Bulletin of Electrical Engineering and Informatics
Vol. 12, No. 2, April 2023, pp. 1209~1218
ISSN: 2302-9285, DOI: 10.11591/eei.v12i2.4682

Reduction of false negatives in multi-class sentiment analysis

Chris Aloysius, P. Tamil Selvan
Department of Computer Science, Karpagam Academy of Higher Education, Coimbatore, India

Article Info

Article history:

Received Aug 31, 2022
Revised Nov 2, 2022
Accepted Nov 17, 2022

Keywords:

Lexicon model
Machine learning
Multiclass analysis
Sentiment analysis
SVM classifier

ABSTRACT

Sentiment analysis classifications are done as positive, negative, as well as neutral ones. The increased usage of social media and its effects on society call for a more thorough, fine-grained explanation than that. In this study, classification is done in five classes-strongly positive, weakly positive, neutral, weakly negative, and strongly negative-in a more precise manner. Instead of using the typical ways of measuring accuracy alone, a novel method to eliminate false negatives (FN) is focused together with a fine-grained categorization. A bigger risk in sentiment analysis is a false negative. FN classification occurs when the context's polarity is identified as True when it is actually false. A complex dataset is used in this research for the experimental study, and the entire dataset is separated into five classes. Each class's FN are assessed using the suggested methodology. Comparing the proposed strategy to other, it was found to achieve about 53% more reduction in FN cases than rule based models and better predictions than compared machine learning models.

Corresponding Author:

Chris Aloysius
Department of Computer Science, Karpagam Academy of Higher Education
Coimbatore, Tamil Nadu, India
Email: chrissnowdon27@gmail.com

1. INTRODUCTION

Due to social media platforms, a significant volume of unstructured data is produced on the internet every second. To understand human psychology, the data must be analyzed as quickly as it is generated . Sentiment analysis, which detects polarity in texts, can help with this [1]. Sentiments are divided into four categories in sentiment analysis: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) [2]. All sections of the community began to be impacted by social media. This necessitates a more in-depth examination of social media content. Despite the categorization of sentiments into good and negative, a more detailed analysis looking at the specifics of feelings needs to be done [3]. "Details are where the devil always lurks." Fine-grained analysis should be used if a more accurate result is required. It not only makes it possible to evaluate comments on the service, but it also significantly aids in determining which component is being discussed. Fine-grained analysis can also be used to determine the intensity of feelings. Instead of labelling situations as positive or negative, it is possible to classify them as strongly positive, weakly positive, neutral, strongly negative, or weakly negative. Another tiresome effort is the detection of FN in sentiment analysis. A type II mistake is what this is also known as. The final outcome of sentiment analysis prediction is typically evaluated using matrices like accuracy, precision, and recall. However, in all of these evaluations, TP are given more weightage than FP, FN, or TN [3], [4]. It is important to give proper consideration to lower FN or type II mistakes. The classification of sentiments into positive and negative cases was the only prime focus of the early studies in sentiment analysis [4]. Later, the research focused on measuring and comparing the accuracy of predictions made using several methods for feature extraction, including n-gram analysis, tf/idf, the bag of words approach, and many rule-based, machine learning and deep learning algorithms for classification [5].

In line with [6], it's predicted that there will be over 2 billion digital purchasers worldwide, that by 2021, global retail e-commerce sales would exceed $4.88 billion, and that by 2040, 95% of all purchases will be made online and 85% of consumers will have done their research online. This indicates that sentiment analysis needs to be more credible. The majority of sentiment analysis research studies used the opinion-lexicon method to evaluate text sentiment in social media. Data were taken from microblogging sites, mostly Twitter, and applications of sentiment analysis may be found in business, politics, healthcare, and the economy [7]. Sentiment analysis in clinical medicine has improved recently as more and more researchers are doing studies with the aid of this beneficial technology after realizing its potential to benefit the field [8]. Sentiment analysis is being used in politics due to the rise in user political engagement in online social networks. Bouazizi and Ohtsuki in [9] examined the problems and various issues associated with multi-classification, where they also presented measures to gauge the disparity across feelings. Even though multi-class analysis is crucial, it was determined that doing a sentiment detection job, in which all of the sentiments present in a text are collected, might be more engaging.

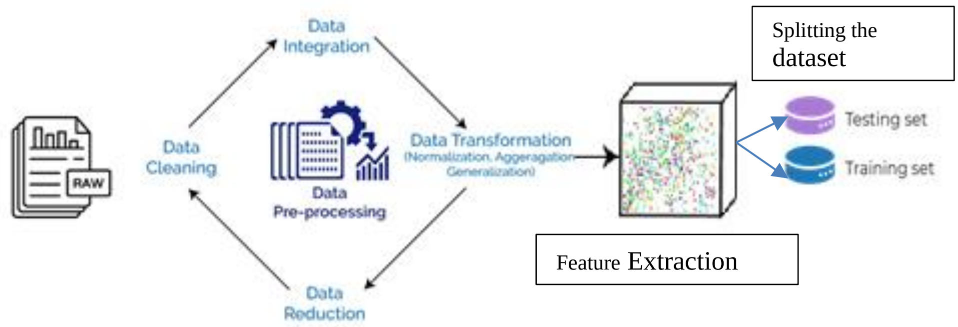
Research by Barbounaki *et al*. [10] conducted a thorough analysis of the various sentiment analysis methods. Advanced research techniques and many sorts of data are introduced, along with some potential restrictions. Despite the expanding significance of sentiment analysis, this research focuses on the area where previous efforts have not been organised in a clear and systematic manner. This study, on the one hand, focuses on outlining standard methodologies in the field of sentiment analysis from three different angles: task, granularity, and method-oriented. The authors reviewed the benchmark datasets currently used in this field and addressed potential future research avenues for multimedia sentiment analysis in [11]. This study analyzed 100 papers from 2008 to 2018 and divided the studies into several categories based on the methodologies they used. The result of prediction accuracy alone was used to compare various models.

