RESEARCH ARTICLE   OPEN ACCESS

Sentiment Analysis of Amazon Customer Product Reviews: A Review

Tayybaha Quyyam*, Dr. Hamid Ghous **
*(Computer Science, Institution of Southern Punjab, Multan)
** (Computer Science, Institution of Southern Punjab, Multan)

--************************--------------------------------

Abstract:

Internet is an easy way to for the customer to exchange information about the product over the shopping websites. It is important for the Shopping websites to understand the sentiment of the customer's reviews. Sentiment Analysis is a type of natural language processing that can be used to understand the sentiments of customer reviews from the online platform such as Amazon. Sentiment analysis is also used for the client reviews which contain the type of texts, words, expressions, and star rating techniques. Sentiment analysis of the customer reviews can increase the sale of products. A variety of research has already been conducted related to the sentiment analysis using machine learning and deep learning methods. In this paper, we have analysed the literature related to sentiment analysis of customers' reviews on Amazon based on Machine learning and Deep learning methods. We have also discussed the limitations of related literature and suggested new directions for future researchers.

Keywords **— Sentiment analysis, Amazon product reviews, machine learning and deep learning**.
--************************--------------------------------

I.   INTRODUCTION

Amazon is the largest shopping platform for the customer's in the world and Sentiment Analysis is a method of analysing the customer's opinions or views by using natural language processing whose information depends on the text layout. Sentiment analysis can help out the customer for good decision making when they purchase items from the internet platform. Sentiment analysis is being used for the customer's opinion that can be positive or negative about the product. Sentiment analysis is a branch of data mining that is used to predict the customer's opinion or emotion from the online platform as "Amazon" by using natural language processing. By using this way it can evaluate the customer's demand about the item and then help the organization to easily improve the sale of the businesses. Sentiment analysis understands and predicts the emotions, feelings and opinions of the consumers that are observing the product whose information consists of in the text format. Difficult to organize a large amount of data that consists of the opinion of the customer can easily predict therefore then provide the benefit of the client and internet platform.

Sentiment analysis for the brand is very important for both parties i.e. organization and client, to increase the sale of products. According to the organization to achieve a lot of profit by understanding the real behaviour of the customer about the sale products. With a vast internet platform, customers continuing to play an important part in the organization, and clients able to express their behaviour more on the online platform. The need for sentiment analysis is to increase the sale of

the product and the customer's good choice of the product provides the benefit. Improvement of the product, quality, and sale of the product can predict the customer reviews that can be positive, negative, or neutral. By using sentiment analysis "to provide the high quality of services and also the improvement of the product's quality" enhances the customer traffic on the selling platform. Amazon product reviews can help the customer about the same product choice and it can also help the customer for better decision making.

The deep learning methods are used for difficult features extracted by natural language processing techniques based on sentiment analysis. The deep learning method is being used for the extraction of information "in a large amount of data" from the online platform. The deep learning method is used to extract complex features and it can also make the pattern from the big information. When the customers "after purchasing the product" leave positive reviews on the platform. If the reviews are positive then the customers can take the product confidently and in this way the organization can easily improve the business. Machine learning technique includes that includes LR, SVM, RF, NB, K-mean and also contains supervised DT, NB, SVM, RF, LR, LR Apriori and unsupervised technique K-men, Apriori Association, clustering. Deep learning technique also includes CNN, LSTM, RNN, RBFNS, supervised Feature engineering is applied that can predict the customer reviews about the product, therefore, thought the increase the traffic on the Amazon.

Significance:

Sentiment analysis is used for the positive-negative and neutral, opinion mining reviews extracts in the type of text that is about the brand of the item. Sentiment analysis is the kind of data mining which procedure preference of the customer reviews through the natural language processing used in the Bag of word or text. There is used the customer opinion extract from the online platform and other sources to enhance the traffic on the website. It is a prevailing text analysis tool that repeatedly used data which includes an online system used for customer reviews about the product, and it used deep learning and machine learning techniques.

This review paper includes section1: Introduction present on the sentiment analysis. In section 2: the Background of the sentiment analysis of customer reviews. In section 3: the literature review. In section 4: discussion about the Amazon product review. In section 5: Conclusion

II. BACKGROUND

As described above about the customer reviews, this part defines the customer review of Amazon product reviews. Amazon is an MNC of America that manages the customer product, sales and manufacture control. The products are present for customers in the market then the customer easily purchase and exchange the information over the Amazon. The AI, cloud system, IT and many other fields which are involved in the United State that focus on online platforms.

Amazon was an adventure by (Jeff Bozos in Bellevue, 1994). He selects the city of Seattle because there was Microsoft available. In 1997, Amazon was published for use publicly. In (1998, the start of the product was sold as music and video. In (2002) organizations become popular which provide the cloud services provider, web services provider, the pattern of the internet. In (2006) the organization was allowed to the customer virtual payment and data store over the internet. In (2012) organization was purchase the company which manufactures mobile robotic   The beginning of Amazon was based on the online system that provides the products in the market for the customer uses as books, foods, movie products, sports, toy, and necklaces. In (2015) the scope of Amazon was high as compared to "Wal-Mart, Big-box-store, and Super-market" in which they provided a big discount and food products for

the customer. In (2017) Amazon has acquired 13.4 billion of all food industry. In (2018) Amazon prime of two days product services was on the top in the world. It is the largest online system that also provides cloud computing and streaming services. Amazon prime was starting salaried payment, delivery of product movie review, books, foods, services providers and different other products. Amazon is the biggest Internet business by income in the world. In (2020) Amazon has become the largest valuable brand worldwide. In (2021) Amazon which is open in Italy two centre which provides the 1100 jobs for the peoples and also invests the 278 million dollars. Amazon acquires the online system of shopping which was used for the sale of different products. Shellfire is also an online shopping platform used for the cloth sale and book and Amazon purchased it.

III.   LITERATURE REVIEW

Literature reviews in the past literature many Deep learning and machine learning methods have been used for sentiment analysis of Amazon customer reviews.

A. *Deep learning*

[1]Stated that customers leave the review on the product when they purchase a product and collected the review of the customers about the products. A voting system was applied to customer's reviews for collecting data. Preprocessing techniques like removing space, basic punctuation marks and characters were used. A classification method with the deep learning model BERT was used to classify the customer's review on online products. The simple linear classifier provided better results that were helpful and predict online customer reviews. The use of classifiers was helpful to test the cross-domain and other domains that were used. The datasets were taken from the five type's category of products from Amazon.com. The results that were obtained from BERT

Classifier f1- score high as 0.777 to 0.885, zero and non zero similar domains 0.732 f1-score, and for All domain 0.550 f1-score. In the future, we will work on the big unlabeled product, product review data, surprised and unsupervised learning methods, and many other approaches. [2] Stated that customer reviews were helpful for decision making when the customer purchases the product on an online platform. Feature engineering and feature extraction were used and semi-supervised neural learning methods were used to predict a helpful review. RNN, CNN, MGCNN, MTN, semi-supervised neural learning method, semi-supervised, and pseudo-labeling review classifier were used for predicting the customer review and applied on the Amazon dataset which provides the best accuracy. Accuracy = 61.28 AUC= 65.35 RNN 58.63 57.25 61.31, CNN 59.21 57.60 61.98, MGCNN 60.10 58.05 62.11.

[3]Examined that online shopping platforms were dedicated to giving the sale of online products. The bulk of websites conveys their skilful opinion on the product to the customer when they bought it. The reviews provided high information about the customer's experiences, which were the best benefit for the product and customer .The Text Preparation, Normalization was implemented to perform the machine learning model RCNN. Two approaches supervised approach and semi-supervised were used to predict the helpfulness review. The SVM classifier was used on the dataset that Dataset was collected from the review of Amazon. Best accuracy results provided were Movies and TV, Fast text= 83.9%, SVM= 76.4%, Bi-LSTM (3 Layers) = 82%, CNN= 79%, RCNN= 84%. In the future, it can improve the technique and DL method, and also more cutting edges can be enhanced to the prediction of the classification results.

[4]Analyzed that the Neural network method was used for the customer review about the product, the AHUPA model was used for prediction of the customer attention, and the attention-based DKN model was used on the

customer news click record. The model involved three major modules like review encoded to define the review of the words, other user product review defined and rating encoder capture the customer attention. The pre-processing technique implemented was normalization and the dataset was collected from Amazon.com. The best outcome provided of model AHUPA 0.8676, 0.9181, 0.8699 and SVD method showed similarities of 0.9185, 0.9543, 0.8949 and also HFT showed the result 0.9073, 0.9475, 0.8920. In the future, we aim to more widely look at user-product attention relations to explore the different products that customers use and develop improved embedded learning methods for extracting users and products using better language models.

[5]Explained that High-level classification was helpful for the customer decision making about the products, customer review. Implemented pre-processing techniques were tokenization, transfer lowercase, specific word count that was used for the length of review, and also performed lemmatization. There were different methods implemented that were SRN, LSTM, CNN, Naïve Bayes, SVM, Decision tree. Four datasets were collected from Amazon, IMDB, Twitter, and Hotel reviews. Best results were provided in the short length and high reliability and also defined the word count accuracy of SRN (%) =78.10, LSTM (%) =85.82, and CNN (%) =87.42. The high readability accuracy LSTM was over 70%, SRN, CNN=60, and Low readability LSTM=50.In the future, Hyperparameter algorithms can Perform on classification and help with guidance and setting.

[6]It helped the customer decision making about the purchase of the products predict the review of the products. Implemented the GOLVE, uni-gram, Bi-gram, and machine learning techniques were used their LTSM, RNN, and Random forest. Dataset was collected from the Amazon product of twenty-four categories. Results were provided of different models random forest= F1 score of 0.565 with an accuracy of 0.619 and LSTM

an F1 score of 0.549 accuracies of 0.65.RNN F1 score of 0.156 accuracies 0.555. In the future work merge social network modelling technique and deep learning to know the drivers for helpfulness ratings.

[7]Stated as the product of movie customer reviews were using sentiment analysis. They were applying Pre-processing used for the customer review HTML tag, Removing stop words, and similar word clusters. The method of RNN and decision tree was applied on the data set of the movie review IMDB of providing result deep learning of the precision= 90.3% is recall= 87.2% and F-measure = 88.7%, accuracy= 89.8%.

[8] The focused opinion of the customer about the product; it expressed the opinion of the customer. Pre-processing implemented the opinion mining like Token, Tagging, (POS) tags, and feature engineering of the word embedding. The different techniques were used RNN, Bi-LSTM, and NN models apply on the dataset of two Amazon. They were provided the result Laptop LSTM-RNN of P= 81.92 of the R= 73.30 and F1= 77.14.

[9]Customer-focused on the product and find the best product on the base of the customer item reviews. It helped the customer for decision making when it purchased the item. Customer-focused positive and negative reviews and this purpose were implemented to the techniques; Tokenization, POS Tagging, Adjectives, Noun, and Adverb. Methods were used their Neural Network and Bayesian Classifier and Dataset were collected from the Amazon.com book reviews. The result showed this method NN=94.9 accuracy and Bayesian Classifier accuracy =84.7. In the Future, labelling can be performing on a different part of speech like adverbs, verbs, and now. That can help to improve the performance of the system.

B. Machine learning

[10]Explained that many online websites generate daily online customer reviews based on purchases. To predict the customer review and feedback on the product, novel

approaches were used for sentiment analysis. The model was performed by the Pre-processing techniques, Tokenization, Stemming, stop word removal, classification, negativity, positivity, and SVM classifier. Dataset was taken from the Amazon customer reviews and three kernels (RBF) graph form was plotted by using a confusion Matrix, curve, and evaluating the MSE accuracy precision-recall. The four classifiers of accuracy (LB-ALSA, BP-ALSA-L, BP-ALSA-P, BP-ALSA-R, BPALSA, and BP-ALSA) cross-validation phase score 97%. It means that only 3% of the information was lost. A vision, spam, fake, and sarcasm will be done in the future.

