Grenze International Journal of Engineering and Technology, Jan Issue

Sentiment Analysis of Online Products using Ratings and Reviews

Mr. T. Praveen Kumar[1] and Mr. Sandeep Pamulaparty[2]
[1]Methodist College of Engineering and Technology/Dept of CSE, Hyderabad, India
Email: pravinthumukunta@gmail.com
[2]CISCO/Santa Clara, California, USA
Email: pamulaparty@gmail.com

***Abstract*—Every second, we can observe a massive surge of data being generated. Data present abundantly needs to be processed to become meaningful information. One way of processing data and knowing the current trend is through product analysis. This concept can be used by organizations to identify their flaws and enhance their productivity. It is often performed on textual data to help businesses monitor brand and product sentiment in customer feedback, and understand customer needs. In this project, we throw light on the aspect that data generated by textual comments along with the rating contribute the efficient classification of data rather than only the rating and obtain various insights as a part of the analysis. In general, when a person wants to buy a product, he checks for the review of the product and then makes a decision whether to buy or not. In more specific terms, the user checks only the rating and makes a decision. We have understood that sometimes rating is misleading the customers and it alone cannot be the sole parameter to take a decision hence we are comparing ratings and comments for better efficiency and authenticity of product reviews.**

***Index Terms*— NLP, SVM, KNN, and Logistic Regression.**

I. INTRODUCTION

Researchers already use social data to analyze the sentiment of users' opinions on a product, event, or setting. Moreover, People utilize online comments to convey their interests or attitude in social networks, thanks to the rapid development of Internet technology. We can utilize these comments to extract vital information, such as consumer sentiment for a particular product. One of the most important research tasks in text mining is sentiment categorization. Researchers are already analyzing the sentiment of consumers' opinions on a product, event, or location using social data. In addition, sentiment analysis is frequently referred to as opinion mining, which is an important NLP task. The direction of the sentiments is determined by this sentiment analysis. In respect to text, as neutral, positive, or negative in addition, text analytics, computational linguistics, and natural language processing (NLP) are all used in sentiment analysis to recognize and understand the text. Organizing the user's points of view. The primary purpose of sentiment analysis is to determine the author's perspective on a similar event or the overall polarity of the material. Views of emotional communication might include a user's judgment or appraisal, emotional state, or deliberative state. It's typically used to identify sentiment information disclosed by people in comments, such as movie reviews. Sentiment analysis, also known as option mining, is the process of extracting a reviewer's attitude from a movie, usually in the form of a sentence or a review. Sentiment analysis

Grenze ID: 01.GIJET.9.1.759

can be thought of as a binary classification of positive and negative sentiment in its most basic form. We know that sentiment analysis analyses people's sentiments or opinions based on their written texts, and that it's important in data mining and natural language processing.

II. Literature Survey

Kushal Dave et. al. [1] They start by defining the special characteristics of the product review issue and create a technique for automatically differentiating between favorable and unfavorable reviews. For feature extraction and scoring, the classifier uses information retrieval techniques, and depending on the testing scenario, different metrics and heuristics produce different results. Performance is constrained when working with individual sentences gathered from web searches because of noise and ambiguity. However, the results are qualitatively highly valuable when used in conjunction with a comprehensive web-based application and helped by a straightforward mechanism for categorising texts into qualities. The results are qualitatively very helpful in terms of qualities. Finn et. al [2] Separating reviews from other sorts of content is a difficult undertaking. Having trouble categorising a genre or style. It entails determining subjectivity.
XING FANG et. al. [3] with thorough process descriptions, a general procedure for sentiment polarity categorization was suggested. Online product reviews gathered from Amazon.com [4] were the source of the data for this study. Both sentence-level categorization experiments and review-level classification experiments were conducted with encouraging results. Finally, provided a preview of our upcoming sentiment analysis work.
NAJMA SULTANA et. al**.** [5] Sentiment analysis's primary goal is to classify online data by determining its polarity. Although sentiment analysis is a text-based process, it might be difficult to determine a sentence's precise polarity. This asserts that a superior solution must be found in order to outperform any prior method or strategy utilised to determine sentence polarity. Therefore, there was a need for automated data analysis approaches to determine the polarity or emotion of a user or client. In addition to a novel strategy that was proposed in this research, a thorough survey of several sentiment analysis methods was included.
JAYAKUMAR SADASIVAM et. a,l [9] improved the accuracy of the provided reviews, an ensemble technique has been used. SVM and the Ensemble algorithm are merged in this study. They put forth the Ensemble technique, which improves upon the current algorithm's accuracy. The user is advised to buy the specific product once the accuracy calculation is done based on the reviews.
YANG et. al. [11] Analyzed that the opinions of the reviews of customers which are useful for consumers as well as the also the shop owners. They improve the service and consumer satisfaction. Zendesk et. al. [10] also agree upon the same concept that the online shopping is greatly effected with the reviews.

