IEC
Institute of Electronics and Computer

Journal of the Institute of Electronics and Computer, 2021, 3, 24-52
https://iecscience.org/journals/JIEC
ISSN Online: 2643-8240

Online Reviews & Ratings Inter-contradiction based Product's Quality-Prediction through Hybrid Neural Network

Nashit Ali[1], Anum Fatima[1], Hureeza Shahzadi[2], Nasrullah Khan[1, 3], Kemal Polat [4, *]

[1] *Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Vehari 61100, Pakistan*
[2] *Department of Computational Science & Engineering, National University of Sciences and Technology, Islamabad.*
[3] *College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[4] *Department of Electrical and Electronics Engineering, Bolu Abant Izzet Baysal University, Bolu 14280, Turkey*
*Corresponding Author: kpolat@ibu.edu.tr

How to cite this paper: Nashit Ali, Anum Fatima, Hureeza Shahzadi, Nasrullah Khan, Kemal Polat (2021). Online Reviews & Ratings Inter-contradiction based Product's Quality-Prediction through Hybrid Neural Network. Journal of the Institute of Electronics and Computer, 3, 24-52. https://doi.org/10.33969/JIEC.2021.31003.

Received: July 28, 2021
Accepted: August 10, 2021
Published: August 12, 2021

Abstract

With the explosive growth of web usage, people feel comfort to use internet for their personal work such as shopping, sharing information etc. People are more interested in "what other thinks" so the comments and reviews on any online product, movie has huge effect on its earning. Sentiment analysis help people to judge quality by analyzing the reviews. In this era where everyone is so busy in their routine work, it feels like very difficult and time consuming task to check all available reviews on a product. As reviews on a product can be written by anyone so their certainty can be doubtful for example people can give fake reviews just for enjoyment. Another problem with reviews is that contradiction which exists between ratings and reviews of a product for example a person gives 5 star to a product but write a lengthy list of problem exists in that product. To deals with this problem, we are introducing a methodology that finds out contradiction of reviews and ratings of a product then finds out the actual score/quality of product which can save people's time to read all the reviews or clear their confusion if they stuck between whether to trust on ratings or reviews. In this research we used CNN Hybrid model which gives 97.5% accuracy which is better than previous models. Dataset is collected from amzon.com. We have also applied CNN Hybrid model on different training and testing dataset ratio in both cases i.e. bigger and smaller dataset. We evaluate CNN hybrid model on both smaller and bigger dataset and experimental results conclude that increasing dataset will increase accuracy of CNN Hybrid Model.

Keywords

CNN Hybrid Model, Contradiction, Online Ratings, Online Reviews, Product Quality, Sentiment Analysis.

1. Introduction

With the explosive growth of Web usage, People feel comfort to use internet for even their personal work such as shopping, sharing information etc. people are more interested in “what other thinks” so the comments and reviews on any online product, movie has huge effect on its earning. Sentiment analysis help people to judge quality by analyzing the reviews. We discussed concept of sentiment analysis online reviews and what benefits it gives to users. We have also explained sentiment analysis, its types and techniques. This research is important because it helps users to make a decision for a product or online deals. Many researcher works on the online reviews of the shopping websites by judging rating or by analyzing overall comments and calculating contradiction between comments and ratings then calculate the actual rating of product for commercial improvement purpose or entrepreneur. In this research our focus is to find contradiction between user comments and ratings and calculating the overall actual rating of product which helps in judgement of the product quality.

Sentiment Analysis is all about people’s views, behavior and reactions toward a thing. Sentiment analysis is the area of NLP that decide user opinion and attitude towards specific product on the base of user written text. Sentiment describes attitude or opinion of any individual person, this opinion could be towards an attitude of a person, regarding any specific belief, regarding any market product or behavior of any person. [1] Every person has different emotions and they react different in different context. Different factor influences the emotional state rather than textual data like body language. Possibly there may be more than one type of emotion involved in any text. Sentiment analysis can be implemented on the binary class of emotions like positive, negative and natural [1]. People start using internet for online shopping and many other purpose. Online shopping gives customer a lot of easiness but also deliver damaged stationary. So people need prove to trust online shopping websites. People will satisfy by knowing what others think so before attempting any purchase, people goes to comment section to know quality of the product. Reading the comments of customer becomes tiring when volume of comments is large. Sentiment analysis gives comfort to users by showing overall polarity of comments. Polarity is degree of sentiments i.e. positive and negative.

From last few years, many researchers are working on sentiment analysis. Sentiment analysis can lead towards different tasks. The dissimilarity between those researchers is the parameters that they are seeing in order to analyze text. SA comprised 3 classification level i.e. sentence, aspect and document-level SA. We used Document_level SA used when we have to classify the whole document as +ve or -ve sentiment. In this, to check the complete document’s polarity, we give document containing text as input and Sentiment analysis techniques are used. Sentence_level SA is used to classify a sentence as +ve or -ve sentiment. For this, first we check if the given sentence is subjective in nature or objective in nature. As we know document consists of both subjective and objective sentences but most of research only focus on subjective sentences. Aspect-level SA is used for sentiment classification w.r.t particular characteristics of entity. In this, Single sentence/phrase is classified on the base of set features. This can be word level and feature based. In word_level, we usually focus the objective part of sentence. Because verbs, nouns and other parts of sentence are there to add some sense and meaning in sentence. To understand concept of this, we will take an example of online product's reviews where negative review mean dislikeness and positive review means likeliness. [2] [3]

The Process of sentiment extraction consists of three parts which is as follows:

• Subjectivity Classification: Subjective sentences are often used to express individual's sentiment or opinion where objective sentences are precise in context and document is collection of both sentences. Subjectivity classification is classification of subjective sentences into two parts: one is sentence that represent opinion and other that does not represent any opinion [3].

• Sentiment Classification: After we get output of subjectivity classification, we need to find polarity of sentences that contain opinion. Sentiment classification classify the sentence in to two parts, +ve polarity and –ve polarity. In some cases, a 5-scale polarity is also used for sentiment classification. This scale contains extreme –ve, -ve, neutral, +ve, extreme +ve. [3]

• Complimentary Tasks: There are two tasks which are considered as complimentary task foe SA. First one is to extract the information of that person whose sentiment is in use for analysis. Different application need different scenario for sentiment analysis, in some cases it is important to identify person behind an opinion. A document may more than one opinion so in this case it is important to identify entity behind each opinion [3]. The other task is to extract feature from text. It is identification of that thing for which peoples are giving opinions. [3]

A sentiment lexicon is known as list of +ve & -ve sentiment words in English language but it does not mean that lexicon can't be generated in other language [4]. Feldman stated that the lexicon is considered as most significant resource for most decisive analysis of algorithms [5]. Weichselbraun, S. Gindl, and A. Scharl elaborated that context is very important in production of lexicon. They also claim that the automated systems have restricted capability to solve ambiguity and process context information. [6] Because System does not have same understanding level as human has so human easily understand opinion but system can't interpret opinion like human. According to B. Liu [7] opinion comprises of the following things: entity along with its features known as Opinion targets, +ve or-ve sentiment, Person who has given opinion, Time of given opinion. [7] In order to do sentiment analysis, the pre-processing is done on input text then the sentiment analysis techniques are used to check polarity of text. R. Dale [8] proposed following steps involved in Pre-processing sentiment analysis of natural language. Tokenization is process of spitting sentence into token where token can be words, symbols or items having some meaning. Lexical analysis is process of creating lexicon and to do parts of speech tagging to tokens. Syntactic Analysis gives information about structure of text or sentences in text. Parsing is also included in Syntactic analysis. Semantic Analysis is used to check the meaning of text where Pragmatic Analysis is used to find meaning of text w.r.t context. [8]

After pre-processing, the methods of sentiment analysis applied to the text. Sentiment classification consists of two types, one is supervised and other is unsupervised classification method. In Supervised method, dataset collected from user text is used for both training and testing purpose which results in classifying text in four classes i.e. constructive, +ve, neutral and _ve. For classification, existing algorithm like SVM or NB can be used. In Supervised learning algorithm, input is labeled and we know the correct answer, machine learn from dataset and then give required output. In Unsupervised machine learning algorithm, input is not labeled and algorithm predicts the next word on basis of input given. k-means Clustering is an example of unsupervised machine learning. [9]

Deep learning algorithms are used to simplify text. In these algorithm, data has been processed through several layers as each layer simplify data and pass it to next layer. Deep learning algorithms are also used for analysis of images. They use edges of images for processing. The benefit of Deep earning is that it can process large no. of feature in unstructured data. Deep learning is used for making artificial applications. These algorithms are used to train machine to act like human. In this research CNN hybrid model (CNN LSTM Layer) will be used. Previous neural network does not cover the sequence data so it does not give good result with previous neural network. This problem has been removed in LSTM. RNN has problem of vanishing gradient error in which machine cannot be learned properly. LSTM is solution of problem faced in RNN. It also gives more accuracy than LSTM. LSTM consists of 3 gates & one cell. The gates are named as input, output and forget gate. These gates are used to store information about dependencies of words on each other in sentence. Input gate is used to control flow of value, forget gate control the time for which value can remain in cell, output gate is used to control the limit at which a value can used for computation of output. Cell is used to store information just like memory. This network consists of connection which go in and out from each gate and the weight of these gates tells about type of operation of gate. Supervised learning is used for training of LSTM.