[11]Stated that customers share opinions on Online shopping platforms about the product. 81% of customers explore online platforms like the Amazon platform based on star rates and customer feedback. This research implemented pre-processing techniques like POS tagging, tokenization, stop word removal, extra spaces removal, HTML tags, special character, special symbol perform, ten adverb classifications, classify the positivity, negativity, and neutral. Different reviews were helpful in sentiment analysis to provide better accuracy. A hybrid technique was used that gather the part of words and adverbs and the dataset was taken from the online website Amazon DVDS musical product. Superlative-adv tenable the maximum value of F was 0.86. In the future, we will use to combine adverb, verb, and adjective to analyze sentiment classification after that this method will be used in other domains.

[12] Highlighted that customer reviews were helpful for other customers to make purchasing decisions on online shopping websites. Amazon.com has the top positive reviews and it helped the customers to make purchasing decisions about the product. In this model preprocessing technique was used POS tagging for sentiment score and noun, verb adjectives were calculated. Two techniques, RF and G-boosting were used and the dataset was taken from Amazon which is

the largest e-commerce website. The Random forest provided the results of Recall of 57.42, Precision was 51.25, and F-measure was 54.15. And Gradient boosting provided the results Recall of 57.21, Precision of 48.75, and F-measure of 52.64. It had shown the result of phone products reviews. In the future, we will apply cognitive behavior and other factors to refine the review helpfulness assessment.

[13]Predicted customer opinion and experience on the shopping platform about the product it helped for the choice of the good items. The main point was to examine the link of the customer review about the product that reviews helpfulness. The pre-processing techniques like removed the fake and duplicate review and perform normalization were implemented. Predictive methods for review helpfulness were used ML Ensemble, Random forest, gradient boosting, linear regression and MLP-BP, MARS, GLM, CART method; and Ensemble model provides good performance than others. The dataset was taken from the real-life review online Amzon.com website. The best results of the Ensemble model were MSE having 0.0199, RAE having 0.3192, RMSE having 0.1408, RRSE having 0.4204, and MAE 0.0991. In the future implement recommendation system span and text on proposed variables.

[14]Predicted the helpfulness review about the product according to the customer's behavior. They had taken reviews of helpfulness can enable the purchasing behavior of the customer. The pre-processing techniques of data mining; Min, Max, Mean Standard Deviation, and Number of Bins Discriminate were implemented in this model. The Naïve Bayesian network was used to guess the helpful review reviews of the item. The accuracy of NBN was predicted by using the KNN method and the NN model. When it was compared, NBN and NN capture the high noisy data and also capture a high relationship between variables. Crawled sample data was collected from Amazon.com. NBN results perform accurately by using the (kNN) method. Its Prediction performances were

helpful. In future studies, we can think about extra variables with important relationships with helpfulness.

[15]Analyzed that nowadays many people pass by existing online customer reviews on the majority of shopping platforms to create an improved buy decision. Pre-processing techniques like POS tagging, Tokenization, and a new method to categorize action verbs were used for the review of the text. Machine learning techniques used there were Random forest, Decision tree, and SVR. Dataset is collected from Amazon.com and the best result was provided by a random forest classifier. The results were Random forest 83.06, Decision trees 72.31, and F-measure 0.96 and 0.82. In the future, the system can be built as an unsupervised model based on the taxonomy category of verbs.

[16] Investigated that the customer leaves an online review in which interesting patterns can be tracked and it can improve the reputation of the business. The reviews were attracted the customer's attention which the majority lead consciously deviating from past rating behavior. User rating behavior was helpful for industry improvement. Implementation of the preprocessing techniques was feature engineering, mean, S.D, Min, and Max. The methods used in this model were linear regression and NBR. Dataset was collected from Yelp.com having number datasets of 90,671, Amazon, and IMDB reviews. The result provides the highest R2 0.450 compared with 0.406, 0.293 low RMSE 0.264, helpfulness reviews of 0.165, the value of R2 was increased to 0.177 and second, the boosted of 0.450. In the future, researchers can think about models by planned features and ML algorithms to predict the usefulness of online reviews.

[17]Explained that Selling websites play an important role in the customer behavior in the purchasing decision on the platform. The present position patterns of the platform were generally on the ratio of the helpful votes. Sometimes a few reviews cannot be helpful

for better rank which was caused this ranking pattern. Pre-processing techniques that were implemented Tokenization, Lowercase Conversion, Stemming, Stop word removal and SVM Logarithm, NB, and RF Algorithm. Dataset was collected about the reviews of mobile Redmi Note 5, Black, 64 GB, 4 GB ram that were from the online website flipkart.com. The best results that were provided by Random forest provide the best accuracy of 94.5, 94.5, 94.5, and 94.5 with the maximum feature. In the future direction, it will test the validity of a review and remove the fake review.

[18] Stated that sentiment analysis was performed on the customer online review about the products and also extracts the narrative form of text. Pre-processing techniques implemented in this mode were Lemmatization, POS tagging, and vectorization for the review of text analysis. The CRISP-DM, LATM, and SVM were used and provide the best results. The Naive Bayes model produced similar results of 0.889 and 0.888. Dataset was collected from the Amazon reviews of UCSD Design Lab. In the future, it can perform narrative for accuracy.

[19] Narrated that the online review provides many benefits for the customer when they purchase the product. The pre-processing technique was used fake review; stop word removal and lemmatization perform and also implement keyword extraction. The methods were implemented SMA-CNN classifier, LSTM, LDA FCM, RF, DT, NB. Dataset was collected from the Amazon.com review of the product which was provided after implemented and provide the results NB book, 84% of accuracy and kindle product, and 79.93% of accuracy, SMA-CNN 92.83% of classification accuracy. In future work, an effective unsupervised system will be industrial to further get better the classification accuracy of sentiment analysis.

[20] The online platform was the main part of the according to the sale point of view they prefer online sale and purchase. The reviews were helpful for possible clients to pick goods

from the vast number of options available and implemented the NLP pre-processing technique word embedding, extract words, word vector, and matching bigrams. Performed a convolution neural network (CNN) used to predict online customer review helpfulness. Star rating accuracy predicted the customer review about the product. Two datasets were collected from the Amazon product of review and the best result was provided from the CNN Text Star= Accuracy(%)76.14 of F1(%)=76 and NN Text= Accuracy(%)74.4 of F1= 74.4.

[21] For choosing a product, a purchaser wanted to go through many reviews to know produce. Implemented the pre-processing techniques Tokenization, removing Stop Words, POS tagging, and also implemented Feature Extraction. The different methods were used their linear SVM Multinomial, NB, Stochastic Gradient Descent, RF, LR, and Decision tree. Dataset was collected from the Amazon product reviews of the large dataset and the best result was provided Linear SVM of the accuracy 10 Fold=93.52 of the accuracy 5 Fold=91.72and Precision=0.98, Recall=0.99 and measure F1 score=0.98. In the future PCA can use to automated data and large dataset can be used local market and text base review and comment generalized

[22] Sentiment analysis used for the prediction of the customer opinion about the product. The pre-processing technique was used their tokenization, stop word removal, Punctuation, Lower case, Stemming, Lemmatizing performed on the customer review about the product. Implemented the different machine learning algorithms there Multinomial NB, Bernoulli NB, LR, SGD, Linear SVM and RF. Two different datasets were collected; one was Amazon book review and the second was crawling from the IMDB movie review site. The best result provided Linear SVM, Amazon results were Matthew association coeff of 0.84364 of average precision: 0.88930 and F1 Score, 0.92332 high accuracy: 92.18% and C=2. In the future financial, political, and social networks can

watch the optimization of parameters how they utilize to increase the accuracy.

[23] Predicted the customer reviews were used sentiment analysis base on the machine learning technique. It was very important for the online platform so there were Pre-processing methods implemented like Stemming, stop words removal, lower case, punctuations removal and white space. Implemented the method SVM, NB, Logistic Regression and Random Forest on the dataset. Dataset was collected from the mobile phone reviews Amazon.com and the best result provided Random Forest accuracy 90.2161 of 90.6622, SVM= 90.0066 and 90.3159. In the future short text and tweet text on apply Emoji2vec, doc2vec l for sentiment classification.

[24]Online plate form helped the customer decision making buying a product. To improve customer comment and selling skills, frequently the sellers gave a platform for the customers to communicate their views. The pre-processing technique was used Tokenizing, stop word removal, and also punctuation removal and word extraction. Implemented the different classification algorithms SVM, NB and Maximum Entropy Classifier. Dataset was collected from the Amazon product review fine food. Result provider of SVM accuracy (%)=82.85 and precision (%)=84.45 of recall (%)=82.13,F-measure (%)=82.38. In the future, it is the need of the hour to improve and detect fake comments. It can be helpful market and quality improvement of products.

[25]Review of the product was helpful for the customer on the online shopping platform about the fake brands of the product. Implement the pre-processing methods on the product review were used Tagging stop words, tokens, single words and bigrams. The use of method SVM NB, BN Decision tree Implemented on the Amazon dataset review of the products. The best result provided accuracy can be approximately 71.7% for 3 and 73%. BN NB provided 0.9. In the future web where tagging is unseen in the web and

we will add the prevision of add other languages terms.

[26]A lot of studies paid attention to the prediction of the helpfulness of customer reviews to discover helpful reviews. Customer opinion helped in the decision making and provided information about the sale and purchase behaviour of the product. Implemented Pre-processing POS tagging, Tokenization and Linear Regression. Dataset collected from the Amazon product review and also batter result provide AIC of Confidence= 83.112764 and DC =71.1737. In the future, it will implement on non-linear regression method.

[11] Strong and weak reviews predicted the customer about the product based on the sentiment analysis. Pre-processing technique there stop words, removed quote, and characters for use the extract the reviews. Implemented the different machine learning methods Bayes net, Naïve Bayes multinomial and SVM, Multilayer Layer. Dataset collected from the movie review and Best accuracy Naive Bayes multinomial= accuracy of 91.1696%. In the future, different techniques use pre-processing and attribute selection use.

[27] Predict the online customer behaviour about the product it helped the customer product and information. The pre-processing, feature selection and machine learning algorithm performed there. Tokenizing, Removing all stop words, lower case and Stemming perform on the base of sentiment analysis. Different machine learning classifier was used their NB, RF, Decision Tree, SVM and LR. Dataset collected from the Amazon customers' product-review data for Android Apps. The best result provided LR= high min =32.43%, Max =58.50 of accuracy and other Decision Tree= min in trigram: 24.10%, max in gram is 34.58%

[28] Predicted the online customer reviews about the product were helpful for the customer and online sailing platform. The pre-processing technique was applied on the text as the customer reviews that the products Part-of-speech, Tagging, Normalization, Stem mining, Tokenization, stop word, and punctuation removes process. The Implemented the LR method on the Dataset collected from the Amazon review of Apps for Android and the result was provided. LR multiclass classification method= average of (3.31%) and LR classifier with POS has higher classification accuracy (~13%).

[29]Online shopping decision for the customer was more difficult and the sale of the product also difficult and it helped them both through the predict the customer review. Implemented the pre-processing technique part-of-speech tagging, word segmentation, and Feature Extraction. The different models used and provided the results Regression Model, SVM, and random forest. Dataset collected from Amazon.com and provided the result average accuracy rate was 75.15% of the nonlinear SVM is 75.46%. In the future experimental method, customer attitude and other factors and helpful reviews can be added and the model can be improved. And also game theory and economic factors understand the customer reviews.

[30]Very bad impact of the customer of the fake review online shopping platform customer ignored the product cause of fake review. Pre-processing technique Implemented tokenization, stemming, tagging, fake review detection and feature extraction were used in the text. Different models implemented SVM, NB, decision tree, RF on the Dataset of Amazon Mechanical Turkers. And best result provide RF of Precisions is 86.01 is Recall = 89.89 %, of F-measure = 87.87 %, and Accuracy = 83.99%. In the Future develop taxonomy of nonverbal behaviours.