III. Proposed System

In the proposed system, we are comparing the rating and review to give a brief overview. Then the prediction of the user rating from the reviews and comparing these predicted review with the user reviews are performed. The results in the next section speak that if the user review and prediction are the same there will be no conflicts and if not, then the user reviews and predicted review conflicts are also determined. In order to perform all the said tasks the architecture is proposed as below

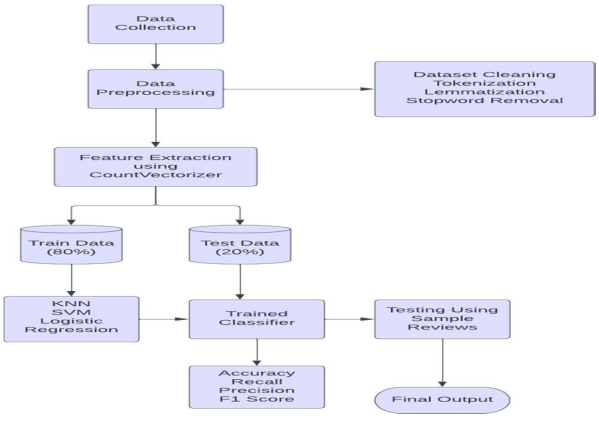

Figure 3.1 Overview of the proposed system architecture

The overall structure is described in the architecture.In the first step obtain the data set. The data set includes several kinds of reviews. The data should go through pre-processing after collection to get rid of unnecessary words, phrases, and symbols. Before performing an analysis, it is crucial to consider the representation and quality of the data. Before spell checking, symbols land Stop Words will be eliminated. A fantastic utility offered by the Python sci-kit-learn module is Count Vectorizer

Usually, training data is larger than testing data. 80% of the data are often used for training and 20% are typically used for testing in machine learning. Here, Logistic Regression, SVM, and KNN classifiers will be applied to collect and analyses training data. Unsupervised and supervised classification are the two main categories in general. Labels will be provided as part of the supervised classification. The unsupervised classification will not be given any set of labels, whereas the training data will consist of a collection of training examples. The usage of supervised classifiers like Logistic Regression, KNN, and SVM is suggested in this work. Finally, by merging all of the proposed techniques, an Ensemble strategy is used. After classification, the group votes for the majority (Accuracy). As illustrated below, the confusion matrix is computed for multiclass classification.