In word embedding, vectors or real number are used to represents words. As compare to Vocabulary size, these vectors have low dimension. Word embedding has also known as vector space model/ semantic space model in which each word is mapped in to vector that collectively make a vector space. Words which have similar context or semantics are positioned nearby in vector space. In NLP, word embedding is used in PoS tagging, Syntactic analysis and Parsing. Word embedding can have used in two ways, one is for counting words and other one is for prediction of next word on basis of context. Different techniques of word embedding i.e. Word2Vec or Glove were used to weighting the words.

This research aims for identification of the contradiction between user's review and ratings on basis of these polarities and measure the actual rating of product. The significance of this study to judge the product quality on basis of finding contradiction between review's and rating. The main contribution of this study is helping users by detecting the product which are not quality product but have good star rating. In article [10] work has been done on detecting unfair reviews on amazon products which helps other buyers in future. In another article [11] a classifier has been design which distribute into two groups, one of comments with star rating and other of comments without star rating but this paper only works on comments without star rating so we plan to research on comments with star ratings. Then the article [12] works on detection of fraud apps in play store using sentiment analysis. With reference to these article, the proposed topic aims to include sentiment analysis of user's review on products and judging the actual rating of product which tells about quality of product which result in detection of fraud or non-quality products.

The Research Objectives is as follows:

- To do sentiment analysis of the user's reviews
- To identify contradiction between user review and ratings
- To find out actual quality of product which will be quality of product.

In this research we considered following research Questions:

Q1: Does previous research worked on sentiment analysis of online reviews of product on basis of reviews and ratings?

Q2: Does previous research find contradiction of reviews and ratings?

Q3: Does previous research predict quality of product after finding contradiction of reviews and ratings?

Q4: How does it better to use neural networks and sentence level analysis for predicting product quality.

Our proposed research is limited to online reviews having star ratings. This research does not cover reviews without rating or completely incorrect reviews. Peoples usually check others comments before buying some new products. This will be a tiring process in case of a huge volume of reviews. This research will help them to judge quality by doing automated analysis of these reviews. People like to write freely so they write positive and negative phrase in same sentence which can change the whole semantics of this text. Our research will work on these types of sentences and show the actual rating of the Product.

2. Literature Review

Nowadays, a large number of user reviews has been available on almost everything that is present on the websites of the e-commerce environment. Reviews may contain user reviews on products that help other in their buying decision making. Huge numbers of reviews exist, which makes it difficult for a buyer to go through all comments and make a choice. Also, if the buyer reads some of the product reviews, it is hard for them to differentiate between fair and unfair product reviews. Similarly, user comments are an important source of information for consumers. However, depending on their authority, they can improve the reputation of products or websites. Elshrif Ibrahim Elmurngi & Abdelouahed Gherbi presented comparison of four supervised machine learning algorithms: NB, DT (J48), LR and SVM for sentiment classification on amazon reviews [10]. Their results indicate that LR performs better than others in term of accuracy [10].

Sasikala P and L.Mary Immaculate Sheela [11] have designed a classifier model which accepts all the online reviews and group them into two categories as reviews with ratings & reviews with blank or empty ratings. They predict the opinion from the reviews without ratings. They use Naïve Bayes and logistic regression model for this prediction. [11] They conclude that in few cases logistic regression performance is better than the Naïve Bayes. [11]

Laila Abd-Elhamid and other co-authors proposed a feature_based SA method for mining user generated reviews in Arabic [13]. With help of Semantic rules and PoS tagging. Features were extracted spontaneously from dataset containing online reviews. To represent the relationship b/w aspects on which review has been given along with their features, features collected from dataset are organized into decision tree. To fulfill this purpose, they have applied 5 rules. At Last, Evaluation of every rule has been done by using lexicon based classification. Their final result represent that the methodology discussed in this research is capable of automatic polarity labeling for a bigger no. of extracting feature and results in increased accuracy. [13]

As online reviews are considered as help to find out if the product is valuable or not so trend of research seems to trending in this era. This article [14] finds out the recommendation out of the reviews to save people's time of reading comments, this result may find if reviews on product suggest to use product or does not suggest the product to buy. They have done sentiment analysis using Binary LR and CHAID DT for classification of reviews and to see if the positive sentiment has better impact on suggestion or negative sentiments have more. [14]

This article [15] has aim to do sentiment analysis on amazon dataset with ensemble method. This method is combination of 5 approaches i.e. "NB, Random Forest, SVM, Boosting and Bagging, testation of proposed model was implied on 6 different situations i.e. stopwords removal with or without using unigram, bigram and trigram and compared all techniques with ensemble method. In case of stopwords removal using unigram, random forest outperforms other approaches with 87% accuracy but in other situations, ensemble method gives the best result. [15]

This article [16] works on real-time SA of user's review on a product. They used amazon dataset and SVM for designing methodology for experiment. The main contribution of this study is development of application which can do real time sentiment analysis and predict percentage of positive comments and negative comments by finding the polarity of user given review. This experiment has been done for both sentence and review level. [16]

This article [17] has done research on implicit sentiment. They have evaluated different deep learning techniques on Chinese dataset to find out their performances. They have used "LSTM and its version, DNN & CNN" for this purpose. This study has concluded that CNN and LSTM gives better performance because of their unique feature extraction deign. DNN shows poor performance w.r.t others. [17]

Reshmi Gopalakrishna Pillai, Mike Thelwall and Constantin Orasan [18] enhanced the existing TensiStrength method to identify stress/relaxation strengths by using a pre-processing phase to remove ambiguous affect uncertainty from content tweets. It uses WSD to increase the accuracy of Tensi-strength. The final conclusion of this study is that including WSD significantly increases output accuracy of the traditional Tensi-strength. [18]

To improve classification accuracy on basis of semantic positioning & ML methods, Ahmed Al-Saffar and other co-authors [19] has offered method named "Malay SA classification model". In first step, total dataset containing 2,478 Malay sentiment_lexicon is assigned score known as polarity. Then a hybrid approach containing combination of supervised ML methods and lexicon_knowledge techniques are used for sentiment classification. Then three individual classifiers (Naïve Bayes, SVM, Deep-belief network) and a combined classifier (Voting-classifier combination) are designed for evaluating the accuracy of classification. They conducted experiments over dataset containing MRC. It concludes that the proposed research increases the accuracy of Malay SA on the basis of combination of classification techniques. [19]

Puspita Kencana and co-authors [20] measured quality of Electronic-Commerce facility on basis of online reviews of users with SA. They used “Tokopedia” as data set. In First step, they collected different reviews set from “Tokopedia” and NB classification method was used in this article. They conclude overall service of Tokopedia as positive. [20]

Online customer reviews have very essential role in electronic-commerce So fake or raw data written in reviews may result in difficulty to judge quality for those customer which make choice on basis of reviews on product. Yoon-Joo Park [21] has done this study to do analysis of the features implanted in online reviews on product of 5 different categories (beauty, mobiles, clothing, grocery, and video) and concludes effects of these features to see at which extent these features of review are helpful to user [21]. 4 data mining approaches (LR, SVR, M5P, and random forest) were compared in this to see the best one. Their result conclude that SVR method outperforms other methods by predicting most accurate helpfulness among the four methods for all five datasets. [21]

