I. J. Engineering and Manufacturing, **2021, 2, 40-52**
Published Online April 2021 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2021.02.04

Sentiment Analysis of Amazon Product Reviews Using Hybrid Rule-based Approach

Anjali Dadhich[1] and Blessy Thankachan[2]
[1]Jaipur National University, Jaipur, Rajasthan, India
[2]Jaipur National University, Jaipur, Rajasthan, India
Email: anjalidadhich01@gmail.com; blessyt218@gmail.com

Received: 26 October 2020; Revised: 25 December 2020; Accepted: 20 January 2021; Published: 08 April 2021

Abstract: In recent years, the retail market industry has taken a broad form to sell the products online and also to give the opportunity to customers to provide their valuable feedbacks, suggestions and recommendations. The opinion summarization and classification systems extract and identify a range of opinions about different online available products in a large text-based review set. This paper addresses and reviews the concepts of automatic identification of the sentiments expressed in the English text for Amazon and Flipkart products using Random Forest and K-Nearest Neighbor techniques. It presents a detailed comparative study of such existing sentiment analysis algorithms and methodologies on the basis of five key parameters. It results in evaluating their performance in terms of parameter usage and contributions. The paper also discusses their experimental results and challenges found. Therefore, this study shows the maximum usage of feature extraction, positive-negative sentiment, Amazon web source, mobile phone for a large set of reviews in the existing algorithms.

Index Terms: Sentiment Analysis, Random Forest, KNN, Opinion Summarization, Online Products.

1. Introduction

Automated sentiment analysis or opinion mining [1-3], [5-6], [8-17] is a prominent sub-field of the Natural Language Processing (NLP) and Machine Learning (ML), which interprets, extracts, summarizes and classifies the text [4], emotions and images [7,18] in the subjective review data. It is also known as the subjective analysis to classify the text according to the polarity and orientation of the opinions. The sentiment polarity is expressed as bi-, tri- or multi-polarity such as positive-negative, positive-neutral-negative, StrongPositive-Positive-Neutral-Negative-StrongNegative, respectively. The positive sentiment polarity occurs if the total number of positive word appearances is greater than the total number of negative word appearances. In the negative sentiment polarity, the total number of negative word appearances is greater than the total number of positive word appearances. The neutral sentiment polarity refers to the equal number of positive and negative word appearances.

SentiWordNet, a lexical resource, is used to provide three sentiment scores called positive, negative, and neutral in opinion mining. The objectivity is used as the opposite of subjectivity. Objectivity is not influenced by the personal opinions or feelings to represent the facts. Therefore, sentiment analysis is the contextual mining of text to identify and extract the subjective reviews in supply material. Furthermore it supports an enterprise to recognize the social sentiments in their brand, products or services while monitoring and managing online conversations and feedbacks [19] [20]. The performance measures in sentiment analysis are precision, recall and F1-measure. They are computed in terms of True Positive (TP), True Negative (TN), False Positives (FP) and False Negatives (FN), where TP, TN, FP and FN represent true classification as positive terms, truly classification as negative terms, false classification as positive terms and false classification as negative terms, respectively.

This paper aims to provide the detailed review and a systematic tour of the existing sentiment analysis and opinion mining methodologies using Random Forest (RF) and K-Nearest Neighbor (KNN) techniques. These algorithms are compared with each other on the basis of some key measuring parameters, which evaluate them for feature reduction, sentiment polarity, online web source, product types and review set size in English language only. The primary concern is to study the sentiment analysis of the online available product reviews of Amazon and Flipkart. Alongside, this paper illustrates the comparison among these existing algorithms in terms of their accuracy results and additional result findings. All these algorithms face many critical issues and challenges, which lead us towards the designing of a new, automated and efficient sentiment review analyzer.

RF is a collection of trees where each tree is different from each other. It is a supervised learning algorithm and is used in regression and classification problems. It constructs multiple Decision Trees (DT) and then it merges them together to get the absolute and stable values for review set training and testing. The reason behind the use of RF is that as it grows many classification trees so it resolves the instability problem of previous decision trees. To classify a new object from an input vector, this input can be run through every tree in the forest [6]. KNN is a simple algorithm which is used for both classification and regression problems. It stores all the available cases during training and then classifies new cases using a majority vote of its k neighbors.

This paper is organized as follows. This paper began with the introduction. Section 2 illustrates various concepts of sentiment analysis in terms of three components. The systematic tour and comparative study of various existing sentiment analysis and opinion classification approaches are described in Section 3. Section 4 depicts the accuracy results and additional findings on them. It also elaborated and compared various challenges, which were faced by each algorithm. Section 5 graphically evaluates the performance metric of these algorithms in terms of % usages. These results and observations are obtained by comparing these existing algorithms for some key parameters. Lastly, section 6 concludes the paper with the future scope.

2. Sentiment Analysis Concepts

Automatic sentiment analysis system is used to extract and classify the sentiments and emotions into different categories. The sentiment classification consists of the subjective part of an opinion, whereas emotion classification includes the projections or display of a feeling expressed in text. The sentiment analysis follows a sequence of steps such as the review collection, the text to lowercase conversion, punctuation and additional white space removal, tokenization, stop word removal, lemmatization, feature reduction and classification.

A few articles described the sentiment analysis and its related concepts as discussed next. Lior Goren Ruck, Senior Data Product Manager, Yotpo discussed about the customer feedback analysis with opinion mining for customer-driven brands in her article "How Customer Reviews Can Be Used for Opinion Mining" [19] on March 14, 2018. She addressed the concepts of opinion mining and customer reviews, the working of opinion mining and sentiment analysis, and topics and opinions extraction from the reviews. According to the recent article "Sentiment Analysis: Aspect-Based Opinion Mining - An investigation into sentiment analysis and topic modelling techniques" [20] in Towards Data Science, posted on October 27, 2020, Lowri Williams focused on the main challenges of growing user base with the large amount of data generated in natural language. She demonstrated the theoretical and implementation concepts in aspect-based sentiment analysis, which extracts both the entity described in the text of a product or service and the sentiment associated with these entities. With this she also discussed the topic modeling for text aspect extraction.

Figure 1 depicts the basic level of components in sentiment analysis, which are categorized as content based, source based and supplementary. The contents in the sentiment analysis process can include the text, images, and emotions, where the text can be monolingual, bi-lingual or N-lingual. The sentiment polarity is bi-polar, tri-polar or N-polar, such as positive-negative, positive-negative-neutral and many more.