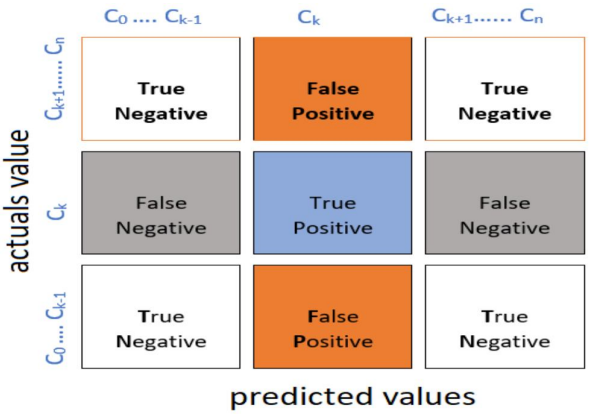

Figure 3.2 : Confusion Matrix for multiclass classification

The performance measure is calculated by precision, recall and F1-Scores. The calculation is described in the equations as follows

$$Precision(Class_i) = \frac{TP_{a_i}}{TP_{a_i} + FP_{a_i}}$$

$$Recall(Class_i) = \frac{TP_{a_i}}{TP_{a_i} + FN_{a_i}}$$

$$F1 - Score(Class_i) = 2 * \frac{Precision(Aspect_i) * Recall(Aspect_i)}{Precision(Aspect_i) + Recall(Aspect_i)}$$

$$Accuracy(Class_i) = \frac{FP_{a_i} + FN_{a_i}}{TP_{a_i} + TN_{a_i} + FP_{a_i} + FN_{a_i}}$$

$$Over\ All\ Accuracy = 1 - \frac{\sum_{i=0}^{a_n} (FP_{a_i} + FN_{a_i})}{\sum_{i=0}^{a_n} (TP_{a_i} + TN_{a_i} + FP_{a_i} + FN_{a_i})}$$

Then providing the sample reviews with rating and checking with the predicted rating. If the sample review and predicted rating doesn't match, we are giving the conflicts otherwise we are providing no conflicts. Lastly, we are testing the algorithms with the X_test reviews and Y_test rating. We are predicting the rating of X_test reviews and comparing it with Y_test rating.

IV. EXPERIMENTATION

The experimental results presented in this section show the performance of the proposed work which involves data preparation, data pre-processing, data labelling, preparation of emoji lexicon, data labelling, feature extraction and classification using machine learning approaches. The experimentation was conducted using a python programming language.

A. Dataset Characteristics

In the case of the suggested technique, the Amazon product reviews dataset is used as the input. And there are approximately 6,000 records in this collection that are freely available online.They have features like Product Name, Brand Name, Price, Ratings, Reviews, and Review Votes in their dataset. The data (i.e., 20%) is used for testing while the remaining data (i.e. 80%) is used for training. Finally, the sentimental analysis based on language includes both ratings and user written evaluations are used. The dataset consists of six columns and 5183 rows. The columns consist of Product Name, Product Brand, Price, Rating, Review, Review Votes.

B. Dataset Cleaning

Data cleansing is performed on data for analysis by Fixing spelling and syntax errors, Standardized records,Fixing errors such as blank fields and identifying duplicate data points.The extraction of reviews and ratings from the dataset is performed. Then the numbers are removed and website URL from the reviews also to look cleaner. The new reviews are stored in a separate column.

C. Tokenization

Tokenization is performed to separate reviews sentenceds, words, letters, or subwords. Text is divided into sentences, called as tokenization of the sentence. Words into word tokenization. After tokenization the words are converted into lower case.

D. Lemmatization

Lemmatization is performed for translating words in the reviews into its basic form. lemmatizinging takes context of reviews into account and transforms a word into a basic meaningful form.

E. Stop words Removal

A stopword is a language word in the review that doesn't make much sense in a sentence. After removing stop words, the result in result in stopWordRemoval column. Then we are adding all words in the stopWordRemoval column into a sentence and storing them in the Final_Review column.

F. Feature Vectorization

Count Vectorizer is a great tool provided by Python's Scikit-learn library. This approach in Machine Learning counts how many times a particular word is found in the reviews. The end result is an encoded in a matrix in which each unique word is represented in a matrix column and each product review in the document represents a row in the matrix.

