

Sentiment Analysis using Machine Learning Based Ensemble Model for Food Reviews

H.S.Priyanka , Dr. R.Ashok Kumar

Department of ISE, B.M.S College of Engineering, Affiliated to VTU
Bangalore, Karnataka

Email : hspriyanka98@gmail.com, ashokkumar.ise@bmsce.ac.in

Abstract: Sentiment analysis or determining the opinion is study about the people's sentiment, emotions and opinion that is expressed in terms of text. It's very popular because it helps us in understanding the opinions of other people. Which is the key influencer and important part of almost all human activities. The opinion of the people can be either positive or negative. The main objective of our work is to build a Machine Learning Ensemble model for Sentiment analysis. Ensemble strategies combine various learning algorithms to gain optimum performance. The Amazon Fine Food Reviews dataset has been used. The proposed model has been evaluated using various performance metrics such as Accuracy, Recall and Precision. Proposed system showed promising results when compared to individual base models.

Keywords: Sentiment Analysis, Machine Learning, Ensemble Model, Precision, Recall, Accuracy, Amazon Fine Food Reviews (AFFR).

I. INTRODUCTION

In today's world because of the advancement of technology, everything is getting digitalized. From entertainment to shopping items for daily needs everything we get online. The reviews and feedback provided by the customer are very important. These reviews and feedback help companies to understand customer needs and customer views. The reviews are associated with a negative opinion or positive opinion or neutral opinion. Generally, feedbacks/reviews are provided in text format. Sentiment analysis technique can be used for text classification. It provides various ways to understand the people's opinions, emotions that are expressed in the text format. Sentiment analysis can be implemented using Machine Learning. There are two types based on which text is classified. i.e, Subject/Object-based, Feature/Aspect based text classification.

Subject/Object-based classification: Here the given text is classified into one of two classes i.e, Subjective or Objective. This method is difficult when compared to the Aspect/Feature-based classification because the context needs to be known before classifying the text.

Feature/Aspect based classification: Here the opinions or the sentiment of the entities is captured based on the aspects or features. This approach is very much useful for identifying the polarity of any given text i.e, Reviews, Feedback etc. In this work Feature/Aspect based classification is used.

Machine learning is used to solve many complex problems. This can be supervised, reinforcement and unsupervised learning.

In supervised learning, the data along with the labels will be given to the machine. The machine will learn from this data and when new data comes it will infer based on its learning. In unsupervised learning, only data will be given to the machine, it will identify a similar pattern and creates different groups. Whenever new data comes it will map the data to one of the groups. Reinforcement learning is like a hit and trial method. agents will take some actions and it will

take the feedback from the environment and learn from the experience.

In this project supervised learning has been used. The machine learning models are trained using AFFR dataset [4][8]. The proposed method mainly focusses on building Machine Learning based ensemble model to achieve greater performance when compared to base models and we have observed promised results from the proposed model.

II. RELATED WORK

Levent Guner, Emilie Coyne and Jim Smit [1] have performed a comparative study on different machine learning algorithms for Amazon dataset. They have compared MultinomialNB, Long-Short Term Memory (LSTM), Linear Support Vector Machine (LSVM) algorithm. They have found LSTM performed well when compared to other model and had achieved the accuracy of about 90% (Amazon dataset) and 92% (for Scraped Amazon dataset).

According to Vijay Bhati, Jayveer Kher [2] sentiment analysis is a powerful technique to infer the opinion or sentiment or emotions associated with the text. It provides a detailed survey report for processing the user data through various steps such as cleaning and pre-processing of the data.

Shah F M, Haque T U, Saber M N [3] have performed sentiment analysis using Amazona dataset for the various product category. They have extracted data for various categories like musical data, cellphones and accessories and electronics. They have proposed supervised ML models using different algorithms for each of the categories of the reviews. They have achieved the highest accuracy of 94% using a linear support vector machine.

According to S Rajaraajeswari, Deepu S, Pethuru Raj [9] currently there are a lot of unstructured data. Preprocessing of these data is required before inferring anything from it. Bag-Of-Words is a suitable approach for obtaining useful insights from the data. The proposed framework for implementation of Bag-Of-Words.

III. DESCRIPTION OF THE DATASET

Our work uses AFFR dataset [4][8] that consists of the customer's reviews on fine food. Dataset consists of nearly 5 lakh reviews, 2 lakh users and 74K products information. The data format of the records in the dataset can be seen in Table.1. The columns used in this work are Score and Text.

Table. 1 Data Format of the records in the dataset.

