Opinion Classification of Product Reviews Using Naïve Bayes, Logistic Regression and SentiWordnet: Challenges and Survey

A Dadhich¹ and B Thankachan¹

¹School of Computer and Systems Sciences, Jaipur National University, Jaipur, Rajasthan, India

E-mail: anjalidadhich01@gmail.com

Abstract. In recent years, automated opinion classification has evolved as one of the most demanding area in natural language processing. Many such systems have been implemented and developed for the summarization and classification of text and reviews of online products. There are many data sources and domains which sells the online products, such as Amazon, Flipkart, Snapdeal etc. In the same direction, this paper is intended to present a detailed review and comparative analysis of various existing sentiment analysis algorithms especially for the Amazon products, which have worked upon the supervised learning techniques called Naïve Bayes, logistic regression and SentiWordNet. Various key parameters and aspects of such a comparative tour are the use of feature reduction method, sentiment polarity, dataset domain and sources, product name, data set size and classifier. Further this paper includes the discussion on their accuracy results; additional results including important findings; and needs, challenges and limitations. Lastly, the performance of these algorithms is evaluated by comparing the % usage of the key parameters.

1. Introduction

Automatic Sentiment analysis or opinion mining [1][2][3][4] is an important and demanding subfield of the Natural Language Processing (NLP) which analyses, compares and recommends various online available product’s reviews from different data sources [5][6][7][8][9][10]. The generic application areas of NLP essentially include the sentiment analysis, text classification, image classification [11][12][13][14], and text mining [2][15][16], and also the recent upcoming sub-fields of Machine Learning (ML), and deep learning. One of such a prominent field, sentiment analysis performs a sequence of steps and identifies the polarity of the reviews. It first acquires the reviews from the web, then performs the pre-processing, and lastly classifies the reviews into categories by detecting their polarity. The sentiment orientation or sentiment polarity generally consists of two types of reviews, called as, positive and negative. However, many other analysis systems use multiple sentiment orientations (N-level sentiment orientations), such as, positive-neutral-negative, strongpositive-positive-weakpositive-neutral-weaknegative-negative-strongnegative and many more. The other ways to find the polarity are by providing the stars to the products from 0 to 5 or by extracting and processing the textual contents of the reviews. On one side, such systems find the results in very effective ways, however on the other side, they face the critical issues and challenges of fake reviews, polarity, cost etc [17][18][19][20].
This paper is designed to provide a tour and detailed study on various sentiment analysis classification systems, which accept and classify the online available product’s reviews of Amazon using the Naive Bayes (NB), Logistic Regression (LR) and SentiWordNet techniques. The NB technique includes the independence among the features found, whereas Bayesian networks are based upon the dependence property of the features [21][22]. LR is a statistical model and its basic form implements the logistic function to model a binary dependent variable [10]. SentiWordNet is an opinion lexical analysis resource which is derived from the WordNet database [1][7][8]. In this method, a numerical or polarity score is provided to each term, so that the reviews can be differentiated into different dimensions of sentiment orientations. So, each term gets a non-zero value for both positive and negative scores [1][7][8].

This paper is organized as follows. The systematic tour and comparative study of various existing opinion classification approaches are illustrated in Section 2. Section 3 depicts various challenges and issues found in these approaches. Section 4 graphically shows the analytical results of these algorithms in terms of % usage. These results and findings are obtained by comparing the algorithms for some key parameters and aspects. Lastly, section 5 concludes the paper.

2. Related Work
This section demonstrates various approaches of sentiment analysis and opinion classification using NB, LR and SentiWordNet techniques. Section 2.1 includes a systematic tour of these approaches and algorithms for the year range of 2009 to 2020. Further, their comparative analysis is tabulated in section 2.2, which is drawn on the basis of their many important and primary parameters.

2.1. A systematic tour of opinion classification approaches
This section illustrates the detailed review of sentiment analysis and opinion classification based algorithms for the online available products. These algorithms primarily follow the steps of web crawling and review collection, pre-processing, feature reduction, classifier training, classification, and sentiment polarity determination. For this, they worked upon the review data sets of many online products, which they collected from many different data domains and obtained the results in terms of accuracy, precision, recall and F1-score. These approaches are demonstrated below one by one [23][24].

The sentiment classification and opinion mining method [1] first applied the SentiWordNet lexical resource to count the positive and negative term scores of a document and then determined the orientations of sentiments. It followed a sequence of steps, such as, pre-term score calculation using SentiWordNet, negation detection, linear scoring, feature selection, and finally classification with 3-fold cross validation. The holistic fully automatic unsupervised learning method [2] summarized various product reviews by first analysing the sentiments and then computing the feature scores using PicAChoo. This method used the positive and negative sentiment dictionaries to get the feature dependent sentiment polarities of opinion words. The feature based and statistical opinion analysis based method [3] summarized the customer reviews by first extracting the customer feedbacks from the websites; then adding these reviews to the review database and using Part Of Speech (POS) to tag these reviews; then extracting all the frequent features and finding the polarity feature wise and also the overall polarity; and finally summarizing all the reviews. For this, it used GO tagger, MS Excel, SQL server 2000, VB.net and other statistical techniques for opinion mining [25][26].

