A Novel Hybrid Lexicon Ensemble Learning Model for Sentiment Classification of Consumer Reviews

Sourav Sinha[1] and Dr. Revathi Sathiya Narayanan[2*]

[1]Research Scholar, Department of Computer Science & Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India.
sourav_cse_phd_17@crescent.education, Orcid: https://orcid.org/0000-0002-3327-8215

[2*]Professor, Department of Computer Science & Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. srevathi@crescent.education,
Orcid: https://orcid.org/0000-0001-9584-5089

Received: June 17, 2023; Accepted: August 05, 2023; Published: August 30, 2023

Abstract

In recent past during the era of consumerism with easy accessibility to social networking world, the consumers usually give comments and opinions on daily usable ingredients, electronic goods and services offered on payments. These comments or opinions are innumerable and huge on each item, hence need the special attention for sentimental value particularly on their text parts. The present study is an attempt to perform sentiment prediction on Amazon Electronic products, gift cards and Kindle dataset. In this paper, the HLESV (Hybrid Lexicon Ensemble based Soft Voting) model is proposed by combining lexicon and ensemble approaches using optimally weighted voting to predict the sentiment, subsequently to evaluate model using various performance metrics like precision, recall, F1-score. This paper computes an additional metric namely subjectivity score along with sentiment score and proposes non-interpretive sentiment class label to evaluate the polarity of the reviews using our proposed HLESV model for sentiment classification. The accuracy score of our proposed HLESV model is evaluated to assess its effectiveness on Amazon consumer product review datasets and observed an increase of 1-6% accuracy over existing state-of-the-art ensemble methodology.

Keywords: Hybrid Learning, Sentiment Classification, Senti-Wordnet, Consumer Reviews.

1 Introduction

The Sentiment analysis on consumer reviews acts with positive significance on various applications in business entities like Collaborative filtering in Recommendation system, Governance Intelligence and Review Summarization and so on. Thus, the results effectively energize all major upstream entrepreneurs like designers, manufacturers, and marketing individuals on their product reviews while consumers would be happier with detail analysis about merits and demerits of items they want to buy. Despite certain contextual ambiguity in usage of meaningful words, sarcasm in text writing or negative handling, conventional methods for sentiment analysis usually guide towards productive outputs. The sentiment prediction results have further improved with advent of Knowledge based and Machine learning based tools and techniques. On outset a brief study has been made to establish a generalized

Journal of Internet Services and Information Security (JISIS), volume: 13, number: 3 (August), pp. 16-30.
DOI: *10.58346/JISIS.2023.I3.002*

*Corresponding author: Professor, Department of Computer Science & Engineering, B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India.

consumer review pattern on textual opinions communicated by customers for the items they purchased online. Those online reviews are made on platforms of recognized business houses forums or social media sites. Primarily the customer reviews text can be classified into two separate entities on basis of viewpoint inclinations namely Polar reviews and non-polar reviews as shown in above Figure -1. Polar reviews include Positive reviews and Negative reviews due to their attitudinal viewpoint inclination while both are interpretive and informative from the text material viewpoint. The Neutral reviews belong to non-polar review class but they have informative text material.

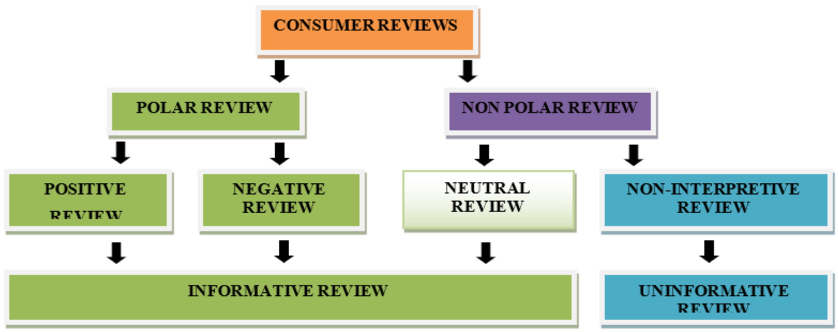

Figure 1: Pictorial Representation of Consumer Reviews Pattern

Hence all three namely Positive, Negative and Neutral reviews are covered in a group designated as Informative reviews and they possess some utility components. Finally, Non-polar reviews possess a new entity first time introduced as non-interpretive reviews only after finding lack of interpretive contents in the review text. This minimum utility valued component is also identified as Uninformative reviews.