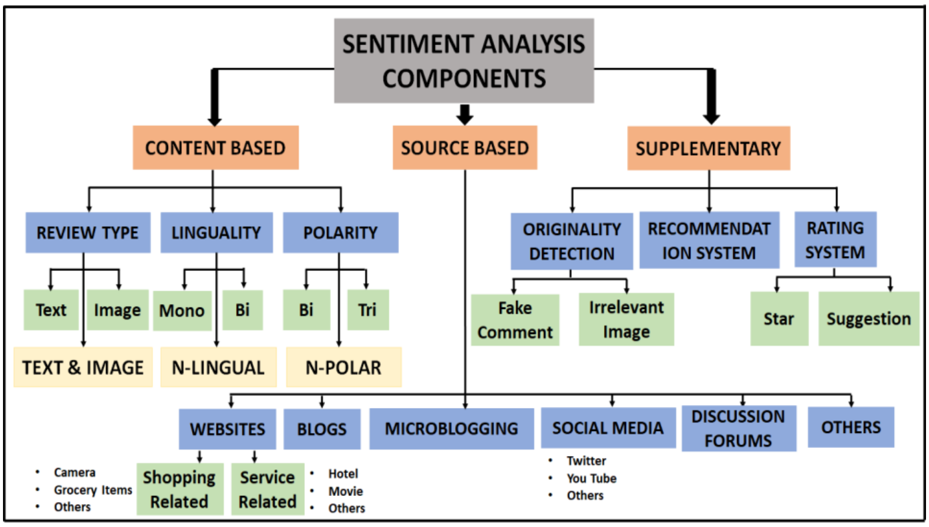

Fig. 1. Sentiment analysis and the components.

The online websites either sells the products or provides the services to the customers. Further the data sources in sentiment analysis are websites, blogs, microblogging, social media, discussion forums and other domains. The third component includes the concepts of originality detection of the reviews, recommendation system, and the star rating system. Many times, reviews are not straightforward, understandable and clear just like black and white, so the star rating system is used.

3. Related Work

This review illustrates the theoretical and algorithmicaspects of recent approaches in opinion summarization, classification and sentiment analysis using the RF and KNN techniques. Their observations and key findings are compared with each other with respect to six primary parameters which measure and evaluate their procedural concerns and practical results. The detailed review of these approaches is given below.

The sentiment expression extraction and sentiment classification approach [1] used the combination of four tree kernels, called Feature Sentiment Tree (FST), Generalized Feature-Sentiment Tree (GFST), and Boundary Marked GFST ($GFST^m$) And Generalized Feature-Sentiment Tree with polarity labels ($GFST^p$). It used three types of feature sets, such as lexicon features, syntactic features and semantic features for feature vector kernel, and then it represented each classification instance as a vector of n-dimensional predefined features. With this, it collected the reviews; split and indexed them into sentences using Indri2; and classified the sentiments using 10-fold cross validation. The automated product feature extraction system [2] used the combined approach of bootstrapping and ID3 along with Hu's method and Popescu's method. It followed the steps of data collection; preprocessing including extraction of consecutive noun sequences as candidate noun sequences and expression of the candidate noun sequences with classification features; feature extraction from candidate noun sequences; and lastly the final and new feature set extraction.

The review in [3] discussed the theoretical and practical concerns for efficient implementation of opinion classification in sentiment discovery using the classifiers such as Latent Dirichlet Allocation (LDA), Support Vector Machine (SVM), Multinomial Logistic Regression (MLR) or maximum entropy, Naïve Bayes (NB), Artificial Neural Network (ANN), DT, RF, and boosting and bagging algorithms with K-fold cross-validation. The empirical study and sentimental analysis method, presented in [5], classified the review text using NB by finding the review's subjectivity and objectivity; by correcting the spellings of the reviews using Word Net dictionary; by finding their semantic meanings; and lastly by calculating their polarity using DT. It followed the steps of review dataset collection; pre-processing including tokenization and stop word removal; feature selection; and then reviews classification and evaluation.

The sentiment analysis method [6] used a supervised lazy learning model along with syntactic rules for subjectivity/objectivity analysis and linguistic patterns identification. This method collected the reviews; pre-processed them with Part-of-Speech (POS) tagging through Penn Treebank annotation scheme with linguistic patterns identification through SentiWordNet; obtained the orientation of opinion sentences based on high score values of opinion words; extracted the product features and opinion words from subjective review sentences; extracted the opinions and summarized the reviews; and lastly classified them. Next sentiment analysis method [8] provided the optimized summary of text and image of customer reviews. Its steps include the data acquisition from input review corpus; the selection of some data from total data; the thread reading and loading into runtime memory; the input corpus size estimation after applying the pre-processing mechanisms; the stop word corpus acquisition including the word list for early elimination; the use of tokenization and stemming on their root words using the Porter stemming algorithm; the computation of token score and overall score; the determination of sentiment depth; the use of polar tag and emotion estimation; and finally the review classification.

The comparative study [9] presented the comparison between lexicon dictionary based approach with N-grams and three ML algorithms. These ML algorithms were RF Learner with word vector and Principal Component Analysis (PCA) features, DT learner with document vector features, and RF with N-grams. The proposed method in [9] collected the data from corpus; divided it using EmEditor; cleaned and pre-processed reviews using punctuation eraser, stop word remover, case converter, and snowball stemming; obtained the features; and then classified the reviews. The sentiment analysis method [10] classified the customer reviews by data preprocessing; data resampling; feature extraction including dictionary based features and 50-d glove dictionary based features; and classification using classifiers such as Multinomial Naive Bayes (MNB), SVM with linear kernel, SVM with Radial Basis Function (RBF) kernel, KNN-4, 5, 6 and Long Short-Term Memory (LSTM) classifiers.

The sentiment analysis method [11] first performed the steps of data preprocessing including removal of stop words, punctuation marks, alpha-numeric character, HTML-tags, de-duplication, stemming and lemmatization; then text feature extraction using Term Frequency/Inverse Document Frequency (TF/IDF); then data splitting into training and testing; and finally the classification using KNN, LR and SVM with 5-fold cross validation.Next method [12] collected the data; computed the co-variance and correlation on it; selected the dependent and independent variables; pre-processed the data; applied the linear transformation, standardization, normalization and data mining; extracted the features; and finally classified them along with the detection of fake reviews.Another method [13] also began with the collection of online product reviews using web scraping, then it pre-processed it, extracted the features, and classified them.It replaced the output layer of the Convolution Neural Networks (CNN) by SVM.The method followed the steps of data collection, pre-processing, feature selection, detection process, and sentiment classification with K-fold cross validation[14].

The sentiment classification method of unlocked mobile reviews [15] predicted the product's customer rating based on its reviews and then classified these reviews using NB-SVM, RF, and MNB, gradient boosting, CNN, LSTM, and Stochastic Gradient Descent (SGD) classifiers. It included the steps of pre-processing including stop word removal, stemming, and POS tagging; level analysis with unigram classifiers; feature extraction; and lastly the classification.

