Opinion Mining on Integrated Social Networks and E-Commerce Blog

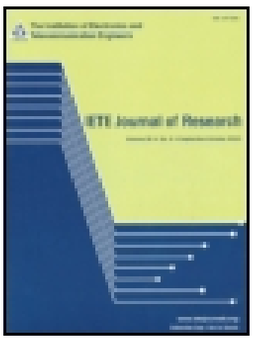
Article *in* IETE Journal of Research · February 2021

DOI: 10.1080/03772063.2021.1886603

2 authors:

S Uma Maheswari
Alagappa University
9 PUBLICATIONS **30** CITATIONS
SEE PROFILE

Dhenakaran S.S
Alagappa University
47 PUBLICATIONS **150** CITATIONS
SEE PROFILE



IETE Journal of Research

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tijr20

Opinion Mining on Integrated Social Networks and E-Commerce Blog

S. Uma Maheswari & S. S. Dhenakaran

To cite this article: S. Uma Maheswari & S. S. Dhenakaran (2021): Opinion Mining on Integrated Social Networks and E-Commerce Blog, IETE Journal of Research, DOI: 10.1080/03772063.2021.1886603

To link to this article: https://doi.org/10.1080/03772063.2021.1886603

Published online: 28 Feb 2021.

IETE JOURNAL OF RESEARCH
https://doi.org/10.1080/03772063.2021.1886603

Opinion Mining on Integrated Social Networks and E-Commerce Blog

S. Uma Maheswari and S. S. Dhenakaran

Department of Computer Science, Alagappa University, Karaikudi 630 003, India

ABSTRACT

The escalation of online shopping trends, social networks, and blogs makes shopping customers to know premier and quality product from customer reviews posted on social networks. Also, the business enterprise can perceive customers like opinion, demand, and expectation in real time. Customers are expressing their opinion, feedback, and emotions of online products through social networks and blogs. The problem focused here is, from customer point of view, identifying trendy and quality product for online purchasing, enterprise seeking customer's opinion, brand reputation based on the reviews in social networks and blogs. The main problem is, for a single product, customer's opinions are vast and vary in different social networks (Twitter, Facebook) and blogs (Amazon, Flipkart). This problem is addressed in this work for extracting useful hidden information by integrating the reviews of social networks and reviews in blogs. The proposed work on opinion mining has developed a new technique with a set of rules for combining social network reviews (Twitter reviews, Facebook reviews) and blog reviews (Amazon reviews). Then the reviews are mined to provide crisp information of reviewer opinion on products to new customers and organization. The new information will assist customer to make a wise decision for buying products. With the extracted information, the enterprises can read the customer's mind to improve their business by offering to customer trending products. The proposed model has achieved more than 80% accuracy and f1-score in sentiment prediction. The result has efficiently proved the performance of the model.

KEYWORDS

Decision support system; information extraction; opinion mining; prediction; reviews; tweets

1. INTRODUCTION

The Opinion Mining (OM) is the study of people's minds, demands, opinions, and emotions. The human language is more informal which leads to more problems for handling and manipulating opinion mining. Opinion mining voices out strongly about the feedback/opinion of customers towards an entity. People often express and shares their opinion of newly launched product via social networks and blogs. People reviews strongly confess whether the product is recommended or not recommended. For online shopping, customer always depends on the reviews of products in social networks and blogs to decide about the product for purchasing. Reviews may be vast and vary according to their budget, taste, area of living, etc. In different social networks and blogs, reviews of particular product vary.

The above significant factor is lacking in existing opinion mining. The lack of efficiency in opinion mining has led to develop a new approach for efficient opinion mining. The proposed approach has collected reviews, preprocessed the reviews, mined the resultant of preprocessing reviews, and classified for predicting the reviews. Due to the vast amount of reviews, distinct sources of social networks, blogs, and informal, unstructured, semi-structured, fusion data, Big Data tools and techniques are used for opinion mining. The mining part is also serviced retailers to take decisions on their business for marketing customer willing products.

Big Data analysis tools is used for storing raw data, result, processing, and mining customer Tweets. In this work, the varying opinion of the product "kindle" available in Twitter, Facebook, and Amazon is mined and then classified its reviews into positive and negative.