As per the existing study, sentiment analysis is a process of analyzing subjective sentences in a review text where opinion of the customer is reflected in subjectivity of sentences. The major outcome from this research study is being mentioned in brief with following points-

- To calculate the subjectivity score for opinionating the unstructured consumer review sentences.
- Define a new sentiment class named as “Non-interpretive Sentiment” on basis of subjectivity score by analysis of reviews to identify subjective expressions.
- Propose a new hybrid model, named HLESV for sentiment prediction by combining lexicon based and ensemble machine learning based techniques using soft voting for sentiment classification.

2 Related Work

Sentiment classification and subjectivity analysis are the focused areas of research works with lexicon based, supervised and unsupervised based, rule-based, deep learning based, ensemble based and hybrid approach which are being implemented in the recent past to achieve better performance and higher accuracy.

(Ohana, 2009) assessed the usage of senti-wordnet opinion lexicon for the sentiment classification of film reviews. The study further suggested that linguistic processing coupled with low dimensionality of data further improves the classification accuracies. (Chathuranga, 2019) proposes a framework using the corpus-based method in construction of corpus lexicon for sentiment analysis for Sinhala language with larger text corpuses. (Alfranzi, 2019) proposes the hybrid semantic knowledge-based Machine Learning approach to overcome the limitations and improve the accuracy of the opinion mining process. (Bhoir, 2015) proposes rule-based system using two methods namely naive-bayes classifier on machine learning and lexicon which is applied on IMDB movie reviews. The paper argues in favor of naïve-bayes classifier. (Cambria, 2013) proposes a model for word level-based sentiment sequence using reinforcement learning approach. His work also explains about importance of concept level sentiment analysis which is a novel approach that helps in efficient transformation of data from unstructured textual information into structured machine knowledgeable format. (Wang, 2019) proposes hierarchical reinforcement learning approach based on document level aspect-based sentiment classification (DASC). These papers (Ravi, 2015) (Medhat, 2014) discussed about various supervised learning classifier like NB, Decision Tree, SVM, Maximum Entropy (ME), Rule based classifier for sentiment classification. (Shelke, 2017) proposes a system to perform sentiment analysis in order to evaluate product review comments using expectation maximization algorithm. Another paper (Thara, 2017) presents with SVD based feature for sentiment analysis of hotel reviews. The paper (Yuan, 2013) discusses about a novel approach for sentiment classification using association rule mining on Amazon reviews. (Rehioui, 2020) proposes a new clustering algorithm using k-means and DENCLUE for sentiment classification of tweets. (Al-Saqq, 2018) presents state-of-the-art methodology with the ensemble of classifiers to determine the opinion polarity of the Arabic text. The author used majority voting based on naïve-bayes, support vector machine, k-nearest neighbor, and decision tree classifiers. (Behera, 2016) discusses about various ensemble learning technique applied to improve the prediction, function approximation, classification, and performance of a model. According to the paper, Adaboost based ensemble learning technique can be implemented in order to address the multiclass and regression problems to achieve better accuracy. (Whitehead, 2008) implements various ensemble learning algorithms like bagging space, random space models and suggest that usage of correct ensemble learning algorithm can drastically increase the accuracy of the ensemble learning models. It also justifies that the computational expense is one-time occurrence and ensemble classifier can be adequately fast in terms of sentiment classification. (Alrehili, 2019) proposed a sentiment classification model to classify positive and negative reviews using ensemble machine learning by combining five classifiers namely Naïve Bayes, Support Vector Machine, Random Forest, Boosting and Bagging. Paper concludes that random forest technique gives higher accuracy while using unigram with stop word removal while voting algorithm showing best performance in other cases. (Permatasari, 2018) performs classification of the movie reviews in Indonesian Language using Naïve Bayes Classifier by employing the ensemble features by combining textual features, part of speech features, lexicon-based features, and bag of words. (Sadhasivam, 2019) proposes the majority voting-based ensemble approach by combining naïve bayes and Support Vector Machine to achieve better accuracy, performance, and speed execution of the algorithm. The paper also suggests that the accuracy of the review classification is based on the number of classifiers combined in the ensemble approach used for the output prediction of the review. (Rajeshwari, 2020) applied hybrid approach using lexicon and machine learning methods to address binary classification problem based on twitter, movie, and product review comments. This author uses machine learning algorithms like Naïve Bayes, Linear Regression, Support Vector Machine and Decision Tree with and without lexicon approach in order to evaluate the metrics like precision, recall and AUC. The paper also states that logistic regression works better as compared to another classifier. It also recommends about usage of heterogeneous features and deep learning techniques can improve the accuracy of the system. (Murni, 2019) applies homogeneous ensemble classifier such as bagging decision tree, bagged multilayer perceptron's, random forest, logistic model tree to determine the sentiment of the TripAdvisor's tourist reviews. According to the paper, hybrid method is more effective and overcomes the limitations of each original method based on lexicon method and ensemble method. (Sridhar, 2020) implements with a hybrid bidirectional LSTM-ANN model based on the hypernym features along with the temporal features. According to the paper, the bi-LSTM part of the model identifies the semantic meaning and the ANN part of the model improves the efficiency and performance of the model by adding the hypernym features. (Mukwazvure,2015) embarks a hybrid method to combine the sentiment lexicon and machine learning algorithm like Support Vector machine (SVM) and K-Nearest Neighbor for sentiment analysis and concludes that SVM outperforms K-NN on news comments.