In order to increase accuracy, Shabeeba T and Aswathi T [22] suggested sentiment-based technique for prediction of rating. First they evaluate sentiment by sentiment algorithm. Then use prediction algorithm for rating prediction. The research concludes that Sentiment algorithm performs better on positive reviews than negative reviews. [22]

Xiaoyi Zhao and Yukio Ohsawa [23] Proposed a new sentiment analysis model of online-shopping reviews based on “hidden Markov model”. The amazon reviews were used as dataset. For training of model, they have used supervised training method then this model is enhanced by using genetic algorithm’s variation. They evaluate performance of this algorithm by doing sentiment classification of Amazon reviews comparing to other methods such as SVM, Logistic Regression, etc. The result show that the adapted hidden Markov model outperforms other. [23]

Swati Redhu, and other co-authors [24] presented a summary of techniques used in SA and text mining explaining on every sub-tasks for example extracting relation, entity-recognition and extracting information etc. [24] They have also worked on tweeter content in Spanish, Arabic and many more languages to find out sentiments. SA consists of 4 important stages i.e. feature-extraction, data-conversion, data-acquirement, feature-representation and other ML algorithms. [24]

For evaluation of Quality and user’s satisfaction for product, Mahboob K and Ali [25] has done sentiment analysis on a dataset containing online reviews of medicines. In this approach, each sentence is converted into lexicon and then polarity has been given to those lexicons. The approach is known as lexicon-based SA. [25]

Xiangfeng Dai, Irena Spasić and Frédéric Andrès [26] has included topic modeling in their research. They presented methodology to find out dominant feature from user's review and then rate these feature on a five-star scale accordingly. This framework consists of 5 segments i.e. topic modelling, Preprocessing, Sentiment analysis, classification of text then star rating. A document consists of large number of sentences. To extract the main feature or concept of that document, topic modeling is being used which results in all possible topics found in document then sentences has been classified according to these topics. [26]. To calculate sentiment score of each sentence, word embedding technique has been used in this research. each sentence consists of two parts, its topic and sentiment. Then the total sentiment score of topic is estimated on to scale having five-star rating. They used Airbnb online reviews as dataset [26].

Ha-Na Kang and his fellows [27] presented sentiment analysis of online reviews on online games. They applied ML methods such as ANN. Then in this research, CART were tested on dataset. Dataset is collected from STEAM games. This research has been done to do analyze those factors which may affect the usefulness of game's review. After the analysis has done, CART was concluded better than ANN [27]

Mohammad Suleiman with co-authors [28] Presented overview of multimodal SA and slightly discussed latest developments made in multimodal SA in different fields such as images, spoken reviews, human_machine, video blogs and human_human interaction [28]. This survey concludes that multi-model SA performs better in most of cases as compared to the unimodel SA. [28]

Dr.B. Radha and V. Meera [29] Presented a detailed study on SA & Opinion mining and their techniques. This study concludes that neither classification model consistently outperforms the other. [29]

Ms. Sneha A. Sahare [30] also presented a study on opinion mining or SA. In this survey, author slightly discussed idea of OM and SA, techniques used in OM, Areas in which OM & SA can be used, challenges faced in way of research of OM & SA and research possibility in OM & SA. [30]

P Deepa Shenoy & other co-authors [31] proposed "Gini Index based feature selection method with "SVM classifier" to do sentiment classification on dataset having enormous number of movie's reviews. Experimental result illustrate that this method gives better outcome with reducing error rate and increasing accuracy. [31]

Elshrif Elmurngi and Abdelouahed Gherbi [32] presented sentiment analysis on movie reviews to detect fake reviews. They have used ML algorithms for classification of movie reviews. Reviews were divided into sets, one with positive polarity and other with negative polarity [32]. For reviews sentiment classification

They have selected 5 supervised ML for comparison i.e. NB, SVM, KNN, K* and DT(J48). 2 different datasets i.e. “movie review dataset V2.0 and movie reviews dataset V1.0” were used in this research and result conclude that SVM perform better than other algorithms. [32]

Dr Rajdev Tiwari and other authors [33] presented survey on OM w.r.t its architecture, different level, techniques applied, tools that can be used for opinion mining, comparison of its methods and challenges.

Gaurav Dubey and other authors [34] provides SA of dataset containing reviews of smart phone. These reviews have been classified as +ve, neutral and –ve behavior. NB and SVM were used as Classifier for analysis of dataset. The result show that SVM performs better than NB as it gives 90% accuracy while NB gives 40% accurate result [34].

V. Varshitha and co-authors [35] discussed about SA in which dataset contains online hotel booking sites, they have proposed calculation of star rating, opinion features, stopwords removal, multi-dimensional trust evaluation and multi-criteria recommender. This recommender consists of internal and external opinion, where internal opinion is customer’s opinion and external opinion is obtained from social media [35]. They proposed a multidimensional system which have internal, external reviews then removed stopwords. Then they classify positive and negative reviews and calculate score based on that gives star rating by which user feel more comfortable in choosing hotel. [35]

Dr. U Ravi Babu [36] compare the services of different E-shopping websites and analyzing which one is the best. They use five large datasets of five different E-shopping website which contains reviews related to the services. Score of words is calculating by using Sentiwordnet. Then classification of sentiment has been done in three categories as +ve, -ve and neutral. They Concluded their findings as preprocessing of data has great effect on detected sentiment’s quality. [36]

K.R. Sharmila and other authors [37] Presented sentiment analysis on online product reviews. Online product reviews from website are selected as dataset. The PoS tagging is used to extract the features to get better results in classifying the sentence as positive or negative. Then separation of positive and negative comments is used to analyze the quality of the online products. [37]

Vidushi and Gurjot Singh Sodhi [38] proposed a novel strategy through which Sentiment analysis of online reviews has been done. The grammatical mistakes are also considered for pre-processing and in this step, reviews were removed from special characters, stop word, tokenized, etc. TF-IDF score based approach was used to calculate score for each review [38]. Then Chi Square Feature Selection method was applied on it. The extracted feature forms a term document matrix which is used in the classification algorithm [38]. Two classification algorithms i.e. Naïve Bayes and KNN are compared and concluded that Naïve Bayes outperforms K-NN. [38]

Sana Prasanth Shakthi and co-authors [39] use NB algorithm and DT for classification where comments of e-commerce websites are taken as dataset and sentiment analysis has been done on dataset. To get the comments from any website required as dataset, web crawler has been used. To extract the exact meaning of words, spelling should be correct so to do the spelling correction, WordNet dictionary has been used. Then stopwords has been removed by stemming. NB algorithm has been used for classification then DT is used to find out final polarity. [39]

S. Muthukumaran and Dr.P. Suresh [40] explains different methods for sentiment analysis. Dataset for this research is collected [40]. The proposed study show that the statistical methods are often combined with traditional linguistic rules and representations. This study also justified that NB and sentiment polarity on online product reviews is analyzed by using HMM classifier w.r.t their computational simplicity. [40]

Norwati Mustapha and fellow authors [41] explain the extent at which preprocessing stages effect the process of SA. Movie's reviews are used as dataset. supervised ML was used for classification of the reviews. "SVM with linear & non_linear kernel" was used for reviews classification [41]. Classifier performance has been evaluated based on factors "recall, f-measure, precision & accuracy" and TF and TF-IDF for feature representation were used. [41] Experimental Results show that the SVM with non-linear kernel provides better results. [41]

El-Din and co-authors [42] conducted a study about challenges faced in SA. This research has study related to SA approaches and methods. this study has been conducted on 2 comparisons which were selected from 47 papers. The first one is made on basis of relation b/w sentiment structure of review & challenges of SA and this comparison results in another challenge named domain dependency [42]. Then 2nd comparison depend upon challenges related to accuracy rate of SA and This results in importance of sentiment challenges for sentiment evaluation and increasing accuracy. [42]

J. Drew Procaccino and his fellow [43] discovers benefits of analysis of those reviews which only contain text by using "text mining and visual analysis for SWOT Analysis". Text mining is used to find hidden information in text and SWOT Analysis was applied on this information of 3 hotels [43]. For dataset used in this research, the analytical power has been doubled when text mining is applied on combination of rating and reviews. On the other main characteristics of hotel and its position in competition is illustrated by visual analysis. [43]