According to Yue *et al*. [12], used Valence Aware Dictionary and sEntiment Reasoner for sentiment analysis to identify the overall attitudes and feelings observed in the dataset. Latent dirichlet allocation was also used for topic modelling to infer the various themes of conversation. To solve the sentiment analysis challenge, the authors of [13] suggested a hybrid model called the hybrid convolutional neural network-long short-term memory (CNN-LSTM) model that combines LSTM and an extremely deep CNN model. They trained the initial word embedding using the word to vector (Word2Vc) technique. The findings demonstrated that in terms of precision, recall, f-measure, and accuracy, the suggested hybrid CNN-LSTM model performed better than conventional deep learning and machine learning techniques. By forecasting a real-valued score between -1 and -+ 1, a fine-grained supervised technique is proposed in [14] to identify bullish and bearish attitudes linked with businesses and equities. A suggested supervised learning method uses many feature sets, including lexical, semantic, and a combination of lexical and semantic information. According to Dahal *et al*. [15], employed support vector machine (SVM) classifiers with information gain (IG) as a filter features selection strategy with WOA to condense the search space searched by WOA. The results of the thorough trials shown that the proposed algorithm performed better than all other algorithms in terms of sentiment analysis categorization accuracy

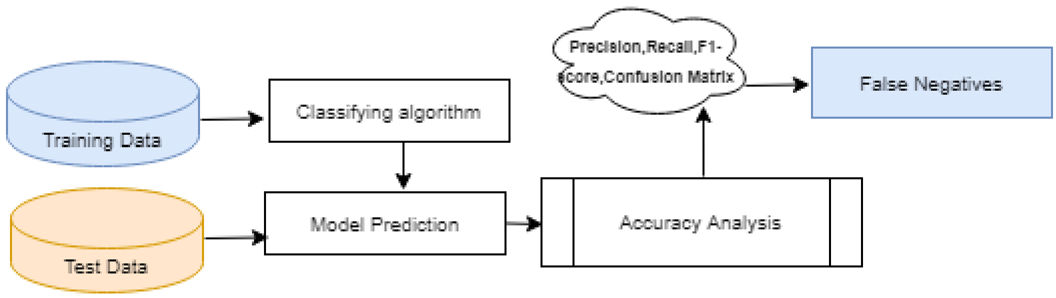
It is clear from looking at the literature reviews done in the area of sentiment analysis that the effort is only concentrated on accuracy. It also takes into account the accuracy of positive, negative, and to some extent, neutral cases. However, no study examined how to lower FN in a sentiment analysis model with finer granularity. The focus should be on developing a more fine-grained model that minimizes FN without sacrificing accuracy as the requirement for in-depth sentiment analysis becomes more and more important. Equal weight should be given to the accuracy score and classifying a true instance as negative at the minority level. By doing a fine-grained analysis, this work focuses on reducing FN in sentiment analysis-related sectors, which can be useful for many real-world applications. The employment of multilingual categorization models and algorithms that can identify language polarity more accurately is suggested. Additionally, a classification method utilizing a polarity-based fine-grained sentiment analysis model is proposed, experimented and analyzed.

2. METHOD

A five-stage fine-grained analysis approach is used to classify and categorize incoming data. Figures 1(a) and (b) flowcharts provide a full breakdown of the process's stages and how they are implemented. A data set of movie reviews that had been manually classified into 5 different sentiment groups was used to conduct the proposed research. The data is cleaned in order to extract the features. The data is cleaned up using natural language processing (NLP) algorithms. A 5 class classification is done in the data set of the proposed system with label 1 as strongly negative, label 2 as weakly negative, label 0 as neutral, label 4 as weakly positive and label 5 as strongly positive, after extraneous information like URLs and email addresses have been removed.



(a)



(b)

Figure 1. Proposed methodology (a) feature extraction and (b) classification and accuracy prediction

2.1. Data collection

For evaluation, the dataset of movie reviews that is divided into five classes manually has been used. Table 1 contains an example of the collected raw data. With the collected corpus a detailed compositional analysis to forecast the sentiment, whether a sentence or review is strongly negative, weakly negative, neutral, weakly positive, or strongly positive is determined. The total number of samples in each class corresponds to the graph in Figure 2. The dataset is made up of around 8,543 samples, which are divided into 5 labels-highly positive to strongly negative-ranging from 5 to 1. 1,624 neutral, 1,092 strongly negative, 2,218 weakly negative, 2,321 weakly positive, and 1,288 strongly positive samples make up the complete corpus.

Table 1. Sample raw dataset

Polarity_Label	Reviews
label__4	An excellent independent film that might use additional trimmings and more chemistry between its leads.
label__2	Never would I have imagined saying this, but I'd much rather see teenagers sticking their genitalia in fruit pies!
label__1	Ham-fistedly performed and tediously derivative.
label__3	Your desire for canned corn strongly influences how full of heaven you feel.

2.2. Preparing data

Following the raw dataset's retrieval, it was prepared for data preparation by applying the first formatting, which resulted in an index and appropriately labelled attitudes for the dataset's use in subsequent processing. Table 2 provides an example snapshot of the dataset following the first formatting. The data is cleaned using techniques for natural language processing [16]. The data are cleaned up by removing duplicate and unnecessary information. The second stage was to correct the structural data issues that might have otherwise compromised classification accuracy. To make the data more process-ready, unwanted outliers are removed from it. Algorithm 1 outline the procedures to be utilised for cleaning the data.