Selected Column	Column-ID	Description
	Product_ID	Unique Id associated with the product in amazon database
	User_ID	Identity of the user
	Profile_Name	Username
	Helpfulness	Proportion of the users who found the reviews were useful
*	Score	Product Rating
	Time	UST timing
	Summary	Summary of reviews given by customer
*	Text	Detail Review of the customer

IV. PROPOSED METHODOLOGY

Sentiment analysis is a study about the people's sentiment, emotions and opinions that is expressed in terms of text. This is very helpful in understanding people's opinions on anything. The opinions can be positive, negative and neutral. The proposed methodology focusses on building the Machine Learning Ensemble model. It is focused on both positive and negative reviews for analysis i.e, Binary Classification. The complete Process Diagram of the proposed system is represented in Fig. 1.

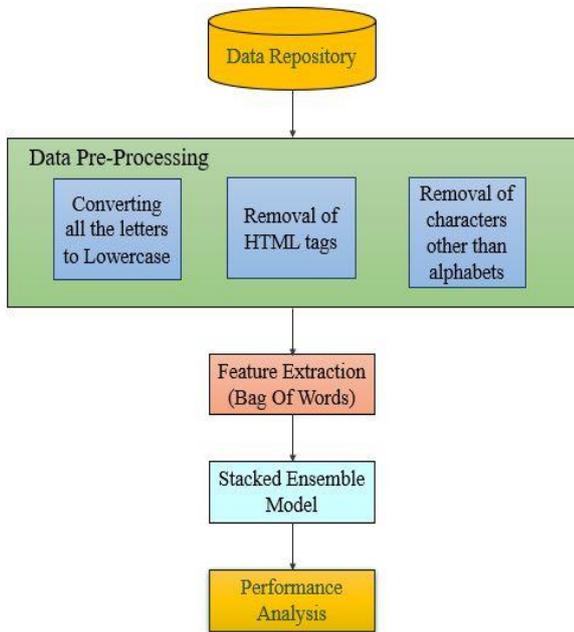


Fig. 1 Process Diagram of Proposed System.

i. Data Repository

This is the initial step of any analysis or research work in this field. The data can be real-time data or simulated and stored. In this paper, stored data i.e, Amazon Fine food Review dataset has been used.

ii. Data Pre-Processing

This is an important step in analyzing any kind of data. It has a powerful impact on the success of the Machine Learning models. There is various kind of Pre-Processing methods and one must choose the methods suitable for their data. In this work, three different preprocessing methods have been used.

- Lowercase Conversion.
- HTML tag Removal.
- Filtering Numbers and Special Characters

Lowercase Conversion.

Since the reviews given by different people is used for analysis purpose, there is a high chance that the same words would be present in a different case. Now, these will be considered as different words by machine and it will have a good impact on the performance of the models. To avoid this kind of scenario, the text is converted into lowercase.

HTML tag Removal.

The reviews have been extracted from the website before storing it into the dataset. It has some HTML tags associated with it, these HTML tags have been removed since it had no impact on identifying the polarity of the review.

Filtering Numbers and Special Characters

To make the data clean and neat, all the unwanted things that are not required for identifying the polarity of the review must be removed. The numbers and special characters have been removed.

iii. Feature Extraction

The machine learning algorithm can understand the data and work on it if it is in a specific format. Since here we have textural data that needs to be converted into numerical feature vectors which are called as vectorization. This strategy of tokenization, normalization and counting is called Bag of Words (BOW).

In this work, CountVectorizer has been used to represent words in terms of BOW. CountVectorizer will allow us to specify n-gram range. N-gram range is a tuple which consists of minimum and a maximum length of the sequence of words that needs to be considered as features.

iv. Stacked Ensemble Model

In Machine Learning, Ensemble model [6][7] is constructed using multiple base models. The output of these models will be fused using a different strategy to get converged output for a given problem. The main aim of the Ensemble model is to build a more optimal model, whose performance is greater than the individual base models.

There are different strategies for building the ensemble model and they are

1. MaxVoting
2. Averaging
3. Weighted averaging
4. Stacking
5. Blending

6. Bagging

In this work, stacking is used for building the Ensemble model, let us understand how the ensemble will work using Fig. 2 Stacking consist of two levels.

Level 0

Here, multiple base models can be used. In this paper Logistic Regression, Multinomial NB, Bernoulli NB is used as base models. These models will be given train data. The predictions from these models will be given as an input to the next level.

Level 1

Here, we build a new model which will be trained based on the predictions from the previous level i.e, Level 0. Now, this model is used to make final predictions on test data.