The information extraction method [4] from product reviews first extracted the information from web pages and then extracted the web page template features. Furthermore, it labelled the seed adjective based opinion words; expanded this set to a word set coverage reaching 90%; identified the comparative sentences, words match and negation patterns; extracted the features; and finally summarized the information. In addition to this, this method also provided the users’ feedback to the product manufacturers and retailers. Next sentiment analysis method [6] collected the online reviews; pre-processed them with stop word removal, stemming and POS tagging; extracted the noun aspects;
and finally identified the orientation of sentences and aspects (or phrases) by counting the number of positive and negative opinions of each aspect. This frequent item set mining method found long and good words for positive opinions, and poor and bad words for negative opinions [27][28].

Another sentiment analysis and opinion mining approach [7] acquired the data from online websites; pre-processed it with Porter Stemmer algorithm, Stanford POS tagger and WEKA tool; generated seven categories of score words using SentiWordNet; trained the model; and then classified the reviews with 10-fold cross validation using NB, Support Vector Machine (SVM) and Multilayer Perceptron (MLP) techniques. Another such method [8] extracted the product reviews from database; determined POS using Stanford POS tagger; split the phrases into sentence vectors and sentences into word vectors; extracted and chose the correct meaning of each word from SentiWordNet; computed sentence and text scores (between 0 and 1); and finally classified the reviews. The algorithm proposed in [10] extracted the text, source code, list of products and display review list from reviews; removed the stop words; extracted the features and then classified the reviews. The unified dynamic opinion mining and sentiment analysis framework [11] analysed and summarized all users’ reviews by extracting them through web crawling; by tokenizing the sentences and words, tagging the POS and extracting and distributing the features; then classifying them on the basis of polarity; and finally producing their critical review summary [29].

Next such algorithm [12] included the steps of text cleaning, white space removal, tokenization and segmentation, stop word removal, stemming, negation handling, noun phrase selection, feature selection, and classification. The feature based opinion mining method [13] of online products collected the data; pre-processed it; segmented the sentences; extracted one, two and three word features; filtered the features using filtering methods such as subset, superset and distance-based; determined the sentence sentiment polarity and opinion word popularity using sentiment lexicon; and lastly summarized the reviews. The sentiment classification method in [14] collected the data; pre-processed it; extracted its features; and finally classified it using NB, SVM, and LR of Apache Spark's scalable ML Library (MLlib). Another classification method of online reviews [15] first extracted and pre-processed the reviews, then used unigrams and weighted unigrams having positive and negative keywords for feature extraction, and finally classified them [30][31].

A comparative study of text sentiment classification using Term Frequency/Inverse Document Frequency (TF/IDF) vectorization [16] was proposed with supervised ML and lexicon-based algorithms. The ML algorithms were LR, SVM, and gradient boosting, whereas lexicon-based algorithms were Valence Aware Dictionary and Sentiment Reasoner (VADER), pattern, and SentiWordNet. It followed the steps of data collection; pre-processing including sentence extraction, un-escaping of HTML escape sequences, expansion of contractions, and tokenization; TF/IDF vectorization; training and classification. An ensemble approach of sentiment analysis [17] performed the steps of data collection and pre-processing; feature extraction and feature selection using union, intersection and revised union methods; and classification of sentiment polarity with 10-fold cross validation using SVM, Multivariate Bernoulli Naïve Bayes (MNB), Random Forest (RF) and LR [32].

The sentiment analysis method on contextual polarity classification [18] of online product reviews used WEKA tool to classify the hybrid set of text and emoticons using four-fold cross validation. It first collected the data; pre-processed, analysed and labelled its text and emoticons; found their polarity using analysis; and then obtained the accuracy. Another analysis method [19] first acquired the data as pool based active learning from reviews; pre-processed it; extracted the features using Bag Of Words (BOG) model, TF/IDF and Chi square approaches; and finally classified the reviews using NB, SVM, Stochastic Gradient Descent (SGD), LR, RF and Decision Tree (DT) with 10-fold cross validation. This method labelled the unlabelled Amazon data using both manual and active learning approaches. Next sentiment analysis method [20] first collected the reviews; pre-processed them with stop word removal, lemmatization, stemming, POS tagging and dependency parser; extracted the
features using double propagation; determined and decided the sentiment orientation using feature and opinion pair data; and finally classified and generated the feature based product summary.

Next sentiment analysis method [21] extracted the online reviews; pre-processed them including tokenization, stop word and noisy data removal, duplicate removal, missing value solving and stemming; extracted the features including TF/IDF, term co-occurrence, POS tagging, chunking, sentiment segmentation, and N-grams; mapped each feature's information to the crawled product features; sentiment classification; polarity detection using SentiWordNet; and result interpretation. Next sentiment analysis model [22] followed the steps of data acquisition; pre-processing including stemming, stop word removal, case conversion, punctuation removal and white space elimination; hyper-parameters selection; representation of word vectors for feature extraction on different aspects by finding cosine distance; sentiment classification using CBOW (Continuous Bag Of Words), and skip-gram models with 10-fold cross validation.

