in collaboration with deep learning methodologies is required.

REFERENCES

1. Alaa Noor, Mohrima Islam, "Sentiment analysis for women's ecommerce reviews using machine learning algorithms", 10th ICCCNT, Kanpur, India, Jul 6-8, 2019.
2. ErfanAhmen, et al., "Challenges, comparative analysis and a proposed methodology to predict sentiment from movie reviews using machine learning", ICBDACI, 2017.
3. AhlamAlrehili, KholoodAlbalawi, "Sentiment analysis of customer reviews using ensemble method", ICCIS, 2019
4. BanukaAthuraliya, Cassim Farook, "Revyew hotel maintenance issue classifier and analyzer using machine learning and natural language processing", IEEE IEMCON, Vancouver, BC, Canada, Nov 1-3, 2018.
5. Shandro Chakraborty, et al., "Rating generation of video games using sentiment analysis and contextual polarity from microblog", CTEMS, 2018.
6. TanushreeDholpuria, Y. K. Rana, Chetan Agarwal, "A sentiment analysis approach through deep learning for a movie review", 8th ICCSNT, 2018.
7. Geetika Gautam, Divakar Yadav, "Sentiment analysis of twitter data using machine learning approaches and semantic analysis", 2014.
8. Nazma Iqbal, AfifaMim Chowdhury, Tanwir Ahsan, "Enhancing the performance of sentiment analysis by using different feature combinations", 2018.
9. Sneh Kalra, Jay Shankar Prasad, "Efficacy of news sentiment for stock market prediction", Com-IT-Con, India, Feb 14-16, 2019.
10. Melody Moh, et al., "On multi-tier sentiment analysis using supervised machine learning", IEEE/WIC/ACM, 2015
11. Nhamo Mtetwa, Awukam Ojang Awukam, Mehdi Yousefi, "Feature extraction and classification of movie reviews", 5th ICSCMI, 2018
12. V. K. Singh, et al., "Sentiment analysis of movie reviews and blog posts", 3rd IEEE IACC, 2013
13. MaisYasen, Sara Tedmori, "Movie reviews sentiment analysis and classification" IEEE JEEIT, 2019

AUTHORS PROFILE

Mir Habeebullah Shah Quadri is a master's degree student in the field of Artificial Intelligence from CVR College of Engineering, Hyderabad, India. He is a researcher in the field of AI and ML. His research and articles have been published in international scientific/tech journals and magazines. He is the founder of Code Latte and educational platform for teaching computer science. He is an open source enthusiast and contributes actively to several open source projects. He is a regular contributor to "Open Source For You" Magazine.

Dr. R. K. Selvakumar is Professor in the Department of Computer Science and Engineering at CVR College of Engineering, Hyderabad, Telangana. He has an M.Sc. in Mathematics from Madurai Kamaraj University. He also received PGDCA from University of Kerala. He received an equivalent M.Tech. in Computer Technology through DOEACC-C Level from DIT, Govt. of India. He hasanother M.Tech. in Computer &IT from Manonmaniam Sundaranar University, Tamilnadu, India. He has a Ph.D. in Computer Science & Engineering from the CITE, ManonmaniamSundaranar University, Tirunelveli, Tamilnadu, India. He is a Professional member in IEEE as a Senior Member (2018), CSI as a member and ISTE as a life member.