G. Classification

In the proposed study, the count vectorizer-acquired training data set of product review words is used to train the classifiers. It is used to turn a text into a vector depending on how frequently (count) each word appears across the entire text. A new column named "sentiment" was added once the reviews had been cleaned up. The ratings (1–5) are converted into Positive, Negative, and Neutral to create this column. Negative ratings fall between 1 and 2 stars, neutral ratings fall between 3 and 5, while positive ratings fall between 4 and 5 stars. Review-level classification is the term for this. The training set of classifiers is applied to the test set of data, and the array is measured for the classifiers that consider Logistic Regression and Support Vector Machine. As a result, the methodology combined some machine learning techniques with one feature extraction method.

V. RESULTS

In this study, sentimental analysis was conducted for product reviews that include it’s both ratings and comments. In the end, the machine learning algorithms were used to assess the data. It was observed that there are 5333 non-recurring words in the dataset which are columns. The 5182 text samples in the dataset each represented as rows of the table. Each cell has a number in it that represents the amount of words in that specific

text. Lowercase letters have been used for all words. The words have been sorted in columns using the alphabet. The ratings are given on a scale of 1 to 5. The diagrammatic representation of the dataset is given below

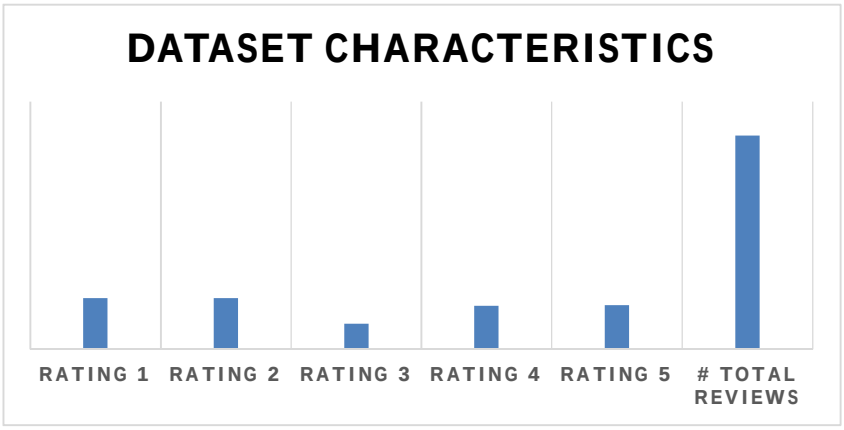

Figure 5.1 Data Set Characteristic

The Data Cleaning, Tokenization and Lemmatization is performed and the output is passed to the machine learning classifiers. Initially, Logistic Regression machine learning algorithm is used for classification. It provides probabilities as it can classify product reviews datasets. After training and testing the dataset with logistic regression algorithm with 80, 20 ratios. The confusion matrix of the proposed modal is as follows

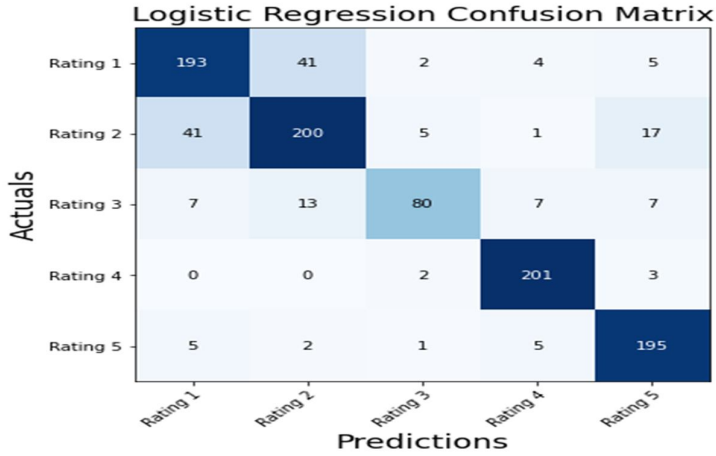

Figure 5.2 : Confusion matrix for Logistic Regression

A abbreviated version of the Logistic Regression model Classification Report achieved during the modelling stage is shown in Table 5.1.