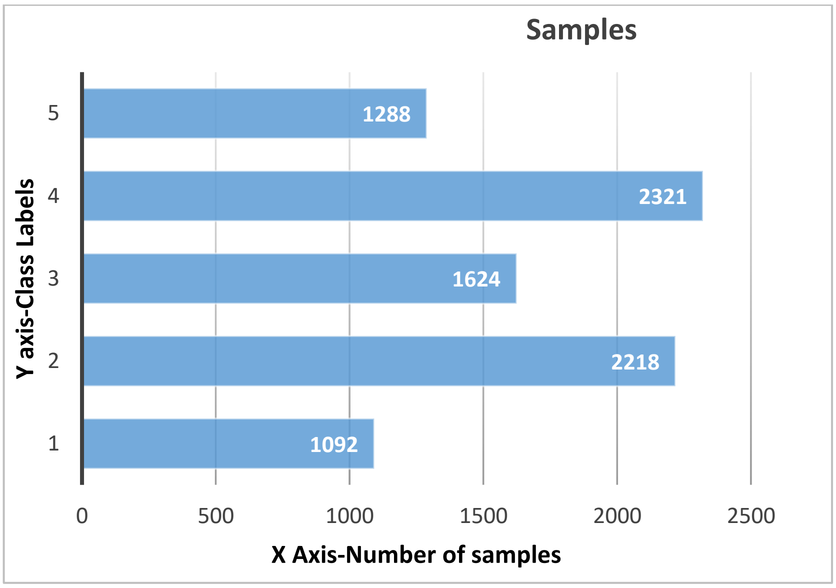


Figure 2. Statistics of input data

Table 2. Indexed and sentiment classified data after initial formatting

Index	Sentiment	Reviews
1	5	Derrida is an undeniably fascinating and entertaining man, whether or not his lectures on the other and the self have helped you learn anything.
2	4	However, this lovely act is still going on.
3	3	You'd think that at this point, America would be sick of the plucky British eccentrics with the good hearts.
4	4	The singer-composer Bryan Adams provides a number of songs, some of which have the potential to be successes and others which are merely unnecessary for the plot, but the whole thing definitely catches the intention or spirit of the work.

Algorithm 1. Preparing the data
review ◄— Each review from the database
for each review in the database
  Change the case to lowercase for every letter
  Remove the email ids
  Remove the URLs
  Eliminate HTML tags
  Avoid all the accented characters
  Remove the special symbols and characters
end for

All letters are transformed to lower case as the algorithm specifies at the outset. Unwanted outliers that won't change the dataset's meaning are deleted once it has been uniformized to make it more manageable. The data is then sent for additional processing after the algorithm has cleaned it. Table 3 depicts the sample of the cleaned dataset that was used in the method. Figure 3 shows the word cloud that was produced for the preprocessed dataset. It provides a summary of the most and least used words in the dataset.

Table 3. Preprocessed data

Index	Sentiment	Processed reviews
1	5	Derrida is an undeniably fascinating and entertaining man whether or not his lectures on the other and the self have helped you learn anything
2	4	However, this lovely act is still going on
3	3	You think that at this point America would be sick of the plucky British eccentrics with the good hearts
4	4	The singer-composer Bryan Adams provides a number of songs some of which have the potential to be successes and others which are merely unnecessary for the plot but the whole thing definitely catches the intention or spirit of the work