Maleerat Sodanil [44] Presented SA in multi-language for Reviews which they get from 2 travel portal (Thai, English). They developed a class of features then used 3 classification method (SVM, DT, NB) to find accuracy of predicting words [44]. This study conclude that SVM perform better in this study. [44]

Harnani Mat Zin and co-authors [45] discussed necessary background and future directions of SA methods used for big data. They discuss Sentiment analysis approaches such as sentiment polarity detection, and to which extent it is suitable for the big data [45]. They also discuss Sentiment features such as explicit and implicit features, sentiment classification practices such as NB, SVM etc. and applications of Sentiment analysis in Big data. [45]

Amal Ganesh and co-authors [46] presented a study having comparison of different methods of SA. They compared "machine learning algorithm (SVM, N-gram SA, NB Method, Maximum Entropy Classifier, K-NN & Weighted K-NN, Multilingual SA, Feature- Driven SA), Rule-Based Approach and Lexical-Based Approach". [46]

Abhinash Singla and his fellow author [47] use NB algorithm and DT for classification where comments of Flipkart websites are taken as dataset and sentiment analysis has been done on dataset. To get the comments from any website required as dataset, web crawler has been used. To extract the exact meaning of words, spelling should be correct so to do the spelling correction, WordNet dictionary has been used. Then stopwords has been removed by stemming. NB algorithm has been used for classification then DT is used to find out final polarity. [47]

P. Ajitha and his fellow author [48] Presented study of SA on dataset containing product reviews. They proposed a technique which automatically find the most frequent word used for any aspect extracted from online review named as "semantic orientation" [48]. Sentiment orientation algorithm includes 2 main approaches i.e. "Dictionary-based approach & Corpus-based approach". This model proposed in this study provides complete result for product review. [48]

Kiran Gawande and co-authors [49] presented a system that performs SA of online reviews to find out sentiment of review then classify these reviews. Combination of SA with Reviews Classification results in increased system's accuracy which means it gives precise reviews to customer [49]

Abinash Tripathya, Ankit Agrawal and Santanu Kumar Rath [50] compare results obtained from classification algorithm of NB & SVM [50]. Both algorithms classify reviews into two classes, one is +ve review or other is -ve review. This study presented SVM algorithm as better classifier which gives better accuracy in case of sentiment prediction [50]

For researcher, it is difficult to choose appropriate articles for literature used in research. Other researcher's reviews on paper are the vital source to get help from these papers for finding the best paper amongst all of them. These reviews consume less time & cost. Osama Ismael and co-authors [51] proposed a technique named SAOOP (SA of online papers). This technique is used to improve performance & accuracy of BoWs model by making enhancement in it. The approach discussed in this article covers two issue i.e. reviews SA and solution of SA challenges [51]. it calculates scientific paper's parameters used in topic domain such as date at which paper is published, place of publication and evaluation of paper score on basis of no. of citation and the result of this research show that Proposed technique has increase accuracy [51]

Dr. S. Koteeswaran and his fellow [52] presented a study on art-of-state of SA. The main idea of opinion used in SA and tasks involved in opinion mining is discussed in this study. SA is used to calculate opinion from text and to assign polarity to it [52]. This research concludes that combination of different algorithms and features result in progressive performance because one method's drawback will be covered by other method's advantage. [52]

Shailesh Kumar Yadav [53] conducted a study on SA & OM. Concepts used in sentiment classification, classification methods i.e. NB, ME, SVM method etc. and tools used for SA i.e. NTLK, GATE etc. and the challenges still faced in Sa such as SA of complex text, identification of implicit feature, features extraction from different datasets, extracting more than one opinion from one document etc. [53]

Xing Fang and Justin Zhan [54] discussed the process of categorization on online reviews on basis of sentiment polarity. online reviews on product on Amazon.com was collected as dataset and review_level & sentence_level categorization were applied on it [54].

Baizhang Ma and co-authors [55] presented new technique for SA which is based on features of product and their dependence. A set including different sentiment lexicon were included in this method [55]. For feature extraction of a product, "PageRank" algorithm was used. Then the polarity was calculated for each review [55]. Model presented in this research gives better output than NB & Opinion-Observer method. [55]

Lu Jingli and other authors [56] presented a "feature-based vector model" [56]. It also proposed "novel weighting algorithm" for SA. Reviews on Chinese product are taken as dataset. Strength of polarity was expressed in terms of degree of punctuation and adverb. and Evaluation of this method has been done on 3 different datasets, which results in improvement of sentiment classification performance [56].

Neha Nehra [57] delivers a complete study on SA or OM. Movie reviews are taken as dataset. Author discussed concept of sentiment analysis along with its types and techniques, also presented detailed available literature on sentiment analysis related to movie review. [57]

Jyotika Yadav [58] also conducted literature survey on sentiment classification of movie reviews. According to their survey, there exists many solutions for sentiment classification of reviews but natural language consists of unstructured data so researcher remain unsuccessful to introduce a a fully automatic system [58]. This research used hidden markov model for automatic classification of reviews by giving sentiment polarity to these reviews. [58]

Qingxi Peng and Ming Zhong [59] use the combination of SA methods with spam detection of reviews. Shallow dependency parser was used for computation of sentiment score from text written in natural language then they established series of rules through observation and then combination of rules with time series used to effective detection of spam review [59]. It was found out by results that model used in this study has better performance as compared to previous methods. [59]

In this article [60], a system is proposed to classify reviews on basis of review's sentiments on 1 to 5 rating scale. This study combined the existing text analyzing packages and used this combination for reviews mining on the basis of sentiment score. The system proposed in this article used score rating and gives good result. The model proposed in this article deals with sentiment polarity classification that is considered as essential problem of sentiment classification. Data that is available online has two faults. To handles these fault this methodology has been proposed. [60]

User can post anything so the quality of their content is not certain. Many fake users can write fake reviews.

Review's polarity cannot be learnt or inaccessible. [60]

The system proposed in this article used score rating and gives good result. This study does not work for sentences having hidden sentiments. It only works for sentence having clear sentiments. This study aims to combine prediction based techniques with existing methods to extract more feature so that it deals with hidden SA in future. [60]

In article [61], This study deals with peer reviews domain of scholarly papers as it is challenging task of SA. This study proposed an abstract_based neural network model. This study helps in predicting decision or recommendation automatically for a paper. This study further tells sentiment (+ve or –ve polarity) in text of a submission paper written by Reviewer. [61]

SA usually covers all areas such as reviews, tweets etc. but the domain discussed in this study is still needs to be covered. This domain includes scholar's paper which can have long text so the text can be mixture of multiple types of sentences that can be sentence having open sentiment or sentence having hidden sentiment or a sentence having no sentiment at all. This study works on this domain and also tells advantages and disadvantages of paper submission. [61]

This article concludes that this study gives stable results as it gives consistent results in case of recommendation decisions and peer reviews text. Because it takes long time of author to do deep study of reply given by reviewers for improvement of their paper so this study helps author by extracting sentiments from reviewer's reply on paper. [61]. This study has plan to collect bigger dataset in this domain for machine training and testing, also plan to experiment different machine learning methods in future. [61].

In Article [62], this study aims to gives a sorted list of recipes on basis of their main ingredient. As there exists thousands of recipes for one item available on internet, peoples got lost while finding best recipe of that item. So this study seems to solve their problem by ranking all recipes on basis of their reviews [62]. Tis application takes main ingredient of recipe from user as input then sort out all the recipes having that ingredient. Further it finds out sentiment score of the reviews on each recipes and then sort the recipes on basis of score [62]. This study has plan to do comparison of its ranking method with other techniques available for this purpose in future. [62]

This article [63] proposed methodology Hybrid CNN-LSTM and then compare it with traditional CNN and LSTM. They have done experiment on two different dataset of movie reviews and conclude that Hybrid CNN LSTM outperforms CNN and LSTM by resulting in 91% accuracy. [63]

3. Proposed Methodology

Previous Section contains necessary information and background required for clear understanding of sentiment analysis and techniques used in SA or deep learning. I have also described the work done so far related to sentiment analysis of online reviews, sentiment classification, opinion mining and related aspects of this area.

This research is purposed to identify the contradiction between rating and reviews on a product then identify the actual rating on this product. This research focuses on finding the contradiction between user given reviews and rating then finding actual rating of product using LSTM. Nltk tool is used for implementation.