TABLE 5.1 : EVALUATION METRICS FOR LOGISTIC REGRESSION CLASSIFIER

class	TP	FN	FP	TN	Accuracies	Precesion	Recall	F1 score
Rating 1	193	52	53	739	0.79	0.78	0.78	0.79
Rating 2	200	64	56	717	0.77	0.78	0.75	0.77
Rating 3	80	34	10	913	0.88	0.88	0.70	0.78
Rating 4	201	5	17	814	0.95	0.92	0.97	0.95
Rating 5	195	13	32	810	0.90	0.85	0.93	0.87

From the Table 5.1 it is observed that the F1 Score for the rating 4 is high of 95% compared with the others. Next another classification algorithm was used and results are described below

SVM is also a supervised machine learning that is used in classification. The SVM will separate with the boundaries which are linearly seperable. So this expression defines the decision boundaries returned by the

SVM. Despite its simplicity, the closest neighbour has been successful in numerous classifications and regression problems, including handwritten numbers and satellite imagery scenes. As a nonparametric technique, it often succeeds in classification situations where the decision boundaries are very irregular. Confusion matrix follows as

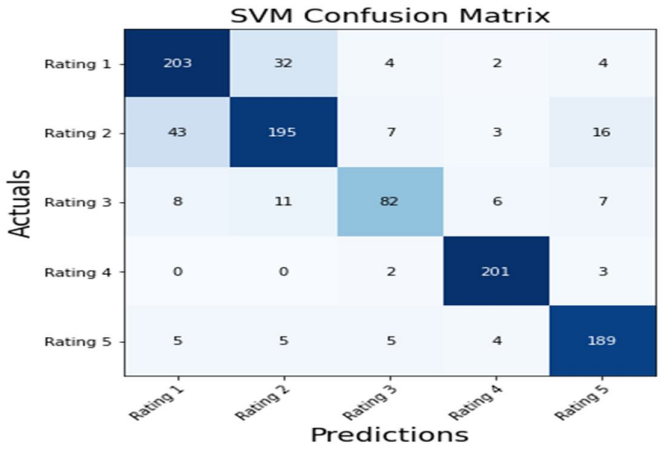

Figure 5.3 : Confusion matrix for SVM

A abbreviated version of the SVM model Classification Report achieved during the modelling stage is shown in Table 5.2.

TABLE 5.2 : EVALUATION METRICS FOR SVM CLASSIFIER

class	TP	FN	FP	TN	Accuracies	Precision	Recall	F1 score
Rating 1	203	42	56	736	0.81	0.78	0.83	0.80
Rating 2	195	69	48	725	0.78	0.80	0.74	0.77
Rating 3	82	32	18	905	0.86	0.82	0.72	0.77
Rating 4	201	5	15	816	0.96	0.93	0.97	0.95
Rating 5	189	19	30	818	0.90	0.86	0.91	0.88

It is observed that the ratings 4 and 5 are having high accuracy above 90% compared with the other ratings. As the accuracies are very similar to each other so the other modalities like Precision, Recall and F1 Score are also analysed for clarity. All the three parameters are high for the rating 4. Then the KNN classification is also performed as explained below

The KNN classifier with the training and testing model is performed. The Confusion Matrix of the KNN model, the predicted sentiments for all 5 labels for ratings 1 to 5.