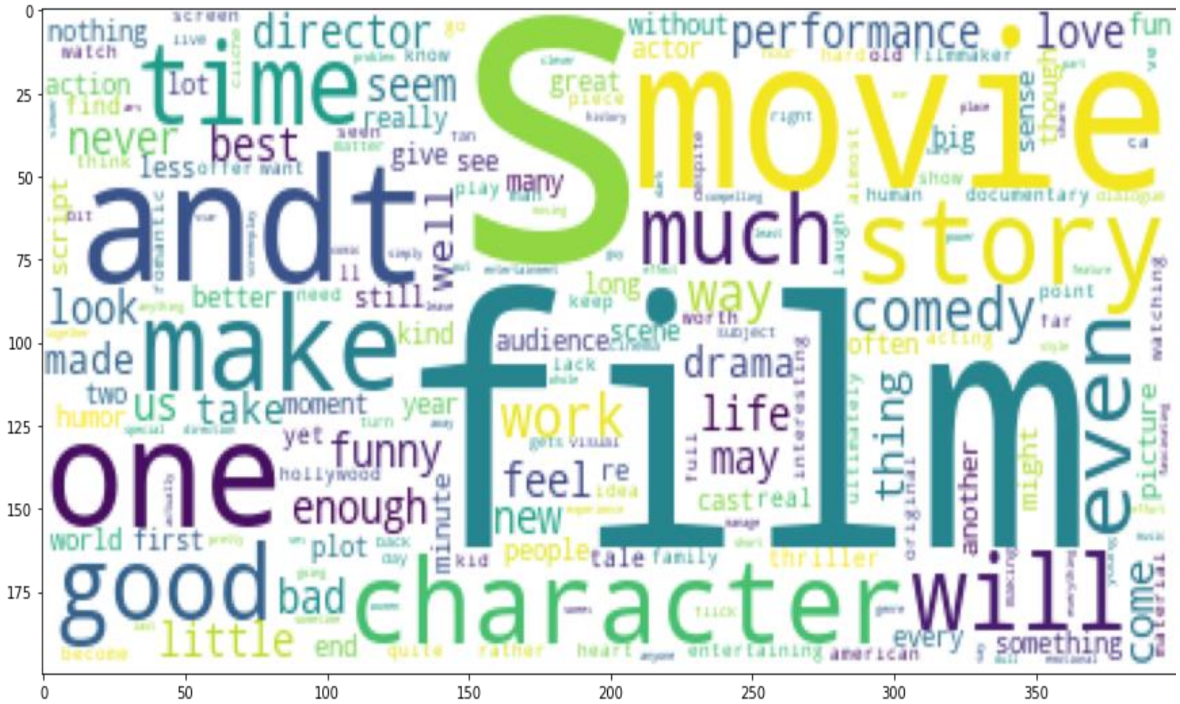


Figure 3. Word cloud of the cleaned dataset

2.3. Word vectorization

The dataset collected and cleaned is subjected to vectorization by implementing the term frequency/inverse dense frequency (TF/IDF) method. TF/IDF vectorization method is used for classifying the texts or by converting words to weighted vector values. The inefficient calculation deformities that can be found in the bag of words are overcome by TF/IDF. In the bag of words method, all words are treated equally without giving importance to specific features [17]. The words are just broken down and number values are assigned correspondingly. So, in the bag of words, it’s likely to miss important keywords as they would be less repeated and insignificant words like “a”, and” on’ will be repeated more. Inverse document frequency in TF/IDF helps to overcome this drawback of the bag of words method. IDF score helps to give importance to the most relevant features in the sentence [17]. TF focuses on the probability of occurrence of a term. Used together TF/IDF vectorizes each word considering the importance of the feature in the sentence which helps to give a contextual analysis [18]. It is calculated using the formula given in (1) and (2). The steps to perform TF/IDF are as shown in Algorithm 2.

$$IDF = Log\ [(Number\ of\ documents)/(Number\ of\ documents\ containing\ the\ word)] \quad (1)$$

$$TF = (No.of\ repetitions\ of\ a\ word\ in\ a\ document)/(No.of\ words\ in\ a\ document) \quad (2)$$

Algorithm 2. Vectorizing the data
words ⟵ Each word from the pre-processed dataset
for each word in the database
  tokenize each word with the value of frequency.
  Calculate Term Frequency for each word
  Evaluate Inverse Domain Frequency for the word
  Vectorize the vocab
end for

Figure 4 represent a sample of the features extracted using TF/IDF. Figure 5 represents a part of the data frame format after TF/IDF vectorization of the whole pre-processed dataset as per the given algorithm, with their scores. After vectorization, classification is modelled with rule-based models and supervised machine learning models. The rule-based approach is a useful tool for analysing text without the need for machine learning models or training. The outcome of this method is a set of guidelines that are used to categorise the text into various sentiments.

'adapted', 'add', 'added', 'addition', 'adds', 'adequate', 'adequately', 'admirable', 'admirably', 'admire', 'admission', 'admit', 'admittedly', 'adolescence', 'adolescent', 'adult', 'adults', 'advance', 'advantage', 'adventure', 'adventures', 'adventurous', 'advertised', 'advice', 'affair', 'affecting', 'affection', 'affleck', 'aficionados', 'afloat', 'afraid', 'african', 'after', 'afternoon', 'afterschool', 'again', 'against', 'age', 'aged', 'agency', 'agent', 'ages', 'aging', 'ago', 'ah', 'ahead', 'ai', 'aimed', 'aimless', 'aimlessly', 'aims', 'air', 'airless', 'akin', 'alabama', 'alas', 'albeit', 'album', 'alert', 'alexandre', 'alfred', 'alice', 'alien', 'alienation', 'aliens', 'alike', 'alive', 'all', 'allegory', 'allen', 'allow', 'allowed', 'allowing', 'allows', 'alluring', 'almost', 'alone', 'along', 'already', 'also', 'alternately', 'alternative', 'although', 'altogether', 'always', 'am', 'amaro', 'amateurish', 'amazing', 'amazingly', 'ambiguity', 'ambiguous', 'ambition', 'ambitious', 'america', 'american', 'americans', 'amiable', 'amid', 'amidst', 'among', 'amount', 'amounts', 'amused', 'amusing', 'amy', 'an', 'anachronistic', 'analytical', 'analyze', 'ancient', 'and', 'anderson', 'andor', 'andt', 'anemic', 'angel', 'angels', 'anger',

Figure 4. A sample of feature extraction

index	10	100	101
7935	0.0	0.0	0.4824659510987902
1416	0.0	0.0	0.4050752623907632
879	0.0	0.0	0.36316108879290526
2589	0.0	0.0	0.32291019787575653
4881	0.0	0.0	0.30197234971929177
3716	0.0	0.43918593766997205	0.0
6739	0.0	0.41596227730482405	0.0
4910	0.0	0.41252283057059713	0.0
440	0.0	0.3233061663597549	0.0
6392	0.0	0.27483008438100043	0.0
5705	0.424480839324193	0.0	0.0
1724	0.41949359474190523	0.0	0.0

Figure 5. TF/IDF vectorization score

2.4. Sentiment classification-rule based methods

Rule-based methods frequently concentrate on pattern matching or parsing. Rule-based predictions have a high sensitivity and specificity and low precision [19]. The classifier here uses the impact of if-then rules to classify the data. If conditions, then the conclusion is sorted out. The overall process goes through three stages. First feature extraction and rule learning are performed. After those relevant words of opinion are extracted from the set and finally, a prediction on the orientation of polarity will be done. Pattern matching is done with the help of pre-trained data which is divided into two categories, positive and negative. This dataset library is called lexicons. Each token will be matched with lexicons after feature extraction, and then classified as positive or negative. Finding the maximum case determines the overall polarity. For example, if the polarity is greater than zero the sentence may be categorized as positive and as negative if less than 0. Two rule-based algorithms Vader and text blob have experimented with the dataset. Text blob uses polarity and subjectivity to determine a subject's sentiment. While subjectivity gauges the objectivity of the subject, polarity provides the measurement of sentiment. In comparison to machine learning models, Vader is a predictive analytic tool that utilizes less computing power. Vader is increasingly widely used in social media analysis because it can recognize the polarity and intensity of emotions [19]. However, the primary flaw in these two approaches was that they skipped contextual analysis.