The sentiment analysis [16] proposed a method Contextual Analysis (CA) to predict the performance of ML models, and to give an early warning to indicate when the performance of ML models were deteriorating with new datasets. This method collected the data from corpus; pre-processed it including tokenization, stop word removal, and normalisation with stemming and lemmatization; constructed a Hierarchy Knowledge Tree (HKT) with CA using unlabelled data; defined the threshold and created the branches for remaining words using the nodes created in a level of the HKT; and then classified the reviews. The multi-class classification [17] of comparative user reviews into different classes first acquired the dataset from YouTube using beautiful soup; pre-processed it including tokenization, POS tagging, spell correction, hash tag removal and other special characters; applied the multiple classifiers; and lastly evaluated the performance to recommend the best classifier.

In this way, this tour described the procedural details of many recent opinion and sentiment analysis based algorithms. Alongside, they show many differences with each other. Table 1 depicts such differences by comparing and evaluating these approaches on the basis of six measuring criteria, such as feature evaluation, sentiment polarity, dataset domain, product corpus, dataset size, and classification method.

Table 1.Comparison among existing proposals using discriminating criteria.

Ref. No.	Problem Focused	Feature Evaluation Method	Sentiment Polarity	Dataset Domain	Product Corpus	Data set size	Classifier
[1]	Tree kernel based opinion mining of online product reviews	Feature Extraction	Positive & Negative	1. Customer review data 2. Epinions	1. Digital camera, cell phone, MP3, DVD player, router, antivirus software 2. No. of sentences	1. Six products with 442reviews 2. 7,101,733 review bodies	SVM with tree kernels & polynomial kernel
[2]	Product feature extraction using a combined approach	Feature Extraction	Positive & Negative	Amazon	Digital camera	One product with1000 reviews	-
[3]	Sentiment discovery using classification techniques	Feature Extraction	Positive & Negative	Amazon, EBay & Twitter	Samsung phone & others	Sufficient number of products & reviews	LDA, SVM, LR, NB, ANN, DT, RF, boosting & bagging
[5]	Sentimental analysis using a hybrid approach	Feature Selection	Positive & Negative	Flipkart	MOTO X Play mobile phone	5 products with 10 reviews	NB & DT
[6]	Summarization of product reviews based on features &opinions	Feature Extraction	Positive & Negative	Amazon	Canon digital cameras 1& 2, Nokia phone, MP3 player & DVD player	5 products & sufficient number of reviews	KNN
[8]	Hierarchical sentiment analysis for automatic review classification	Feature Extraction	Positive & Negative	Flipkart, Snapdeal&BestBuy	Nokia: Lumia-920 & Lumia-1020, Samsung Edge: S6 & A5, & Sony-Xperia: Z1 & M4	6 products & sufficient number of reviews	RF
[9]	Study on opinion mining on Amazon product reviews	Feature Selection	Positive & Negative	Amazon	Random selection	Many products. 57.3% positive, 25.56% negative & 17.14% neutral review	RF with word vector & N-gram, & DT with document vector
[10]	Sentiment analysis for Amazon product reviews	Feature Extraction	Positive & Negative	Amazon	Random Selection	34627 reviews	MNB, SVM, KNN & LSTM
[11]	Sentiment analysis using ML	Feature Extraction	Positive & Negative	Amazon (Kaggle)	Food-based products (Total 74258 products)	Few products.Sets of 568454 & 100,000 reviews	KNN, LR & SVM
[12]	Sentiment analysis of customer reviews on social network.	Feature Extraction	Positive & Negative	Flipkart & Amazon	Mobile phone (Honor 7A, Redimi 6, Realme 2, MotoX4 & Nokia 6.1)	Five Products with sufficient number of reviews	-

[13]	Sentiment analysis of product reviews scraping from web pages	Feature Extraction	Positive, Neutral & Negative	Amazon, Flipkart&Snapdeal	Random Selection	Forty-five products & 43777 reviews	KNN, SVM, RF, CNN &SVM-CNN
[14]	Sentiment analysis of social media network using RF algorithm	Feature Selection	Positive, Neutral & Negative	Flipkart	Clothes, toys, tools & hardware, school supplies, baby care, mobile & accessory	Six products with good number of reviews	RF & SVM
[15]	Sentiment analysis on mobile phone reviews using supervised learning	Feature Extraction	Positive & Negative	Amazon (Kaggle)	Unlocked mobile phone	One product with more than 413,840 reviews	Gradient boosting, SGD, MNB, LSTM, RF, NB-SVM & CNN
[16]	Supervised ML performance prediction for sentiment analysis using contextual methods	Feature Extraction	Positive & Negative	Amazon with four different domains	Book, DVD, kitchen, electronics and movie	Five products & 13000 reviews	MNB, RF & SVM
[17]	Sentiment classification of user reviews using supervised learning with comparative opinion mining	Feature Extraction	Positive, Negative & Neutral	You Tube	Business-related comparative reviews on 1. iOS vs Android 2. Microsoft vs Google	Three products.1. 6000 reviews 2. 3000 reviews	NB, SVM, RF, KNN, DT, LR &gradient boosting

In this way, it is seen that this study not only compared various opinion mining, sentiment analysis and feature based methodologies on the basis of six key parameters, but also elaborated their concepts and practical significance very clearly. Most of them worked upon the successful implementation of classification systems with a large set of product reviews, but still there is the need of an efficient sentiment analysis classification system for multiple online available products of multiple data domains and sources.

4. Discussion on Results and Challenges

Section 3 described and distinguished many algorithms of opinion classification and sentiment analysis. Their accuracy results, additional result findings, needs, challenges and limitations are compared with each other and are tabulated in Table 2. These systems evaluated their performance in terms of accuracy, precision, recall and F1-score. Many systems were implemented either with multiple classifiers or the hybrid classifier of two classifiers. Their results were also compared with each other and performance was evaluated. Further the future needs, challenges and limitations of these systems present the critical issues and scope of these existing algorithms.

Table 2. Comparison of accuracy results, additional result findings along with needs, challenges and limitations.