[31]Predicted the customer information online platform positive negative neutral reviews about the products .Pre-processing

implemented enhances the reviews of the Tokenized POS tagger, single word noun and Sentence Extraction for the customer reviews. The different methods implemented NB, SVM, Corpus and Ling Pipe Classifier. Data was collected from amazon.com using Soup API. The best result provides. Future work on preparation the classifier extra and building the sharing balance to minimize efficiency.

[32]customer opinion predicted on the base of sentiment analysis about the online purchased products. Stop word and tagging implemented different classification technique was used here NB, DT, NN and SVM applied on product movie review dataset. Results were provided accuracy=.7072, F1=.5984 with NN and accuracy=.7063, F1=.6033 with SVM of NB= 60% accuracy and accuracy=.7534 and F1=.6361.Future sentiment analysis increase to the different online website and health checks like Twitter and Face book.

[33]Online shopping helped the customer easily purchase the food from the Amazon platform. Predicted the customer review when they shop for any other product and different methods were used there. Pre-processing technique implanted there. Stop word removed and Remove Sparse Terms. The use of Latent Dirichlet Allocation and Regression analysis showed the customer review hidden factor. Dataset had taken from the Amazon grocery product online shopping customer reviews of coffee. The result provided Regression analysis a mean of 4.40 and an S.D of 1.08 and d increase by 57.9% if decreased by one unit. In the future other food type experiment.LDA is a powerful survey for the hidden text and finds necessary text.

[34] Predicted the customer reviews from the comment that helped the sentiment analysis. Sentiment Analysis is the method that defined the positive and negative or neutral reactions. Implemented the pre-processing method which was URL and HTML tag removal and also punctuation and number removed and included stop word removal and stemming. The different methods were used their KNN, (RF), (NB), Decision Trees, and also include

SVM. Dataset was taken from the Amazon book reviews. The result provided the All method accuracy of KNN was 90.00, DT of 88.44, NB of 94.62, SVM= 94.60, and RF= 94.40. In the future understand review of the product Hybrid model of SVM is used for the enhancement of accuracy.

[35] Online shopping helped customers and the organization of different opinions about the product. It provided the customer and manufacturers a large amount of review different and previous comments about the product. The pre-processing technique was used there by using sentiment method Features from Reviews, Tokenize, POS Tagger and also implemented normalization. The machine learning technique used their SVM and the base of the used SMO algorithm. Dataset was taken from the product of customer reviews from Amazon.com. The result provider of this model accuracy of 68.73339% and Rating helpfulness, SVM= accuracy of 68.7%.

[36] Predicted the customer reviews from the online Amazon electronic product. The reviews were considered helpful or unhelpful contain raw text on the product. The pre-processing technique implemented and feature selection method and they described the Anatomical features, Length Sentence count, Character count, Number of question marks, words in all caps and token-based, Lexical features, and also stop word. The methods implemented there NB and SVMs and dataset collected from Amazon electronic product reviews. The results were provided their SVM 70% predict accuracy, higher accuracy of 80%, and sigmoid SVM was of 50% accuracy and NB of 9.7%accuracy.In the future, Apply SVM linear kernel particular order of reviews rising specialized classifiers for special types should be best.

[37] The online platform the growing toward online shopping and it provided the help of customers and manufacturers through customer opinion. Pre-processing and NLP Feature extraction were used for the customer opinion and also used their POS tagged. The

different methods were implemented but there was used and provided the best result NB, DT J48 algorithm, MLP, and Bagging. Dataset had been collected from the Amazon review of product cell phone iPod laptop. The best result provided by Bagging of sentiment analysis was the Precision of 0.952, Recall of 0.951 and F-score was 0.942, and second was DT J48= Precision of 0.885, Recall of 0.872, and F-score of 0.855. In the future work improve opinion base on the anaphora resolution and this method deafening vague anaphora-antecedent necessary.

[38]Customer review helped when the purchase of the product and it helped any organization to increase the customer based on the point of sale. Customer review informed the interest of the customer choice about the product. Feature extraction helped the online customer review predict and also used their POS Tagging, Synonym Lexicon extension, Mean, and SD.LDA method used for the online customer review. Data set collected from the Amazon website of review of products. The results were provided LDA was F-Measures of Mean=72.25, SD=7.15, Recall of Mean=67.94,SD=9.87of the Precision=Mean=78.01 of SD=7.10 and Association Rule Mining of Precision was the Mean of =70.44 SD=9.31, Recall of Mean =56.54, SD=13.02, F-Measures of Mean =61.96, SD=9.66. Compare the other feature extraction method in the future.

[39] Sentiment Analysis used their predict the online customer reviews it helped off the customer for best decision making. The pre-processing technique was used uni-gram and bigram, POS, feature extraction based on the adverb and adjective. IG used for the famous feature extracts and also used the Method of Machine learning SVM, BMNB Algorithm. Dataset collected from the Amazon product review books, DVDs, and electronic. The result showed their F-measure (%) BMNB of F1= 82.7, PIGF1= 89.2, PmRMRF1= 90.2 and SVM of F1= 84.2, PIGF1= 85.8, PmRMRF1= 87.1, Feature vector size= F1= 9045, PIGF1 and PmRMRF1 =600.

[40]Predict the customer review based on the star rating and it helped them understand the customer and online platform of the decision when they manufacture and purchase the products. Sentiment Analysis used to predict the text and different method used Feature extraction, Stop word and Feature filtering. There were two methods Implemented RST-based algorithms ExhaustiveLEM2 and Association rule mining. The dataset from the online product review from the Amazon.com website. The result showed the Algorithm and high accuracy of the RST-based algorithms ExhaustiveLEM2.Association rule mining of the 192, 0.138, 0.700, 8.42, 0.004, RST exhaustive of 26, 0.068, 0.481, 81.4, 0.023 and RST LEM2 23, 0.028, 1.000, 36.7 0.027. Future work can consider increasing the review to face special types of product, trying other methods.

[41] Customer reviews helped when they purchase the product and online sale platform. Customer reviews helped the business platform they improve the manufacture good product and increased the sale of the product. Sentiment Analysis was used to predict the customer's positive and negative reviews, Stop words, Positive and negative tags, Positive POS and word tokens. Implemented the Method Sentiment-LDA and Dependency Sentiment-LD there. Dataset was taken from the Amazon product reviews Books and DVDs. The result provided the accuracy for the Sentiment-LDA of 60%~65%Dependency Sentiment LDA of 69.0%. In the future develop mutual sentiments and topic methods for many difficult in sentiment analysis, like as rating and sentiment.

[42]The customer left the review in the farm of the text and it helped the customer with the purchase of the products. The review was articulated on the farm of the text consist of the positive-negative and neutral. Sentiment Analysis was used to predict the customer reviews POS, Tokens, Verbs, Adjective Nouns and other features extracted involved. Machine learning method implemented there NB and Bayes I, II, and III. Dataset collected from the website www.amazon.co.uk movie

review and Books reviews. The results provide NB accuracy 75.0 % of the 76.3 % ,71.0 % ,70.3 % ,74.3 %and the average ,Std Err of 73.4 % ,1.16 % and also this Bayes 1,2,3, accuracy of 73.4 %, 72.1 %, 80.7 %.

[43] Reviews of the Customer were helpful for the organization and the large impact of the online platform. Sentiment reviews of the product helped the decision making both online platform and customer. Implemented the pre-processing technique Tokens, Stemming, stop words, punctuation, numbers and non-alphabet characters. The methods were used for the purpose to predict the review of the JST model, Reverse-JST and model LSM. Dataset collected from the Amazon.com books, DVDs, electronics, and kitchen. The result provides LSM achieved 72.5% and 74.1% accuracy. In the future semi-supervised learning of the JST method, other supervised and parameter opinion and methods like product reviews or the text label involuntarily

[44]Sentiment Analysis was used to predict the customer reviews positive negative and neutral. Customer opinion helped in the decision making of the customer and manufacturer about the product. This method POS Tagging, Word Level Features Vector, noun phrases, adjective and n-grams used for the extract the text. Implemented this technique to predict the online customer reviews (NB) classifier and (SVM) used. Dataset collected from different movie review product from Amazon.comcnet.com, epinions.com.This method after the experiment showed the accuracy of the product reviews, NB classification of the Recall 89% accuracy78%, NB-SVM sentences polarity of the Recall =90% accuracy=80% and Also ML sentences level polarity of the Recall=88% accuracy=82%. In the future Machine learning technique provide high results in our technique.

[45] Online review of the customer based on the sentiment analysis it can easily predict and it can help the customer to purchase products.

The pre-processing method was used to predict the opinion from the sentence about the product. Easily applied to stop word removal, stemming, Feature selection and uni-gram. These Methods were implemented NB and boosted SVM for the prediction of online customer reviews. Dataset collected from the Product reviews Amazon movie and book reviews. After the experiment results SVM with boosting the accuracy by 93%. In the future for the sentiment, Text blog Apply the NN.

[46] Customers left the opinion on the shopping web site then easily predict the customer positive review through the sentiment analysis. The pre-processing technique involved and Part-Of-Speech tagging apply to the Text and extracted features given data. Implemented the SVM method Dataset is collected from the reviews from Amazon. In and after the experiment provide the result accuracy rate of 77.98%.In the future work Named Entity Recognition can take the enhancement in the work.

[47] Customer reviews showed the online shopping system of the increased product quality. Customer review was very important for both customers and the online system then easily took better decision making. Predict the customer review used this technique Feature Extraction, Tokenization nouns, verbs, adjectives, adverbs HTML formatting tags applied on the dataset to predict the helpful review. SVM regression RBF kernel is used the automatically access the helpful reviews on the data set. Dataset was taken by the review of Amazon.com showed that after the experimental result of SVM Average 82%, 95% confidence. In the future, it can be done to summarize and extract the reviews which automatically generated.

[48]Customer opinions were taken by the company by hand then it too difficult to predict. Customer reviews were in the large amount and it was very helpful to purchase the products. Pre-processing technique Applied on the data set POS tagging, Remove HTML, Token, Removes punctuations used

for the classify the positive and negative reviews. The Methods were used machine learning rule-based, baseline system bootstrapping for the extracted review of large explanation about the products. Dataset was taken from the reviews of Camera Amazon and result in the show after experiments opinion sentence extraction sentence level R=91.41 (%), P=85.81(%), F=88.52(%).In the future can achieve the batter result pronoun resolution in perform sentence.

C. Hybrid

[49] Predicted the product features by using the helpfulness review that was increasing on online platforms day by day. Pre-processing in data mining was used the techniques of z-score, normalization, Blank text, or review removal. Machine learning models were used MAR, CART, RF, NN, and DNN. Best results were provided by Deep NN and the dataset was taken from the two real-life Amazon item reviews. The best results provided by the Deep NN DS2 dataset were (0.06 MSE, 0.237 RMSE, and 0.478 RRSE). In future studies, Hybrid evolutionary algorithms can be used and get better prediction accuracy of helpfulness to predict models.

[50]states that many business websites analyzed the interest of selling online products and predict the customer reviews that were left there. Implemented pre-processing techniques were normalization and text classification. There were three approaches used which were supervised learning, semi-supervised and pre-trained model. Supervised learning was done by using ML and Deep learning was performed by semi-supervised combining with DL, word embedding, and pre-trained data. Dataset was collected from the Amazon.com website and the Best Results provided by supervised and semi-supervised models were RCNN (T1) 83.25 %, RCNN (T2) 88.75 %, BERT 86.16 %, and Roberta 84.12%. In the future, we can explore more cutting edges and improve the predicted classification result.

[51]analyzed that the helpfulness review showed the relation word of mouth. It predicted the review with the help of the movie based on the director or power of the actor through the DEA. It showed the review sentiment positivity and negativity with competence to power review helpfulness. The pre-processing method implemented Mean SD, F, and Significance and showed the result. To predict, k-nearest neighbor and automatic neural network were used for providing the better result for the review helpfulness. Dataset was collected from the Naver Movies website. Best results provided for Helpfulness Efficient was 437, Inefficient was 437, t-value, p-value, RMSE = 31, 31 k-nearest 0.205 and NN 0.185, incompetent 437, 31 0.226 0.199 t-v and p-v 1.98, 0.026, 1.428, 0.080. In the future, this model can implement competence using unusual earnings to scale in the DEA model. The results of this study can hand out as the basis of a successful opinion web design that promotes review helpful.