2.5. Sentiment classification-machine learning methods

Two machine learning classifiers, SVM and LR were constructed and evaluated further to check the accuracy and false negative reduction. SVM is applied to text data where texts are scattered. SVM can be implemented for both classification and regression challenges. Classification in SVM is done using a hyper plane, which helps to divide data into different categories [20]. The hyper plane is constructed by SVM with the help of "Kernels". Kernels can be both linear and nonlinear. All data values in SVM are represented as points in a plane that corresponds to specific coordinates. Classification is done on these points by creating a decision line which is normally called a hyper plane. The points which are close to the hyper plane are considered for the evaluation process in SVM algorithms. These points are called support vectors. The margin, which is normally the area between the support vectors and hyper plane, should be measured and compared. The algorithm extracts the hyper plane with maximum margin value as the optimal hyper plane. SVM can be both linear and nonlinear. Dividing the entire dataset into two categories by clearly drawing a decision boundary or hyper plane is called linear SVM [21]. In those cases, a linear SVM classifier has to be implemented. If data samples are scattered in such a way that it cannot be categorised with one hyper plane, it’s a type of nonlinear SVM classifier [22]. In the case of nonlinear classification problems, the data is converted into linearly separable data in a higher dimension, by adding one more dimension as in (3).

$$z = x\hat{}2 + y\hat{}2 \qquad (3)$$

The statistical algorithm linear regression attempts to predict Y given X [23]. The data sets are thoroughly examined to find a relationship [24]. LR concentrates on how the X input, common phrases, and words are related to the Y output polarity [25], [26]. This will help to scale up the words between the maximum (really positive) and minimum (really negative) [27].

2.6. Accuracy evaluation matrices

For evaluating the classification accuracy of experimented methods precision, recall and F-measure matrices are used [28]. Precision is a value of the number of correct predictions done. For multiclass classification, it is calculated using (4). The precision value ranges from 0.0 to 1, where 1 shows the perfect accuracy.

$$Precision\ =\ sum\ c\ in\ C\ TruePositives\_c\ /sum\ c\ in\ C\ (TruePositives\_c\ +\ FalsePositives\_c) \qquad (4)$$

Recall gives the measure of correct predictions against all those predictions that could have been made. Recall gives an indication of the missed positive predictions [29]. Recall for multiclass classification can be calculated using (5).

$$Recall\ =\ sum\ c\ in\ C\ TruePositives\_c\ /sum\ c\ in\ C\ (TruePositives\_c\ +\ FalseNegatives\_c) \qquad (5)$$

After calculation precision and recall, F-measure is also calculated. F-score is the widely accepted measure to verify the accuracy of imbalanced classifications [30]. The harmonic mean of the two fractions precision and recall give the F1-score. F1-score can be calculated using (6).

$$F - measure\ =\ (2\ *\ Precision\ *\ Recall)\ /\ (Precision\ +\ Recall) \qquad (6)$$

Additionally, the confusion matrix is also calculated to get a fine-grained analysis of the predicted output. Confusion matrices while applying to an imbalanced multiclass classification will give a clear picture of which classes are correctly predicted and which classes are least accurate. Confusion matrices are taken for all the feasible methods for an in-depth analysis. FN are calculated with the help of confusion matrix.

2 RESULTS AND ANALYSIS

The experimental study carefully analyzed various methods for determining the polarity in a dataset and gauging the sentiment of a text. Data is categorised into five categories using the proposed system: very positive, positive, neutral, negative, and extremely negative. The entire procedure entails gathering a dataset of movie reviews, cleansing the data using NLP techniques, vectorizing the corpus, identifying the polarity of each review, and then calculating accuracy. Algorithms for rule-based and supervised machine learning were put into practice, evaluated and analyzed for accuracy and FN.

Each class's precision, recall, and F1 score are identified separately. The outcomes are shown in Figure 6. Figure 7 depicts the weighted macro average of all five classes that implemented the models to provide a better analytical perspective. It is clear from the analysis that the accuracy matrices are stronger when employing the suggested SVM, TF/IDF model. But by itself, that does not fulfil the goals of the suggested methodology. Analytical study for false negative values is performed on each fine-grained class.

Figure 8 plots the total number of incorrect negative predictions to provide a deeper understanding. The analysis shows, without compromising accuracy values, that the suggested approach of SVM implementing TF/IDF reduced FN. Figure 9 analyses the computation of false negative values for each of the five classes using the four models from the confusion matrices. The analysis shows that rule-based models have a very high percentage of false negative predictions.