Another sentiment classification method [23] classified the reviews at the sentence level for electronic device category and at the review level for mobile phone accessory category. It followed the steps of data collection using Raku-review software; pre-processing; statistical information extraction; and classification along with the recommendation. Next method [24] analysed the sentiments by taking the raw reviews; cleaning the words to remove the stop words; stemming; creating the features; and finally predicting and recommending the polarity of reviews. Next approach [25] followed the steps of data collection; of advanced pre-processing including transformation, tokenization, stop word removal, token filtration, and stemming; of TF/IDF vectorization; and of sentiment classification.

The hybrid and ensemble sentiment classification system [26] included the steps of data set preparation; pre-processing; information indexing; feature reduction; sentiment classification with 10-fold cross validation using combined (NB, SVM and Genetic Algorithm) approach; and finally the computation of classification results. Another research work is the feature-based dynamic sentiment analysis system [27], which was designed for summarization of online product reviews. Its steps included the review extraction using PostgrSQL database; the text formatting including fuzzy matching and domain grouping of synonym words; the feature extraction including POS tagging, association mining and probabilistic approach; the opinion extraction including words extraction, opinion polarity identification, feature-opinion pair formation, negation words identification, and final polarity detection for feature-opinion pair; the feature based summarization; and finally the placement into their respective feature-based cluster.

The dual sentiment analyser [28] followed the steps of data collection and pre-processing; use of data expansion techniques including creation of reversed reviews opposite to that of the actual reviews for every training and test review; comparison of both types of reviews; use of the classification algorithms to compare the results; and finally prediction result generation. Another analysis approach [29] applied the steps of the pre-processing including stemming, stop words removal, and N-gram; the feature extraction; and the classification with 10-fold cross-validation using WEKA tool. This ensemble ML or voting method combined all five classifiers to work upon six scenarios of unigram, bigram and trigram using each one with and without stop word removal. Next approach [30] performed the steps of the pre-processing, the sentence extraction, the BOW extraction, the sentiment score computation, and the sentiment classification. The method in [31] collected the tweets, pre-processed them, extracted the features using TF vector, trained the classifier, tested the system with unknown tweets, and finally labelled them. Next algorithm [32] collected the unstructured reviews, converted them into understandable form, determined the tokens, removed the stop words, tagged the POS, searched the particular words, and classified the reviews by computing their score, and finally obtained the polarity [33].
2.2. Comparative Analysis of Various Existing Algorithms

Section 2.1 illustrated a systematic tour of the existing sentiment analysis and opinion mining algorithms. These algorithms and approaches are compared in this section on the basis of six discriminative parameters and aspects as depicted in Table 1. These key parameters are the feature reduction, sentiment polarity, dataset domains and sources, product name, data set size and classifier. Therefore, these algorithms used the feature reduction methods such as feature extraction, feature selection and feature filtering; the sentiment polarity types such as positive-negative, positive-neutral-negative, and Strong Positive, Positive, Weak Positive, Neutral, Weak Negative, Negative, Strong Negative (SP-P-WP-N-WN-SN); the online data domains and sources such as Amazon, Twitter, movie, Epinions, Mouthshut, Ebay, CNET, Rakuten, Kaggle, US Airline, and other random selection; the online available products such as electronics and software products, mobile phones, cameras, Internet Movie Database (IMDB), clothes and wearables, kitchen appliances, and other supplementary products; and the classifiers such as NB, SVM, LR, RF, DT, SentiWordNet, feature-based, gradient and Adaboost, K-Nearest Neighbour (KNN), frequency distribution, MLP, maximum entropy, VADER, pattern, bagging, genetic algorithm, SGD, and linear regression classifiers.