3 Research Work

The present paper proposes a HLESV (Hybrid Lexicon Ensemble based Soft Voting) model using hybrid approach by combining lexicon approach and ensemble machine learning approach (Srinadi, N.L.P., 2023). During the ensemble learning process both supervised and unsupervised learning are combined to achieve better performance in terms of classification and prediction. The open-source library named NumPy, scikit learn (Buitinc, 2013) (Pedregosa, 2011), senti-wordnet, Natural Language Toolkit are also installed to perform this simulation.

Datasets

The present work is done on datasets namely Consumer Electronic Product Reviews (CEPR), which had been downloaded from kaggle.com website. The datasets for Gift cards based on Amazon reviews are hosted in GitHub website (McAuley, 2017) (Gift card dataset). The CEPR (Kindle dataset) dataset contains list of 34661 reviews of Amazon electronic consumer products namely Kindle; Fire TV etc. in comma separated value format. The content of dataset includes product information, review comments text, ratings. Similarly, (Electronic Products dataset) and (Gift cards dataset) contain 2972 and 2375 reviews respectively.

Proposed Algorithms

A hybrid innovative model named HLESV has proposed by combining the lexicon-based techniques using senti-wordnet and ensemble-based machine learning techniques. The below Figure 2 is architecture diagram of proposed HLESV model using soft voting in this paper for analysis and prediction of opinions associated to the reviews. The proposed HLESV model achieves better performance with higher accuracy which is broadly explained in two phases as follows:

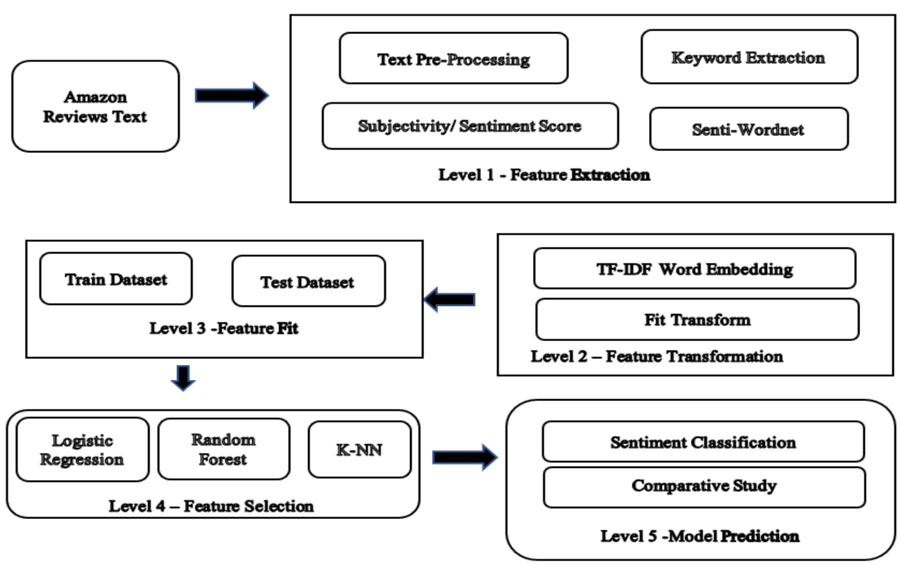

Figure 2: Architecture Diagram for Proposed HLESV Model

- **Phase 1: Applying Lexicon based Technique**

During this phase the given datasets are pre-processed to extract the keywords using tokenization and lemmatization for identification of keywords and removing unnecessary stop words, determinants, prepositions etc. The POS tagging is computed using the NLTK library based on the natural language processing.

Finally, the sentiment and subjectivity score are being computed based on the semantic features obtained from the senti-wordnet dictionary. In below Algorithm 1 describes the algorithmic steps used to compute the overall sentiment score and subjectivity score respectively based on Amazon review comment. The dataset is initially divided into train and test data in the different ratio 90:10 and 80:20 respectively. We try to use the senti-wordnet dictionary to determine sentiments for each of these keywords. The individual sentiment score is predicted on basis of semantic features of keywords internally present in the senti-wordnet dictionary.

This train dataset is used to compute sentiment and subjectivity score for determining the opinion polarity and identify subjective sentences which can highly influence overall sentiment score. The training data is being used to compute the sentiment score to determine the sentiment polarity of these review sentences.

Overall Sentiment Score per review sentence is computed as represented using below mathematical equation:

$$\textbf{Sentiment Score}_{\textbf{Sentence}} = \sqrt[3]{\frac{\sum_{w=1}^{n}(\text{postive\_score})_w^3 - \sum_{w=1}^{n}(\text{negative\_score})_w^3}{\text{numberofwordsinsentence}(n)}} \text{---------------------------- (Equation 1)}$$

positive_score$_{\mathbf{w}}$- Positive sentiment score as computed from the senti-wordnet dictionary

negative_score$_{\mathbf{w}}$-Negative sentiment scores as computed from the senti-wordnet dictionary

Similarly, the subjectivity score is computed based on objective score using semantic features of the keywords available in senti-wordnet dictionary.

Overall Subjective Score per review sentence

is computed:

$$\textbf{Subjectivity Score}_{\textbf{Sentence}} = \sum_{\boldsymbol{words=1}}^{\boldsymbol{n}} \frac{\mathbf{1}-(\boldsymbol{objective\_score})_{\boldsymbol{words}}}{\boldsymbol{n}} \text{ ------------- (Equation 2)}$$

Similarly, the subjectivity score is computed based on objective score using semantic features of the keywords available in senti-wordnet dictionary. Overall Subjective Score per review sentence is computed:

The aggregation of this individual score to compute the overall sentence sentiment score which helps in determining if the sentiments associated to the review sentence are positive, neutral, and negative for each of the review sentences. The sentiment is determined as positive, neutral, negative, non-interpretive sentiment respectively for the training dataset based on the computed value of the sentiment and subjectivity score. This train dataset is being used to train the proposed HLESV model and performance of the proposed model is being evaluated on test dataset.

- **Phase 2: Applying Ensemble Learning Technique**

In this phase, the proposed HLESV model uses the ensemble learning based algorithm to make sentiment predictions. The HLESV model determines the sentiment prediction on comments using the Amazon reviews dataset of consumer electronic products by computing the weighted average voting of individual classifier. The below Algorithm 2 describes the algorithmic steps used to evaluate the performance of proposed ensemble learning model by various performance measures like precision, recall, F1-score, and accuracy score re