[52] Said that helpfulness online reviews were predicted based on quality. The accessible techniques were mostly joint with opinion texts and ratings for helpful prediction. Deep learning used for the rating and TRI provided a better predictor of customer helpfulness review of the text. Pre-processing techniques that were implemented on blank and English review was filter and tokenization and stop word removal. The Reviews of customers on the online platform improve the product. Classification and regression algorithms were used, therefore, final prediction of the text review. The SVM classifier was used there. SVM and TFIDF were used for review of unigram, CNN was used for sentence classification, RAT and SVM were used for star rating information, EG-CNN was used for word-level embedding, CM1 was used for raw rating, LIWC and SVM were for word count dictionary-based, GI and SVM pragmatic part of speech and TRI add and TRI Concat both were used for the enhancement of the star rating method that can improve the rating. Dataset was

collected from the Amazon 5-Core. The best results of the star rating enhancement were TRI Add= 72.24*** 79.00*** 80.06** 87.01*** 83.58*** 80.45*** and TRIConcat= 72.04*** 78.37 80.22*** 87.22*** 83.50*** 80.57***. Further study will be based on the gating behaviors based to predict the star rating and review of the customer about the product.

[53] Explained that online customer reviews helped the online shopping platform to reduce the risk. These reviews helped the customer and retailer to suggest the product and it also helped the customer for the purchase of the good and the retailer to improve the product. Pre-processing techniques used in this model were Feature Engineering, non-negative matrix, factorization, feature extraction, and linear regression. Support vector classifier, KNN, Gaussian NB, and DT, two DL technique; CNN-1,2 CNN-GRU DL technique were used to predict opinion helpfulness. Dataset had collected two community Amazon product review datasets of House and Kindle. The best accuracy provided was 82% accuracy of Kindle" and dataset CNN-GRU 75% kitchen accuracy. In the future, the outcome of p-trained Text embeddings can be implemented on top of the appearance of the model.

[54]stated that the positive and unlabeled instance was the problem of binary classification like the review of helpfulness classification. The implemented pre-processing techniques were tokens, in the lexical feature using uni-gram and also implement syntactic feature technique. NCWS approach was used for the prediction of positive instances and unlabelled instances of the user review which were dependent on the customer opinion and reputation. SVM Classifier was used for user review, CNN prediction of the customer reviews and Deep CoNN was used for a low rate of prediction error. BERT classifier provides a high f1 score. There were three datasets collected; one was from Yelp and two were from Amazon. In the electronic datasets provided the result of BERT IS of F1= 0.5823,

Recall =0.7127 and Precision =0.4923. In the future, NCWS can be used for other binary classification techniques.[55] Assumes the problem the customer review played a vital role in the online platform. Products orders were generated on the base of the sentiment analysis .it also implemented the pre-processing, word embedding used for the bag of words and also CNN and LTSM used and RNN for the vector of word base of the online customer review. Dataset collected from the Amazon food review and the best result provided of the CNN-LSTM 0.8345. In the future apply deep network and feature matrix factorization to learn a function.[56] Online customer review helped the customer and online sealing platform of the best decision making. Predicted the online customer reviews of positive and negative and the pre-processing implemented for the customer positive and negative review URL, hyperlink removed, Tokenization word count and Punctuation removed. The supervised and unsupervised techniques were used for unable product five supervised and three lexicons used NB, LR, K-NN, RF, and SVM. Dataset collected from the phone 5s reviews from Amazon, Reevoo, and Face book. The best result provides Amazon dataset SVM accuracy =90.45 and precision=95.5 of Recall= 97.38. In the future predict the Majority review positive.

[57] Predicted the online customer reviews about the Amazon products and sentiment analysis performed on the unlocked mobile review. Pre-processing was used to enhance the accuracy of the classifier some were used there, Remove punctuations, Stop-Word Removal, Stemming, POS-Tagging, feature extraction, Word Cloud and also perform Data visualizations. Sentiment and Different classification algorithm were implemented NB-SVM, CNN, NB, LSTM, RF and Gradient Boosting Classifier and also performed Stochastic Gradient Descent (SGD). Dataset was collected from the Amazon .com unlocked mobile review and the Best result provided the RF accuracy= 85.50 of precision= 86, Recall= 86, F1-

Measure= 85.in the future can predict the rating of the text and achieve the best accuracy using the feature extraction on the cuprous based.

[58] Predicted customer opinion about the unlocked mobile phone product. It helped the customer and online platform decision making of the product. Implemented Pre-processing technique on the text, Tokenization, Spell check, Stop word, Lemmatization, Lower case, Feature extraction used TF, TF-IDF, Glove, Word2vec used for the positive and negative review. Machine learning algorithm was used their stochastic gradient descent (SGD) and NB, CNN. LR, GD and SVM. Dataset was collected from the Amazon mobile review unlocked system of reviews. And the best results provided CNN, Word2vec= accuracy (92.72%).In the future work include LSTM, KNN, and Maximum entropy.

[59]Predicted the customer review online shopping platform with the product of the star rating. Sentiment analysis defined the positive and negative reviews about the product when purchased. The pre-processing techniques were used their Remove hyperlink, Space removed, Punctuation and tokenization. Implemented the deep learning sentiment analysis, (RNN) with Gated Recurrent Units (GRU) SVM. Dataset was collected from the product of reviews from Amazon.com and the best result provided. SVM=accuracy of 81.29% and also increased the accuracy to 81.82%.In the future, a deep learning model based on sentiment analysis can build and enhance customer reviews.

[59] Predicted the customer opinions about the product, it depends on customer review which was the help of the customer and online platform decision making. For sentiment analysis performs the pre-processing technique, Tokenizing, Stop Words. Stemming and also Feature selection and

D. Tables of Deep learning methods

I.   TABLE

different classifiers were used NB and AdaBoost. Dataset was collected from the Twitter and Amazon product review and the best results were provided NB =86% of accuracy and t-score and z-score=90% accuracy.

[60] Predict customer opinion, attitude based on the sentiment Analysis about the products. Pre-processing methods were implemented Tagging, Keyword Extraction, POS, Unigrams, Bigrams, and URL. The different techniques were used NB and SVM. Dataset was collected Product of reviews from the Amazon and provided the result Sentiment Analysis of accuracy 80.4% and precision 80.3% of Neg. Rec. (%)=80.5 ,Pos .P. (%)=80.5 ,Neg .P. (%)=80.3.

[61] Customer Opinion, attitude, predicted to the based on the sentiment Analysis about the product takes best decision making. Data pre-processing, Stop words removal, Stemming and Pos tagging, and method NB implemented. Dataset collected from the www.amazon.com review Provided the best result Naive Bayesian of the accuracy of 80.36% and 92.37% of accuracy Precision, =75 and Recall =85.71, F-measure=80.36. . In the future, sum up the aspects based Extract the review can be examined the client's attractive aspects on goods.

Author Publication year	Pre-processing	Methodology	Datasets	Future direction	Results
Shih-Hung Wu and Yi-Kun Chen 2020	➤ space and basic punctuations marks ➤ characters	➤ Deep learning model BERT	Data set taken from the five product category from the Amazon.com	In the future, we will work on a big unable product, product review data, surprised and unsupervised learning method, and many other approaches.	BERT Classifier f1- score high as 0.777 to 0.885 zero and non zero similar domains 0.732 f1-score All domain 0.550 f1-score
Yue Feng, Miao Fan, Mingming Sun, and Ping Li 2020	➤ feature extraction ➤ feature engineering • LIWC, INQUIRER	➤ RNN ➤ CNN ➤ MGCNN ➤ MTN ➤ the semi-supervised neural learning method ➤ semi-supervised pseudo-labeling review classifier	datasets collect to form the garments and Electronic of Amazon	N/A	Cloth RSSNL F1 = 63.16 Accuracy = 61.28 AUC= 65.35 RNN 58.63 57.25 61.31 CNN 59.21 57.60 61.98 MGCNN 60.10 58.05 62.11
Abdalraheem Alsmadi, Shadi AlZu bi, Mahmoud Al-Ayyoub and Yaser Jararweh 2020	➤ Preprocessing ➤ Text classification ➤ Text preparation ➤ Normalization	➤ Supervised Learning ➤ RCNN ➤ Bi-RNN ➤ Fast text ➤ SVM ➤ Bi-LTSM ➤ CNN ➤ RCNN ➤ Semi-Supervised ➤ RCNN ➤ Word embedding ➤ Pre-trained ➤ BERT ROBERTa	Dataset collect Amazon product reviews	In the future, we can explore the more cutting edge and could improve the predicted classification result.	RCNN (T1), 83.25 %RCNN 88.75 % (T2), BERT, 86.16 % RoBERTa 84.12 %
Zhongqin Bi, Shuming Dou, Zhe Liu, and Yongbin Li 2020	➤ Removed the punctuation, stop words ➤ feature vectors ➤ feature extraction • LDA	➤ (AHUPA)model ➤ Neural network methods ➤ long short-term memory ➤ RNN	Dataset collect from the three real-world datasets from Amazon	In the future, we aim to more widely look at user-product attention relations to explore the different products that customer and develop improved embedded learning methods for extracting user and product and used their better language models.	AHUPA 0.8676 0.9181 0.8699 SVD 0.9185 0.9543 0.8949 HFT 0.9073 0.9475 0.8920

Lin Li, Tiong-Thye Goh, and Dawei Jin 2020	➤ Tokenization
	➤ transfer
	  lowercase
	➤ specific
	  word count
	➤ lemmatization
		➤ SRN
		➤ LSTM
		➤ CNN
		➤ Naïve Bayes
		➤ SVM
		➤ Decision tree
			datasets are collected from Amazon, IMDB	In the future, Hyperparameter algorithms can Perform on classification and help with guidance and setting.	LSTM is over 70% and SRN and CNN=60 and Low readability LSTM=50

Bobby Nguy 2016	➤ Pre-processing ➤ Segmenting ➤ Tokenizing ➤ Stop Words ➤ FEATURE SELECTION ➤ t-score ➤ Z-score ➤ Top unigram ➤ top bigrams	➤ Random forest ➤ RNN ➤ LSTM ➤ SVM ➤ NB ➤ AdaBoost	Dataset collect from the Amazon review data 24 types of product	In the future work social network and deep learning making helpfulness ratings.	random forest= F1 score of 0.565 accuracy of 0.619 LSTM an F1 score of 0.549 accuracies of 0.65. RNN F1 score of 0.156 accuracies 0.555
Arman S. Zharmagambetov and Alexandr A. Pak 2015	➤ HTML tags ➤ Removing stop words ➤ similar words clusters ➤ normalizing ➤ Feature extraction	➤ K-Means ➤ Neural Network ➤ Random Forest ➤ RNN ➤ Decision tree ➤ SVM ➤ BN ➤ Maximum entropy	Dataset collect from the movie review IMDB	future works enhancement homonymy and researching syntax arrangement of the text	deep learning precision= 90.3% recall= 87.2% F-measure = 88.7% Accuracy= 89.8%
Pengfei Liu , Shafiq Joty and Helen Meng 2015	➤ Preprocessing ➤ Token ➤ Stop Word Removal ➤ Tagging ➤ (POS) tags ➤ feature engineering ➤ word embedding ➤ Extraction feature ➤	➤ RNN. ➤ Bi-LSTM ➤ NN models ➤ NB	Dataset collect from the Amazon	fine-GD method opinion with term finding strength of sentiment of the opinion expressions. The RNN-based multi-task method used	Laptop LSTM-RNN P= 81.92 R= 73.30 F1= 77.14
Patrawadee Tanawongsuwan 2010	➤ Tokenization ➤ POS Tagging ➤ Adjectives ➤ Noun ➤ Adverb	➤ Neural Network ➤ Bayesian Classifier	Data set collected from the Amazon .com book reviews	In the Future, labelling can be performed on a different part of speech like adverbs, verbs, and now. That can help to improve the performance of the system.	NN=94.9 Accuracy Bayesian Classifier Accuracy =84.7