The proposed methodology consists of following steps:

- Dataset Collection
- Preprocessing
- Word Embedding
- Implementation and results

The proposed methodology is shown in Figure 1 and consist of steps as follows:

- **Dataset Collection**: For this research, dataset is collected from amazon.aws.com. Dataset is collection of label, rating and reviews given by user on a product. This dataset contains 2000 numbers of reviews and ratings. This dataset is divided into three categories i.e. label, reviews and rating. After preprocessing reviews, dataset is now divided into parts i.e. reviews with labels and ratings. [64]

• **Preprocessing**: Preprocessing is used to clean data from those words which does not have any meaningful information. Our research focused on reviews which is user—generated text in which text is written in informal style so there exists great chance to existence of ambiguity in text. Preprocessing is use for removal of ambiguity and to extract meaningful data.

Preprocessing is divided into following four tasks:

Stopwords Removal: Stopwords are those words which does not play any role in calculation of sentiment score of any text. These words were removed from text so that we can save our time and avoid extra calculation.

Lemmatization: Lemmatization is used to normalize text. It is used to get the actual root word by removing suffixes or prefixes. For example, after lemmatization, “goes” turns in to “go”.

PoS Tagging: PoS tagging refers to Parts of Speech Tagging. In PoS tagging, each word is tagged with parts of speech. The tags are used with each word so we can identify type of each word. “/” is used to distribute PoS tag and Word.

Tokenization: Division of each sentence into words is known as tokenization. For calculation of sentiment score, it is important to convert sentence into tokens i.e. words.

• **Word Embedding**: In word embedding, vectors or real number are used to represents words. As compare to Vocabulary size, these vectors have low dimension. Word embedding has also known as vector space model/ semantic space model in which each word is mapped in to vector that collectively make a vector space. Words which have similar context or semantics are located near to each other in vector space. In NLP, word embedding is used in PoS tagging, Syntactic analysis and Parsing. Word embedding can have used in two ways, one is for counting words and other one is for prediction of next word on basis of context. Different techniques of word embedding i.e. Word2Vec or Glove can be used to weighting the words but we have used Word2Vec technique in this research which results in weighted words that were used in sentence creation.

• **Implementation and Results**: We have used CNN LSTM layer for calculation of score of a product on basis of reviews. We have implemented this in nltk and used CNN LSTM Layer for machine training and testing in which 80% of dataset is used for training and 20% is for testing. Then the score of product on basis of ratings has calculated. We have compared both results of ratings and reviews of products to show the contradiction between rating and reviews. Then the actual score on basis of both reviews and ratings has been calculated to show the actual quality of product. In this phase, we have further evaluate performance of CNN LSTM layer on basis of accuracy, recall, f-measure and precision. The results will be discussed in next Section.

A. Algorithm

1- Input
  - Read Text File

2- Split the data in following:
  a) Reviews
  b) Ratings
  c) Labels

3- Apply Text Preprocessing on Reviews
  i. StopWords Removal
  ii. Lemmatization
  iii. Punctuation
  iv. Tokenization

4- Return Tokens

5- Import Vector Files of meaningful Words

6- Converting Words into Vectors
  For w in tokens
    Assign index to each word

7- Embedded Words
  For i in token
    Compare vector with each word
    Assign weight

8- Creation of Sentences which was Split in Step 2
  For j in tokens
    s=s + token

9- Attach corresponding label with each sentence

10- Return Sentences

11- Apply CNN HYBRID on Sentences

12- Return result of Reviews

13- Input ratings

14- Extract the rating score by applying mathematical formulas

15- Return result of rating

16- Apply contradiction formula on reviews result score and rating score

17- Return actual result/score which is product quality.

B. Proposed Model

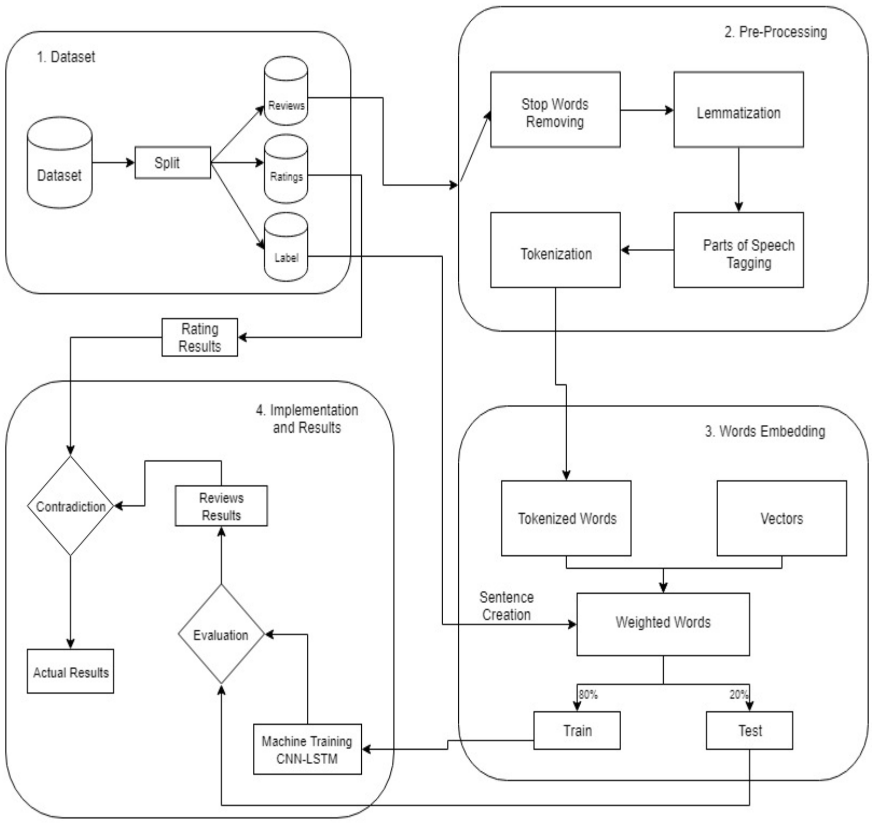

Figure 1. Proposed Methodology

4. Results

In previous Section, we have discussed methodology to judge the actual quality of product on basis of online reviews and ratings of product. We have used CNN LSTM Layer algorithm for machine training and testing. The results of our implementation will be discussed in this Section. The results of sentiment score of product on basis of reviews and ratings is as shown in figure 2 which is actually showing the contradiction between score of reviews and ratings.

Figure 3 shows following results

- Score on basis of Product reviews
- Score on basis of Product ratings
- Actual Score on basis of both reviews and rating after finding contradiction b/w both