The remaining part of this work is organized into the following sections. Related work on opinion mining and sentiment analysis is presented in Section 2. In Section 3, proposed methodology for opinion mining is discussed. Section 4 has described the experimental results. Finally, Section 5 has stated the conclusion of this work.

2. RELATED SURVEY

The following research work is studied, analyzed to better understand opinion mining on social networks such as Twitter, Facebook, Amazon, and Big Data techniques used for opinion mining.

Ashwini Patil *et al.* [1] have proposed public sentiment variation on Facebook [15,16]. The authors have considered Tweets of "iPhone" products and "Kindle" of Facebook products over a period of time. The authors' analysis and classification have increased the percentage of positive or negative tweeting products by 50%. Kindle product on Facebook produced 63.16% accuracy [17,18]. Akhil Sharma *et al.* [2] have suggested, "Hybrid Classifier for Sentiment Analysis Using Effective Pipelining". Observation says Hybrid Classifier has produced a better performance of classification than other classifiers [19].

Hongyu *et al.* [3] have recommended, "Improving the performance of lexicon-based review sentiment analysis method by reducing additionally introduced sentiment bias". Here, the authors have experimented a novel sentiment bias strategy for the lexicon-based sentiment analysis [20]. Amazon dataset is collected and preprocessed for each word and score retrieved by SentiWordNet lexicon. Synset score (sentiment score) is calculated by poscore – negscore [21].

This literature survey has made clear that no research work has analyzed the public opinion by considering integrated reviews of the same product on Twitter, Facebook and Amazon websites. This problem is focused in this proposed work for collecting reviews of products on distinct social networks and blogs to study the mindset of customers and sellers. Also noted that very few research work only focused on the meaning of various wiring styles (love, loveeee, LOVE, veryyyy), emoji, slang words. And most research work have accomplished under opinion mining or sentiment analysis on existing supervised machine learning techniques, which needs training and testing process and also time-consuming. Few research works have developed a new unsupervised classification technique/model for opinion mining. These limitations are eliminated in the proposed model.

3. PROPOSED METHODOLOGY

The objective of the work is to develop a novel approach for opinion mining with Hadoop, Python and NLP to combine real-time product reviews of social networks and blogs related to online product buying.

This approach has performed deep insights on customer's opinions, expectations, demands for products and quality, in addition to measure the reputation of the enterprise.

Table 1: Dataset attributes

S. No.	Attributes	Count
1	Emoji	751
2	Negation	104
3	Slang word	290
4	Intensifier	Based on reviews
5	Modifier	92
6	Conjunction	4
7	AFINN	18,543
8	ReTweet	Based on Tweets
9	Likes	Based on Tweets
10	Overall rating	Based on Amazon reviews
11	Tweet's text review	13,000
12	Facebook's text review	1000
13	Amazon's text review	13,000

Proposed Work Model

Extracted tweets and reviews by Twitter API and tweepy library are stored in HDFS, and preprocessed for the removal of stop word, tokenization, stemming and POS-Tagged by available efficient libraries in Python language. Sentiment analysis techniques are applied on the result of preprocessing with emoji handling, negation words handling, slang words, intensifier words, modifier words handling, conjunction words handling, ReTweets & likes count, ratings count handling methods. Tweets and reviews are classified and predicted by the proposed technique. In MongoDB, the outcome of sentiment classification and prediction is stored.

Methods and Material of Proposed Work

Dataset Description

(1) **Twitter:** Nearly, 4000 Tweets on Kindle product collected in real time [5].
(2) **Facebook:** 1000 users Facebook posts of Amazon "Kindle" product collected [4].
(3) **Amazon:** Amazon data around 1.5 million reviews available freely. Here 4000 reviews involved [6,7,14].

The proposed system runs on Tweets, Facebook dataset, and Amazon dataset, which has covered around 10,000 reviews on Kindle product. The words of expressing satisfaction, interest, taste, dislike, and feedback are collected as attributes or variables to define the values in the database. The attributes of the dataset are defined in Table 1.