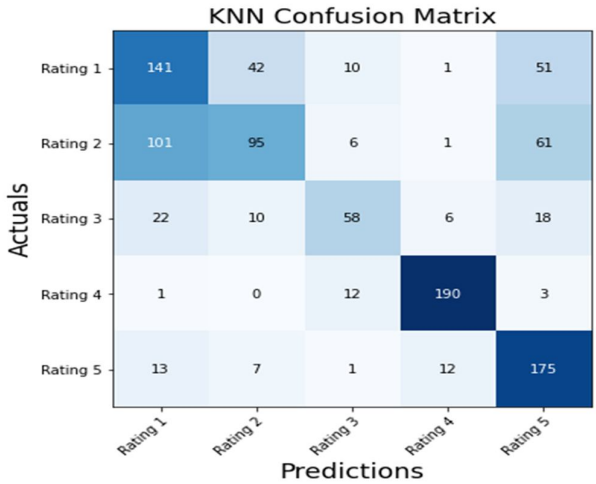

Figure 5.4 : Confusion matrix for KNN

A abbreviated version of the KNN model Classification Report achieved during the modelling stage is shown in Table 5.3.

TABLE 5.3 : EVALUATION METRICS FOR KNN CLASSIFIER

class	TP	FN	FP	TN	Accuracies	precision	Recall	F1 score
Rating 1	141	104	137	655	0.56	0.51	0.58	0.54
Rating 2	95	169	59	714	0.57	0.62	0.36	0.45
Rating 3	58	56	29	894	0.77	0.67	0.51	0.58
Rating 4	190	16	20	811	0.92	0.90	0.92	0.91
Rating 5	175	33	133	729	0.69	0.59	0.84	0.68

It is observed that the ratings 4 are having high accuracy above 90% compared with the other ratings. As the accuracies are very less except Rating 4 so the other modalities like Precision, Recall and F1 Score are also analyzed. The ratings of 4 have all three parameters high. The classification algorithms performance is evaluated for thorough revision.

In this step, we are comparing the accuracy of Logistic Regression,SVM and KNN algorithms using matplotlib. pyplot. The X-axis consists of scores and Y-axis consists of algorithms. The model accuracies of all the classifiers have been calculated. The score ranges from 0.0 to 1.0, this represents the percentage of accuracy of algorithms.

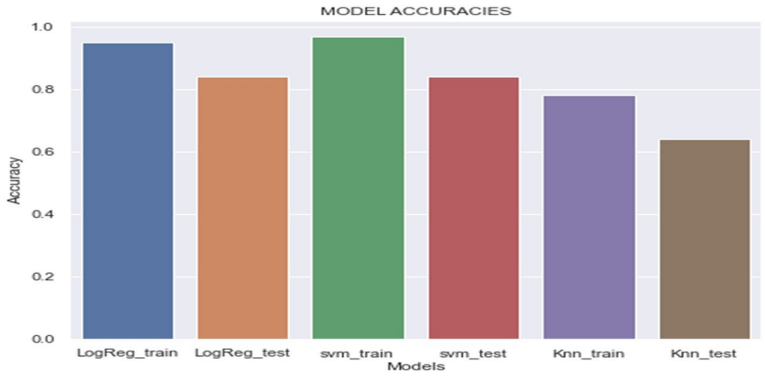

Figure 5.5 : Classification Model Performance accuracies

The accuracy of the SVM classifier is 96.6% accuracy. The Logistic regression is also nearly 95.6%.

In this next step, the checking of the test data with predicted data is performed. Here the storing of the predicted ratings of X_test data in the array predicted_values. The original ratings of the Y_test data are stored in the array original_values. The comparison is performed on the predicted ratings of the X_test data with the original ratings of the Y_test data to check the conflicts. The comparison of predicted ratings of the X_test data with the original ratings of the Y_test data with all three algorithms and counting the non-conflicts and conflicts values. Lastly, we are printing the conflicts and non-conflicts values at the end. As the dataset increase the values of conflicts and non-conflicts will increases. The graph depicting conflicted and non-conflicted reviews is presented below.