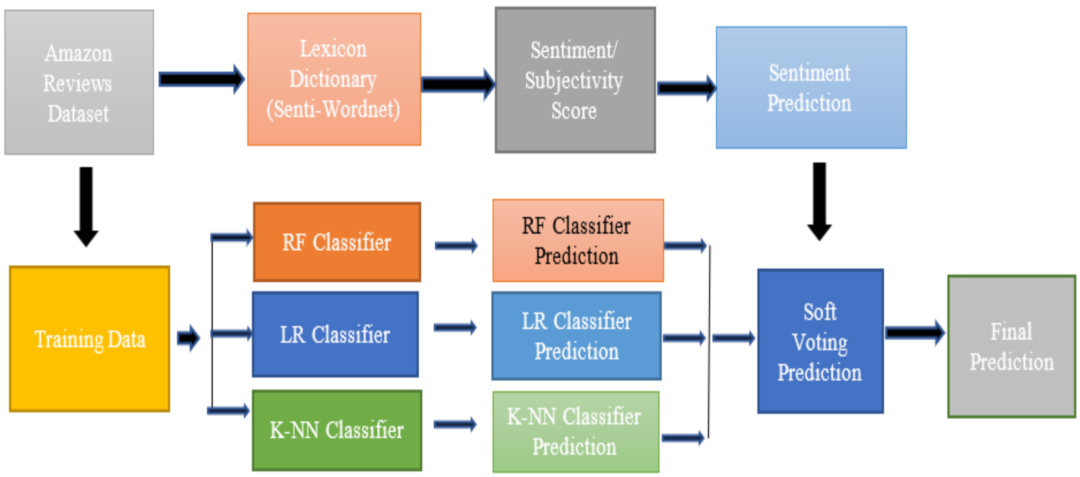

Figure 3: Data Flowchart for Proposed HLESV Model

Random Forest Classifier combines multiple classifiers in form of decision trees to solve the complex issue and improve the overall performance of our model. It relies on multiple decision tress and makes predictions based on majority voting of greater number of trees in the forest. The higher number of decision tress helps to prevent overfitting problem and improve the accuracy on large datasets with less training time.

For logistic regression classifier, Activation function used by the logistic regression classifier is based on logistic function as shown below:

$$\sigma(z) = \frac{1}{1+e^{-z}} \text{ --(Equation 3)}$$

Now the equation for logistic regression classifier can be retrieved by substituting the linear regression in the logistic function as shown below:

$$\sigma(z) = \frac{e^{\alpha_0 + \alpha_1 x}}{1+e^{\alpha_0 + \alpha_1 x}} \text{ --(Equation 4)}$$

Log-Likelihood Function is performance evaluation metric for any model used to evaluate how good the model fts with the given data and provide support for different values of the parameters. Log-likelihood function for the logistic regression classifier is mathematically represented as below:

$$L$$

$$(\alpha_0, \alpha_1) = \prod_{i=1}^{t} p(x_i)^{n}(1-p(x_i))^{1-n} \text{ --------------------------------------(Equation 5)}$$

$$\text{L } ((\alpha_0, \alpha_1) = \sum_{i=1}^{n} n \, logp(x_i) + (1-n) \, log(1-p(x_i)) \text{-----------------(Equation 6)}$$

K-NN classifier is supervised non-parametric which identifies k-nearest neighbor based on various distance measures namely Euclidean distance, Manhattan distance, Minkowski distance. K-NN algorithm is often called as lazy learner.

Algorithm 1: Proposed Lexicon-Based Algorithm to Compute the Sentiment and Subjectivity Score

Input: Review Comment

Output: Sentiment and Subjectivity

Definition: Stopwords, Tokenizer: {Whitespace Tokenizer}, Lemmatizer: {Wordnet Lemmatizer}

Library: Pandas, Nltk(Natural Language Tool Kit), Pos Tagging, Senti-wordnet, String

Function Remove_Punctuations:/* Module used to remove punctuations from
given sentence */ punctuations = extract punctuations from string library/*
identify punctuations marks*/ return puntuations
End Function