It is well known that the attribute is a small piece of information making sense to the entire comment of customers. In social network also, the customers express their views of online products by certain pieces of

Figure 1: Model of opinion mining on integrated reviews

information. In Amazon, Twitter, and Facebook, the customers make use of slang words or Emoji, Intensifier (attributes), etc. to represent their satisfaction or dissatisfaction on online products. This work has collected various such representations (attributes) to measure the value of customer Tweets. Hence the attributes are words of people commenting online products in social network. The work has considered all kinds of such words, called as attributes, expressing customer opinion on products.

Opinion Mining recognizing text or piece of information can be positive or negative. Opinion Mining being assistance for buyers helps to decide product purchase through online. OM can tell whether the particular product recommended or not recommended by previous buyers. OM assists also organizations and enterprises to realize feedback and demand about the newly launched products (Figure 1).

3.1 Extracting Tweets, Facebook reviews, Amazon reviews, and Storing

Tweets

Real-time tweets are obtained by python library Tweepy and Twitter API. Tweets may contain many attributes, for this work, the extracted attributes are

- Tweet id
- Tweet date
- Text
- Retweet
- Likes

Facebook

In order to be confidential and for security reason, Facebook restricts to access and retrieve real-time data.

Due to this problem/reason, freely available Facebook reviews about the "Kindle" product are used in this work. From this Facebook review, Text content only is retrieved for the work.

Amazon

In Amazon.com, million numbers of product reviews are freely available. For collecting variety and vast amount of reviews, Big Data Hadoop Distributed File System is used. The Amazon.com reviews also have many attributes. For the proposed work, the extracted attributes are

- Overall Ratings
- Text.

3.2 Data Preprocessing

In order to mine the reviews, it needs meaningful contents. Due to this necessity, the reviews are preprocessed, for eliminating irrelevant data, analyzing the meaningful content. The preprocess methods are

- Special symbol removal, Spelling corrections, URL removal, HashTag removal
- Stop words removal, User mentions removal
- Tokenization and Stemming.

3.2.1 POS – Tagging

Part Of Speech – Tagging is grammatical words labeling/tagging. POS – Tag method is applied to tokenize sentences. POS – Tag can find phrases, adjacent words, and grammar relations presented in sentence/reviews. For the purpose of this work, POS – Tag retrieves the presence of Noun, Adjective, Verb, and Adverb from the sentence/review.

3.2.2 Word List Dictionary

Fill For the intention of performing opinion mining, it is required opinion/sentiment words with positive and negative words with associated polarity/sentiment score. To compute accurate sentiment score of the review, the proposed work has focused on customer review writing manner. Customer's writing manner often informal but distinguishes meaning to every writing. Example: Loveeeee, WOW, OMG, WORST, WONDERFUL, nice!!!..., Very much, all these word expressions are different.

This kind of attitude of writing not only produces the sentiment word, special symbols, upper or lower case but also produces more crucial on the reviews. These words can impact on sentiment score computation. These types of words are named Intensifier, Modifier, and Degree Modifier.

This proposed approach has created different types of dictionaries *viz.* Emoji dictionary, slang word dictionary, Stop words list, Negation words list, and modifier word list (very, most, ugly, truly) with Python programming. Also, exclamatory symbols, words with repeated characters like "loveeee" and capitalized words like "FAMOUS" are represented as intensifiers [12].

3.2.3 Lexicon Method

Emoji is identified and transformed into Unicode [8]. Each Unicode is depicted with its polarity scores to build a dictionary. Emoji score span from -1 to $+1$. With the support of the lexicon method, each word from the Reviews/Tweets scrutinizes matching with AFINN word. If the word matches with AFINN word dictionary, then the associated polarity score is retrieved.

3.2.4 Negation Words

To manipulate negation words, the polarity shifting method is employed [9]. Action of identifying the presence of slang word recognizes slang word with an associated elongated word in the dictionary, and then elongate slang word using the slang word dictionary [10].