Table 1. Comparative Analysis of Existing Algorithms Using Different Parameters

| Ref no. | Problem focused | Feature reduction method | Sentiment polarity | Dataset domain and sources | Product name | Data set size | Classifier |
|---------|----------------|--------------------------|--------------------|----------------------------|--------------|--------------|------------|
| [1]     | Automatic sentiment classification of reviews | Feature Selection | Positive & Negative | Movie | Online available movies | Many movies with sufficient number of reviews | SVM |
| [2]     | Product review summarization with customized feature extraction tool | Feature Extraction | Positive & Negative | Random selection | Digital camera | 100 products & 6000 reviews | Feature based classification |
| [3]     | Opinion mining of online free format customer reviews | Feature Extraction | Positive & Negative | Electronic & software products | Hitachi router, Nokia 6610, Norton antivirus, Apex player, Nikon 4300, Micro mp3 & Creative IPOD | 7 products with sufficient number of reviews | Frequency distribution & Bayesian statistics |
| [4]     | Product review information extraction based on adjective opinion words | Feature Extraction | Positive & Negative | Epinions | Digital camera | One product with sufficient number of reviews | Feature based summarization |
| [6]     | Aspect | Feature Extraction | Positive & Negative | Amazon & Canon camera | One product | NB |
| Extraction and opinion mining of product reviews | Epinions & 100 reviews |
|-------------------------------------------------|------------------------|
| **Sentiment analysis of web data with SentiWordNet** | **Feature Extraction** |
| Sentiment Feature SP-P-WP-N WN-N-SN analysis | Bollywood 2. Mobile Phone Movie Samsung Galaxy-Grand-2 Mouthshut |
| [7] | 1. NH-10 2. Samsung Galaxy-Grand-2  |
| 1. Many reviews. 2. 20 reviews | Two products NB, SVM & MLP |
| **Lexicon based sentiment analysis from product reviews using SentiWordNet** | **Feature Extraction** |
| Lexicon based Feature Positive, Negative & Neutral analysis from product reviews using SentiWordNet | Amazon Electronic devices Many SentiWordNet products with 300 reviews |
| [8] | Amazon Apple Iphone 5S, Three Samsung J7 and Redmi Note 3 |
| 1. Many reviews. 2. 20 reviews | NB, LR & SentiWordNet |
| **Customer reviews summarization of products using NLP** | **Feature Extraction** |
| Customer reviews Feature Positive, Negative & Neutral analysis of online product reviews | Amazon Shoes, shirt, belt, Five products with watch, jeans, and bag sufficient number of reviews |
| [10] | Amazon, Ebay & Epinions |
| **Sentiment analysis of product reviews** | Canon EOS40D, Nikon coolpix 4300 & Nikon D3SLR digital cameras |
| Sentiment Feature Positive & Negative Selection | Amazon, Ebay & Epinions |
| [12] | Three products with more than 5.8 million reviews |
| **Feature based online product opinion summarization** | **Feature Extraction** |
| Feature based Feature Positive, Negative & Neutral analysis of product reviews using NLP | Apple iPhones (iphone 6s and 6s with 80 plus) |
| [13] | Two products SentiWordNet 3.0 |
| **Sentiment classification of product reviews using Apache Spark** | Amazon review polarity |
| Sentiment classification Feature Positive & Negative Extraction | Amazon products Many products with 4000000 reviews |
| [14] | Amazon, Ebay & Epinions |
| **Study of Amazon product reviews with ML** | Amazon Television & other products Many products with maximum entropy |
| Study of Amazon product reviews with ML | Amazon |
| [15] | Many products with 24,500 reviews |
| **Comparative study of sentiment** | Amazon 44 different products of Amazon and 1000 products LR, SVM, gradient |
| Comparative study of sentiment Feature Positive & Negative Extraction (Amazon) | Amazon |
| [16] | Amazon and 43,620 reviews. Text boosting,
### Sentiment Analysis Models

| Model                                                                 | Feature Extraction | Dataset/Source          | Techniques                                                                 |
|----------------------------------------------------------------------|--------------------|-------------------------|---------------------------------------------------------------------------|
| Multi-domain sentiment analysis using ensemble approach              | Positive, Negative, Neutral (SentiWordNet) | IMDB 1. Movies 2. Electronics 3. Kitchen | SVM, GB & LR, & movie reviews for VADER, pattern & SentiWordNet |
| Sentiment analysis of polarity in online product reviews of social media | Positive, Negative & Neutral | Twitter Tweets 1. Unilever 2. Procter & Gamble 3. Samsung 4. GlaxoSmithKline 5. Mobilink | NB, SVM & LR reviews |
| Sentiment analysis on large scale Amazon product reviews              | Positive & Negative | Amazon Cell phones and accessories, musical, & electronics | NB, SVM, SGD, LR, RF & DT reviews |
| Sentiment analysis application on product reviews                    | Positive, Negative & Neutral | Random selection Online available products Many products with 2229 reviews | NB & SVM |
| Aspect based sentiment analysis of product reviews                   | Positive & Negative | Amazon Online available products | MNB, RF & SVM products with sufficient number of reviews |
| Sentiment classification of online consumer reviews using word vector | Positive & Negative | Amazon Mobile phones | SVM, NB, LR & RF One product with more than 4,00,000 reviews |
| User reviews classification at sentence and review levels             | Positive & Negative | 1. Amazon 1. Electronic 2. Rakuten devices 2. Mobile phone accessories | Two products NB with Japanese datasets 1. Random selection 2. 4904 reviews |
| Comparison of prediction methods for                                  | Positive & Negative | Kaggle Woman clothing | NB, DT, RF, KNN, LR & Adaboost |

---

1. Multi-domain sentiment analysis using ensemble approach 
2. Sentiment analysis of polarity in online product reviews of social media 
3. Sentiment analysis on large scale Amazon product reviews 
4. Sentiment analysis application on product reviews 
5. Aspect based sentiment analysis of product reviews 
6. Sentiment classification of online consumer reviews using word vector 
7. User reviews classification at sentence and review levels 
8. Comparison of prediction methods for
| Performance analysis of supervised algorithms for sentiment classification | Feature Extraction | Positive & Negative | 1. Twitter | 1. Tweets | Two products | NB, RF, SVM, DT & KNN |
|---|---|---|---|---|---|---|
| Restaurant reviews | Online blogs, datasets, review sites & micro blogging | Positive & Negative | 2. US airline | 2. Text reviews | 1. 89743 tweets | 2. 14641 reviews |
| Many products with sufficient number of reviews | | | | | | |
| Feature based summarization of customers’ reviews | Feature Extraction | Positive & Negative | Amazon | Cannon G3 camera, iPhone 4s & MP3 player | Three products | Feature based classification |
| Feature based classification | Online blogs, datasets, review sites & micro blogging | Positive & Negative | | | | |
| Many products with sufficient number of reviews | | | | | | |
| Sentimental analysis on product reviews using ML | Feature Extraction | Positive, Negative & Neutral | Amazon | Mobile phones | One product | SVM & NB |
| Feature Selection | Positive, Negative & Neutral | Positive & Negative | | | | |
| One product | | | | | | |
| NB, SVM, RF, bagging & boosting features | | | | | | |
| Comparative sentiment analysis of sentences using ML | Feature Extraction | Positive & Negative | Twitter | Tweets including group of words, emoticons, symbols, URLs & references | Many products with 1.6 million tweets | MNB, SVM, & LR |
| Feature Selection | Positive, Negative & Neutral | Positive & Negative | | | | |
| One product with 2000 reviews | Amazon | Laptops | | | | |
| NB & semantic DT | | | | | | |