Function WordLemmatizer(sentence): /* Module used to lemmatize words in
given sentence */ stopwords = extract list of stopwords from nltk
library/* retrieve predefined stopwords*/ new_words = add the new
stopwords which are not present in stopwords list
for each word in new_words
add the new words to the stopwords as extension
tokenized_words = extract tokens using whitespace tonizer/* retrieve tokens*/
lemmatize_Words = lemmatize words for each W in tokenized words if not in stopwords

return lemmatize_Words
End Function

Function PosTagger(pos_tagging_text):/* Module used for POS tagging for
each input word*/
for each word in pos_tagging_text
return pos_tag
End Function

Main Function Senti&Subj_Score(rvw_cmt):/*Prime functional module to compute subjectivity score*/
sentence_overall_sentiment_score = 0.0 /* Initialize overall sentiment score */
sentence_overall_subjective_score = 0.0 /* Initialize overall objective score */
word_sentiment_score = 0.0 /* Initialize individual word sentiment score */
word_objectivite_Score = 0.0 /* Initialize individual word objective score*/
cleaned_sentence= Remove_Punctuations(rvw_comment)/* clean-up irrelevant text */
lemmmatized_sentence = WordLemmatizer(cleaned_sentence)
sysnetWord = extract synsets based using pos_tag for each synset in
synsetWordword_sentiment_score+ = difference of postive and negative score
using senti-wordnet word_objective_score+ = objective score of
synset/* Sum of all objective score */ sentence_overall_sentiment_score
= weighted average of word_sentiment_score
sentence__overall_subjectivity_score = 1 - weighted average of
word_objective_score
End Function

Algorithm 2: Proposed HLESV Ensemble Learning Algorithm Using soft Voting by Assigning Weights

Input: Training and Test Dataset

Output: Precision, Recall, F-Score, Confusion Matrix

Library: Voting Classifier, TrainTest Split, Classification Report, Confusion Matrix, Accuracy Score

Function TrainingData(punctuation_text): /*Module to split the dataset into training and test data */
tfidf = use the TFIDF vectorizer for extraction /* Using word embedding on term and inverse document frequency*/ X_tfidf = tfidf.fit_transform(punctuation_text) /* Data transformation */
(X_train, X_test, Y_train, Y_test) = split the dataset
End Function

Main Function Soft Voting (): /* Main Module to perform the soft voting and compute accuracy score */
clf1 = LogisticRegression /* Second ensemble classifier initialized withlogistic regression classifier */
clf2 = RandomForestClassifier/* Third ensemble classifier initialized with random forest classifier */ clf3 = kNearestNeighborClassifier /* Fourth ensemble classifier initialized with K-NN classifier */
eclf = VotingClassifier(estimators=[('LR', clf1),('RFC',clf3),('KNN', clf3)), weights=(w1,w2,w3)])
eclf.fit(X_train, Y_train) /*Fit the proposed model */
y_pred = Prediction of HLESV /* Prediction outcome of HLESV*/
print("Accuracy: ",accuracy_score(Y_test,y_pred)*100) /* Compute accuracy score */
End Function

Adaboost is boosting ensemble-based machine learning technique that aggregates multiple machine learning models into single composite machine learning models. It prevents machine learning models from underfitting of data and normally applied on two-class or multi class classification problems.

Boosting Ensemble Classifier = $Sign \sum_{i=1}^{m} WL_i C_i$--------------------------(Equation 7)

WL = Weak Learner Classifier

C = Coefficient for each weak learner

Bagging Ensemble is ensemble-based machine learning technique used to improve the accuracy and performance of machine learning model. It is often applied on classification and regression models and prevents machine learning models from overfitting of data.

Bagging Ensemble Classifier = $\frac{1}{n}\sum_{i=1}^{n} h_i(x)$------------------------------- (Equation 8)

Soft Voting classifier computes the decisions by searching the most potential suitable candidate. It calculates the probabilistic value which is used to predict target class. In soft voting, the predictions are weighted with importance of different classifier and target class label is chosen argmax of the total of probabilities predicted from the individual classifier that forms the ensemble. The generalized equation representing the soft voting classifier is shown below:

Soft Voting (Z) = $argmax \sum_{i=1}^{n} \frac{PC_i W_i}{W_i}$--------------------------------------- (Equation 9)