Ref. No.	Accuracy	Additional results and findings	Needs, challenges and limitations
[1]	89.56% Accuracy, 91.53% Precision, 93.1% Recall, &92.31% F1-score	Complexity reduction due to the use of tree kernels. Best performance of $GFST^m$, &$GFST^m$+Poly with 14.51% &$GFST^p$+Poly with 15.79% improvement over baseline.Obtained better results with feature vector kernel than traditional flat features. Achieved best performance of tree kernels with feature vector kernel.	Challenge to design proper tree kernel spaces for similarity evaluation. Worst performance of opinion mining for negative opinions. Issue of data sparsity. Impractical to use syntactic structure features directly. No need to extract each individual feature from each parsing tree.
[2]	Recall: 81%&Precision: 70%	Used a combined method of bootstrapping & ID3 for feature extraction.	Precision decreased due to increased recall. Avoided textual pattern structure & similarity function of textual pattern.
[3]	Recall: 0.0654 in NB, 0.8 in RF, 0.6 in LR, 0.6 in DT, 0.8 in Bagging, & 0.2 in SVM.	Achieved highest recall accuracy with RF. Found mean accuracy of cross validation: 0.4625 in DT, 1 in LR, 0.875 in SVM, & 0.825 in RF.	Challenges were co-reference resolution, sarcastic sentences, linguistic issues, opinion spamming, &volatility over time.
[5]	High Accuracy.	Better performance of Moto X Play (16 GB) compared to Moto E & Moto G.	Need to improve the accuracy and system performance.
[6]	Average recall: 0.967, precision: 0.761,& F-Score: 0.849.	Training accuracy was 81.37 (+/- 0.77%).	Need to improve implicit features and opinion strength. Got low precision values than recall. Ignorance of some opinion sentences and product features due to grammatical errors in review sentences.
[8]	Average accuracy: 94.71% Accuracy: 96.23% without & 93.75% with manual compression. Positivity: 64.1% in Samsung & 54.93% in Sony.	Provided the image summary of reviews. Found 94.71%Porter stemming accuracy. Existing model outperformed the syntactic & algorithmic error categories.	Need to improve the accuracy and system performance.

[9]	0.896 (Best) using lexicon dictionary based approach with N-grams, & 0.5 (worst) using RF with word vector and PCA.	Achieved better performance with lexicon based approach than other ML techniques. Achieved accuracy of 70% in DT with document vector features &of 85% in RF with N-grams.	Used only 500 MB data. Trained classifier did not work with other domains. Failed to predict reviews having spelling mistakes and slangs. Failed to process ambiguous and contradicting reviews.
[10]	Good Accuracy.	Average training accuracy: 68.02% & testing accuracy: 65.6%. Obtained best results with LSTM. KNN-5 outperformed other two KNN models. SVM with linear kernel slightly outperformed SVM with RBF Kernel. Increase in dictionary's length did not have too much effect on accuracy	Obtained worst result using glove mean than normal word count due to weak individual word feature set. Low accuracy due to data imbalance. Did not find high no. of data points from other resources. Had over fitting problem in SVM with linear kernel. KNN required much higher computation complexity than NB & SVM during training time. Issue of curse of dimensionality.
[11]	Good Accuracy.	-	Need to maintain the balanced proportion of positive and negative reviews.
[12]	Good Accuracy.	Had positive reviews with greater & negative reviews with less than 3.5 rating.	Variation in customer rating for same product. Problem of fake reviews.
[13]	94.2% precision, 92.4% recall, 90.7% F-score, & 92.4% accuracy with Hybrid SVM-CNN.	Best performance with hybrid algorithm. Time taken (milli-seconds): 1400 in KNN, 1317 in SVM, 1429 in RF, 1284 in CNN, & 1102 in hybrid SVM-CNN.	Need to improve the accuracy and system performance.
[14]	97% accuracy (Best) with RF &92% with SVM.	-	Need to improve the accuracy and system performance.
[15]	Average accuracy: 73.85%, precision=72%, recall=74.28%, F1-measure=69.14%	Obtained better prediction in RF than LSTM & CNN. Not found good results with gradient boosting, MNB & SGD. Maximum accuracy: 85% with RF.	Need to improve the accuracy and system performance.
[16]	57% to 75% prediction accuracy with SVM.	Estimated error within 2.75 to 3.94 & 2.30 for 3.51 for average absolute differences. Lower results of SVM than MNB & RF.	Need to improve the accuracy and system performance.
[17]	95% accuracy, 96% precision, 95% recall & 95% F-measure.	Best result with RF in case 1 & with DT in case 2. Also tested the system for Facebook vs Twitter.	Need to improve the accuracy and system performance.

It can be seen that most of the existing algorithms and systems achieved high accuracy, precision, recall and F1-score using single, multiple or hybrid classifiers. The hybrid classifiers achieved high performance results compared to the results of the single classifier. Additionally, these systems achieved and compared such results with a good set of online customer reviews for multiple online available products and services. Although these systems have implemented their algorithm with high level of efficiency and performance, yet they faced many critical issues and challenges.

These primary challenges include the need of improved algorithm; precision decrease with increased recall; curse of dimensionality; handling of fake, negative, ambiguous, contradicting and erroneous reviews; handling the large data set; issues in customer star rating and many others. Therefore, it can be stated that there is a high need of an efficient opinion classification system using multiple classifiers, which can result in very high performance with a large data set and can handle the problems of review types very effectively.

5. Pcsa Design and Architecture

The proposed Product Comment Summarizer and Analyzer (PCSA) system design is a generic, robust and fast system, which classifies online English comments collected from Flipkart shopping websites using five different supervised learning classification techniques. These techniques are Naïve Bayes, logistic regression, SentiWordNet, random forest and K-Nearest Neighbor. The PCSA system is designed in the training and testing phases.

During the system training, the user login registration form and credentials data base is set. It accepts multiples product URLs (Uniform Resource Locators) from Flipkart .It pre-processes the comments, segments the sentences, extracts their features, summarizes them or lastly classifies them using any one classification techniques. The system is trained with all five classification techniques. It is designed to classify the comments of multiple products through any one technique at a time. During the system testing, this proposed PCSA system is tested for a different unknown set of product comments collected from both websites. It goes through all these steps one by one and the chosen trained classifier categorizes these comments efficiently.

The entire PCSA model is designed using two stages called the training stage and the testing stage. With this, the system defines two storage media called Natural Language tool Kit based Corpus (NLTK based Corpus) and class repository. The NLTK based corpus includes the English dictionary, list of stop words, and WordNet information. The class repository includes three predefined classes called positive, negative and neutral. The PCSA system is trained on the known set of comments collected from Flipkart.

The trained classifiers classify the unknown set of comments. During the training stage, these comments are pre-processed for stop word removal, stemming and lemmatization. It makes use of the NLTK based corpus. It identifies the sentences on the basis of delimiters and segments the sentences. The system extracts all the important and relevant words, and it discards all the irrelevant words.

The system extracts the features from the important information. Further, these features train all the five classifiers. Then these classifiers perform the classification and classify all the comments into positive, negative and neutral comments. On the other side, all the comments are summarized and the system provides the rating to the each individual product in terms of the stars. The star rating uses the summarized results from all three classes and then provides the star rating to the chosen products of Amazon. During the testing stage, the system accepts the unknown set of comments. The system performs all the steps on this set of comments. Lastly, it chosen classifier classifies the comments into categories and summarizes all the class results to provide the start rating.