E. Tables of Machine learning methods

II.   TABLE

Author Publication year	Pre-processing	Methodology	Datasets	Future direction	Results
Neha Nandal, Rohit Tanwar, and Jyoti Pruthi 2020	➢ Preprocessing ➢ Vectorization ➢ Tokenization ➢ Stop word removal ➢ Pos TAGINING ➢ Stemming and lemmatization ➢ Bipolar word	SVM classifier used ➢ RBF kernel ➢ BP-ALSA-L, P, R.	Data set taken from Amazon customer reviews	A vision, spam, fake, and sarcasm will be done in future	BP-ALSAr cross-validation phase 97% score meaning that about only approximately 3% of information loss
Ummara Ahmed Chauhan, Muhammad Tanvir Afzal, Abdul Shahid, Mold Abdar, Mohammad Ehsan Basir and Xujuan Zhou 2020	Pre-processing POS tagging Extraction Adverb Classification • Positive negative neutral • Tokenization • Stop word • spaces • Html tagging • fresh line • font • Emotion • extraordinary symbol	Sentiment classification Hybrid approaches NP • Feature • (RRT) • (RT) • (RG) • (RG-WH) • RR • (RGT) • RL	Data set taken from the online website Amazon DVDS musical product	In the future, We will use to combine adverb verb and adjective to analysed sentiment classification after that this method will be used in other domains	Musical DVDs 75% review 4,5 star and adverbs have provided an F of 0.80. Sup-adverbs protected the high value of F- 0.86 "RRT'' of precision =0.89, recall, =0.84, F-measure=0.86
Ziming Zeng, Zhi Zhou and Xiangming Mu 2020	➢ Preprocessing • POS tagging • word segment	➢ Gradient boosting ➢ random-forest	Dataset collect form the Amazon largest e-commerce website	In the future, we will apply cognitive behaviour and other factors to refine review helpfulness assessment	Apple phone 6 Plus Classification method Random forest Recall 57.42 Precision 51.25 F-measure 54.15 Gradient boosting Recall 57.21 Precision 48.75 F-measure 52.64
M. S. I. Malik and Ayyaz Hussain 2018	➢ data cleaning ➢ remove duplicate	Ensemble method Random forest Gradient boosting Linear Regression	Dataset collect from the real-life Amazon.com website used.	In the future the implement recommendation	Ensemble MSE 0.0199 RAE 0.3192 RMSE 0.1408

	➤ reviews ➤ text removed ➤ Normalizatio n	MLP-BP MARS GLM CART		system span and text on proposed variables	RRSE 0.4204 MAE 0.0991
Sangjae Lee, Kun Chang Lee and Joon Yeon Choeh 2020	➤ Min, Max, Mean Standard Deviation, Number of Bins Discriminate	➤ Naive ➤ Bayesian Network ➤ NN	crawled sample data collected from the Amazon.com website	In the future studies can think about extra variables with important relationships with the helpfulness	NBN perform accuracy using (kNN) method Prediction performance helpfulness
Mina Akbarabadi and Monireh Hosseini 2020	➤ Preprocessin g ➤ POS tagging ➤ Tokenized	➤ Random forest ➤ Decision tree ➤ The new model categorizes action verbs. SVR	Dataset collect from the real-life Amazon datasets	future work can be built on the unsupervised model on the bases of the taxonomy category of verb	Book reviews Accuracy Random forest 83.06 Decision trees 72.31 F-measure 0.96 0.82
MUHAMMAD BILAL, MOHSEN MARJANI1, MUHAMMAD IKRAMULLAH LALI , NADIA MALIK , ABDULLAH GANI and IBRAHIM ABAKER TARGIO HASHEM 2020	➤ Preprocessin g technique ➤ Feature engineering ➤ Mean, Std. Dev, Min, Max ➤ Feature extraction • Feature mapping	➤ LR ➤ NBR	Dataset collect from the Yelp.com 90,671and Amazon, review IMDB	future will be thought by planned feature and ML algorithms to expect the usefulness of online OPINION	maximum R2 0.450 compare to 0.406 and 0.293 low RMSE 0.264 e helpful of reviews is 0.165 value of R2 is improved of 0.177 R2 is boosted to 0.450
Paritosh Tripathi a, Sonu Singh b, Pragya Chhajer b, Munesh Chandra Trivedi b, and Vineet K. Singh a 2020	➤ Preprocessin g ➤ Tokenization ➤ Lowercase ➤ Stemming ➤ Stop word removal	➤ SVM ➤ RF ➤ NB	Dataset collect from the review of mobile Redmi N from online-web flipkart.com	In a future way will be to check the validity of an assessment and remove the fake review	Random Forest 94.5 ,94.5 94.5 maximum feature
Vineet Jain And Mayur Kambli 2020	➤ POS tagging ➤ vectorization	➤ CRISP-DM ➤ LATAM ➤ SVM	Dataset is collected from the Amazon reviews of UCSD	In the future PCA model provide sentiment on the base of data understanding	NB model produces similar results of 0.889 and 0.888
Trupthi Mandhula, Suresh Pabboju, and Narsimha Gugulou 2019	➤ Lemmatizati on ➤ URLs ➤ stop words ➤ spam detection ➤ Feature extraction • (BOW)	➤ SMA-CNN classifier ➤ LSTM ➤ LDA -FCM ➤ RF ➤ DT ➤ NB	Data collected from the Amazon.com review of a product	In future work, an unsupervised method more enhances the accuracy of sentiment analysis.	NB book, 84% A kindle product, and 79.93% of A SMA-CNN 92.83% of classification accuracy
Xianshan Qu,	➤ word	➤ CNN	Two datasets	N/A	accuracy of

Xiaopeng Li, and John R. Rose 2018	embedding ➢ extract the words ➢ word vector ➢ matching bigrams ➢ feature maps ➢ bag-of-words	➢ SVM regression ➢ (RNN) ➢ (GRU)	collect from the Amazon product of review		76.14% the best accuracy of 71.9% CNN Text Star= Accuracy(%)76.14 F1(%)=76 CNN Text= Accuracy(%) 74.4 F1= 74.4
Mohammad Mohaiminul Islam and Naznin Sultana 2018	➢ Preprocessing ➢ Tokenization ➢ Remove stop word ➢ Punctuation and stop word ➢ Lower case ➢ Stemming ➢ Lemmatizing ➢ Feature extraction ➢ Feature map	➢ Multinomial NB ➢ Bernoulli NB ➢ LR ➢ SGD ➢ Linear SVM ➢ RF	Two different data set collect one is Amazon book review second is crawling from IMDB movie review site	In the future Financial, political, and social networks can be watch optimization of parameters how they utilize to increase the accuracy.	Linear SVM association co-eff of 0.84364 A P=0.88930 F1=0.92332 A= 92.18% C=2
Barkha Bansal and Sangeet Srivastava 2018	➢ Pre-processing • Stem • Stop-words • Lower-C • Punctuation • white-space ➢ Feature Extraction • PCA	➢ SVM, ➢ NB, ➢ LR ➢ RF	Dataset collect from the mobile phone review Amazon.com	In the future Short text and tweets text on apply Emoji2vec ,doc2vec l for sentiment classification.	Random Forest Accuracy 90.2161 90.6622 SVM= 90.0066 90.3159
Chetana Pujari, Aiswarya and Nisha P. Shetty 2018	➢ Tokenizing ➢ stop word removal ➢ Punctuation removed. ➢ Word extraction ➢ Bag of word	➢ SVM ➢ NB ➢ MNB ➢ LR ➢ Linear SVM ➢ SGD ➢ RF	Dataset collects from the Amazon product review fine food.	In the future need to improve detect fake comment. it can be helpful market and quality improvement of product	SVM Accuracy (%)=82.85 Precision (%)= 84.45 Recall (%)=82.13 F-measure (%)=82.38
Mohan Kamal Hassan, Sana Prasanth Shakthi, and R Sasikala 2017	➢ preprocessing ➢ Tagging ➢ stop words ➢ tokens ➢ single words ➢ bigrams, ➢ Feature extract	➢ Naive Bayes ➢ SVM ➢ decision tree ➢ BN	Dataset collect from Amazon product review	In the future can be hidden tags in the browser and Prevision can add other languages for better results.	Naive Bayes Accuracy=0.9 accuracy can be approximately 71.7% for 3

	• Bag of word				73%. BN
Shih-Hung Wu and Yi-Kun Chen 2020	➢ space and basic punctuations marks ➢ characters	➢ Deep learning model BERT	Data set taken from the five product category from the Amazon.com	In the future, we will work on a big unable product, product review data, surprised and unsupervised learning method, and many other approaches.	BERT Classifier f1- score high as 0.777 to 0.885 zero and non zero similar domains 0.732 f1-score All domain 0.550 f1-score
Erfan Ahmed, Md. Asad Uzzaman Sazzad, Md. Tanzim Islam, Muhitun Azad, Samiul Islam, and Dr. Mohammad Haider Ali 2017	➢ Preprocessing ➢ stop words a ➢ removed quote characters ➢ remove stop words	➢ Bayes net ➢ Naïve Bayes multinomial ➢ SVM ➢ Multilayer Layer	Dataset collect from the product of movie review	The different technique used pre-processing and attribute selection used in future	Naïve Bayes multinomial= accuracy of 91.1696%
Tomas PRANCKEVICIUS and Virginijus MARCINKEVICIUS 2017	➢ Preprocessing ➢ Tokenizing ➢ Removing all stop words ➢ lower case ➢ Stemming	➢ NB ➢ RF, ➢ DT ➢ SVM, ➢ LR	Amazon consumer' product-review data for Android	N/A	Logistic Reg= highest A of min ,max=32.43%, max =58.50% Decision Tree= min in trigram: 24.10%, max in gram is 34.58%)
Tomas Pranckeviþius and Virginijus Marcinkeviþius 2016	➢ Part-of-speech Tagging ➢ Normalization ➢ Stem mining ➢ Tokenization ➢ stop word ➢ punctuation ➢ feature extraction ➢ Bag of word	➢ LR ➢ SVM ➢ NB ➢ RF ➢ DT	Dataset collect from the Amazon review of Apps for Android	N/A	LR multiclass classification method= average 3.31 LR by POS high A of ~13%
Xu Chen, Jie Sheng, Xiaojun Wang, and Jiangshan Deng 2016	➢ part-of-speech tagging ➢ word segmentation ➢ Feature Extraction ➢ feature dimension ➢ bag of word	➢ Regression Model ➢ SVM ➢ random-forest ➢ LSTM ➢ RNN	Dataset collect from Amazon .com	In the future experimental method, customer attitude and other factor and helpful reviews can be added then improve the model. and also the game theory and economic factor understand the customer reviews	The average accuracy rate is 75.15% nonlinear SVM 75.46%