3.2.5 Modifier, Degree Modifier, Intensifier, Conjunction Words

Process of recognizing the existence of Modifier, Intensifier, and Degree Modifier words spots the mentioned words with correlated polarity score and fetched from the AFINN dictionary. Then polarity score is doubled for intensifier and degree modifier being this category of words are much expressive than the ordinary words. Modifier words fixed with score 5 [12]. Dataset may existence with some Reviews/Tweets with two or more sentences. These sentences of Reviews/Tweets are concatenated by the words such as "but while, although". These types of words much influence Sentiment Analysis, so these words are associated with fixed scores 4.

3.3 Twitter Tweets Handling

The proposed approach has included ReTweet and likes count retrieved from the tweet [11]. When the positive tweet, Likes & ReTweet count increase then Likes & ReTweet polarity score are also increased. When Likes & ReTweet count increases the negative tweets, then Likes & ReTweet polarity score are decreased.

For each tweet/review, based on the generated rules, all the emoji's score, negation word's score, modifier word's score, intensifier word's score, degree modifier word's score, concatenation word's score are computed and aggregated to find the total score of a review/tweet.

3.4 Facebook Reviews Handling

From the Facebook review, sentiment words are retrieved, and polarity scores are computed for each review.

3.5 Amazon Reviews Handling

Overall ratings and text reviews being crucial, the meaningful attributes are retrieved from Amazon Reviews. Then sentiment words are retrieved, analyzed, and polarity scores are computed. To compute the total sentiment score of a review, polarity score, and overall rating score are used.

Proposed Technique
Start
Step 1: Collect the Corpus of Tweet & Reviews (CTR)
Step 2: Preprocessing
Input: Corpus of Tweets & Reviews
Process: Removal of Hash Tag, Special Symbols, Spelling Correction, Stemming and Tokenization.
Output: Tokens of Words
Step 3: POS – Tagging
POS = Find_PosTag (Tokens of Words)
Step 4: Sentiment Analysis
Input: Corpus of Tweets & Reviews (CTR)
Notation: EU: Emoji Unicode, SW: Slang Word, NW: Negation Word, IW: Intensifier Word, MW: Modifier Word, CW: Conjunction Word, LC: Likes Count, RTC: ReTweets Count, TPS: Total Polarity Score, TS: Tweet's Score.
Output: Classify the Each Review Ti as Positive and Negative.
Begin: For each T_i in CTR do
EU = Find_Emoji's Unicode (EU)
If EU ∈ EU-List then
S (EU) = Score (EU)
If SW ∈ SW-List then
S (SW) = Score (SW)
If NW ∈ NW-List then
S (NW) = Score (NW)
If IW ∈ IW-List then
S (IW) = Score (IW)∗2

If MW ∈ MW-List then
S (MW) = Score (MW)∗2
If CW ∈ EU-List then
S (EU) = Score (EU)
LC = Find Likes_Count
RTC = Find ReTweets_count
Find Total Polarity Score for Each Tweet Ti

$$TPS = \sum \begin{matrix} S(EU) + S(SW) + S(NW) + S(IW) \\ +S(MW) + S(CW) + LC + RT \end{matrix}$$

Find the sentiment score of Tweets
TS = TPS/ ∑(Words)
Step: 5 Classify the Tweets/Reviews
If TS > 0 then TS & Ti ∈ Positive
Else If TS < = 0 then TS & Ti ∈ Negative
Stop

3.5.1 Opinion/Sentiment Classification

The Opinion/Sentiment classification is the process of seeking customer information in tweets and reviews whether sentiment is positive or negative or neutral.

Generally classification is carried out by the machine learning techniques. But in the proposed technique, automated classifier is built for classification and then categorizes the positive and negative opinionated reviews separately for further processing. In the process of opinion mining, the main task is to build a new classifier technique for opinion/sentiment reviews classification. A new efficient technique is built for the classification and prediction of opinionated reviews after sanitizing the unfavorable words from the collected reviews. Further, it is analyzed to find out only most impacting and meaningful words to calculate the sentiment scores for each Tweet and Review of Facebook, and Amazon dataset. Tweets and Reviews are classified as positive and negative. Finally the sentiment score of each Tweet/Review is classified based on the following rules.