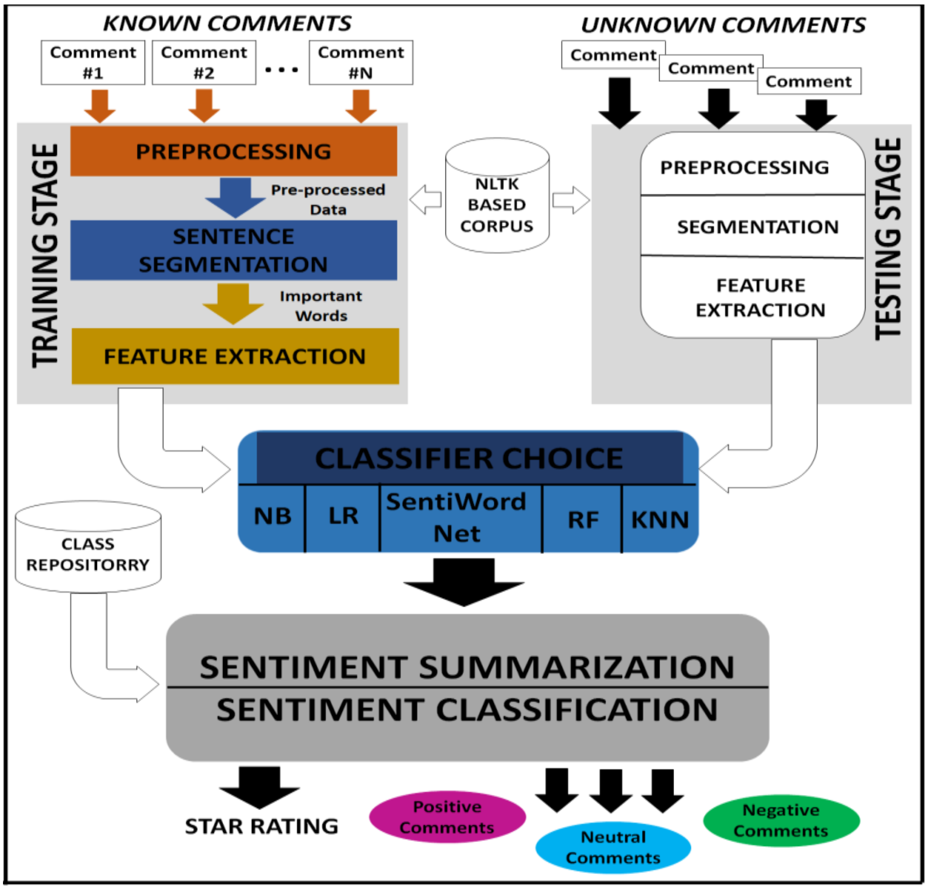

Fig.2. Detailed PCSA system Design

5.1 System Description

The proposed PCSA system follows a sequence of steps, and then summarizes and classifies the results efficiently. This section discusses the entire process of PCSA system in detail. These phases are user registration and login process, comment pre-processing, sentence segmentation feature extraction, training, comment summarization and classification.

5.1.1 Registration and Login

The system design has the homepage which is used to render the first page. It begins with the login details. If a user needs to get registered into the system and needs to create his/her own account. Then the registered user uses his/her login credentials to get access into it. Then the system asks the user to enter the number of URL links and the set of URL links of the online shopping website. In the browser, when a user enters a URL then it checks if the URL is present on the server side or not. If URL is not present, then the browser gives an error otherwise it redirects it for the comment collection. Therefore, the system gets both total number of links and product links in the array. It gets the number of links which the user wants to analyze, say, 1, 2, and 3 ... so on.

When the user writes the Flipkart website URLs, the system does the web crawling and finds the website. Further it shows the list of the products to the user which is available on the website. These websites include all types of products; say wearables, grocery items, household items etc. The user needs to select the products which should be processed by the system.

5.1.2 Comment Pre-Processing and Segmentation

The system collects all the comments of the product from the webpage one-by-one. During the comment pre-processing, it pre-processes all the comments. These comments are pre-processed to get the important words and to discard the unimportant words. These unimportant words consume the space unnecessarily and increase the size of the database. For this it also uses NLTK based corpus. It segments all the sentences of the comments and retrieves only textual information from them. The system removes all the stop words, performs the stemming through the Porter stemmer, and does the lemmatization. The extracted important words are sent for the further processing. These words are the adjectives, nouns, adverbs and verbs.

5.1.3 Feature Extraction and Training

After performing the pre-processing and segmentation, the system extracts the features from the comments. The product features are battery life, resistance, looks etc. These features are used to train all five classifiers. During training, the data gets trained. The NB, LR, RF and KNN algorithms are simpler implementation than SentiWordNet algorithm. The structure of SentiWordNet algorithm is found complex and needs more efforts. The training data are made fit in the pipeline and pass it the CSV file of the comments method of the comment analyzer.

5.1.4 Comment Summarization

In the comment summarization, the system first gets the positive, negative and neutral reviews and then provides the summary of them. On the basis of these summarized reviews, the system is trained to provide the star rating for the particular product. The system functions read the CSV file created by the web scraper and predict the rating using pipeline. For this, it finds the average of the comments in each of its three categories. In this way, it gets the average of the total number of positive, negative and neutral product comments, and then finds the rating of the product.

This system also takes the actual star rating of the products from both websites. Each website provides the rating to each of its product available on its website, which is called the actual rating of the product. The proposed PCSA system provides the predicted rating to each product. Further, the PCSA system compares both actual and predicted ratings of the product.

5.1.5 Comment Classification

Along with the prediction process, the PCSA system categorizes the comments into positive, negative and neutral categories. It makes use of the class repository for this. This method also calculates the precision, recall and F1_score of the each chosen product. These three operations are performed using four variables true_positive, true_negative, false_positive, and false_negative.

After all this, all three values are stored in the class variables and are passed in the review scraper of Amazon depending upon the URL link. The review scraper stores all three values along with the average rating, title, image, and original rating. The system finds the product reviews for the selected website, discriminates the comments into three classes, finds the average rating and separates the rating predicted by the pipeline. Lastly, it performs the operations to get the precision, recall and F1_score results.

6. Experimental Result and Discussion

To perform these experiments, a large set of comments are collected for ten product categories from Amazon. Total three-fourth product sub-categories were used for training stage and one-fourth sub-categories were kept for testing stage. These results are obtained for categorical reviews, summarization and rating, and review classification.

Table 1.Training and testing data from Flipkart

S.N.	**Flipkart**		
	Training Reviews	**Testing Reviews**	**Total Number of Reviews**
C-I	12675	6210	**18885**
C-II	273	597	**870**
C-III	256	3217	**3473**
C-IV	885	559	**1444**
C-V	13043	3718	**16761**
C-VI	1402	520	**1922**
C-VII	469	238	**707**
C-VIII	16616	212	**16828**
C-IX	44	30	**74**
C-X	2295	630	**2925**
Total Reviews	**47958**	**15931**	**63889**

The review data set in PCSA system is collected for ten categories, such as mobile phones, cameras, laptops, Air Conditioners (AC), routers, televisions, books, kitchen items, furniture, and clothes and wearable items. These categories are called as C-I to C-X respectively. Each of categories has three sub-categories for training stage and one sub-category for testing stage. It depicts the total number of Flipkart training reviews for each sub-category along with its sub-category number, brand and model number.