Dongsong Zhang, Lina Zhou, Juan Luo Kehoe & Isil Yakut Kilic 2016	➢ tokenization, ➢ stemming, ➢ tagging ➢ POS ➢ fake review detection ➢ Uni-gram ➢ Bi-gram ➢ Tri-gram ➢ feature extraction	➢ SVM, ➢ NB ➢ DT ➢ RF	Dataset collect from Amazon Mechanical Turkers	expand a categorization of nonverbal attitude	RF of Perception = 86.01 Recall = 89.89%, F = 87.87 % A = 83.99 %
Bryan Nii Lartey Laryea, Chi-Hwan Choi, In-Sun Jung, Kyung-Hee Lee, and Wan-Sup Cho 2015	➢ Tokenized ➢ POS tagger ➢ single word noun ➢ Sentence Extraction	➢ NB ➢ SVM ➢ Ling- Pipe Classifier	Data collected from amazon.com using Jsoup API	Future work on preparation the classifier extra and building the sharing balanced to minimize efficiency.	accuracy =(70%)
Yoosin Kim, Do Young Kwon and Seung Ryul Jeong 2015	➢ stop-words ➢ tagged ➢ selecting features	➢ NB ➢ DT, ➢ NN, ➢ SVM	Data collected from a product movie review.	Future sentiment analysis increase to the different online websites and health checks like Twitter and Face book.	(accuracy=.7072, F1=.5984 with NN (accuracy=.7063, F1=.6033 with SVM NB= 60% accuracy a (accuracy=.7534, F1=.6361
Yan Heng, Zhifeng Gao, Yuan Jiang, Xuqi Chen 2018	➢ Stop word removed	➢ Latent Dirichlet Allocation ➢ Regression analysis	Dataset is taken from the Amazon grocery product online shopping customer reviews of coffee.	In the future other food category test.LDA is a controlling implement to discover the unseen text. find required text	Regression analysis a Mean of 4.40 and an S.D of 1.08 d enhance by 57.9% if the decrease by 1 unit
K. S. Srujan , S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish, and H. M. Keerthi Kumar 2018	➢ Pre-processing ➢ Url and ➢ HTML tag remove ➢ punctuation ➢ stop word removal and stemming	➢ KNN ➢ (RF) ➢ (NB) ➢ Decision Trees ➢ (SVM)	Data set taken from the Amazon book reviews dataset	In the future Understand the review of the product Hybrid model of SVM used then enhance accuracy.	Accuracy of KNN= 90.00 Decision Trees = 88.44 Naive Bayes= 94.62 SVM= 94.60 Random forest= 94.40
Yadong Zhang and Du Zhang 2014	➢ Features from Reviews ➢ Tokenize ➢ POS Tagger ➢ normalization	➢ SVM ➢ (SMO) algorithm	Data set collected from the product of customer reviews from Amazon.com	N/A	Accuracy= 68.73339% Rating ,helpfulness SVM= accuracy of 68.7%
Jordan Rodak, Minna Xiao and Steven Longoria 2014	➢ Pre-processing ➢ Anatomical features • Len-Sentence • Char add • Number of?	➢ Naive Bayes ➢ SVMs	Dataset collect from the Amazon product reviews	in the future, Apply SVM linear kernel particular order of reviews rising specialized classifiers for special types should be best.	SVM 70% prediction accuracy higher accuracy around 80% sigmoid

	marks CAPS ➤ token-based ➤ Lexical features ➤ stop word				SVM=50% Accuracy Nave Bayes =9.7%Accuracy
Ahmad Kamal and Muhammad Abulaish 2013	➤ Feature extraction ➤ POS-tagged ➤ embedded	➤ NB ➤ Decision Tree J48 algorithm ➤ MLP ➤ Bagging	Dataset collect from Amazon review of product cell phone iPod laptop	In the future work improve opinion base on the anaphora resolution This method deafening vague anaphora-antecedent necessary	Bagging sentiment P = 0.952 R = 0.951 F= 0.942 D.T J48= P=0.885 R=0.872 F=0.855
Baizhang Ma, Dongsong Zhang, Zhijun Yan, Taeha Kim 2013	➤ Feature extraction ➤ PoS Tagging ➤ Synonym Lexicon extension ➤ Mean ➤ SD	➤ LDA	Data set collect from Amazon website of review of the product	Compare the other feature extraction method in the future	LDA= F-Measures Mean=72.25 SD=7.15 Recall=Mean =67.94 SD=9.87 Precision=Mean=78.01 SD=7.10 Association Rule Mining Precision Mean =70.44 SD=9.31 Recall= Mean =56.54 SD=13.02 F-Measures Mean =61.96 SD=9.66
Basant Agarwal and Namita Mittal 2013	➤ Features Extraction ➤ pre-processing ➤ POS ➤ Unigrams ➤ bi gram ➤ adverb and adjective	➤ BMNB ➤ SVM	Dataset collect from the Amazon product review books, DVD, and electronic	N/A	F-measure (%) BMNB F1= 82.7 PIGF1= 89.2 PmRMRF1= 90.2 SVM F1= 84.2 PIGF1= 85.8 PmRMRF1= 87.1 Feature vector size= F1= 9045 PIGF1 and PmRMRF1 =600
Wingyan Chung and Tzu-Liang (Bill) Tseng 2012	➤ Feature extraction ➤ Stop word ➤ Feature filtering	➤ RST-based algorithms ExhaustiveLEM 2 ➤ Association rule	The dataset from the online product review from the Amazon.com website	The future job can think rising the evaluation to features unique types of produce, difficult other	Association rule mining 192 0.138 0.700 8.42 0.004

		mining ➤		methods	RST – exhaustive 26 0.068 0.481 81.4 0.023 RST – LEM2 23, 0.028 ,1.000 ,3 6.7 0.027
Fangtao Li, Minlie Huang, and Xiaoyan Zhu 2010	➤ Stop words ➤ Positive and negative tags. ➤ positive (Pos) ➤ SentiWordN et ➤ word token	➤ Sentiment-LDA ➤ Dependency Sentiment-LDA	Data set taken from the Amazon product reviews. Books and DVDs	I n the future develop mutual sentiment and topic method for many difficult in sentiment analysis, like as rating and sentiment	Accuracy of the Sentiment-LDA is approximately 60,~65 percent Dependency Sentiment LDA= 69.0%
Jeremy Reffin, Taras E. Zagibalov, and Ekaterina O. Belyatskaya 2010	➤ POS ➤ Tokens ➤ Verbs Adjecti ve Nouns Other ➤ features extracte d	➤ Naive Bayes ➤ Bayes I, II, and III	Data set collected from the website www.amazon.co.u k movie review and Books review	N/A	Naive Bayes Accuracy 75.0 %76.3 % 71.0 % 70.3 % 74.3 % Average Std Err 73.4 % 1.16 % Bayes 1,2,3, 73.4 %, 72.1 %, 80.7 %; the Naive Bayes
Chenghua Lin, Yulan He and Richard Everson 2010	➤ preprocessin g ➤ Tokens ➤ Stemming ➤ stop words ➤ punctuation ➤ numbers ➤ non-alphabet characters	➤ JST model ➤ Reverse-JST ➤ model LSM	Data set collect from the Amazon.com books, DVDs, electronics, and kitchen	In the future semi-supervised learning of the JST method, other supervised and parameter opinion and methods like product reviews or the text label involuntarily	LSM get 72.5% and 74.1% Accuracy
Aurangzeb khan, Bharum Baharudin and Khairullah khan 2010	➤ POS Taggin g ➤ Word Level Feature s Vector ➤ noun phrases, adjectiv e ➤ n-grams	• (NB) classifier • (SVM)	Dataset collect from different movie review product from Amazon.com cnet.com, epinions.com	Machine learning technique provides high result our technique.	NB classification Recall 89% Accuracy78% NB-SVM Sentences Polarity Recall =90% Accuracy=80% ML Sentences level Polarity=Recall =88% Accuracy= 82%
Anuj Sharma and Shubhamoy Dey 2013	➤ pre-process ed ➤ stop	➤ NB ➤ boosted SVM	Dataset collect from Product reviews Amazon movie and books	In the future for the sentiment, Text blog Apply the NN	SVM with boosting the accuracy of 93%

	➢ word removal ➢ stemming ➢ Feature selection ➢ N-gram		reviews		
Raisa Varghese and Jayasree M 2013	➢ Part-Of-Speech tagging ➢ extracted features noun phrases	➢ SVM	Data set collected from the reviews from Amazon. in	Named Entity Recognition can take the enhancement in the work	the accuracy rate of 77.98%
Soo-Min Kim, Patrick Pantel, Tim Chklovski and Marco Pennacchiotti 2006	➢ Feature Extraction ➢ Tokenization ➢ nouns, verbs, adjectives and adverbs ➢ HTML formatting tags ➢	➢ SVM regression ➢ RBF kernel	the dataset is taken by the review of Amazon.com	In the future can be done summarized and extract the reviews which automatically generated	SVM Average 82% 95% confidence
Wei Jin, Hung Hay Ho and Rohini K. Srihar 2009	➢ Pre-processing ➢ POS tagging ➢ Remove HTML ➢ Token ➢ Tagging ➢ Removes punctuations	➢ rule-based baseline system bootstrapping	Dataset is taken from the reviews of Camera Amazon	In the future can achieve the batter result pronoun resolution in perform sentence.	Opinion sentence extraction sentence level R=91.41 (%) P=85.81(%) F=88.52(%)

F. Tables of Hybrid Methods

III. TABLE

Author Publication year	Pre-processing	Methodology	Datasets	Future direction	Results
Muhammad Shahid Iqbal Malik 2020	➢ Preprocessing ➢ z-score ➢ normalization ➢ Blank text or review remove ➢ Feature extraction • feature matrix	➢ Machine learning model ➢ MAR, ➢ CART, ➢ RF ➢ N-Net ➢ DL NN	Data set in use from the two real-life Amazon item review	in future studies. Hybrid evolutionary algorithms used, get the better P of the accuracy of a helpful predict of model	Deep NN DS2 dataset 0.06 MSE, 0.237 RMSE and 0.478 RRSP
Abdalraheem Alsmadi, Shadi AlZu bi, Mahmoud Al-Ayyoub and	➢ Preprocessing ➢ Text classification ➢ Text	➢ Supervised Learning ➢ RCNN ➢ Bi-RNN ➢ Fast text	Dataset collect Amazon product reviews	In the future, we can explore the more cutting edge and could improve the predicted classification result.	RCNN (T1), 83.25 %RCNN 88.75 % (T2), BERT, 86.16 %

Yaser Jararweh 2020	➢ preparation ➢ Normalizatio n	➢ SVM ➢ Bi-LTSM ➢ CNN ➢ RCNN ➢ Semi-Supervised ➢ RCNN ➢ Word embedding ➢ Pre-trained ➢ BERT ROBERTa			RoBERTa 84.12 %
Sangjae Lee and Joon Yeon Choeh 2020	➢ Preprocessin g ➢ Mean, ➢ SD, F, Significance	➢ k-nearest ➢ NN ➢ RMSE ➢ decision tree	Data collected from the Naver Movies website	Can implement the competence using unusual earnings to the extent in the DEA technique	k-nearest and NN 0.205 0.185 Inefficient (437) 31 0.226 0.199 t-v ,p-v 1.98 ,0.026 1.428 ,0.080,
Jiahua Du Liping Zheng,Jiantao He, Jia Rong, Hua Wang and Yanchun Zhang 2020	➢ Pre-processing ➢ empty and non-English ➢ Tokenized ➢ stop word removal ➢ word embeddings ➢ feature EXTRACTI ON • TRI • feature maps	➢ linear SVM classifiers ➢ CNN Classification regression algorithms ➢ TFIDF + SVM ➢ SIGNS + SVM ➢ GV + SVM ➢ DS + SVM ➢ LIWC + SVM ➢ GI + SVM ➢ RAT + SVM ➢ EG-CNN ➢ CM1 ➢ CM2 ➢ MTNL ➢ TRIAdd ➢ TRIConcat	Dataset collect form the Amazon 5-Core dataset	Further study on the gating behaviours base predict the star rating review of the customer about the product	TRIAdd= 72.24*** 79.00*** 80.06** 87.01*** 83.58*** 80.45*** TRIConcat= 72.04*** 78.37 80.22*** 87.22*** 83.50*** 80.57***
Mohammad Ehsan Basiri Shirin Habibi 2020	➢ Feature Engineering ➢ non-negative matrix ➢ factorization ➢ feature extraction • reduce the dimensionali ty • (LSA) • (NMF)	➢ LR ➢ SVM ➢ K-n ➢ Gaussian (NB), ➢ (DT), ➢ DL, CNN-1 and 2 ➢ LSTM ➢ RNNs	Two community Amazon review dataset House and Kitchen and Kindle	the outcome of pre-trained word embeddings on the presentation of the method	CLASS NEGATIVE AND POSITIVE precision = 0.90 P Recall= 0.73 f1-score= 0.81 NEGATIVE Class p = 0.77 R= 0.92 f1= 0.84 (CNN-GRU) 82 ACCURACY