Nevertheless, LR beat the rule-based algorithms, but it excels at doing short-term predictive analysis. However, when applied to a broader classification of predictive fine-grained analysis, the hyper plane classification concept implemented in SVM is demonstrated to provide accurate and fine-grained analysis while also lowering false negative values or Type II Error. The precision, recall, and accuracy of LR are 0.41, 0.33, and 0.33, respectively, while that of SVM is 0.41, 0.35, and 0.35. The number of FN in LR (1550) was high than that of SVM (1527). Due to its classification property based on the closeness to the hyper-plane notion, the proposed methodology using SVM with a TF-IDF-based model is the more accurate. The proposed methodology allows for the derivation of more classes with more precise predictions. The suggested method achieves its ultimate goal of reducing FN without sacrificing accuracy. When using the suggested model with SVM implementing TF/IDF, the false negative values in nearly all five classes of polarity are significantly reduced.

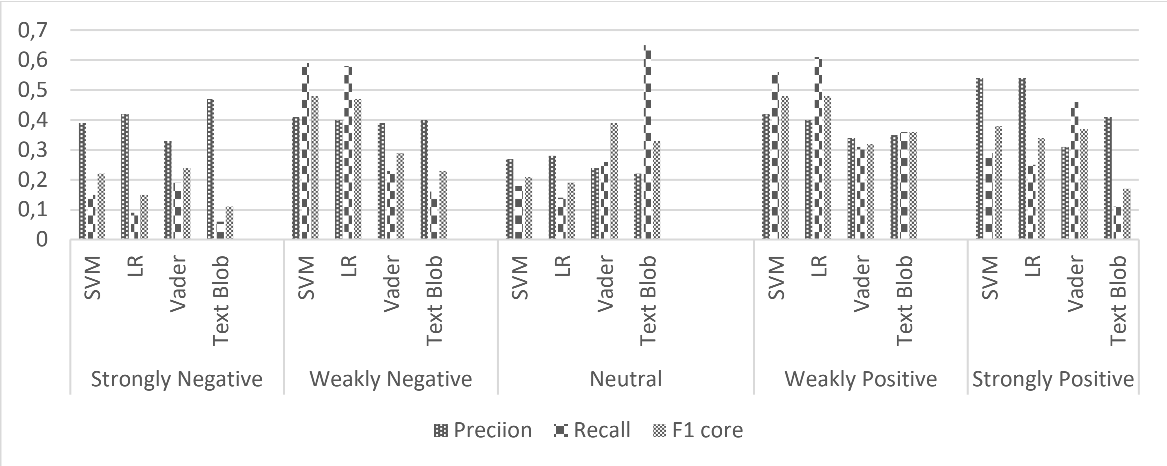


Figure 6. Precision, recall and F1 measures

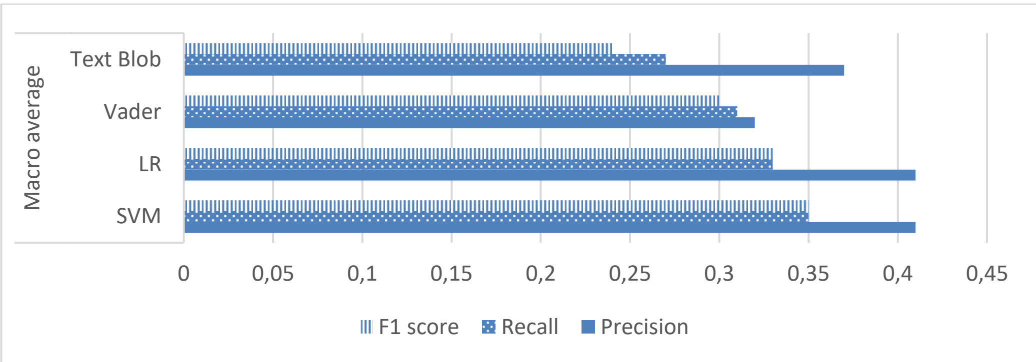


Figure 7. Weighted macro average of precision, recall and F1 measures

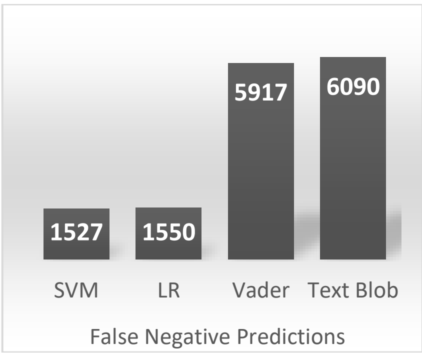


Figure 8. False negative prediction

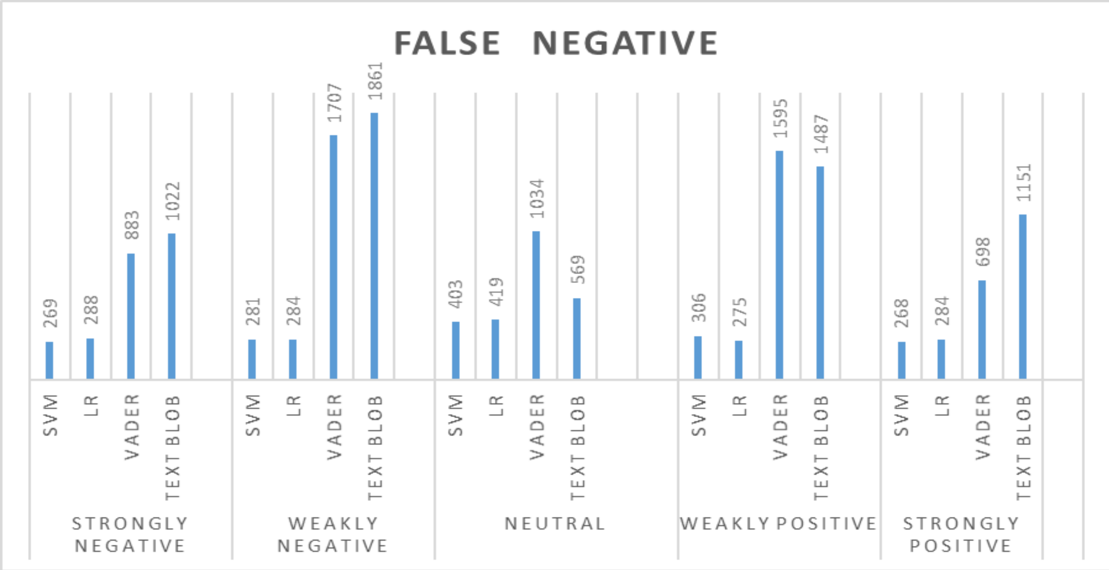


Figure 9. Analysis of false negative values

3 CONCLUSION

With the development of technology, the requirement to gain deeper insights into consumer opinions and product reviews is becoming increasingly mandatory. The significance of sentiment analysis and several approaches to producing a fine-grained, definitive output on opinion mining are examined in this study paper. Several supervised machine learning methods were analysed and evaluated alongside rule-based models. Algorithm application is discovered to be dependent on the focus area and type of conclusive judgment required. Unlike the existing studies, the proposed system makes a fine-grained classification of the dataset. Without referring solely to total precision, recall, and the F measure, false negative values of each class are identified separately, experimented and analyzed. The increase in accuracy is given equal importance with the reduction in FN, which is totally a novel method in sentiment analysis. Experimented quantified results shows that the proposed method gave a good accuracy in fine grained analysis by simultaneously reducing FN.

REFERENCES

[1] J. Shobana and M. Murali, "An efficient sentiment analysis methodology based on long short-term memory networks," *Complex Intell. Syst*., vol. 7, no. 5, pp. 2485–2501, Oct. 2021, doi: 10.1007/s40747-021-00436-4.