- If Sentiment score of the Tweet or Review is > 0, then Tweets/Reviews Positive
- If Sentiment score of the Tweet or Review is ≤ 0, then Tweets/Reviews Negative.

3.5.2 Opinion/Sentiment Prediction

The Opinion/Sentiment Prediction is the process of defining futuristic classification from the past observation of customer's reviews. The proposed classifier model is designed to predict unseen data, whereas customer opinions predicted based on the rules and Threshold values.

Here, it is forecasting reviews when the existing classified reviews and newly predicted reviews match then the prediction is correct else prediction wrong. The proposed approach thinks of all Tweets and Reviews of Amazon and Facebook Tweets which are integrated as a single dataset. This integrated review dataset consists of classified Tweets & Reviews with associated sentiment score and opinion/sentiment (+1 for +ve/−1 for -ve) of each Review/Tweet. Then all the Review's sentiment scores are aggregated to calculate sentiment score. Based on the rules and Threshold value, sentiment of Tweets & Reviews is predicted. If the sentiment score is greater or equal to the threshold value, then the review is predicted as positive else if sentiment score is less than the threshold value, then the review is predicted as negative.

4. EXPERIMENTAL ANALYSIS

An efficient technique is built in the process of opinion mining, tested on vast and different volume of dataset. The Facebook dataset has a constant volume for all test cases which has covered nearly 1000 reviews.

4.1 Experiment – I on Tweets, Amazon, Facebook

This proposed method is experimented by classification algorithm considering Tweets, Amazon reviews, and Facebook posts of "Kindle" product with distinct volumes of reviews separately. The result is tabulated the sentiment/opinion of customers about the "Kindle" product.

The tweets/reviews/posts not holding sensational words are dropped by experimentation is shown in Table 2. So out of Total Reviews, the preprocessing process has generated reviews with sensational words. When comparing experimental results on Tweets, the results on Amazon Reviews, and results on Facebook Reviews, all furnished the highest volume of positive data.

This extracted knowledge has commented that the "Kindle" product from Amazon is much recommended by the online customers to purchase (Figure 2).

Table 2: Experimental results on kindle Tweets, Amazon, and Facebook reviews

S. No.	Sources	Tweets	Amazon	Facebook
1	Volume	4000	3000	1000
2	Preprocessed reviews	3294	2289	747
3	Positive reviews	1777	2253	565
4	Negative reviews	1515	36	182
5	% of positive reviews	53.95	98.43	75.64
6	% of negative reviews	45.99	1.57	24.36

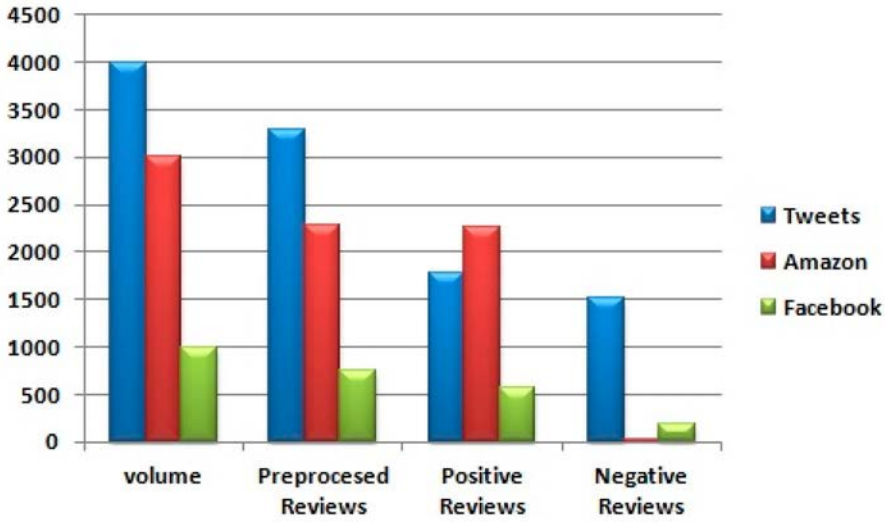


Figure 2: Reviews of Tweets, Amazon reviews, Facebook posts

Table 3: Experimental results of customer comments on Kindle Product

S. No.	Tweets	Amazon reviews	Facebook reviews	Total tweets/reviews	Positive tweets/reviews	Negative tweets/reviews	% of positive tweets/reviews	% of negative tweets/reviews
1	4000	4000	1000	7034	5274	1760	74.98	25.02
2	4000	3000	1000	6369	4765	1604	74.82	25.18
3	4000	2000	1000	5714	4401	1313	77.02	22.98
4	4000	1000	1000	4967	4013	954	80.80	22.75