					Kindle" dataset (CNN-GRU) 75% KITCHEN ACCURACY
Xi Wang, Each Ounis and Craig Macdonald 2020	➢ Tokens ➢ Lexical Feature, Unigram ➢ Syntactic Feature ➢ Structural feature • ASL ➢ Syntactic Feature • syn	➢ CNN ➢ BERT ➢ SVM ➢ NCWS supervised approach ➢ Deep CoNN	Three data sets used 1 is Yelp and the second is Amazon Kind, Elec reviews	NCWS used another binary classification technique	SVM-ALL Yelp= 0.6340 Kindle= 0.5877 Electronics= 0.6336
Bui Thanh Hung 2020	➢ Preproc essing ➢ Word Embed dings ➢ Vactor	➢ CNN-LSTM ➢ RNNs	Dataset collect from the Amazon food review	. In the future apply deep network and feature matrix factorization to learn a function	Con -NN =0.8276 Long-STM= 0.8254 CNN-Long-STM= 0.8345
Waqar Muhammad, Maria Mushtaq, Khurum Nazir Junejo, and Muhammad Yaseen Khan 2020	➢ Tokenized ➢ Word Count ➢ Punctuation ➢ Classificatio n ➢ URLs removed ➢ UNI-Gram ➢ Bi-Gram	➢ NB ➢ LR ➢ K-NN ➢ RF ➢ SVM ➢ Average	Dataset collect from the I phone 5s reviews from Amazon, Reevoo, Face book	Majority review Semi-supervised Approach get better the result	Amazon dataset svm Accuracy =90.45 Precision=95.5 Recal= 97.38
Momina Shaheen, Shahid M. Awan, Nisar Hussain, and Zaheer A. Gondal 2020	➢ Preprocessin g • Remove punctuation • Stop-Word Removal • Stemm • POS • Tagg • feature extract • Feature map • Word Cloud • Data visualization s • word embedding	➢ Classification ➢ NB-SVM ➢ CNN ➢ NB, ➢ LSTM ➢ RF ➢ Gradient Boosting Classifier ➢ Stochastic Gradient Descent (SGD)	Dataset collect from the Amazon .com unlocked mobile review	in the future can predict the rating of the text and achieve the best accuracy using the feature extraction on the cuprous based.	Random Forest Accuracy= 85.50 Precision= 86 Recall= 86 F1- Measure= 85
Sara Ashour Aljuhani and Norah Saleh Alghamdi	➢ Preprocessin g ➢ Tokenization	➢ (SGD) ➢ NB ➢ CNN	Dataset collect from the Amazon mobile review	future work includes LSTM, KNN, and Maximum entropy.	CNN Word2vec= accuracy

2019	➢ Spell check ➢ Stop word ➢ Lemmatization ➢ Lower case ➢ Feature extraction • Unigram • Bi-gram • Word-2vec	➢ . Logistic regression LR gradient descent (GD) ➢ SVM	unlocked.	Increase the accuracy	(92.72%) word2vec With CNN LL= 0.52 F = 79.57 RC= 79.57 PR = 79.55 ACC= 79.60
Nishit Shrestha and Fatma Nasoz 2019	➢ Preprocessing • Remove hyperlink • Space removed • Punctuation • Tokenization ➢ feature extraction • PCA	➢ (RNN)with Gated Recurrent Units (GRU) ➢ SVM	Product reviews collected from Amazon.com	In the future, a deep learning model on the base of sentiment analysis can build and enhance the customer review	SVM=accuracy of 81.29% increases the accuracy to 81.82%
Sudarshan S. Sonawane and Satish R. Kolhe 2017	➢ preprocessing ➢ Segmenting and Tokenizing ➢ Stop Words ➢ Stemming ➢ Feature extraction	➢ SVM ➢ NB ➢ AdaBoost	Dataset collect from the Twitter and Amazon product review	N/A	NB =86% of accuracy n t-score and z-score=90% accuracy
Jesus Serrano-Guerrero, Jose A. Olivas, Francisco P. Romero, Enrique Herrera-Viedma 2015	➢ Tagging ➢ Keyword Extraction ➢ POS ➢ Unigrams ➢ Bigrams ➢ URL ➢ Feature Extraction	➢ Naive Bayes ➢ Support Vector Machines ➢	Dataset Product reviews from the Amazon dataset	N/A	Sentiment Analysis accuracy 80.4% precision 80.3% Neg. Rec. (%)=80.5 Pos.Prec. (%)=80.5 Neg.Prec. (%)=80.3 Neutral=0
A.Jeyapriya (P.G Scholar) and C.S.Kanimozhi Selvi (Associate Professor) 2015	➢ Data pre-processing ➢ Stop-words remove ➢ Stem ➢ POS ➢ Tag ➢ Feature extract ➢ T-test	➢ Naive Bayesian ➢ RF ➢ SVM ➢	Dataset collects from the www.amazon.com review	In the future sum up the aspects based Extract the review .can be examine the client's attractive aspects on goods.	Naive Bayesian accuracy of 80.36% 92.37% of accuracy Precision, = 75 Recall =85.71 F-measure=80.36

IV.   DISCUSSION

Most of the methods of sentiment analysis depend upon machine learning and deep learning which target the increase of sale through customer reviews. During the study focus on the enhancement of the sale of customer's, there is a target domain of the sentiment analysis using machine learning and deep learning techniques. Since discussion in this relative research, a few problems have come to any time of the large amount of data which is based on customer's reviews. Based on the literature review the authors can organize the sentiment analysis techniques, which are using the predict the customer opinion about the product. Following explore and investigative the limitation a few removals are capable of being helpful to get better accuracy and extract the online customer reviews using sentiment analysis.

Limitation:

The use of supervised and unsupervised techniques will help the sentiment analysis of customer's reviews and can also predict the use of spam, fake, and sarcasm for the customer reviews. Some gaps are also left in this literature review, in the future it will combine adverb, adjective, adverb to analyse sentiment analysis which will create batter results, and will also implement the span and text on proposed variables for the removal of a customer product review of online shopping organization. Implementation of the unsupervised data to analyse taxonomy category, Hybrid evolutionary technique of the verb, and cutting edge then will enhance the results about the shopping platform of the customer reviews. Machine learning algorithms will utilize to predict the usefulness of online customer reviews then enhance the accuracy of the model. There are different models still necessary, which will gain better results of the customer reviews deep network, feature matrix factorization, Semi-supervised Approach, LSTM, KNN, and Maximum entropy Increase the accuracy.

V. CONCLUSIONS

More than the days, sentiment analysis has gained usually a bundle of interest from experimenters effective in the area of using machine learning, deep learning extracts opinion form the text. Sentiment analysis achieves much interest in the Amazon product of customer reviews. This research provides the different methods of Amazon product review of customer behaviour. Most researchers look at many challenges, such as the same product choice then it is very difficult to predict the customer's review predict on the online platform of Amazon. Starting research since they require of enhance the sentiment online customer reviews, on the other area that provides the skill to remove the opinion, judgment, and customer reaction that provides via customer. Therefore, sentiment analysis is used to predict the customer's opinion from the Amazon product review since the explore issue via a lot of research person explains the difficulty by providing a technique for the experiment based on the different technique applied then the enhance the accuracy of the method of sentiment analysis. Perform the literature review on the sentiment analysis Amazon product review on the earlier research associate the similar to facilitate and the sentiment analysis provides a technique for customer reviews to achieve the batter result of the model. Because for each of these researches no method or technique however future provides the ideal explanation other than short method planned are for all time improved than the technique planned in conditions of accuracy.

REFERENCES

[1]   S. H. Wu and Y. K. Chen, "Cross-Domain Helpfulness Prediction of Online Consumer Reviews by Deep Learning Model," *Proc. - 2020 IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. IRI 2020*, pp. 412–418, 2020, doi: 10.1109/IRI49571.2020.00069.

[2]   Y. Feng, M. Fan, M. Sun, and P. Li, "A Reinforced Semi-supervised Neural Network for Helpful Review Identification," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 2021–2024, 2020, doi: 10.1145/3340531.3412101.

[3]   A. Alsmadi, S. Alzu'bi, B. Hawashin, M. Al-Ayyoub, and Y. Jararweh, "Employing Deep Learning Methods for Predicting Helpful Reviews," *2020 11th Int. Conf. Inf. Commun. Syst. ICICS 2020*, pp. 7–12, 2020, doi: 10.1109/ICICS49469.2020.239504.

[4]   Z. Bi, S. Dou, Z. Liu, and Y. Li, "A recommendations model with multiaspect awareness and hierarchical user-product attention mechanisms," *Comput. Sci. Inf. Syst.*, vol. 17, no. 3, pp. 849–865, 2020, doi: 10.2298/CSIS190925024B.

[5]   L. Li, T. T. Goh, and D. Jin, "How textual quality of online reviews affect classification performance: a case of deep learning sentiment analysis," *Neural Comput. Appl.*, vol. 32, no. 9, pp. 4387–4415, 2020, doi: 10.1007/s00521-018-3865-7.

[6]   B. Nguy, "Evaluate Helpfulness in Amazon Reviews Using Deep Learning," pp. 1–7.

[7]   A. S. Zharmagambetov and A. A. Pak, "Sentiment analysis of a document using deep learning approach and decision trees," *Proc. 2015 12th Int. Conf. Electron. Comput. Comput. ICECCO 2015*, 2016, doi: 10.1109/ICECCO.2015.7416902.

[8]   P. Liu, S. Joty, and H. Meng, "Fine-grained opinion mining with recurrent neural networks and word embeddings," *Conf. Proc. - EMNLP 2015 Conf. Empir. Methods Nat. Lang. Process.*, no. September, pp. 1433–1443, 2015, doi: 10.18653/v1/d15-1168.

[9]   P. Tanawongsuwan, "Product review sentiment classification using parts of speech," pp. 424–427, 2010, doi: 10.1109/iccsit.2010.5563883.

[10]   N. Nandal, R. Tanwar, and J. Pruthi, "Machine learning based aspect level sentiment analysis for Amazon products," *Spat. Inf. Res.*, vol. 28, no. 5, pp. 601–607, 2020, doi: 10.1007/s41324-020-00320-2.

[11]   M. S. I. Malik and A. Hussain, "Exploring the influential reviewer, review and product determinants for review helpfulness," *Artif. Intell. Rev.*, vol. 53, no. 1, pp. 407–427, 2020, doi: 10.1007/s10462-018-9662-y.

[12]   Z. Zeng, Z. Zhou, and X. Mu, "User review helpfulness assessment based on sentiment analysis," *Electron. Libr.*, vol. 38, no. 2, pp. 337–351, 2020, doi: 10.1108/EL-08-2019-0200.

[13]   U. A. Chauhan, M. T. Afzal, A. Shahid, M. Abdar, M. E. Basiri, and X. Zhou, "A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews," *World Wide Web*, vol. 23, no. 3, pp. 1811–1829, 2020, doi: 10.1007/s11280-020-00785-z.

[14]   S. Lee, K. C. Lee, and J. Y. Choeh, "Using bayesian network to predict online review helpfulness," *Sustain.*, vol. 12, no. 17, 2020, doi: 10.3390/su12176997.

[15]   M. Akbarabadi and M. Hosseini, "Predicting the helpfulness of online customer reviews: The role of title features," *Int. J. Mark. Res.*, vol. 62, no. 3, pp. 272–287, 2020, doi: 10.1177/1470785318819979.

[16]   M. Bilal, M. Marjani, M. I. Lali, N. Malik, A. Gani, and I. A. T. Hashem, "Profiling users' behavior, and identifying important features of review 'helpfulness,'" *IEEE Access*, vol. 8, pp. 77227–77244, 2020, doi: 10.1109/ACCESS.2020.2989463.

[17]   P. Tripathi, S. Singh, P. Chhajer, M. C. Trivedi, and V. K. Singh, "Analysis and prediction of extent of helpfulness of reviews on E-commerce websites," *Mater. Today Proc.*, vol. 33, no. xxxx, pp.