The observations obtained from Table 1 state that these existing algorithms performed sentiment analysis and opinion mining on a diverse set of online available products and reviews using a good set of classifiers. They worked primarily on feature extraction, positive-negative sentiment polarity, Amazon domain, electronic and software products, and NB classifier.
3. Challenges and Critical Issues

The Sentiment analysis and opinion classification techniques face many issues and challenges, when they process online available product reviews. These issues are related to the review quality, sentiment orientation, imaged or non-imaged dataset, feature-based concerns, and other related aspects which may lead the system towards the review misclassification and thereby may degrade the overall system performance. Some of these issues are listed below.

- Presence of fake, spam or junk reviews.
- Different and multiple meanings of the same words in different domains.
- Change in behaviour and performance of feature set in different environments and domains.
- Change in sentiment polarity with more number of negative reviews.
- Difficult filtering of sentiment classification methods.
- Incapability in processing the natural language data sets.
- Cost of the online available software and tools.

Along with all these issues, the careful analysis of detailed review also highlights the critical issues and challenges where these systems and algorithms did not perform well. They either faced some challenges with some specific datasets or could not be applied in certain environment and circumstances. They are illustrated below in Table 2 in terms of the accuracy results, additional results and findings, and needs, challenges and limitations.

### Table 2. Accuracy, Additional Findings, and Challenges and Issues Found in Existing Algorithms.

| Ref. no. | Accuracy | Additional results and Needs, challenges and limitations |
|----------|----------|-------------------------------------------------------|
| [1]      | 65.85% Accuracy.  
SentiWordNet features (no refinement): 67.4%; with linear weight adjustment to scores: 68%; & positive class.  
with negation detection and linear weight scoring: 68.5%. | Achieved better recall for negative class as compared to the positive class.  
Low level result gain with high expectations in film reviews. Got some inaccurate SentiWordNet scores due to reliance on glosses to determine term orientation. Overall classification accuracy reduction due to SentiWordNet scores on term glosses. Use of colloquial language & expressions where no opinion information exists. Disambiguation of WordNet terms with more than one meanings. Inaccuracies in assignment of POS tags. Correct detection of named entities led to misclassifications. |
| [2]      | Good accuracy. | - | Need to modify sentiment analysis method. |
| [3]      | Good accuracy. | Improved performance of 20% in proposed method than existing one. Got high-scored reviews 5 to 10 times more than low-scored ones. | Difficulty in finding out the negative features. |
| [4]      | Average Accuracy: 71.82%. Other accuracies are Hitachi Router: 50% to 57% accuracy with features | Achieved 50% to 57% accuracy with features | Need of automatic word classifier for polarity determination & of
88.23%, Nokia 6610: 76.09%, Norton Antivirus: 57%, Apex player: 67%, Nikon Coolpix 4300: 70%, Micro MP3: 74.46% & Creative IPOD: 70%

(< 250) & high accuracy with features (> 250).

opinion strength improvement.

Dependency of system accuracy on the number of iterations a feature commented.

[6] Opinion sentences: 84.75%, aspect extraction: 80.36%, & sentiment orientation: 92.37%.

Precision, recall, true positive rate, & f-measure rate depict high efficiency rate & lower false positive rate for NB & MLP. NB & MLP outperformed the SVM.

[7] Accuracy: 77.7%. Precision: Recall: F-Measure = 0.847:0.782:0.772 (in NB), 0.225:0.423:0.263 in SVM, & 0.862:0.876:0.867 in MLP.

Precision, recall, true positive rate, & f-measure rate depict high efficiency rate & lower false positive rate for NB & MLP. NB & MLP outperformed the SVM.

[8] 61%.

75% to 100% with 117 reviews, 40% for short reviews (< 50 words).

[9] 61%.

10% to 83%.

Capable to handle noisy & uninformative data.

[10] Recall: 70.2%, Precision: 64.16%, & Best performance of NB among all other algorithms.

Precision, recall, true positive rate, & f-measure rate depict high efficiency rate & lower false positive rate for NB & MLP. NB & MLP outperformed the SVM.

[11] Good accuracy.

8.88% feature set size reduction of its original size.

[12] 60% to 83%.

Capable to handle noisy & uninformative data.

[13] Average accuracy: 87.03%. Aspect extraction accuracy: 72% & sentiment polarity detection accuracy: 85%. Precision: 82.9% & recall: 90.55%.

Achieved 7% higher recall & 11.8% higher precision of sentiment determination than of feature extraction evaluation.

[14] NB: 85.4%, SVM: 86%, & LR: 81.4%.

SVM outperformed NB & LR.