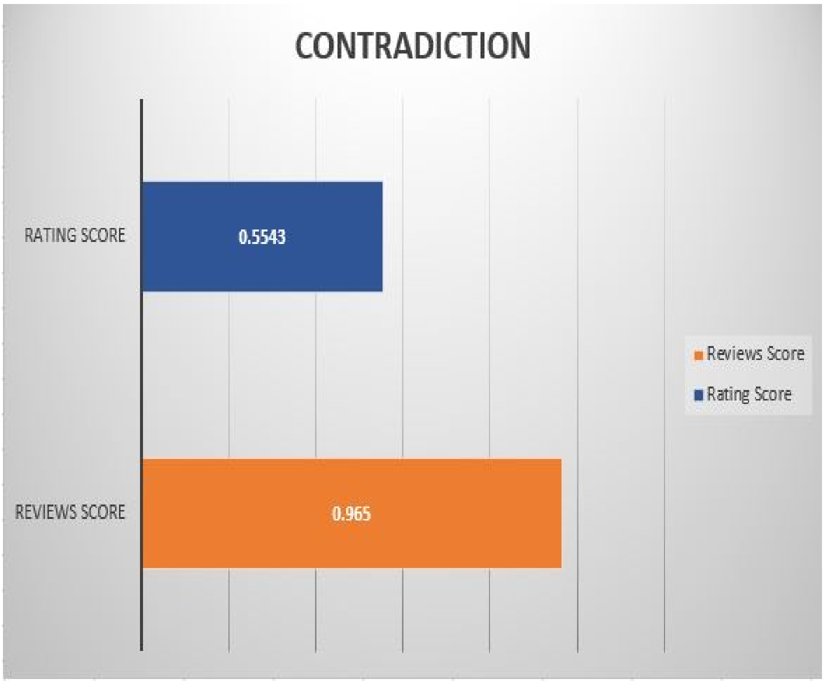

Figure 2. Contradiction between Ratings and Reviews

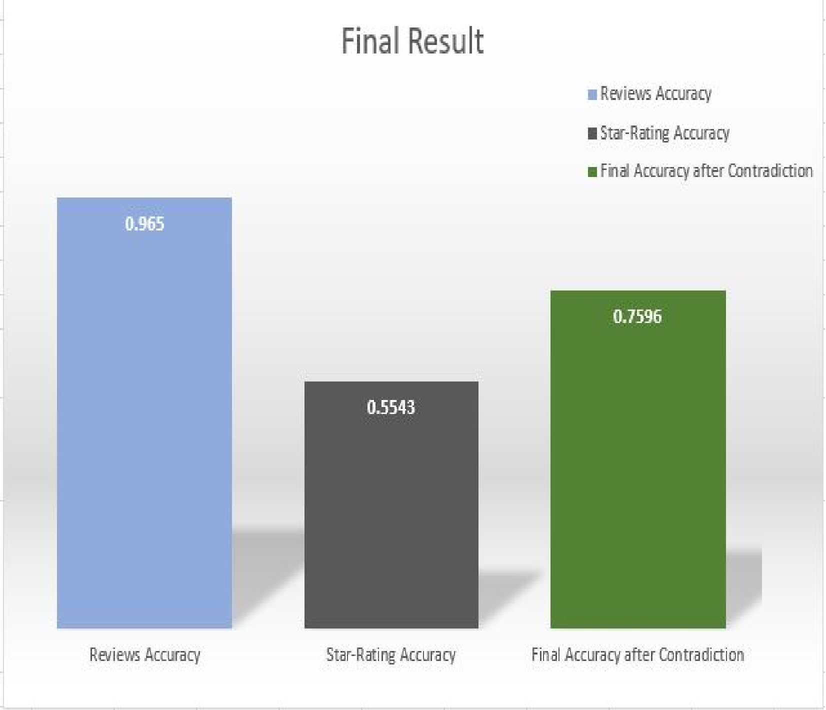

Figure 3. Final Results: shows the rating and reviews score of product and actual rating after contradiction which we have find using CNN hybrid

5. Evaluation

In previous section, we have discussed evaluation of algorithm now we have discussed its results in this section. CNN LSTM layer has been evaluated on f-measure, recall, precision and accuracy. For this purpose, we have used following formulas.

Recall= TPR = TP/RP [65] [66]
Precision= TPA = TP/PP [66] [67]
INVERSE RECALL = TNR = TN/RN [66] [68]
INVERSE PRECISION = TNA = TN/PN [66] [69]
ACCURACY=TER=PP * TPA + PN * TNA [66] [70]
F-MEASURE = 2·TPR/[TPR+C·FPR+1] [66] [71]

The results of evaluation are shown in figure 4 and table 1.

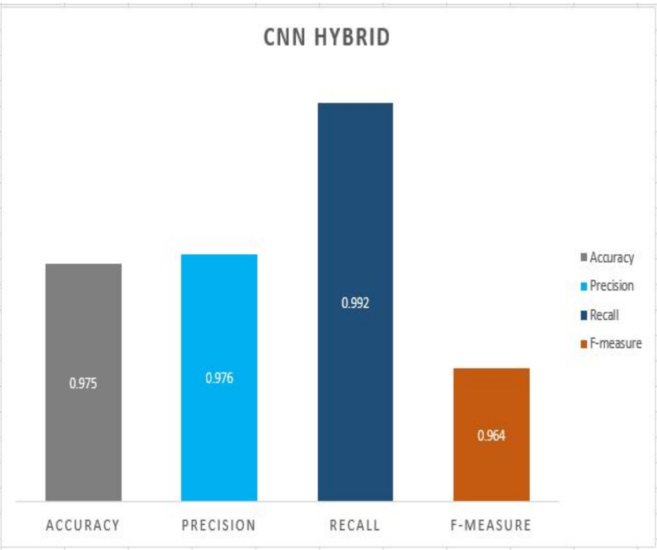

Figure 4. CNN Hybrid Evaluation

Table 1. CNN Hybrid Evaluation

MODEL	ACCURACY	RECALL	F-MEASURE	PRECSION
CNN HYBRID	0.975	0.992	0.964	0.976

Moreover, we have used different ratios of dataset used for training and testing as elaborated in table 2 by which we have come to conclude that dataset ratio used for training directly proportional to accuracy of algorithm.

Table 2. **CNN** Hybrid **Evaluation (Bigger dataset)**

TRAINING% -TESTING% RATIO	ACCURACY	RECALL	F-MEASURE	PRECISION
80%-20%	0.975	0.992	0.964	0.976
70%-30%	0.955	0.950	0.962	0.955
60%-40%	0.949	0.942	0.955	0.947
50%-50%	0.922	0.903	0.932	0.914
40%-60%	0.892	0.929	0.860	0.889
30%-70%	0.879	0.815	0.930	0.865
20%-80%	0.880	0.914	0.854	0.880

Moreover, we have used different size of dataset and evaluate it by using different training and testing ratio shown in table 3 and conclude that CNN LSTM layer gives better result on bigger dataset.

Table 3. CNN Hybrid Evaluation (Smaller Dataset)

TRAINING%-TESTING% RATIO	ACCURACY	RECALL	F-MEASURE	PRECISION
80%-20%	0.845	0.792	0.881	0.831
70%-30%	0.863	0.797	0.922	0.847
60%-40%	0.843	0.856	0.836	0.842
50%-50%	0.816	0.843	0.803	0.818
40%-60%	0.812	0.770	0.844	0.798
30%-70%	0.813	0.773	0.849	0.803

20%-80%	0.786	0.727	0.819	0.764

Bigger dataset consists of 2000 reviews and 2000 ratings of a product where smaller dataset consists of 1000 reviews and 1000 ratings of a product.
This evaluation comparison has concluded that

- More percentage of dataset used for training results in better results.
- CNN hybrid gives better result on bigger dataset as compared to smaller dataset.

We have also compared our results of CNN Hybrid model with Previously used machine learning techniques used for sentiment analysis of online reviews discussed in research article [11]

Table 4. Comparison of Machine Learning Techniques

ALGORITHM'S NAME	ACCURACY
Multinomial Algorithm	0.928
Bernoulli Algorithm	0.923
Logistic Regression Algorithm	0.933
Naïve Bayes Algorithm	0.553
CNN LSTM Layer Algorithm (Proposed Model)	0.975

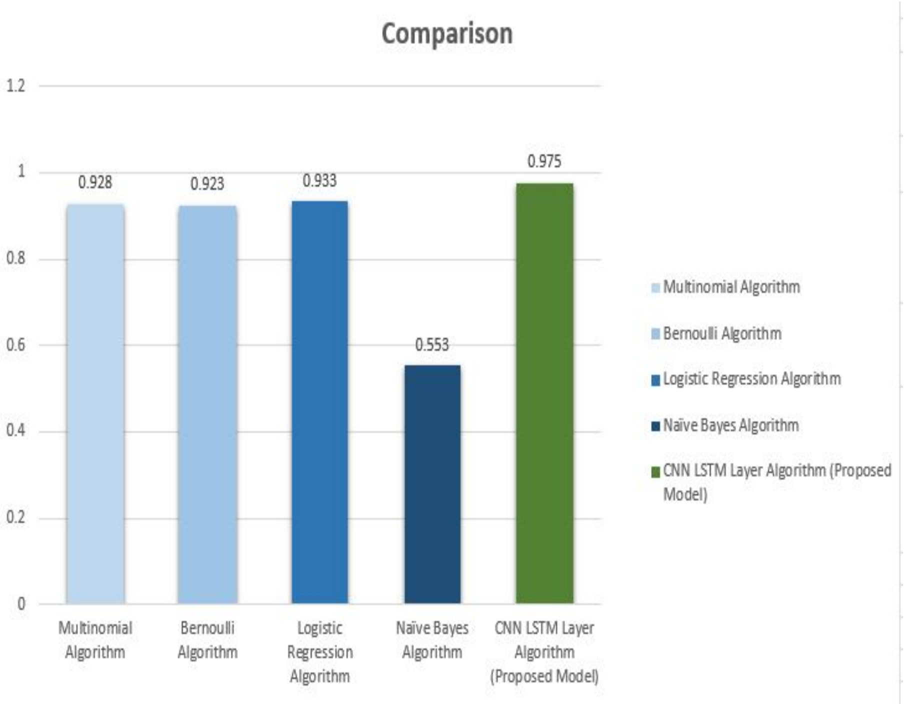

Figure 5. Comparison of Machine Learning Algorithms

From table 4 and Figure 5, it concludes that CNN hybrid model outperforms previously used techniques foe sentiment analysis of online reviews in terms of accuracy

6. Discussion

How can we say it better to use proposed approach rather than any other approach? To answer this question, we compare results of proposed method with other proposed methodology in article [72]. This article's methodology has some similarity with our proposed methodology. They apply LSTM-CNN on Arabic text as we used but used FastText word embedding method where we used word2vec. Below comparison show that our methodology works better.

