

REVIEW ANALYSIS SYSTEM USING COLLECTIVE BLENDED MECHANISM

Dr. Janga Prasad¹, Ms. D. Nishitha², Ms. V. Bhavani³, Ms. P. Yamini⁴

1. **Associative Professor, Department of Electronics and Communication Engineering, CMR Institute of Technology, Medchal, Hyderabad.**
2. **Bachelor's Student, Department of Electronics and Communication Engineering, CMR Institute of Technology, Medchal, Hyderabad.**
3. **Bachelor's Student, Department of Electronics and Communication Engineering, CMR Institute of Technology, Medchal, Hyderabad.**
4. **Bachelor's Student, Department of Electronics and Communication Engineering, CMR Institute of Technology, Medchal, Hyderabad.**

ABSTRACT Actually, most of businesses remain as failures due to lack of sufficient profits, lack of proper improvement measures. Specifically, the restaurant owners are facing a lot of problems and difficulties to improve their businesses due to lack of productivity in their business. Based on a report, the number of restaurants that have been shutting down in the first year are upto 60% and the restaurant businesses which are stopping their services within the first 5 years are about 80% in the world. So, this is becoming a universal problem now. There are various reasons involved in the failure of restaurant businesses, but serving low-quality food and lack of proper taste in the food items they are serving is a major one. To improve their business, they should know drawbacks of their food items and improve them by taking quality measures. To resolve the problem of business loss which is due to various drawbacks of food items, a solution

is to implement a “review analysis system using collective blended mechanism”. Each food item in the restaurant is assessed through the text review given by the customer where the text review is verified as a positive review or a negative review by the implementation of collection of classification models (Collective ML Model) in machine learning after handling textual data with natural language processing (to convert textual data to numerical data). And this data is stored into the database where each food item consists of number of customers, number of positive reviews, number of negative reviews, positive rate and negative rate. Now, the owner checks the least positively rated food item from the database, and takes necessary quality measurements which involve replacement of chefs or updating the ingredients used etc.

Keywords - Lemmatization, Stemming, Count vectorization, Universal Sentence

encoder layer, Back propagation, Blending, Meta classifier, Collective Blended Model.

INTRODUCTION The great opportunities and challenges for textual research has been opened up by the substantial increase in the volumes of unordered or unstructured textual data accompanied by the proliferation of tools to analyse them. Significantly, In the last few years, the scientific researches and study of the sentiment analysis has become very famous. Now a days, Business managers not only want to know about their product marketing strategies and the profit measures based on the number of transactions completed with customers but also want to know about the reviews/feedbacks and thoughts of the people based on their mindsets on using these products. The reviews they receive via social media and other internet services has become a key factor and a very important aspect to measure the quality of a product they are providing. Textual Sentiment analysis is a specific area or domain where the analysis is primarily aims on the extraction of feedback/review and thoughts of the users towards a specific aspect from a structured, or unstructured textual data. In this paper, we try to focus our effort on resolving the problems mentioned above and on implementing textual sentiment analysis on restaurant

review databases. We used the textual sentiment expression to classify the reviews from the customers of the restaurant business whether it is negative or positive and perform identifying least positively rated food items and use these details for updating, maintenance and successful running of the restaurant business. The proposed application involves the working process as follows, each food item in a restaurant is assessed through the text review given by the customer where the text review is verified as a positive review or a negative review by the implementation of collection of classification models (Collective ML Model) in machine learning after handling textual data with natural language processing (to convert textual data to numerical data). And this data is stored into the database where each food item consists of number of customers, number of positive reviews, number of negative reviews, positive rate and negative rate. Now, the owner checks the least positively rated food item from the database, and takes necessary quality measurements which involve changing of chefs or updating the ingredients used etc. The main sections present in our proposed method in analysing various reviews from the customers on food items involves data pre-processing, Building and implementing a collective model. The data is collected from the reviews given by the customers on

different online portal and sentiment depicting English words. The data pre-processing involves the implementation of removal of stop words, applying stemming and lemmatization of words using natural language processing techniques. The Collective ML Model is built using 5 classification mechanisms i.e. Logistic Regression (LR), Decision Trees(DT), Support Vector Machine (SVM), Artificial Neural Network(ANN), XGBoost(XGB).

LITERATURE REVIEW

- Nina Lao et.al, This paper focused on building a useful app that compliments the most popular dishes with the help of the Yelp database. Whenever a user chooses something they want in a particular restaurant, that review is collected on their website. Now, these updates are sorted by Machine Learning Algorithms to classify reviews as positive or negative. A high-quality meal is the most recommended .Also, these highly recommended food items are returned to the user for further exploring. But there is no need for food recommendations to identify drawbacks of food items in the restaurant
- AlhassanMabrouk et.al, compared the estimation results of greater than hundred Deep Learning based

textual Sentiment analysis Classification approaches by using some of the public datasets of feedbacks given by consumers or customers within three particular areas or application domains (products, movies and restaurants). These datasets have various features (balanced/imbalanced, size, etc.) which lead to give a universal vision for our research study. Also, explained the various kinds of literature based comparisons to identify the most significant key factors in the three phases such as: feature representation, data preparation and classification techniques. The comparison demonstrates how the proposed data factors and other measures quantitatively affect the performance or estimation ability of the studied Deep Learning based textual Sentiment Classification approaches.