Our proposed HLESV model takes advantage of soft voting ensemble techniques along with lexicon-based technique which helps in depicting the target class labels. We attempted to use three classifiers namely Random Forest, Logistic Regression and K-NN classifier respectively in our proposed model. In our model weights are being assigned to three classifiers and then the weighted average of probability voting is being computed by multiplying weights with predicted class probabilities of each classifier as presented in the below mathematical equation:

$$\text{HLESV Weighted Probability} = argmax \frac{w1pc1+w2pc2+w3pc3}{w1+w2+w3} \text{--------------- (Equation 10)}$$

pc1 = Predicted Probability determined by Random Forest Classifier (Supervised Learning)

pc2 = Predicted Probability determined by Logistic Regression Classifier (Supervised Learning)

pc3 = Predicted Probability determined by K-Nearest Neighbor Classifier (Unsupervised Learning)

w1 = Weight assigned for Logistic Regression Classifier

w2 = Weight assigned for Random Forest Classifier

w3 = Weight assigned for K-Nearest Neighbor Classifier

The main advantage of proposed HLESV model is that it depends on hybrid approach to get merits from lexicon and machine learning techniques for making predictions using group of classifiers based on weighted average. Another advantage of proposed HLESV model is that it makes appropriate correction for the error made by various participating classifier.

4 Results and Discussion

The primary objective of proposed HLESV model is to predict better sentiment and subjectivity scores for sentiment classification. The proposed model has been applied to compute sentiment analysis for Amazon's consumer products like magazines, gift cards, kindle, fire tv etc. using training data and subsequent calculations of subjectivity score.

Simulation Results

The simulation was performed on two different system hardware configurations as Intel Celeron N4000 processor with 4 GB RAM. The results are already tabulated with the comparison of sentiment and subjectivity scores in the below Table 1. Python with version 3.7.9 were installed on these systems with above hardware configurations. The proposed HLESV model implements the ensemble learning algorithm by combining three classifiers namely Logistic Regression (LR), Random Forest (RF), K-Nearest Neighbor (K-NN) respectively.

Table 1: Comparison of Sentiment/Subjectivity Score Computed Using HLESV and Text Blob

Review Comments	**Senti Score HLESV**	**Sub Score HLESV**	**Senti Score Text blob**	**Sub Score Text blob**
This is a great tablet for price would recommend it to anyone looking for a basic tablet	0.2988	0.6406	0.4	0.4375
Great e-reader especially love backlight. Recommend to others	0.4472	0.7343	0.6	0.75
Tablet stopped working after owning for two weeks. Came back on after an hour.	0.0429	0.375	0.1666	0.3273
Absolute junk, don't waste your money. Bought for wife for a Christmas gift. Would not connect to Wi-Fi at my house. Useless!	-0.0036	0.4519	0.1666	0.3666

The subjectivity and sentiment score are being computed in the proposed HLESV model and the result obtained has been compared with text blob library which also provides native API to calculate the opinion and subjectivity score. Text blob (Bonta,2019) is popular python library designed on lexicon-based approach and is often used for text mining and processing of the textual data.

In this paper, we evaluate the information retrieval metric for our HLESV model along with most common ensemble classifier like Bagging Ensemble, Boosting Ensemble, Decision Tree respectively and the comparative analysis has been tabulated in below Table 2:

Table 2: Comparative Analysis Using Precision, Recall and f1-Score for HLESV

DATASET	ALGORITM	PRECISION	RECALL	F1-SCORE
Kindle	Boosting	0.52	0.42	0.46
	Bagging	0.57	0.41	0.48
	Decision Tree	0.61	0.40	0.49
	HLESV	**0.79**	**0.42**	**0.55**
Gift-cards	Boosting	0.73	0.39	0.51
	Bagging	0.84	0.57	0.68
	Decision Tree	0.71	0.38	0.49
	HLESV	**0.91**	**0.56**	**0.69**
Electronic Products	Boosting	0.72	0.52	0.62
	Bagging	0.61	0.54	0.57
	Decision Tree	0.59	0.51	0.55
	HLESV	**0.74**	**0.52**	**0.63**

Subjective score is a metric which is often used to measure such emotions in subjective sentences. The problem is that very few research works has been carried out based on subjectivity score for sentiment prediction. Also, existing state of the art methods of sentiment analysis broadly divides the sentiment into three categories namely positive, negative, and neutral. Hence the present study emphasizes on analysis of subjectivity score of consumer review data after considering text review content. Review Comments Sentiment.