In PCSA system, the categories used to test the system are Samsung Galaxy A50 (White, 64 GB) - 4 GB RAM, Canon PowerShot SX430 IS (20 MP, 45x Optical Zoom, 4x Digital Zoom, Black), Apple MacBook Air Core i5, LG 1.5 Ton 5 Star Split Dual Inverter AC, D-Link DIR-615 Wireless N 300 Router, Sony Bravia X7002G 108cm, Object-Oriented Programming with C++, Kitchen Burner- Sunflame - Smart Stainless Steel Manual Gas Stove, Table-WFH DeckUp Versa Engineered Wood, and HONOR Band 5 Watch. The reviews for these ten categories were collected and called C-I to C-X respectively.

Table 2. Positive, negative and neutral reviews for Flipkart

S.N.	Category Number	Flipkart Polarity			
		+ve Reviews	**-ve Reviews**	**Neutral Reviews**	**Total Reviews**
1	C-I	**Positive Reviews**	**Negative Reviews**	**Neutral Reviews**	**Total Reviews**
2	C-II	28204	2679	167	**31050**
3	C-III	2528	398	59	**2985**
4	C-IV	13683	2318	84	**16085**
5	C-V	2339	404	52	**2795**
6	C-VI	16664	1755	171	**18590**
7	C-VII	2410	166	24	**2600**
8	C-VIII	950	214	26	**1190**
9	C-IX	803	233	24	**1060**
10	C-X	117	29	4	**150**
Total Polarity		3008	133	9	**3150**
		70706	**8329**	**620**	**79655**

Table 3. Product rating by PCSA and Flipkart along with the rating difference.

S.N.	Category Number	PCSA System Predicted Rating							
		NB	**LR**	**SentiWordNet**	**RF**	**KNN**	**Average Rating**	**Actual Rating from Flipkart**	**Rating Difference**
1	C-I	5	5	5	5	5	**5**	4.3	**-0.7**
2	C-II	4.9	5	4.8	5	5	**4.94**	4	**-0.94**
3	C-III	4.1	4.2	4.1	4.1	4.5	**4.2**	4.2	**0**
4	C-IV	4	4.1	4	4	4.2	**4.06**	4	**-0.06**
5	C-V	3.9	4	3.9	4.1	4	**3.98**	3.8	**-0.18**
6	C-VI	4.8	4.9	4.8	5	4.9	**4.88**	4.3	**-0.58**
7	C-VII	4.2	4.5	4.6	4.5	4	**4.36**	4.2	**-0.16**
8	C-VIII	4.1	4.1	4.3	4.4	3.8	**4.14**	3.9	**-0.24**
9	C-IX	4.5	4.5	4.4	4.4	4.5	**4.46**	5	**0.54**
10	C-X	4.8	4.8	5	4.4	4.4	**4.68**	4.2	**-0.48**
Final Total		**4.43**	**4.51**	**4.49**	**4.49**	**4.43**	**4.47**	**4.19**	**-0.28**

6.1 Results on Review Classification

The PCSA system performed the comment classification very efficiently and achieved high accuracy with overall testing dataset. This accuracy is computed by averaging the results of precision, recall and F1-Score.Table 4 depict the system classification results in terms of precision, recall and F1-Score for Amazon products, respectively. They also show the final total of precision, recall, F1-score and accuracy for Amazon. The PCSA system achieved total 0.942 precision, 0.8816 recall, 0.9104 F1-score and 0.91133 accuracy with Flipkart.

Table 4. Precision, recall and F1-Score results with all classifiers for Flipkart products.

Category Number	Naïve Bayes			Logistic Regression			SentiWordNet			Random Forest			K-Nearest Neighbor			Total Precision	Total Recall	Total F1-Measure	Total Accuracy
	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure	Precision	Recall	F1-Measure				
C-I	0.83	0.71	0.77	0.72	0.76	0.74	0.83	0.52	0.64	0.78	0.74	0.76	0.72	0.59	0.65	0.776	0.664	0.712	0.71733
C-II	0.96	0.78	0.86	0.92	0.86	0.89	0.96	0.81	0.88	0.96	0.86	0.91	0.96	0.83	0.89	0.952	0.828	0.886	0.88867
C-III	0.91	0.88	0.9	0.9	0.87	0.88	0.88	0.65	0.75	0.9	0.92	0.91	0.82	0.74	0.78	0.882	0.812	0.844	0.846
C-IV	0.98	1	1	1	0.98	0.97	1	1	1	0.96	0.98	1	1	0.97	1	0.988	0.986	0.994	0.98933
C-V	0.95	0.95	0.95	0.95	0.95	0.95	0.95	1	0.97	0.84	0.94	0.89	1	0.96	1	0.938	0.96	0.952	0.95
C-VI	1	0.9	0.95	1	0.9	0.95	0.94	0.89	0.92	1	0.9	0.95	0.9	0.98	0.95	0.968	0.914	0.944	0.942
C-VII	1	0.85	0.92	1	0.79	0.88	0.91	0.77	0.83	0.91	0.83	0.87	1	0.79	0.88	0.964	0.806	0.876	0.882
C-VIII	1	1	1	1	0.73	0.85	1	1	1	1	1	1	1	1	1	1	0.946	0.97	0.972
C-IX	1	0.93	0.97	1	0.93	0.97	0.93	0.93	0.93	0.93	0.93	0.93	1	0.93	0.97	0.972	0.93	0.954	0.952
C-X	1	0.95	0.97	1	0.95	0.97	1	1	1	1	0.95	0.97	0.9	1	0.95	0.98	0.97	0.972	0.974
Total	0.963	0.895	0.929	0.949	0.872	0.905	0.94	0.857	0.892	0.928	0.905	0.919	0.93	0.879	0.907	0.942	0.8816	0.9104	0.91133

7. Evaluation of Performance Metrics

The section 3 illustrated the detailed review and comparative analysis of many existing sentiment analysis and opinion classification algorithms for recent years. Such comparison discriminated their methodologies on the basis of five primary parameters, such as, feature reduction, sentiment polarity and orientation, data domains and sources, product types and classifiers. Their performance is measured and analyzed on the basis of these parameters and is shown graphically through subsections 5.1 to 5.4.