Table 5. Comparison of Methodologies

Methodology	Accuracy
Word2vec before LSTM-CNN (Proposed)	97.05
FastText before LSTM-CNN	90.75

Why we don’t use GloVe instead of word2vec. As per our research for finding the suitable approach we come to conclusion that word2vec is better as said in article [73]. This article compares both word2vec and GloVe on text and results in finding word2vec better.

Last but not the least, Reason of LSTM-CNN is better is already shown in Table 5.

7. Conclusion

In this research, our main objective is to predict quality of a product on basis of both reviews and ratings given on product. We have collected dataset from amazon.com. Our dataset consists of label, reviews and ratings which we got after sentence splitting, the dataset consists of 2000 reviews and 2000 ratings of same product. 80% of dataset is used for machine training while 20% is used for machine testing. We have design a methodology in which CNN hybrid model is used to calculate sentiment score of reviews. Firstly, the score of product on basis of reviews has been calculated separately then score of product on basis of ratings has been calculated. Then contradiction has been shown by comparing both results. Finally, the actual result has been calculated on basis of both reviews and ratings which shows actual quality of product. We have also evaluate our methodology on smaller dataset which consists of 1000 reviews and 1000 ratings. The CNN hybrid model gives accuracy of 0.87 in case of smaller dataset and 0.975 in case of bigger dataset. This study has concluded that More percentage of dataset used for training results in better results and CNN hybrid gives better result on bigger dataset as compared to smaller dataset as in this research, we have used CNN hybrid model. In future, we focus on trying different deep learning algorithm to achieve high accuracy. We have also plan to implement this methodology on bigger dataset.

References

[1] K. R. Scherer, "Vocal communication of emotion," A review of research paradigms, p. 227–256, 2003.

[2] A. H. ,. H. K. Walaa Medhat, "Sentiment analysis algorithms and applications:A survey," Ain Shams Engineering Journal, vol. 5, 2014.

[3] F. C. J. C. Orestes Appel, "Main Concepts, State of the Art and Future Research Questions in Sentiment Analysis," Acta Polytechnica Hungarica, vol. 12, no. 3, 2015.

[4] M. H. a. B. Liu, "Mining and Summarizing Customer Reviews," ACM SIGKDD, pp. 22-25, 2004.

[5] R. Feldman, "Techniques and Applications for Sentiment Analysis," Communications of the ACM, vol. 56, no. 4, pp. 82-89, 2013.

[6] S. G. a. A. S. A. Weichselbraun, "Extracting and Grounding Contextualized Sentiment Lexicons," IEEE Intelligent Systems, vol. 28, no. 2, pp. 39-46, 2013.

[7] B. Liu., "Sentiment analysis tutorial," AAAI, 2011.

[8] R. Dale, "Classical Approaches to Natural Language Processing," in Natural Language Processing , 2010, pp. 3-7.

[9] L. S. R. G. R. M. A. M. A. A.-G. GHULAM MUJTABA, "Email Classification Research Trends: Review and Open Issues," Digital Object Identifier, 2017.

[10] E. I. E. a. A. Gherbi, "Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques," Journal of Computer Science, 2018.

[11] L. I. S. Sasikala P, "Sentiment Analysis and Prediction of Online Reviews with Empty Ratings," International Journal of Applied Engineering Research, vol. 13, no. 14, pp. 11525-11531, 2018.

[12] G. Rao, "Detection of Fraud Apps using Sentiment Analysis," International Journal of Innovative Science and Research Technology, vol. 3, no. 5, pp. 91-94, 2018.

[13] D. E. ,. A. S. E. Laila Abd-Elhamid, "Arabic Feature-Based Level Sentiment Analysis Using Lexicon-Based Approach," Journal of Fundamental and Applied science, vol. 10, no. 4, pp. 143-148, 2018.

[14] P. R. João Guerreiroa, "How to predict explicit recommendations in online reviews using text mining and sentiment analysis," Journal of Hospitality and Tourism Management, 10 July 2019.

[15] K. A. Ahlam Alrehili, "Sentiment Analysis of Customer Reviews Using Ensemble Method," in 2019 International Conference on Computer and Information Sciences, 2019.

[16] I. U. W. J. N. A. Jahanzeb Jabbar, "Real-time Sentiment Analysis On E-Commerce Application," in Proceedings of the 2019 IEEE 16th International Conference on Networking, Sensing and Control,, China, May 2019.

[17] J. Y. L. L. D. S. Donghang Pan, "Deep neural network-based classification model for Sentiment Analysis," IEEE, 2019.

[18] M. T. C. O. Reshmi Gopalakrishna Pillai, "Detection of Stress and Relaxation Magnitudes for Tweets," The Sixth International Workshop on Natural Language Processing for Social Media, pp. 23-27, 2018.

[19] S. A. H. T. N. O. W. A.-S. M. A.-b. Ahmed Al-Saffar, "Malay sentiment analysis based on combined classification approaches and Senti-lexicon algorithm," PLoS ONE, vol. 13, no. 4, 2018.

[20] A. A. a. S. W. Puspita Kencana Sari, "Measuring e-Commerce service quality from online customer review using sentiment analysis," in IOP Publishing

International Conference on Data and Information Science, Indonesia, 2018.

[21] Y.-J. Park, "Predicting the Helpfulness of Online Customer Reviews across Different Product Types," Sustainability, vol. 10, 2018.

[22] A. T. Shabeeba T, "Rating Prediction using review texts based on Sentiment Analysis," International Journal of Innovative Research in Science, Engineering and Technology, vol. 7, no. 5, pp. 4714-4718, 2018.

[23] X. Z. a. Y. Ohsawa, "Sentiment Analysis on the Online Reviews Based on Hidden Markov Model," Journal of Advances in Information Technology, vol. 9, no. 2, pp. 33-38, 2018.

[24] S. S. B. B. G. G. Swati Redhu, "Sentiment Analysis Using Text Mining: A Review," International Journal on Data Science and Technology, vol. 4, no. 2, pp. 49-53, 2018.

[25] M. K. a. Ali, "Sentiment Analysis of Pharmaceutical Products Evaluation Based on Customer Review Mining," Journal of Computer Science & Systems Biology, vol. 11, no. 3, pp. 190-194, 2018.

[26] I. S. F. A. Xiangfeng Dai, "A Framework for Automated Rating of Online Reviews Against the Underlying Topics," in ACMSE, USA, 2017.

[27] H.-R. Y. a. H.-S. H. Ha-Na Kang, "A Study of Analyzing on Online Game Reviews using a Data Mining Approach: STEAM Community Data," International Journal of Innovation, Management and Technology, vol. 8, no. 2, pp. 90-94, 2017.

[28] D. G. B. J. B. S. S.-F. C. M. P. Mohammad Soleymani, "A Survey of Multimodal Sentiment Analysis," Image and Vision Computing, 2017.

[29] V. Dr.B.Radha, "A Survey on Research Issues in Opinion Mining," International Journal of Innovative Research in Computer and Communication Engineering, vol. 5, no. 2, pp. 1849-1853, 2017.

[30] M. S. A. Sahare, "A Survey Paper on Opinion Mining and Sentiment Analysis," IJARIIE, vol. 3, no. 2, pp. 5075-5078, 2017.