- RubaObiedat et.al, explained the textual sentence based sentiment analysis of customers feedbacks or reviews using a hybrid evolutionary SVM based approach in an unstructured an nonuniform Data Distribution of the information. The data is collected from the reviews for some restaurants from the Jeeran

website. After successfully completing the data preparation process, four different kinds which are in the form of versions of the dataset are presented using the different textual tokenization and embedding methods on the data. Earlier, individual random k value and a random weights for the sampling parameter are developed and initialised. By applying a technique called PSO optimization to determine the best k value and the best possible weights for the dataset attributes for every oversampling technique, after that, applied the SVM classification mechanism to the weighted and oversampled dataset to determine and extract the best possible sentiment results of the customer reviews of a specific restaurant.

- ShilpaShendre et.al, suggested a way to classify business reviews by using Sentiment analysis. Sentiment analysis focuses on the output of user's feedback and opinions on a particular topic from formal, or informal text data. In this case, major focus is on sentiment analysis at the restaurant review site. By testing the sentiment phrase, restaurant business reviews as distinguished pretty much as good

or bad and make an output feature and use these features to review and maintain the business with Naïve Bayes and Linear Support Vector Classification (SVC)

- MarouaneBirjali et.al, described the categorization of the foremost used Sentiment analysis approaches briefly to possess a overview of accessible techniques and approaches like Machine Learning, Lexicon based and hybrid approach. Semantic orientation of words which are present in a text are used to extract sentiments from them in the lexicon based approach. Three key mechanisms for developing, annotating and creating sentiment lexicons are manual approach, corpus based approach, and dictionary based approach. Collection of common lexicons utilized by many researchers are WordNet, SenticNet, Sentiment WordNet, MPQA based subjectivity lexicon. Classification mechanism with Naive Bayes and SVM algorithms are taken as key factor and base features to compare newly proposed techniques

PROPOSED METHODOLOGY Our Proposed Methodology involves the mechanism of the Collective Blended

Model, which takes the use of the blending ensemble method. Every individual model provides results to some extent. But, the results provided by an individual model might not be optimal. To acquire optimal or better accurate results, there is a need to use a model which combines the properties of all individual models i.e., an ensemble model. So, we introduce a “Collective blended model” which involves the mechanism of the Blending ensemble method. The prediction of customer’s textual review as whether +ve/-ve can be done by our collective blended model which involves the following phases.

- Data Pre-processing
- Prediction using individual models
- Developing blended model
- Performing final prediction
- Data Pre-processing

An ML model can’t make predictions on textual data as it operates only on numerical data. So, to transform a textual data to the numerical data, different Natural Language Processing techniques were used. Firstly, the removal of stop words (like ‘the’, ‘and’, ‘an’,....) from the customer’s textual review gets happened followed by stemming the each word in the review. Finally, all the selected words from the review can undergo count vectorization technique to transform into the numerical data. The numerical data is formed on the

basis of relative positions and corresponding frequencies of all words with respect to all other words in English language Stemming is widely used for text processing. Stemming is the process of finding root of words by removing either prefix or suffix. It links words with similar meanings to one word. Text before stemming: amazing Text after Stemming: amaze

machine learning model can’t understand the textual input data on which it has to perform computations to make prediction results. So, we can’t transfer textual data directly to any ML model. Thus, we have to convert the textual data into the numerical form. So, to perform this process we need to use some of the proper NLP (natural language processing) techniques. The key mechanism involved in doing this is to use a bag-of-words (BOW) model. The bag-of-words model removes or discards the all information involved in the textual data and position of the textual characters present in each word of the sentence and only considers the frequent occurrences of the word. It converts specific kind of textual documents to the specific length list of numbers. A global vector is implemented by considering all possible words that can occur in English Language, and each word is assigned with a unique number (commonly index of an array) besides the

frequency count which demonstrating the number of occurrences of that specific word in the given sentence. This kind of encoding mechanism on the textual data, in which the key focus is to demonstrate the representation of the data but not on the position of the data. Count Vectorizer is responsible for the tokenization of the textual sentences. This breaks the each textual sentence into collection of key words after performing the important data pre-processing techniques which involve removal of special characters, transforming the all words into lowercase, removing stopwords, applying stemming and lemmatization which find root words of the given word which are further referred to as keywords, etc.