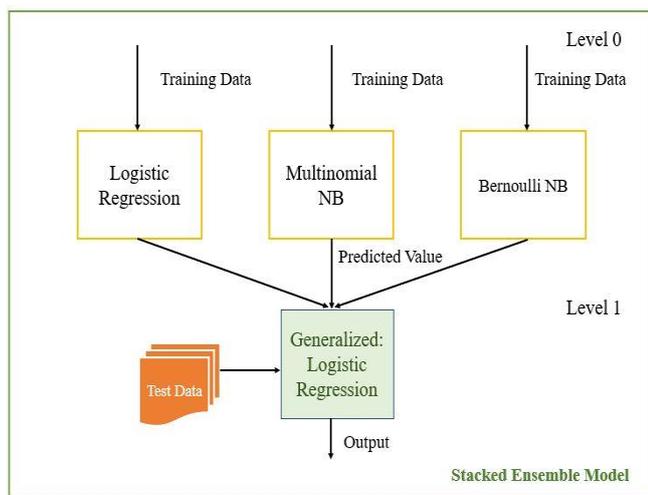


Fig. 2 Stacked Ensemble model.

Now let us see the Machine Learning based models used in this work. Three base models have been used.

A. Logistic Regression

Logistic regression is one of the supervised machine learning algorithms widely used for classification problem. This can be used for binary classification and multiclass classification. Since we are focused on only positive or negative emotions attached to the review. We have used binomial Logistic Regression classifier. Here, we have achieved an accuracy of 93.49%.

B. Naïve Bayes (NB)

NB classifiers are designed based on Bayes theorem. This is also the Machine learning algorithm used for classification problem and can be used for classification of the text. There are different variants of NB Classifiers. We have used Multinomial NB and Bernoulli NB.

Multinomial NB

This algorithm is very much suitable for document/text classification and the technique behind this is based on the word count and occurrence of the words in the text. We have trained the models using train data and validated using test data and achieved the accuracy of 92.28%.

Bernoulli NB

This is also used for classification of the text/document, but the difference is in the underlying technology used to do classification. It works well for text classification with BOW model. Here, the occurrence of the word is used i.e, whether the word is present in the text or not. We have achieved an accuracy of 87.99%.

v. Performance Analysis

Various models such as Logistic Regression, Multinomial NB, Bernoulli NB are used against our proposed system to evaluate the performance. The proposed system has performed well when compared to other individual models and achieved accuracy of 94%.

V. PERFORMANCE METRIC

The measure used to estimate the performance [5] of the proposed system is described in this section. Confusion matrix (Table. 2) is one of the powerful and easiest metrics used to identify the correctness of the classification model. It consists of four measures.

Table. 2 Confusion Matrix

	PV	Predicted-Negative	Predicted-Positive
AV			
Actual Negative		TN	FP
Actual Positive		FN	TP

- **TP-True Positive:** The food reviews that are positive, classified correctly as positive.
- **FP-False Positive:** The food reviews that are negative, classified wrongly as positive.
- **TN-True Negative:** The food reviews that are negative, classified correctly as negative.
- **FN-False Negative:** The food reviews that are positive, classified wrongly as negative.

The metric that is important is Accuracy, Precision, Recall, F1-Score. These metrics can be defined using the above four measures.

Accuracy(A)

Accuracy can be characterized as a proportion of accurately anticipated occurrences to a total number of occurrences. A is given by,

$$A = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision(P)

Precision can be characterized as, among all the predicted positive occurrences how many are actual positives. Hence, P is given by,

$$P = \frac{TP}{TP + FP}$$

Recall(R)

Recall can be characterized as, among the actual positives how many did the model predict as positive. R is given by,

$$R = \frac{TP}{TP + FN}$$

F1-Score(F1)

F1 score is the fraction of harmonic mean of precision and recall. F1 is given by,

$$F1 = \frac{2 * (P * R)}{(P + R)}$$

These metrics provide probabilistic values as output, where 1 represents that the model is accurate and 0 represents that model is not accurate.

VI. RESULTS AND ANALYSIS

This section describes the results of the proposed system and it is based on the performance metrics as described in the previous section. The Confusion metrics values of the different Machine Learning models can be seen in the below Table. 3

Table. 3 Confusion Matrices values of all the Models

Model Name	TP	TN	FP	FN
Logistic Regression	16673	16984	1140	1203
MultinomialNB	16628	16596	1528	1248
BernoulliNB	16995	14683	3441	881
Proposed System	16647	17219	905	1229

Here we compare the efficiency of our model with the other base models such as MultinomialNB, BernoulliNB and Logistic Regression.

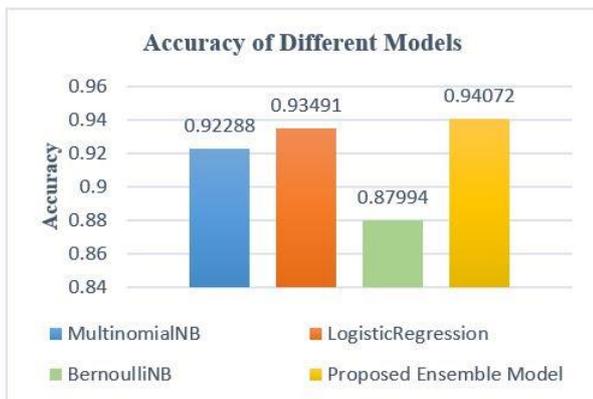


Fig. 3 Accuracy comparison of other base models with proposed ensemble model.