The main aim of the proposed research is to integrate the Tweets, Amazon Reviews, and Facebook Comments. This process of integration is carried out and tested on the distinct volume of reviews about the "Kindle" product. The experimental result has produced overall sentiment/opinion of customers on the "Kindle" product.

4.2 Experiment – II on Tweets + Amazon Reviews + Facebook Posts

Table 3 shows the experimental results on integrated reviews of Tweets, Amazon Review, and Facebook Posts. Clearly, the positive reviews hold the highest volume of data. Opinion mining on integrated reviews states the "Kindle" product is highly recommended by the customers to purchase (Figure 3).

Performance Analysis

The proposed classification algorithm's is efficiently evaluated by confusion matrix and compared the result of matrices, Accuracy, Precision, and F1-Score with existing machine learning technique namely Naïve Bayes, Support Vector Machine and Decision Tree Algorithms.

Confusion Matrix

The confusion matrix has provided a summary of algorithm's performance result. Accuracy, Precision, Recall, and F1-Score have been calculated by the confusion matrix. Generally, the confusion matrix has actual values that has been used for information extraction and also has the predicted value from this analysis [13].

The results of Table 4 make clear and prove that the proposed technique's Accuracy is 91.57% which has only 0.3% difference in accuracy with the SVM technique but the accuracy is better than NB and DT techniques. The proposed technique's F1-Score is 94.04% which is closed to F1 score of other methods. When comparing the overall performance of Accuracy, Precision, and F1-Score, the proposed technique is better than the existing machine learning technique. Proposed technique's performance matrices have obtained above 90% value. The accuracy in 100%–90% classification is excellent, 80%–90% classification is good, 70%–80% classification is fair, 60%–70% classification is poor and if 50%–60% then the classification is failure [21,22]. Being this proposed technique is produced above 90% result, the good and reasonable performance it proves proposed classification algorithm outstand performing (Figure 4).