A. Prediction using individual models

Each individual model which is already trained on training data from the dataset make predictions on review given by the customer after performing NLP techniques on review. The following are the algorithms involved in our Collective Blended Model.

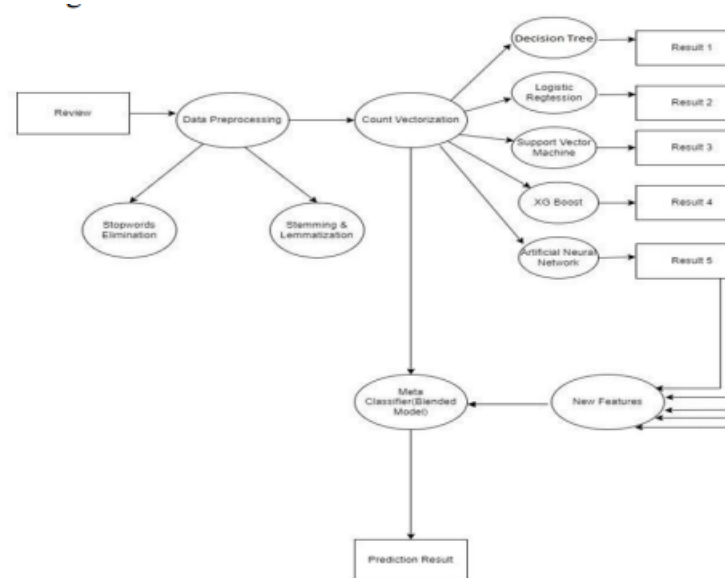


Fig1 : Mechanism for predicting customer re

- Decision Tree
- Logistic Regression
- Support Vector Machine
- XG Boost
- Artificial Neural Networks

RECOMMENDING TOP RELEVANT

REVIEWS The process of recommending top relevant positive and negative reviews based on the reviews given by the customers to our system. The important and key words present in each review are extracted and corresponding frequencies are maintained in a database. Now, the frequency cost is calculated for each review as the sum of the frequencies of all keywords present in the customer review. And finally the positive reviews with highest frequency costs are recommended as top positive relevant reviews and the negative reviews with highest frequency costs are recommended as top negative

relevant reviews to the owner of the restaurant through our system which gets displayed on the dashboard. $\text{freqCost}[\text{review}] = \sum \text{freq}[\text{keyword}_i]$ where, keyword_i is the i th key word present in the given customer review

CONCLUSION This application perfectly helps to the business managers who wants or needs to improve their productivity, which in turn indirectly improves their business standards and significantly lead to good profits. By taking reviews from customers, the business problems got diminished through the development of customer-friendly service standards and productivity. This system can be kept and used in any food hotel or restaurant. It perfectly plays an effective role in taking food improvement measurements which directly lead to improvement one's business. It is Consumer-friendly application. The feasibility of the project lies in business performance of restaurants because this system provide better results and make a greater impact on one's business if and only if number of visiting customers are more. Also, as long as owner want to know the drawbacks of his own restaurant food items, the system is used. Finally, the overall impact made by the application to a restaurant is relatively greater as compared to manually maintaining the data.

REFERENCES

- [1] Nina Lao et.al, Analyzing and filtering food items in Restaurant reviews, <https://aircconline.com/csit/papers/vol10/csit101208.pdf>
- [2] Alhassan Mabrouk et.al, Deep Learning Based Sentiment Classification, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9085334>
- [3] Ruba Obiedat et.al, Sentiment Analysis of Customer Reviews Using a Hybrid Evolutionary SVMBased Approach in an Imbalanced data distribution, <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9706209>
- [4] Shilpa Shendre et.al, Classification of Business reviews using Sentiment Analysis, <https://www.irjet.net/archives/V6/i6/IRJE-T-V6I6571.pdf>
- [5] Maroune Birjali et.al, A Comprehensive Survey on Sentiment Analysis, <https://www.sciencedirect.com/science/article/pii/S095070512100397X>
- [6] Cui, H., V. Mittal and M. Datar. Comparative experiments on sentiment classification for online product reviews. In Proceedings of AAAI-2006, 2006.28.
- [7] Das, S. and M. Chen. Yahoo! For Amazon: Extracting market sentiment from

stock message boards. In Proceedings of APFA-2001, 2001.

[8] Alm, C.O. Subjective natural language problems: motivations, applications, characterizations, and implications. In Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: short papers (ACL-2011), 2011.

[9] Andreevskaia, A. and S. Bergler. Mining Wordnet for fuzzy sentiment: Sentiment tag extraction from WordNet glosses. In Proceedings of Conference of the European Chapter of the Association for Computational Linguistics (EACL-06), 2006.

[10] Daniel Ellis, Using Sentiment Analysis to explore emotions within text methods, towardsdatascience.com.