Fig. 3 represents the accuracy comparison of different models with the proposed model. It can be clearly seen that the accuracy of the proposed model is greater than the other models.

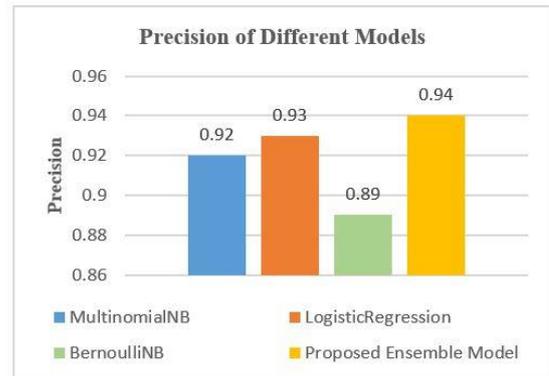


Fig. 4 Precision comparison of other base models with the proposed ensemble model.

Fig. 4 represents the precision comparison of different models with the proposed model. The precision of the proposed model is more than that of other models.

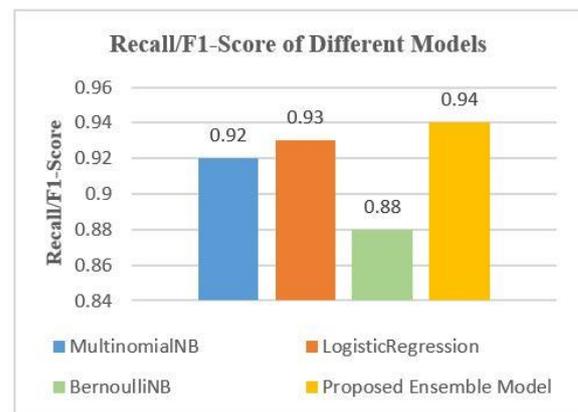


Fig. 5 Recall/F1-Score comparison of other base models with the proposed ensemble model.

Fig. 5 represents the Recall/F1-Score comparison of different models with the proposed model. The performance of the proposed model is greater than the performance of other models.

The text data can be represented and visualized pictorially using Word cloud package in python.



Fig. 6 Word cloud representation of complete text data.

Word cloud is an image composed of different words present in the text, in which the size of each word shows its importance and frequency.



Fig. 7 Word Cloud representation of words associated with a positive review.

Fig. 6 shows a visual representation of all the reviews present in the AFFR dataset using Word Cloud python package.



Fig. 8 Word Cloud representation of words associated with a negative review.

The word cloud representation of the words associated with a positive and negative review is shown in Fig. 7 and Fig. 8 respectively.

CONCLUSION

Sentiment analysis has been done by building Machine learning-based Ensemble model, which is a combination of Logistic regression, Multinomial NB and Bernoulli NB. The strategy used for combining different base model is stacking. The proposed model has achieved an accuracy of 94% which is greater than other base models.

In future, the features extraction can be enhanced, and advanced Machine Learning techniques can be applied so that the accuracy can be further improved.

REFERENCES

- [1] Güner, Levent, Emilie Coyne, and Jim Smit. "Sentiment analysis for Amazon. com reviews." (2019).
- [2] Bhati, Vijay, and Jayveer Kher. "Survey for Amazon Fine Food Reviews." (2019).
- [3] Haque, Tanjim Ul, Nudrat Nawal Saber, and Faisal Muhammad Shah. "Sentiment analysis on large scale Amazon product reviews." In *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, pp. 1-6. IEEE, 2018.
- [4] McAuley, Julian John, and Jure Leskovec. "From amateurs to connoisseurs: modeling the evolution of user expertise through online reviews." In *Proceedings of the 22nd international conference on World Wide Web*, pp. 897-908. 2013.
- [5] Powers, David Martin. "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation." (2011).
- [6] Huang, Faliang, Guoqing Xie, and Ruliang Xiao. "Research on ensemble learning." In *2009 International Conference on Artificial Intelligence and Computational Intelligence*, vol. 3, pp. 249-252. IEEE, 2009.
- [7] Dietterich, Thomas G. "Machine-learning research." *AI magazine* 18, no. 4 (1997): 97-97.
- [8] <https://www.kaggle.com/snap/amazon-fine-food-reviews>
- [9] Deepu, S., Pethuru Raj, and S. Rajaraajeswari. "A Framework for Text Analytics using the Bag of Words (BoW) Model for Prediction." In *Proceedings of the 1st International Conference on Innovations in Computing & Networking (ICICN16), Bangalore, India*, pp. 12-13. 2016.