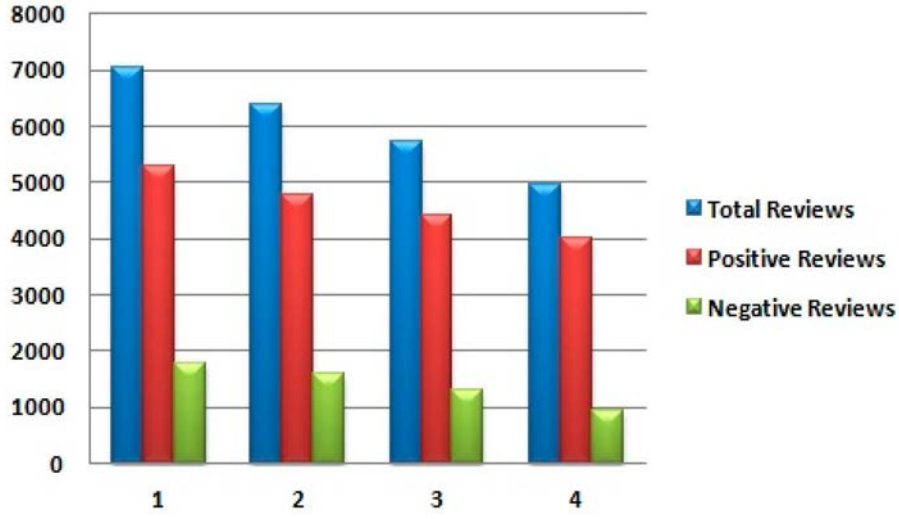


Figure 3: Integrated reviews on Kindle product

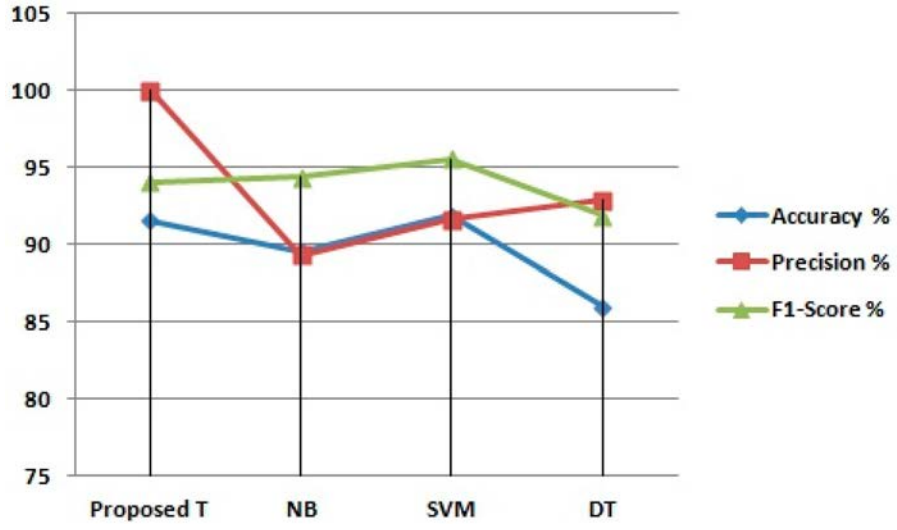


Figure 4: Performance analysis

Table 4: Performance comparison

Techniques	Accuracy %	Precision %	F1-Score %
Proposed	91.57	100	94.04
NB	89.57	89.32	94.36
SVM	91.85	91.65	95.53
DT	85.98	92.88	91.88

5. CONCLUSION

In this proposed work, customer opinions of Amazon "Kindle" product on social networks and E-Commerce Blog are collected and kept as a key dataset. To explore real opinion of customers, Amazon product Tweets, Amazon Reviews, and face book Reviews are integrated. The proposed technique has performed varying public sentiment/opinion mining, classification, and prediction. The results have proved that the unsupervised technique has performed well and reasonable than existing supervised machine learning techniques. This new technique has generated concise information/message to new buyers whether the intended product is recommended or not recommended by the previous customers. Also suggestion is provided to business enterprise/organization to be aware of the customer's demand.

ACKNOWLEDGMENT

This Proposed Research Work "**Opinion Mining on Integrated Social Networks and E-Commerce Blog**", has been done with the financial support of RUSA – Phase 2.0 grant sanctioned vide Letter No. F. 24-51 / 2014-U, Policy (TNMulti-Gen), Dept. of Edn. Govt. of India, Dt. 09.10.2018.

FUNDING

This Proposed Research Work "Opinion Mining on Integrated Social Networks and E-Commerce Blog", has been done with

the financial support of Rashtriya Uchchatar Shiksha Abhiyan (RUSA - Phase 2.0) grant sanctioned vide Letter No. F. 24-51 / 2014-U, Policy (TNMulti-Gen), Dept. of Edn. Govt. of India, Dt.09.10.2018.

ORCID

S. Uma Maheswari http://orcid.org/0000-0001-7650-2240

REFERENCES

1. A. Patil, and S. Thakare, "Analyzing public sentiment variations on and Facebook," *Int. J. Adv. Res. Comput. Sci. Technol.*, Vol. 5, no. 2, pp. 30–3, April–June. 2017.

2. A. Sharma, A. Sharma, R. K. Singh, and M. D. Upadhayay, "Hybrid classifier for sentiment analysis using effective pipelining," *Int. Res. J. Eng. Technol.*, Vol. 4, no. 8, pp. 2276–81, August 2017.

3. H. Han, Y. Zhang, J. Zhang, J. Yang, and X. Zou, "Improving the performance of lexicon-based review sentiment analysis method by reducing additionally introduced sentiment bias," *Plos One*, Vol. 13, February 2018.