7.1 Based on feature evaluation and polarity

Figure 3 depicts the % usage of the various feature reduction techniques and the sentiment polarity and orientation. On one side, it is observed that most of the research works implemented the feature extraction method in their sentiment analysis model as their most preferred method. On the other side, these works designed bi-polar sentiment analysis systems as their primary choice. In both the cases, they contributed 80% with the feature extraction and also 80% with the positive and negative reviews. The second and remaining 20% contribution include the feature selection method in reduction techniques and the positive, negative and neutral review category in sentiment polarity.

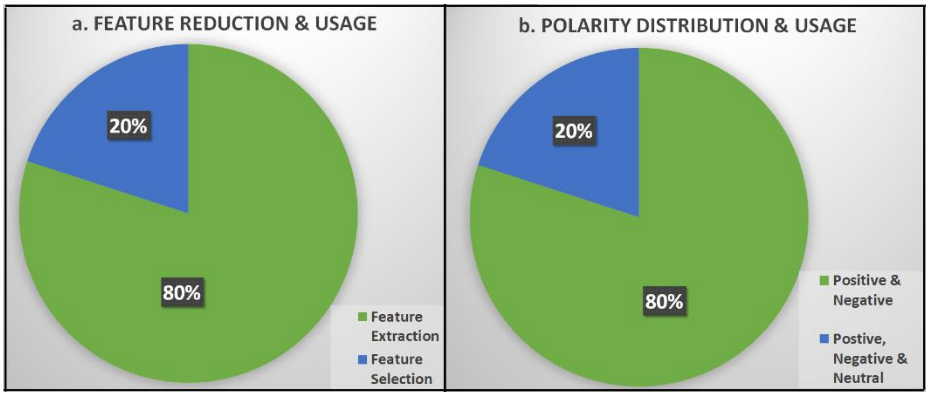

Fig. 3.% Usage of a. feature reduction techniques b. sentiment polarity categories in existing systems.

7.2 Based on data domain and sources

Figure 4 depicts the % usage of various data sources and domains in existing systems which were used to collect the different online product reviews and data sets. These systems worked with the online data sources and websites such as Amazon, Flipkart, Snapdeal, Epinions, Ebay, Twitter, Bestbuy, You Tube, and Customer review data. It is observed that most of these contributors used the Amazon, with 66.67% usage, as their primary source of review collection for online products. The range of % contributions varies from 6.67% to 33.33% for other sources and websites.

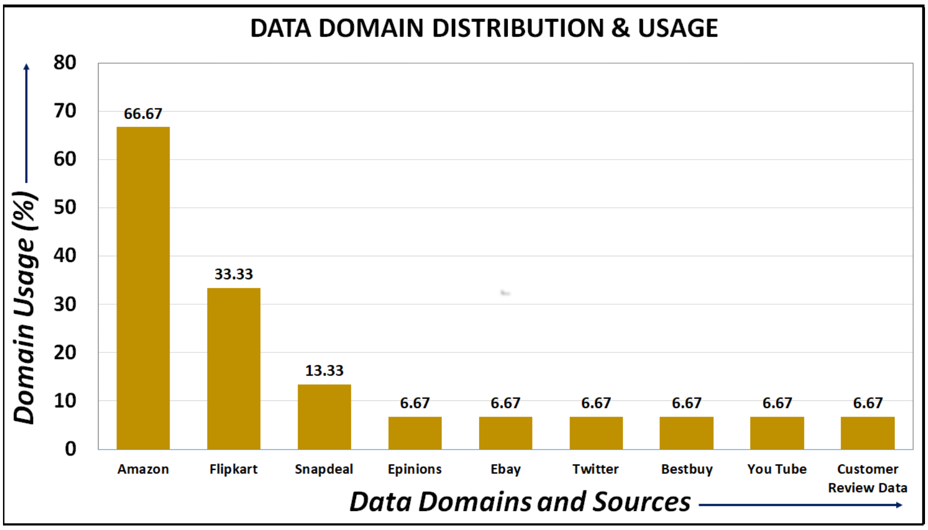

Fig. 4. % Usage of data domains and sources in existing systems.

7.3 Based on product type

Figure 5 depicts the % usage of various types of products under different data sources, domains and websites for the existing systems. These systems found the results using many products such as Mobile phone, DVD player, camera, MP3 player, router, antivirus, food items, movie, book, business related, sentence based, other electronics, and clothes & wearable's. It is observed that most of these contributors used the product, mobile phone, with 53.33% usage, as their first choice. The range of % contributions for other products varies from 20% to 6.67% as shown in Figure 5.

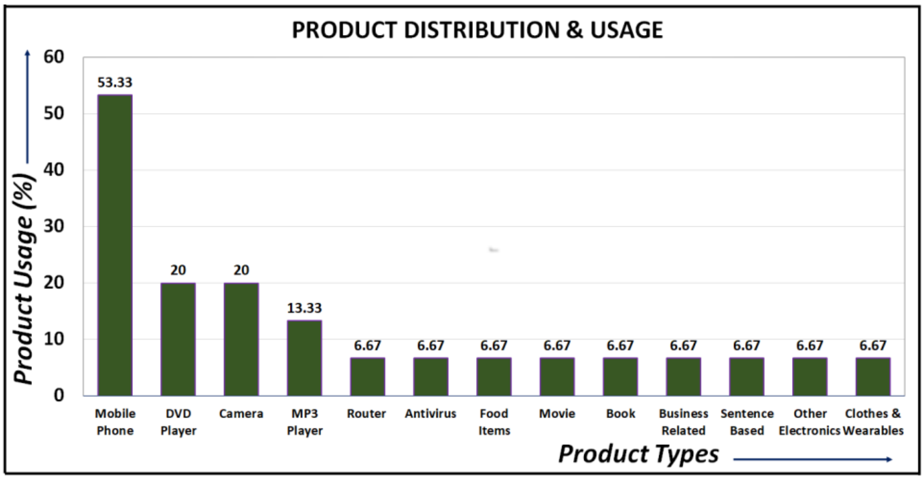

Fig. 5.% Usage of online available products in various sources.

7.4 Based on classifier used

Figure 6 depicts the % usage of various classifiers in the existing systems. These systems implemented their algorithms with many different single and hybrid classifiers such as SVM, RF, NB, ANN, DT, KNN, LR, gradient boosting, LDA, bagging, SVM-CNN, SGD and NB-SVM. It is observed that most of these contributors used SVM and RF classifiers, with 53.33% usage, as their first choice. The range of % contributions of other classifiers vary from 6.67% to 40% as shown in Figure 6.