[31] P. D. S. M. C. M. a. V. Asha S Manek, "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier," World Wide Web, vol. 20, p. 135–154, 2017.

[32] A. G. Elshrif Elmurngi, "Detecting Fake Review through Sentiment Analysis Using Machine Learning techniques," in The Sixth International Conference on Data Analytics, 2017.

[33] D. R. T. a. D. R. P. Surya Prakash Sharma, "Opinion Mining and Sentiment Analysis on Customer Review Documents- A Survey," International Journal of Advanced Research in Computer and Communication Engineering, vol. 6, no. 2, pp. 156-159, 2017.

[34] G. D. a. A. R. Shilpi Chawla, "Product Opinion Mining Using Sentiment Analysis on Smartphone Reviews," in 6th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO, India, 2017.

[35] V. V. J. V. a. N. D. B. Vaishnavi, "Sentiment Analysis in Online Hotel Booking," International Journal of Engineering Science and Computing,, vol. 7, no. 4, pp. 10725-10728, 2017.

[36] D. U. R. Babu, "Sentiment Analysis of Reviews for E-Shopping Websites," International Journal Of Engineering And Computer Science, vol. 6, no. 1, pp. 19965-19968, 2017.

[37] K. T. S. E. Raheesa Safrin, "SENTIMENT ANALYSIS ON ONLINE PRODUCT REVIEW," International Research Journal of Engineering and Technology, vol. 4, no. 4, pp. 2381-2388, 2017.

[38] V. a. G. S. Sodhi, "Sentiment Mining of Online Reviews Using Machine Learning Algorithms," IJEDR, vol. 5, no. 2, pp. 1321-1334, 2017.

[39] S. P. S. a. R. S. Mohan Kamal Hassan, "Sentimental analysis of Amazon reviews using naïve bayes on laptop products with MongoDB and R," in IOP Conf. Series: Materials Science and Engineering, 2017.

[40] D. S.Muthukumaran, "Text Analysis for Product Reviews for Sentiment Analysis using NLP Methods," International Journal of Engineering Trends and Technology (IJETT), vol. 47, no. 8, pp. 474-480, 2017.

[41] N. M. M. A. A. M. a. N. M. S. Harnani Mat Zin, "The effects of pre-processing strategies in sentiment analysis of online movie reviews," in AIP Conf. The 2nd International Conference on Applied Science and Technology, 2017.

[42] D. M. E.-D. M. Hussein, "A survey on sentiment analysis challenges," Journal of King Saud University – Engineering Sciences, 2016.

[43] J. D. P. William J. Amadio, "COMPETITIVE ANALYSIS OF ONLINE REVIEWS USING EXPLORATORY TEXT MINING," Tourism and Hospitality Management, vol. 22, no. 2, pp. 193-210, 2016.

[44] M. Sodanil, "Multi-Language Sentiment Analysis for Hotel Reviews," in MATEC Web of Conferences ICMIE, 2016.

[45] H. M. Z. a. S. N. Nurfadhlina Mohd Sharef, "Overview and Future Opportunities of Sentiment Analysis Approaches for Big Data," Journal of Computer Sciences, vol. 12, no. 3, pp. 153-168, 2016.

[46] S. C. A. G. Devika M D, "Sentiment Analysis:A Comparative Study On Different Approaches," Procedia Computer Science, vol. 87, p. 44 – 49, 2016.

[47] A. S. Gurneet Kaur, " Sentimental Analysis of Flipkart reviews using Naïve Bayes and Decision Tree algorithm," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), vol. 5, no. 1, pp.

148-153, 2016.

[48] R. J. S. a. P. Ajitha, "Survey of Product Reviews using Sentiment Analysis," Indian Journal of Science and Technology, vol. 9, no. 21, 2016.

[49] A. P. H. C. K. Ga. Aashutosh Bhatt, "Amazon Review Classification and Sentiment Analysis," (IJCSIT) International Journal of Computer Science and Information Technologies, vol. 6, no. 6, pp. 5107-5110, 2015.

[50] A. A. S. K. R. Abinash Tripathy, "Classification of Sentimental Reviews Using Machine Learning Techniques," Procedia Computer Science, vol. 57, p. 821 – 829, 2015.

[51] H. M. M. O. I. Doaa Mohey El-Din, "Online Paper Review Analysis," (IJACSA) International Journal of Advanced Computer Science and Applications, vol. 6, no. 9, pp. 220-229, 2015.

[52] D. S. K. Jeevanandam Jotheeswaran, "Sentiment Analysis: A Survey of Current Research and Techniques," International Journal of Innovative Research in Computer and Communication Engineering, vol. 3, no. 5, pp. 3749-3757, 2015.

[53] S. K. Yadav, "Sentiment Analysis and Classification: A Survey," International Journal of Advance Research in Computer Science and Management Studies, vol. 3, no. 3, pp. 113-121, 2015.

[54] X. F. a. J. Zhan, "Sentiment analysis using product review data," Journal of Big Data, vol. 2, no. 5, 2015.

[55] M. X. D. Z. B. M. Zhijun Yan, "A Context-Dependent Sentiment Analysis of Online Product Reviews based on Dependency Relationships," in Thirty Fifth International Conference on Information Systems, Auckland, 2014.

[56] S. W. W. H. L. C. L. J. Liu Lizhen, "A Novel Feature-based Method for Sentiment Analysis of Chinese Product Reviews," China Communications, pp. 154-164, 2014.

[57] N. Nehra, "A SURVEY ON SENTIMENT ANALYSIS OF MOVIE REVIEWS," INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN TECHNOLOGY, vol. 1, no. 7, pp. 36-40, 2014.

[58] J. Yadav, "A Survey on Sentiment Classification of Movie Reviews," International Journal of Engineering Development and Research, vol. 3, no. 1, pp. 340-343, 2014.

[59] Q. P. a. M. Zhong, "Detecting Spam Review through Sentiment Analysis," JOURNAL OF SOFTWARE, vol. 9, no. 8, pp. 2065-2071, 2014.

[60] L. I. S. Sasikala P, "Sentiment Analysis of Online Food Reviews using Customer Ratings," International Journal of Pure and Applied Mathematics, vol. 119, no. 15, pp. 3509-3514, 2018.

[61] K. W. a. X. Wan, "Sentiment Analysis of Peer Review Texts for Scholarly

Papers," Association for Computing Machinery., 2018.

[62] S. T. A. J. A. F. Kunal Varshney, "Sentiment Analysis of Application Reviews on Play store," International Journal for Research in Applied Science & Engineering Technology (IJRASET), vol. 6, no. 3, 2018.

[63] A. K. M. B. R. W. A. Anwar Ur Rehman, "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis," Multimedia Tools and Applications, 2019.

[64] "Amazon Customer Reviews Dataset," Amazon, [Online]. Available: https://s3.amazonaws.com/amazon-reviews-pds/readme.html.

[65] J. Berger, Statistical Decision Theory and Bayesian Analysis, New York, 1985.

[66] D. M. W. Powers, "Evaluation: From Precision, Recall and F-Factor to ROC, Informedness, Markedness & Correlation," School of Informatics and Engineering, Australia, 2007.

[67] Bonett DG & RM Price, "Inferential Methods for the Tetrachoric Correlation Coefficient," Journal of Educational and Behavioral Statistics, vol. 30, no. 2, pp. 213-225, 2005.

[68] C. J., "Assessing agreement on classification tasks: the kappa statistic," Computational Linguistics, vol. 22, no. 2, pp. 249-254, 1996.

[69] C. J., "A coefficient of agreement for nominal scales," Educational and Psychological Measurement, pp. 37-46, 1960.

[70] C. J., "Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit.," Psychological Bulletin, 1968.

[71] J. M. A. &. P. F. Lafferty, "Conditional Random Fields: Probabilistic Models for Segmenting and Labeling Sequence Data. Proceedings of the 18th International Conference on Machine Learning," in Morgan Kaufmann, San Francisco, 2001.

[72] Alimi, A. H. (2020). Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks. Social Network Analysis and Mining, 10(53).

[73] Pickard, T. (2020). Comparing word2vec and GloVe for Automatic Measurement of MWE Compositionality. Proceedings of the Joint Workshop on Multiword Expressions and Electronic Lexicons.