[2] P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min*., vol. 11, no. 1, p. 81, Aug. 2021, doi: 10.1007/s13278-021-00776-6
[3] S. Palit and S. Ghosh, "Real Time Sentiment Analysis," *International Journal of Synthetic Emotions (IJSE)*, vol. 11, no. 1, pp. 27–35, 2020, doi: 10.4018/IJSE.2020010103.
[4] N. Capuano, L. Greco, P. Ritrovato, and M. Vento, "Sentiment analysis for customer relationship management: an incremental learning approach," *Appl Intell*, vol. 51, no. 6, pp. 3339–3352, Jun. 2021, doi: 10.1007/s10489-020-01984-x.
[5] A. Ligthart, C. Catal, and B. Tekinerdogan, "Systematic reviews in sentiment analysis: a tertiary study," *Artif Intell Rev*, vol. 54, no. 7, pp. 4997–5053, Oct. 2021, doi: 10.1007/s10462-021-09973-3
[6] M. A. C.-Valadez, A. F.-Mendoza, C. E. M.-Rodríguez, R. C.-Miranda, N. K. G.-Rosales, and J. R. L.-Robles, "The synergic relationship between e-commerce and Sentiment Analysis: A content analysis of published articles between 2007 and 2020," *in 2020 International Conference on Data Analytics for Business and Industry: Way Towards a Sustainable Economy (ICDABI)*, Oct. 2020, pp. 1–5, doi: 10.1109/ICDABI51230.2020.9325689.
[7] Z. Drus and H. Khalid, "Sentiment Analysis in Social Media and Its Application: Systematic Literature Review," *Procedia Computer Science*, vol. 161, pp. 707–714, Jan. 2019, doi: 10.1016/j.procs.2019.11.174.
[8] S. G. Barbounaki, K. Gourounti, and A. Sarantaki, "Advances of Sentiment Analysis Applications in Obstetrics/Gynecology and Midwifery," *Mater Sociomed*, vol. 33, no. 3, pp. 225–230, Sep. 2021, doi: 10.5455/msm.2021.33.225-230.
[9] M. Bouazizi and T. Ohtsuki, "Multi-class sentiment analysis on twitter: Classification performance and challenges," *Big Data Mining and Analytics*, vol. 2, no. 3, pp. 181–194, Sep. 2019, doi: 10.26599/BDMA.2019.9020002.
[10] L. Yue, W. Chen, X. Li, W. Zuo, and M. Yin, "A survey of sentiment analysis in social media," *Knowl Inf Syst*, vol. 60, no. 2, pp. 617–663, Aug. 2019, doi: 10.1007/s10115-018-1236-4
[11] Z. Li, Y. Fan, B. Jiang, T. Lei, and W. Liu, "A survey on sentiment analysis and opinion mining for social multimedia," *Multimed Tools Appl*, vol. 78, no. 6, pp. 6939–6967, Mar. 2019, doi: 10.1007/s11042-018-6445-z.
[12] B. Dahal, S. A. P. Kumar, and Z. Li, "Topic modeling and sentiment analysis of global climate change tweets," *Soc. Netw. Anal. Min*., vol. 9, no. 1, p. 24, Jun. 2019, doi: 10.1007/s13278-019-0568-8.
[13] A. U. Rehman, A. K. Malik, B. Raza, and W. Ali, "A Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis," *Multimed Tools Appl*., vol. 78, no. 18, pp. 26597–26613, Sep. 2019, doi: 10.1007/s11042-019-07788-7.