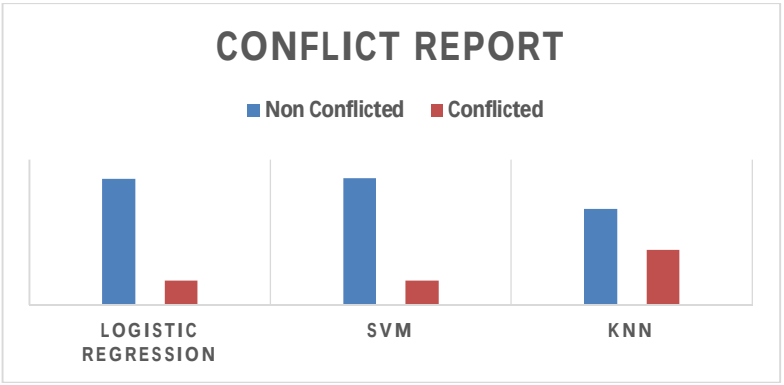

Figure 5.6 : Conflict Report

The number of total reviews on testing SVM data is 1037. In this the Non-Conflicted between rating and review is 870 and Conflicted is 167. In similar manner the other two classifier algorithms are also performed. This study is giving us the inference that the ratings and reviews always need not be on the same lines as either positive, negative or neutral. So, during sentiment analysis the conflicted reviews have to be further analysed for the accurate results.

VI. Future Scope and Conclusion

This is a comparative study uses a combination of reviews and ratings as input and provide the polarity of the data by analyzing and interpreting the data. The system does pre-processing operations like tokenization and removing stop-words from the dataset to extract the important features as a part of feature extraction using vectorization. Each product review and Amazon customer review has been assessed. In this dataset, we used 5182 user reviews. Each product has a distinct rating and review. For measuring the Amazon product review analysis, we developed a system that would compare the user rating with the rating from the trained model and we represent the number of conflicting and non-conflicting reviews from a trained dataset as an efficient measure for the user to take a better decision.TAs far as future scope is concerned, we can upscale the model using Deep Learning Algorithms like LSTM, RNN and also work on data that contains a combination of textual data and emoji data.

References

[1] Dave, Kushal et al. “Mining the peanut gallery: opinion extraction and semantic classification of product reviews.” The Web Conference (2003).

[2] Adan Finn, Nicholas Kushmerick, and Barry Smyth. Genre classification and domain transfer for information filtering. In Fabio Crestani, Mark Girolami, and Cornelis J. van Rijsbergen, editors, Proceedings of ECIR-02, 24th European Colloquium on Information Retrieval Research, Glasgow, UK. Springer Verlag, Heidelbe

[3] Fang, Xing and Justin Zhijun Zhan. “Sentiment analysis using product review data.” Journal of Big Data 2 (2015): 1-14.

[4] https://www.kaggle.com/snap/amazon-fine-food-reviews

[5] Sultana, Najma & Kumar, Pintu & Patra, Monika & Chandra, Sourabh & Alam, Sk. (2019). SENTIMENT ANALYSIS FOR PRODUCT REVIEW. International Journal of Soft Computing. 09. 7. 10.21917/ijsc.2019.0266.

[6] http://cs229.stanford.edu/proj2011/MehtaPhilipScariaPredicting%20Star%20Ratings%20from%20Movie %20Review%20Comments.pdf

[7] http://marcobonzanini.com/Callen Rain,”Sentiment Analysis in Amazon Reviews Using Probabilistic Machine Learning”

[8] Sadhasivam, Jayakumar & Babu, Ramesh. (2019). Sentiment Analysis of Amazon Products Using Ensemble Machine Learning Algorithm. International Journal of Mathematical, Engineering and Management Sciences. 4. 508-520. 10.33889/IJMEMS.2019.4.2-041.

[9] Zendesk. (2020). Zendesk Customer Experience Trends Report 2020. https://d1eipm3vz40hy0.cloudfront.net/pdf/cxtrends/cx-trends-2020-full-report.pdf

[10] Yang, L., Li, Y., Wang, J., & Sherratt, R.S. (2020). Sentiment analysis for E-commerce product reviews in Chinese based on sentiment lexicon and deep learning. IEEE Access, 8, 23522-2353