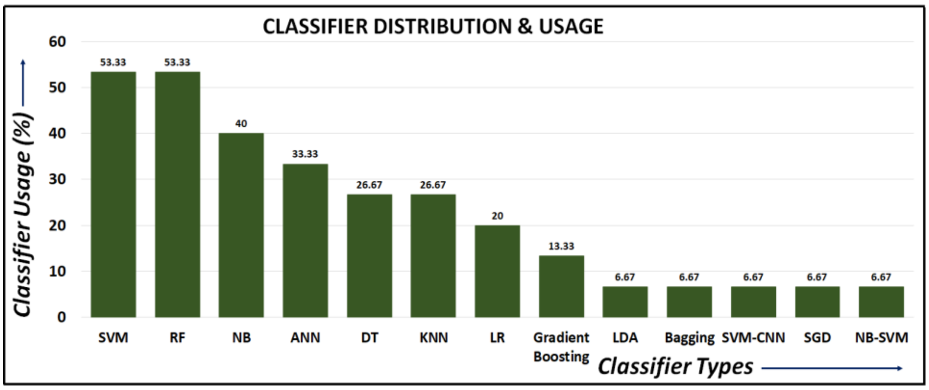

Fig. 6.% Usage of different classifiers involved in existing systems.

8. Conclusion and Future Scope

The proposed PCSA system summarized and classified the comments obtained from the data domains of Amazon. The system polarity was checked for positive, negative and neutral comments using supervised learning algorithms such as NB, LR, SentiWordNet, RF and KNN techniques. This system found the final inclination towards the positive comments. The detailed review and survey provided the strong base towards the need of the PCSA system using multiple classifiers. Future scope will be implementation of this system using the hybrid and multiple supervised learning techniques.

References

[1] Jiang, P., Zhang, C., Fu, H., Niu, Z., Yang, Q: An approach based on tree kernels for opinion mining of online product reviews. In: IEEE International Conference on Data Mining, pp. 256-265. IEEE Press (2010).

[2] Li, Z.: Product feature extraction with a combined approach. In: Third International Symposium on Intelligent Information Technology and Security Informatics, pp. 686-690. IEEE Press (2010).

[3] Adinarayana, S., Ilavarasan, E.: Classification techniques for sentiment discovery-a review. In: International Conference on Signal Processing, Communication, Power and Embedded System, pp. 396-400. IEEE Press (2016).

[4] Puri, S., Kaushik, S.: An enhanced fuzzy similarity based concept mining model for text classification using feature clustering. In: Students' Conference on Engineering and Systems, pp. 1-6. IEEE Press (2012).

[5] Kaur, G., Singla, A.: Sentimental analysis of Flipkart reviews using Naïve Bayes and decision tree algorithm. International Journal of Advanced Research in Computer Engineering & Technology 5(1), 149-153 (2016).

[6] Khan, J., Jeong, B. S.: Summarizing customer review based on product feature and opinion. In: Proceedings of the 2016 International Conference on Machine Learning and Cybernetics, pp. 158-165. IEEE Press (2016).

[7] Puri, S., Singh, S. P.: A technical study and analysis of text classification techniques in N-lingual documents. In: International Conference on Computer Communication and Informatics, pp. 1-6. IEEE Press (2016).

[8] Kaur, J., Bansal, M.: Hierarchical sentiment analysis model for automatic review classification for E-commerce users. In: Banati H., Bhattacharyya S., Mani A., Köppen M. (eds) Hybrid Intelligence for Social Networks, pp. 249-267. Springer, Cham (2017).

[9] Ejaz, A., Turabee, Z., Rahim, M., Khoja, S.: Opinion mining approaches on Amazon product reviews: a comparative study. In: International Conference on Information and Communication Technologies, pp. 173-179. IEEE Press (2017).

[10] Tan, W., Wang, X., Xu, X.: Sentiment analysis for Amazon reviews. In: International Conference, pp. 1-5. (2018).

[11] Diwakar, D., Kumar, R., Gour, B., Khan, A. U.: Proposed machine learning classifier algorithm for sentiment analysis. In: Sixteenth International Conference on Wireless and Optical Communication Networks. IEEE Press (2019).

[12] Sumedha, Johari, R.: SARPS: Sentiment analysis of review(s) posted on social network. In: Singh M., Gupta P., Tyagi V., Flusser J., Ören T., Kashyap R. (eds) Advances in Computing and Data Sciences. Communications in Computer and Information Science, vol. 1045, pp. 326-337. Springer, Singapore (2019).

[13] Suganya, E., Vijayarani, S.: Sentiment analysis for scraping of product reviews from multiple web pages using machine learning algorithms. In: Abraham A., Cherukuri A., Melin P., Gandhi N. (eds) Intelligent Systems Design and Applications. Advances in Intelligent Systems and Computing, vol. 941, pp. 677–685. Springer, Cham (2019).

[14] Karthika, P., Murugeswari, R., Manoranjithem, R.: Sentiment analysis of social media network using random forest algorithm. In: International Conference on Intelligent Techniques in Control, Optimization and Signal Processing, pp. 1-5. IEEE Press (2019).

[15] Shaheen, M., Awan, S. M., Hussain, N., Gondal, Z. A.: Sentiment analysis on mobile phone reviews using supervised learning techniques. International Journal of Modern Education and Computer Science 7, 32-43 (2019).

[16] Aziz, A. A., Starkey, A.: Predicting supervise machine learning performances for sentiment analysis using contextual-based approaches. IEEE Access 8, 17722-17733(2020).

[17] Khan, A. et. al.: Sentiment classification of user reviews using supervised learning techniques with comparative opinion mining perspective. In: Arai K., Kapoor S. (eds) Advances in Computer Vision. Advances in Intelligent Systems and Computing, vol. 944, pp. 23-29. Springer, Cham (2020).
[18] Puri, S., Singh, S. P.: Advanced applications on bilingual document analysis and processing systems. International Journal of Applied Metaheuristic Computing 11(4), 149-193 (2020).
[19] YOTPO Blog, https://www.yotpo.com/blog/opinion-mining/, last accessed 2020/12/1.
[20] Towards Data Science, https://towardsdatascience.com/%EF%B8%8F-sentiment-analysis-aspect-based-opinion-mining-72a75e8c8a6d, last acc0

Authors' Profiles

Anjali Dadhich received her master's degree in Computer Science from Banasthali Vidhyapeeth, Jaipur, India in 2014 and bachelor's degree from Mewar Girls College, Rajasthan, India in 2011.She is doing Ph.D. from JNU, Jaipur, and Rajasthan in the department of computer science. she published 10 papers in various international journals attended many conferences. She has very much interest to do research on Data mining and NLP.

Dr. Blessy Thankachan, received her Ph.D. in the field of Software engineering, University of Rajasthan, Jaipur, India. Currently she is Assistant Director in Jaipur National University, India. Her areas of specialization include Software engineering and optimization. She has published many research articles in International and National Journals and Conferences.

How to cite this paper: Anjali Dadhich, Blessy Thankachan, " Sentiment Analysis of Amazon Product Reviews Using Hybrid Rule-based Approach ", International Journal of Engineering and Manufacturing (IJEM), Vol.11, No.2, pp. 40-52, 2021. DOI: 10.5815/ijem.2021.02.04