[15] SVM, maximum entropy & NB unigrams of accuracy: 62.86%, 60.21% & 66.84%; & with weighted unigrams of accuracy 81.2%, 70.35% & 77.42%, respectively.

Obtained best accuracy with SVM. Worked well on weighted unigrams. 8.88% feature set size reduction of its original size. NB performed better than maximum entropy.

[16] SVM, gradient boosting & LR: accuracy of 89%, 87% & 90%; precision of 90%, 88% & 91%; recall of 98%, 98% & 97%; & F1-accuracy of 96%, 95% & 96%.

Found slightly better results than existing works & worst lexicon-based results than other compared to classifying positive

Discarded neutral reviews in 0 to 5 star Amazon rating. Difficulty in negative reviews classification as compared to classifying positive reviews.
score of 94%, 92%, & 94%, respectively. Pattern, VADER, & SentiWordNet: accuracy of 69%, 83%, & 80%; recall of 72%, 89%, & 88%; precision of 88%, 90%, & 90%; & F1-score of 79%, 89%, & 88%, respectively.

lexicon-based works. Performed well due to large data set. Achieved better results of lexicon-based models. Could classify true negative reviews. Found higher accuracy, recall & F1-scores in ML than lexicon-based models. 

Accuracy: 92.31% in SVM with combined (IG, CHI, GI) method. F1-Score: 89.77%, & IG: 88.62% in SVM. 90.53% (best) in MNB with electronics reviews. 88.47% (best) in LR with kitchen reviews using combined (IG, CHI, GI) method. 87.73% (best) in RF with kitchen domain. 0.94 RUC in SVM for movie reviews. SVM performed better than MNB (86.38%), RF (83.45%) & LR (86.88%) movie inreviews. Better results in Combined (IG, CHI, GI) approach than in each individual one. SVM performed best in linear SVM (F-score: 92.31%) with combined (IG, CHI, GI). MNB also resulted well.

Achieved best accuracy with SVM. Got low precision even with the 100% recall. Achieved highest accuracy with SVM in every dataset without overfitting. Execution time: 1363.8838 sec. & 19.2162 sec. in NB & SVM, respectively. Got higher accuracy in less time with SVM than NB.

Achieved maximum accuracy with RF using CBOW. Fastest SVM. Got 95% confidence interval. Achieved higher accuracy of CBOW as compared to skip-gram model. Found computationally expensive concatenating word vectors. Did not include rating system. Got lower classification accuracy for balanced dataset than unbalanced one due to the use of big data.

Achieved accuracy of Used small scale data set. Difficult
50% at review-level classification, & 75%:25% (training: testing) at sentence-level classification. Achieved high accuracies at review-level classification. Achieved higher classification accuracy than IMDB. 

Achieved highest F1-score with LR & lowest with NB. Due to computational constraints, only 10000 reviews out of 28227 reviews were used.

[24] Accuracy: 88% (Best) of LR, 86% with RF and Adaboost, and 43% (Least) with NB among all. Precision: 93% (Best) with RF & 83% (Least) with KNN among all others. 99% recall (Best): KNN. 

Achieved improved performance of 8% over all the methods. Most of the researchers worked upon SVM, and movies and social network based applications. Average increase of 8% and 6% in overall accuracy and precision respectively.

[25] Accuracy(%): precision(%): recall (%) for
1. Twitter= 17:78.83:77.37 in NB; 83.67:80.6:79.09 in SVM; 81:79.48:74.24 in KNN; 71.5:75.69:61.98 in DT, & 68.53:37.87:56.22 in RF. 
2. US Airlines - 63.33:64.15:61.9 in NB; 69:68.41:66.59 in KNN; 65.33:67.28:64.52 in SVM, 56.58:97:52.21 in decision tree; & 56.33:61.91:49.67 in RF.

[26] NB: 85%, SVM: 85.2%, genetic algorithm: 85.3%, & Hybrid: 93%. Better accuracy & performance of genetics than SVM & NB.

[27] Sentence polarity detection accuracy: 80% in iphone 4s, 76% in Cannon camera, & 82% in MP3. - Need to improve the accuracy of opinion polarity detection & feature extraction algorithms. Need to extend upon adverbs, verbs and nouns.

[28] 91% with SVM & 66% with NB. Used threshold (0 to 1) for polarity detection. Need to improve the system accuracy & performance.

[29] 89.87% (best) in RF using unigram with stop word removal. Ensemble approach obtained better performance than a single classifier except RF. Need to improve the system accuracy & performance.

[30] Accuracy: Precision: F-Score = 93.28%:93.13%:98.19% with NB & 87.76%:87.57%:93.3% with SVM, respectively. Got best results with ML methods. 98.17% with NB & 93.54% with SVM for camera reviews. Need to improve the system accuracy & performance.

[31] 86.23% for LR, 85.69% for SVM & LR achieved maximum Need to improve performance of
4. Performance Evaluation

It is seen that section 2 illustrated the detailed review and comparative analysis of existing sentiment analysis and opinion algorithms for recent years. Such comparison discriminated the algorithms on the basis of five key parameters, such as, feature reduction, polarity and orientation, data domains and sources, product types and classifiers. Their performance is measured and analysed on the basis of these parameters and is shown graphically through subsections 4.1 to 4.5.