4520–4525, 2019, doi: 10.1016/j.matpr.2020.08.012.

[18]   S. Brownfield and J. Zhou, "Sentiment Analysis of Amazon Product Reviews," *Adv. Intell. Syst. Comput.*, vol. 1295, no. October, pp. 739–750, 2020, doi: 10.1007/978-3-030-63319-6_68.

[19]   T. Mandhula, S. Pabboju, and N. Gugulotu, "Predicting the customer's opinion on amazon products using selective memory architecture-based convolutional neural network," *J. Supercomput.*, vol. 76, no. 8, pp. 5923–5947, 2020, doi: 10.1007/s11227-019-03081-4.

[20]   X. Qu, X. Li, and J. R. Rose, "Review helpfulness assessment based on convolutional neural network," *arXiv*, 2018.

[21]   T. U. Haque, N. N. Saber, and F. M. Shah, "Sentiment analysis on large scale Amazon product reviews," *2018 IEEE Int. Conf. Innov. Res. Dev. ICIRD 2018*, no. June 2019, pp. 1–6, 2018, doi: 10.1109/ICIRD.2018.8376299.

[22]   M. Mohaiminul and N. Sultana, "Comparative Study on Machine Learning Algorithms for Sentiment Classification," *Int. J. Comput. Appl.*, vol. 182, no. 21, pp. 1–7, 2018, doi: 10.5120/ijca2018917961.

[23]   B. Bansal and S. Srivastava, "Sentiment classification of online consumer reviews using word vector representations," *Procedia Comput. Sci.*, vol. 132, pp. 1147–1153, 2018, doi: 10.1016/j.procs.2018.05.029.

[24]   C. Pujari, Aiswarya, and N. P. Shetty, "Comparison of classification techniques for feature oriented sentiment analysis of product review data," *Adv. Intell. Syst. Comput.*, vol. 542, pp. 149–158, 2018, doi: 10.1007/978-981-10-3223-3_14.

[25]   M. Kamal Hassan, S. Prasanth Shakthi, and R. Sasikala, "Sentimental analysis of Amazon reviews using naïve bayes on laptop products with MongoDB and R," *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 263, no. 4, 2017, doi: 10.1088/1757-899X/263/4/042090.

[26]   S.-H. Wu, Y.-H. Hsieh, L.-P. Chen, P.-C. Yang, and L. Fanghuizhu, "Temporal Model of the Online Customer Review Helpfulness Prediction," pp. 737–742, 2017, doi: 10.1145/3110025.3110156.

[27]   T. Pranckevičius and V. Marcinkevičius, "Comparison of Naive Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification," *Balt. J. Mod. Comput.*, vol. 5, no. 2, pp. 221–232, 2017, doi: 10.22364/bjmc.2017.5.2.05.

[28]   T. Pranckevicius and V. Marcinkevicius, "Application of Logistic Regression with part-of-the-speech tagging for multi-class text classification," *2016 IEEE 4th Work. Adv. Information, Electron. Electr. Eng. AIEEE 2016 - Proc.*, pp. 1–5, 2017, doi: 10.1109/AIEEE.2016.7821805.

[29]   X. Chen, J. Sheng, X. Wang, and J. Deng, "Exploring Determinants of Attraction and Helpfulness of Online Product Review: A Consumer Behaviour Perspective," *Discret. Dyn. Nat. Soc.*, vol. 2016, 2016, doi: 10.1155/2016/9354519.

[30]   D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, "What Online Reviewer Behaviors Really Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews," *J. Manag. Inf. Syst.*, vol. 33, no. 2, pp. 456–481, 2016, doi: 10.1080/07421222.2016.1205907.

[31]   B. N. L. Laryea, C. H. Choi, I. S. Jung, K. H. Lee, and W. S. Cho, "Web application for sentiment analysis using supervised machine learning," *Int. J. Softw. Eng. its Appl.*, vol. 9, no. 1, pp. 191–200, 2015, doi: 10.14257/ijseia.2015.9.1.17.

[32]   Y. Kim, D. Y. Kwon, and S. R. Jeong, "Comparing machine learning classifiers for movie WOM opinion mining," *KSII Trans. Internet Inf. Syst.*, vol. 9, no. 8, pp. 3178–3190, 2015, doi: 10.3837/tiis.2015.08.025.

[33]   Y. Heng, Z. Gao, Y. Jiang, and X. Chen, "Exploring hidden factors behind online food shopping from Amazon reviews: A topic mining approach," *J. Retail. Consum. Serv.*, vol. 42, no. October 2017, pp. 161–168, 2018, doi: 10.1016/j.jretconser.2018.02.006.

[34]   K. S. Srujan, S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish, and H. M. Keerthi Kumar, *Classification of amazon book reviews based on sentiment analysis*, vol. 672, no. March. Springer Singapore, 2018.

[35]   Y. Zhang and D. Zhang, "Automatically predicting the helpfulness of online reviews," *Proc. 2014 IEEE 15th Int. Conf. Inf. Reuse Integr. IEEE IRI 2014*, no. 1, pp. 662–668, 2014, doi: 10.1109/IRI.2014.7051953.

[36]   J. Rodak, M. Xiao, and S. Longoria, "Predicting Helpfulness Ratings of Amazon Product Reviews," 2014, [Online]. Available: http://cs229.stanford.edu/proj2014/Jordan Rodak, Minna Xiao, Steven Longoria, Predicting Helpfulness Ratings of Amazon Product Reviews.pdf.

[37]   A. Kamal and M. Abulaish, "Statistical features identification for sentiment analysis using machine learning techniques," *Proc. - 2013 Int. Symp. Comput. Bus. Intell. ISCBI 2013*, pp. 178–181, 2013, doi: 10.1109/ISCBI.2013.43.

[38]   B. Ma, D. Zhang, Z. Yan, and T. Kim, "An LDA and synonym lexicon based approach to product feature extraction from online consumer product reviews," *J. Electron. Commer. Res.*, vol. 14, no. 4, pp. 304–314, 2013.

[39]   B. Agarwal and N. Mittal, "LNCS 7817 - Optimal Feature Selection for Sentiment Analysis," pp. 13–24, 2013.

[40]   W. Chung and T. L. Tseng, "Discovering business intelligence from online product reviews: A rule-induction framework," *Expert Syst. Appl.*, vol. 39, no. 15, pp. 11870–11879, 2012, doi: 10.1016/j.eswa.2012.02.059.

[41]   F. Li, M. Huang, and X. Zhu, "Sentiment analysis with global topics and local dependency," *Proc. Natl. Conf. Artif. Intell.*, vol. 3, pp. 1371–1376, 2010.

[42]   J. Reffin, T. E. Zagibalov, E. O. Belyatskaya, S. House, B. Bn, and U. Kingdom, "Comparable domain dependency in sentiment analysis," *Журнал Сибирского Федерального Университета. Гуманитарные Науки*, vol. 3, no. 5, pp. 764–775, 2010.

[43]   C. Lin, Y. He, and R. Everson, "A comparative study of bayesian models for unsupervised sentiment detection," *CoNLL 2010 - Fourteenth Conf. Comput. Nat. Lang. Learn. Proc. Conf.*, no. July, pp. 144–152, 2010.

[44]   A. Khan, B. Baharudin, and K. Khan, "Sentence based sentiment classification from online customer reviews," *Proc. 8th Int. Conf. Front. Inf. Technol. FIT'10*, pp. 1–6, 2010, doi: 10.1145/1943628.1943653.

[45]   A. Sharma and S. Dey, "A boosted SVM based ensemble classifier for sentiment analysis of online reviews," *ACM SIGAPP Appl. Comput. Rev.*, vol. 13, no. 4, pp. 43–52, 2013, doi: 10.1145/2577554.2577560.

[46]   R. Varghese and M. Jayasree, "Aspect based Sentiment Analysis using support vector machine classifier," *Proc. 2013 Int. Conf. Adv. Comput. Commun. Informatics, ICACCI 2013*, pp. 1581–1586, 2013, doi: 10.1109/ICACCI.2013.6637416.

[47]   S. M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, "Automatically assessing review helpfulness," *COLING/ACL 2006 - EMNLP 2006 2006 Conf. Empir. Methods Nat. Lang. Process. Proc. Conf.*, no. July, pp. 423–430, 2006, doi: 10.3115/1610075.1610135.

[48]   W. Jin, H. H. Ho, and R. K. Srihari, "OpinionMiner," p. 1195, 2009, doi: 10.1145/1557019.1557148.

[49]   M. S. I. Malik, "Predicting users' review helpfulness: the role of significant review and reviewer characteristics," *Soft Comput.*, vol. 24, no. 18, pp. 13913–13928, 2020, doi: 10.1007/s00500-020-04767-1.

[50]   A. Alsmadi, S. AlZu'bi, M. Al-Ayyoub, and Y. Jararweh, "Predicting helpfulness of online reviews," *arXiv*, pp. 1–10, 2020.

[51]   S. Lee and J. Y. Choeh, "The moderating effect of movie production efficiency on the relationship between eWOM and review helpfulness," *Manag. Res. Rev.*, 2020, doi: 10.1108/MRR-01-2020-0057.

[52]   J. Du, L. Zheng, J. He, J. Rong, H. Wang, and Y. Zhang, "An Interactive Network for End-to-End Review Helpfulness Modeling," *Data Sci. Eng.*, vol. 5, no. 3, pp. 261–279, 2020, doi: 10.1007/s41019-020-00133-1.

[53]   M. E. Basiri and S. Habibi, "Review Helpfulness Prediction Using Convolutional Neural Networks and Gated Recurrent Units," *2020 6th Int. Conf. Web Res. ICWR 2020*, pp. 191–196, 2020, doi: 10.1109/ICWR49608.2020.9122297.

[54]   X. Wang, I. Ounis, and C. MacDonald, "Negative Confidence-Aware Weakly Supervised Binary Classification for Effective Review Helpfulness Classification," *Int. Conf. Inf. Knowl. Manag. Proc.*, pp. 1565–1574, 2020, doi: 10.1145/3340531.3411978.

[55]   B. T. Hung, "Integrating Sentiment Analysis in Recommender Systems," *Springer Ser. Reliab. Eng.*, no. March, pp. 127–137, 2020, doi: 10.1007/978-3-030-43412-0_8.

[56]   W. Muhammad, M. Mushtaq, K. N. Junejo, and M. Y. Khan, "Sentiment analysis of product reviews in the absence of labelled data using supervised learning approaches," *Malaysian J. Comput. Sci.*, vol. 33, no. 2, pp. 118–132, 2020, doi: 10.22452/mjcs.vol33no2.3.

[57]   M. Shaheen, "Sentiment Analysis on Mobile Phone Reviews Using Supervised Learning Techniques," *Int. J. Mod. Educ. Comput. Sci.*, vol. 11, no. 7, pp. 32–43, 2019, doi: 10.5815/ijmecs.2019.07.04.

[58]   S. A. Aljuhani and N. S. Alghamdi, "A comparison of sentiment analysis methods on Amazon reviews of Mobile Phones," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 6, pp. 608–617, 2019, doi: 10.14569/ijacsa.2019.0100678.

[59]   N. Shrestha and F. Nasoz, "Deep learning sentiment analysis of Amazon.com reviews and ratings," *arXiv*, vol. 8, no. 1, pp. 1–15, 2019, doi: 10.5121/ijscai.2019.8101.

[60]   J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, and E. Herrera-Viedma, "Sentiment analysis: A review and comparative analysis of web services," *Inf. Sci. (Ny).*, vol. 311, pp. 18–38, 2015, doi: 10.1016/j.ins.2015.03.040.

[61]   A. Jeyapriya and C. S. K. Selvi, "Extracting aspects and mining opinions in product reviews using supervised learning algorithm," *2nd Int. Conf. Electron. Commun. Syst. ICECS 2015*, pp. 548–552, 2015, doi: 10.1109/ECS.2015.7124967.