[14] A. Dridi, M. Atzeni, and D. R. Recupero, "FineNews: fine-grained semantic sentiment analysis on financial microblogs and news," *Int. J. Mach. Learn. & Cyber*, vol. 10, no. 8, pp. 2199–2207, Aug. 2019, doi: 10.1007/s13042-018-0805-x.
[15] M. Tubishat, M. A. M. Abushariah, N. Idris, and I. Aljarah, "Improved whale optimization algorithm for feature selection in Arabic sentiment analysis," *Appl Intell*., vol. 49, no. 5, pp. 1688–1707, May 2019, doi: 10.1007/s10489-018-1334-8.
[16] S. Yoo, J. Song, and O. Jeong, "Social media contents based sentiment analysis and prediction system," *Expert Systems with Applications*, vol. 105, pp. 102–111, Sep. 2018, doi: 10.1016/j.eswa.2018.03.055.
[17] S.-W. Kim and J.-M. Gil, "Research paper classification systems based on TF-IDF and LDA schemes," *Hum. Cent. Comput. Inf. Sci*., vol. 9, no. 1, p. 30, Aug. 2019, doi: 10.1186/s13673-019-0192-7.
[18] S. Sarlis and I. Maglogiannis, "On the Reusability of Sentiment Analysis Datasets in Applications with Dissimilar Contexts," *in Artificial Intelligence Applications and Innovations*, Cham, 2020, pp. 409–418, doi: 10.1007/978-3-030-49161-1_34.
[19] T. Pano and R. Kashef, "A Complete VADER-Based Sentiment Analysis of Bitcoin (BTC) Tweets during the Era of COVID-19," *Big Data and Cognitive Computing*, vol. 4, no. 4, Dec. 2020, doi: 10.3390/bdcc4040033.
[20] S. Yi and X. Liu, "Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review," *Complex Intell. Syst*., vol. 6, no. 3, pp. 621–634, Oct. 2020, doi: 10.1007/s40747-020-00155-2.
[21] F. Ouatik, M. Erritali, F. Ouatik, and M. Jourhmane, "Predicting Student Success Using Big Data and Machine Learning Algorithms," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 17, no. 12, Jun. 2022, doi: 10.3991/ijet.v17i12.30259.
[22] S. Kim and J. Choi, "An SVM-based high-quality article classifier for systematic reviews," *Journal of Biomedical Informatics*, vol. 47, pp. 153–159, Feb. 2014, doi: 10.1016/j.jbi.2013.10.005.
[23] M. A. Khan, R. Khan, F. Algarni, I. Kumar, A. Choudhary, and A. Srivastava, "Performance evaluation of regression models for COVID-19: A statistical and predictive perspective," *Ain Shams Engineering Journal*, vol. 13, no. 2, p. 101574, Mar. 2022, doi: 10.1016/j.asej.2021.08.016.
[24] K. Kumari and S. Yadav, "Linear regression analysis study," *Journal of the Practice of Cardiovascular Sciences*, vol. 4, no. 1, pp. 33–36, 2018, doi: 10.4103/jpcs.jpcs_8_18.
[25] S. Ghosal, S. Sengupta, M. Majumder, and B. Sinha, "Linear Regression Analysis to predict the number of deaths in India due to SARS-CoV-2 at 6 weeks from day 0 (100 cases-March 14th 2020)," *Diabetes Metab Syndr*., vol. 14, no. 4, pp. 311–315, Aug. 2020, doi: 10.1016/j.dsx.2020.03.017.
[26] M. D. Coria, J. C. M. Arevalo, J. L. Hilario-Rivas, and J. R. H.-C., and J. I. Prado-Juscamaita, "Supervised Sentiment Analysis Algorithms," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 14, Aug. 2021.
[27] H. Xie, W. Lin, S. Lin, J. Wang, L.-C. Yu, "A multi-dimensional relation model for dimensional sentiment analysis," *Information Sciences*, vol. 579, pp: 832-844, 2021, doi: 1016/j.ins.2021.08.052.
[28] F. Tang, L. Fu, B. Yao, and W. Xu, "Aspect based fine-grained sentiment analysis for online reviews," *Information Sciences*, vol. 488, pp. 190–204, Jul. 2019, doi: 10.1016/j.ins.2019.02.064.
[29] B. Juba and H. S. Le, "Precision-Recall versus Accuracy and the Role of Large Data Sets," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, no. 1, Jul. 2019, doi: 10.1609/aaai.v33i01.33014039
[30] L. He, T. Yin, and K. Zheng, "They May Not Work! An evaluation of eleven sentiment analysis tools on seven social media datasets," *Journal of Biomedical Informatics*, vol. 132, p. 104142, Aug. 2022, doi: 10.1016/j.jbi.2022.104142.

BIOGRAPHIES OF AUTHORS

Chris Aloysius is an Associate Professor at the Department of Computer Science, Don Bosco College (Calicut University), Thrissur, Kerala, where she has been a faculty member since 2009. From 2012-2018, she was also the Head (Undergraduate) of the department and after that from 2019 onwards she was the Academic Coordinator of the entire postgraduate and undergraduate courses in the college. She received the degree in computer science from Calicut University in 2005 and the Master degree in Computer Application from Bharathiar University in 2009. Currently she is working towards the Ph.D Degree in the Department of Computer Science at Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. Her research interests include sentiment analysis, computer-aided medical diagnosis, and data mining. She can be contacted at email: chrissnowdon27@gmail.com.

Dr. P. Tamil Selvan is working as an Associate Professor in the Department of Computer Science at Karpagam Academy of Higher Education, Coimbatore, Tamil Nadu, India. He has more than 14 years of teaching and research experience. He received his Master in Computer Application from Bharathiar University and Doctorate from Karpagam Academy of Higher Education. His research interests include data mining, WSN, and text categorization. He can be contacted at email: tamilselvancs@kahedu.edu.in.