4.1. Based Upon Feature Reduction Methods

Figure 1 depicts the % usage of the various feature reduction techniques. It is observed that most of the research works implemented the feature extraction method in their sentiment analysis model as their primary choice. They contributed 86.67%, 10% and 3.33% for feature extraction, feature selection and feature filtering techniques, respectively.

![Feature Reduction & Usage](image)

**Figure 1.** Depicting the % usage of Feature Reduction Methods in Algorithms.

4.2. Based Upon Polarity and Orientation

Figure 2 depicts the % usage of three types of sentiment polarity and orientation. It is observed that most of the research works designed bi-polar sentiment analysis systems as their primary choice. So that their methods could be able to classify the product reviews as positive and negative categories only. 70%, 26.67% and 3.33% contributions are found for three different sentiment polarity types, such as, Positive-Negative (P-N), Positive-Neutral-Negative (P-N-N), and SP-P-WP-N-WN-N-SN, respectively.
4.3. Based Upon Data Domains and Sources

Figure 3 depicts the % usage of various data sources and domains which were used to collect the different reviews and data sets for the algorithm implementation. It is observed here that many contributors used the Amazon as their primary source of review collection for online products. Some other online web sites and blogs were also used as the other sources for data acquisition. It can be seen that 53.33% research works implemented their systems with Amazon database.

4.4. Based Upon Product Types

The findings of section 2.2 state that many researchers found the results for the electronics and software products of some specific domains. It can be seen in Figure 4 that 40%, 23.33%, 20%, 20%,...
20%, 6.67% and 3.33% research works obtained the results for electronics and software products, mobile phone, camera, others, IMDB, clothes and wearables, and kitchen appliances, respectively.

![Figure 4. Depicting the % Usage of Various Types of Products in Algorithms.](image)

4.5. Based Upon Classifiers Used

It is observed that the sentiment analysis algorithms used a diverse set of classifiers to implement their systems. Most of them used Naïve Bayes as the primary classifier. They not only implemented their system with only one classifier rather than it opted to work upon a set of classifiers in the system. Figure 5 depicts that 76.67% and 60% contributions are found for NB and SVM classifiers. Similarly, other % usages range between 3.33% and 26.67% for remaining classifiers.

![Figure 5. Depicting the % Usage of Different Types of Classifiers in Algorithms.](image)
5. Conclusion
This paper demonstrated the detailed review and challenges in opinion classification of product reviews using NB, LR and SentiWordNet techniques. The review and comparative analysis showed that various sentiment analysis and opinion classification systems have been developed using different classifiers till now which classified the online available product reviews and also detected their orientation and polarities. This paper has been designed to analyze such contributions with NB, LR and SentiWordNet algorithms for Amazon products. This review will further be extended to include some more supervised learning algorithms, such as, RF and KNN techniques. The authors have started the design and implementation of a new hybrid system using these techniques.

References
[1] Bruno O, Brendan T 2009 Sentiment classification of reviews using SentiWordNet 9th IT & T Conf. (Technological University: Dublin) pp 1-8.
[2] Jung-Yeon Y, Jaeseok M and Sang-goo L 2009 A holistic approach to product review summarization Software Technologies for Future Dependable Distributed Systems (Japan: IEEE) pp 150-4.
[3] Naveed A, Ayesha R and Syed H 2010 Feature based opinion mining of online free format customer reviews using frequency distribution and Bayesian statistics The 6th Int. Conf. on Networked Computing and Advanced Information Management (Seoul, South Korea: IEEE Press) pp 57-62.
[4] Yan L and Wei H 2011 Product review information extraction based on adjective opinion words Fourth Int. Joint Conf. on Computational Sciences and Optimization (China: IEEE Press) pp 1309-13.
[5] Shalini P and Sonya K 2012 An enhanced fuzzy similarity based concept mining model for text classification using feature clustering Students’ Conference on Engineering and Systems (India: IEEE Press) pp 1-6.
[6] Jeyapriya A and Kanimuthi S C S 2015 Extracting aspects and mining opinions in product reviews using supervised learning algorithm 2nd Int. Conf. on Electronics and Communication Systems (India: IEEE Press) pp 548-52.
[7] Shoaib A and Ajit D 2015 A novel approach for sentimental analysis and opinion mining based on SentiWordNet using web data Int. Conf. on Trends in Automation, Communications and Computing Technology (India: IEEE Press) pp 1-5.
[8] Alexandra C, Valentin S and Bogdan M 2015 Sentiment analysis from product reviews using SentiWordNet as lexical resource 7th International Conference on Electronics, Computers and Artificial Intelligence (Bucharest, Romania: IEEE Press) pp 1-4.
[9] Arora, Amandeep Singh, Liness Raja, and Barkha Bahl. "Data centric security approach: A way to achieve security & privacy in cloud computing." Proceedings of 3rd International Conference on Internet of Things and Connected Technologies (ICIoTCT). 2018.
[10] K L Santosh K, Jayanti D and Jharna M 2016 Opinion mining and sentiment analysis on online customer review Int. Conf. on Computational Intelligence and Computing Research (India: IEEE Press) pp 1-4.
[11] Akkamahadevi R H, Mayur M and Priyadarshini M P 2016 Summarization of customer reviews for a product on a website using natural language processing Int. Conf. on Advances in Computing, Communications and Informatics (Jaipur, India: IEEE Press) pp 2280-85.
[12] Eswaraprasad, Ramkumar, and Linesh Raja. "Improved intelligent transport system for reliable traffic control management by adapting internet of things." 2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS). IEEE, 2017.
[13] Nilanshi C and Pardeep S 2017 Feature based opinion summarization of online product reviews Third International Conference on Science Technology Engineering & Management (India: IEEE Press) pp 1-7.
[14] Samar A-S, Ghazi A-N and Arafat A 2018 A large-scale sentiment data classification for online reviews under apache spark Procedia Computer Science: The 9th Int. Conf. on Emerging Ubiquitous Systems and Pervasive Networks vol 141 (Elsevier) pp 183–9.