4. K. Zhang, Y. Yu Cheng, A. Agrawal, D. Palsetia, K. Lee, and A. Choudhary, "SES: sentiment elicitation system for social media data," *Icdmsentire*, pp. 129–36, 2011.

5. ttp://www..com Last accessed on 12 November 2019.

6. http://cucis.ece.northwestern.edu/projects/Social Last accessed on 14 September 2019.

7. http://jmcauley.ucsd.edu/data/Amazon/ Last accessed on 22 November 10 October 2019.

8. A. Tupsoundarya, and P. S. Dandannavar, "Sentiment expression via emoticons on social media," *Int. J. Res. Appl. Sci. Eng. Technol.*, Vol. 6, no. 6, pp. 2404–8, June 2018.

9. W. Sharif, N. A. Samsudin, M. M. Deris, R. Naseem, and M. F. Mushtaq, "Effect of negation in sentiment analysis," *Int. J. Comput. Ling. Res.*, Vol. 8, pp. 47–56, June 2017.

10. L. Wu, F. Morstatter, and H. Liu. "SlangSD: building and using a sentiment dictionary of slang words for short-text sentiment classification," *arXiv: 1608.05129v1* [cs.CL] 17 August 2016.

11. R. S. Perdana, and A. Pinandito, "Combining likes-retweet analysis and naive Bayes classifier within twitter for sentiment analysis," *J. Telecommun., Electron. Comput. Eng.*, Vol. 10, no. 1–8, pp. 41–6, April 2018.

12. S. Uma Maheswari, and S. S. Dhenakaran, "Sentiment analysis on social media big data with multiple tweet words," *Int. J. Innovative Technol. Exploring Eng.*, Vol. 8, no. 10, pp. 2514–8, August 2019.

13. A. Sharma, and S. Dey, "Artificial neural network based approach for sentiment analysis of opinionated text," *ACM Transaction*, pp. 37–42, 2012.

14. www.Amazon.in Last accessed on 22 August 2019.

15. P. Turney, "Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews," *Proc. Assoc.r Comput. Ling. (ACL)*, 417–24, 2002.

16. B. Pang, and L. Lee, "Opinion mining and sentiment analysis," *Found. Trends Inf. Retrieval*, Vol. 2, no. 1–2, pp. 1–135, 2008.

17. S. Tan, Y. Li, H. Sun, Z. Guan, X. Yan, J. Bu, C. Chen, and X. He, "Interpreting the public sentiment variations on twitter," *IEEE Trans. Knowl. Data Eng.*, Vol. 26, no. 5, pp. 1158–70, May 2014.

18. P. B. Filho, L. Avanço, T. Pardo, and M. D. G. V. Nunes, "NILC_USP: an improved hybrid system for sentiment analysis in Twitter messages," in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval2014)*, 2014.

19. P. Gonçalves, M. Araújo, F. Benevenuto, and M. Cha. Comparing and combining sentiment analysis methods. 2014.

20. M. Iqbal, A. Karim, and F. Kamiran, "Bias-aware lexicon-based sentiment analysis," in *ACM Symposium on Applied Computing*, 2015, pp. 845–50.

21. N. A. Vidya, M. I. Fanany, and I. Budi, "Twitter sentiment to analyze net brand reputation of mobile phone providers," *Procedia. Comput. Sci.*, Vol. 72, pp. 519–26, 2015.

22. F. Gorunescu. *Data mining concepts, models and techniques, BV. Wowk, machinery vibration, measurement and analysis.* New York: McGraw-Hill, 1991.

AUTHORS

S. Uma Maheswari is presently pursuing research in the Department of Computer Science, Alagappa University, Karaikudi 630 003, Tamil Nadu, India. Her areas of interest are Opinion Mining, Big Data Analytics and Sentiment Analysis.

Corresponding author. E-mail: 17umeshrani@gmail.com

S.S. Dhenakaran, Professor, is currently working in the Department of Computer Science, Alagappa University, India. He has rich experience in teaching and research with dual PG degrees *viz.* Mathematics and Computing and Doctoral degree in computer Science. His main areas of research are Information Security, Data Mining and Mathematical Algorithms.

E-mail: ssdarvind@yahoo.com