[15] Abhilasha S R, Amit A and Preeti D 2018 Comparative study of machine learning approaches for Amazon reviews Procedia Computer Science: Int. Conf. on Computational Intelligence and Data Science ed V Singh, V K Asari, R B Patel, P Sidike vol 132 (Elsevier) pp 1552-61.

[16] Heidi N, Aravind V, Mamadou D and Rashed I 2018 Comparative study of sentiment analysis with product reviews using machine learning and lexicon-based approaches SMU Data Science Review vol 1 pp 1-22.

[17] Monalisa G and Goutam S 2018 An ensemble approach to stabilize the features for multi-domain sentiment analysis using supervised machine learning J Big Data 5 (Springer) pp 1-25.

[18] Marium N, Hafsa D, Ikram U L and Salman T 2018 Sentiment analysis of polarity in product reviews Social Media 14th Int. Conf. on Emerging Technologies (Islamabad, Pakistan: IEEE Press) pp 1-6.

[19] Tanjim U H, Nudrat N S and Faisal M S 2018 Sentiment analysis on large scale Amazon product reviews Int. Conf. on Innovative Research and Development (Bangkok, Thailand: IEEE Press) pp 1-6.

[20] Yuniarta B, Harris V S, Sinta I P S and Jen P S 2018 Application of sentiment analysis on product review e-commerce J of Physics: Conference Series, 1st Int. Conf. on Advance and Scientific Innovation vol 1175 (Medan, Indonesia: IOP Publishing Ltd) pp 1-8.

[21] Sindhu Chandra S 2018 Aspect based sentiment analysis of amazon product reviews Int. J of Pure and Applied Mathematics vol 118 pp 151-7.

[22] Barkha B and Sangeet S 2018 Sentiment classification of online consumer reviews using word vector representations Procedia Computer Science: Int. Conf. on Computational Intelligence and Data Science vol 132 (Elsevier) pp 1147–53.

[23] Yoichi S and Vitaly K 2019 Classifying user reviews at sentence and review levels utilizing Naive Bayes 21st Int. Conf. on Advanced Communications Technology (PyeongChang Kwangwoon_Do, Korea (South), Korea (South): IEEE Press) pp 151-6.

[24] Umid K D and Alaa N 2019 Comparative exploration of prediction algorithms for sentiment analysis using NLP Int. Conf. on Computer Communication and Informatics (India: IEEE Press) pp 1-6.

[25] Sunitha P B, Joseph S and Akhil P V 2019 a study on the performance of supervised algorithms for classification in sentiment analysis TENCON 2019 - 2019 IEEE Region 10 Conf. (India: IEEE) pp 1351-56.

[26] Nisha Y, Rajeev K, Bhupesh G and Asif U K 2019 Extraction-based text summarization and sentiment analysis of online reviews using hybrid classification method Sixteenth Int. Conf. on Wireless and Optical Communication Networks (India: IEEE Press) pp 1-6.

[27] Kushal B and Durga T 2019 Feature based summarization of customers’ reviews of online products Procedia Computer Science: 17th Int. Conf. in Knowledge Based and Intelligent Information and Engineering Systems (Elsevier) pp 142-51.

[28] Stephina R D and Kavita S 2019 Sentiment analysis based on multiple reviews by using machine learning approaches Proc. of the Third Int. Conf. on Computing Methodologies and Communication (Erode, India: IEEE Press) pp 188-93.

[29] Ahlam A and Kholood A 2019 Sentiment analysis of customer reviews using ensemble method Int. Conf. on Computer and Information Sciences (Sakaka, Saudi Arabia: IEEE Press) pp 1-6.

[30] Rajkumar S J, Vishal S S and Sachin N D 2019 Sentiment analysis on product reviews using machine learning techniques Cognitive Informatics and Soft Computing: Advances in Intelligent Systems and Computing ed Mallick P., Balas V., Bhoi A., Zobaa A. vol 768 (Singapore: Springer) pp 639-47.

[31] Poornima A and K S Priya 2020 Comparative sentiment analysis of sentence embedding using machine learning techniques 6th Int. Conf. on Advanced Computing & Communication Systems (India: IEEE Press) pp 493-96.
Acknowledgments
We express our sincere gratitude to all the area experts for giving their guidance, support, constant encouragement and valuable domain knowledge to carry out this research. We would also like to thank the reviewers for their critical comments which helped us to improve the manuscript.