The proposed HLESV model is based on ensemble machine learning algorithms which are designed to achieve better results in collaboration with of multiple classifiers. The ensemble set of classifiers used belongs to both supervised and unsupervised machine learning algorithms. Now the respective weights are being assigned to the classifiers based on performance of the respective individual classifier with more weights assigned to best performing classifier. Opinion polarity is calculated on review comments present in the training dataset and is passed as input to train the proposed HLESV model. Precision, Recall, F1-score are some of the existing standard performance measures for evaluation of the information retrieval system. The outcome of the HLESV has been evaluated using the Precision, Recall, F1-score respectively.

The precision is an IR measure of ratio of accurate predictions against total positive predictions. As referred in the above Table 2 that there is slight improvement in precision value for certain class labels on various classes. There is significant variation across on various dataset but it should be considered that such variation depends on the nature and size of data present in the dataset. The recall is an IR measure of ratio of accurate predictions against total predictions for respective class. It can be inferred from above Table 2 while evaluating with HLESV existing model, recall value has shown significant improvement for “negative” and “non-interpretive sentiment” label across all classes. The proposed HLESV model also overcomes the limitations of lazy learning due to weak classifiers by efficient management of assigning weights to the individual classifiers.

F1 score is an IR measure which is computed as weighted average of precision and recall score. As shown in above Table 2, of F1-score it is observed that our proposed HLESV model based on soft voting performs optimally over existing state of art ensemble model. The proposed HLESV model uses the "non-Interpretative sentiment" to classify the review comments which do not carry any meaningful opinion in review sentence. However, our proposed HLESV model needs to be trained with large number of datasets for improving more in its performance and achieve better accuracy in sentiment classification. It becomes necessary to use the ensemble learning with weighted voting so that we can assign less weight to the lazy and poor performing learners. The weighted average returns class label with maximum average probabilities from the ensemble of classifiers.

Sentiment Classification Outcome

Our proposed HLESV model was able to predict the sentiment associated to various review comments on Amazon Kindle and Electronic Product dataset respectively as shown below in Table 3.

Table 3: Example of Informative Consumer Review

Review Comments	Sentiment Classification
Works great. Love the portability of books. Overall great product	Positive
Great tablet for the price. It is easy to navigate.	Positive
Worse graphics, won't keep a wireless connection, overall, not satisfied	Negative
I regretted buying this item. I will never recommend this to anyone	Negative
Very slow device for most adults. probably fine for a child	Neutral
Very basic. If you need more than the basics, look for an Android tablet.	Neutral

The subjective matter of products reviews should always be logically pertaining to quality or utility value from consumer's viewpoint. It may also reflect his/her assessment on merits and demerits with ranking. Incidentally the extreme polar review comments are also eligible for good score of subjectivity and our model was also able to classify non-Interpretive reviews in Table 4.

Table 4: Example of Non-Interpretive Consumer Review

Review Comments	Sentiment Classification
Sometime it just stays in sleep mode and you have to restart it	Non-Interpretive
I had to buy it, what else is there to say.	Non-Interpretive
Couldn't figure it out	Non-Interpretive

But the reviews mentioned in Table 4 are totally deviated from the subjective matter. They hardly carry any sentiment opinion on product. Hence these natures of reviews are proposed in this research paper to be included a newly added class of "non-interpretive" form of reviews.

Performance Evaluation

The evaluation process of proposed HLESV model had been done using mainly by two metrics namely Accuracy and Receiving Operating Characteristic Curve to assess overall performance and effectiveness on various datasets.

- **Accuracy**

During evaluation of performance for proposed HLESV model in comparison to existing ensemble-based mode, it is observed that there is marginal increase in the accuracy score. As referred in the below Figure 4, our proposed HLESV model has better accuracy score of more than 70% across various datasets. Existing ensemble classifier used for comparison is based on majority voting, this classifier usually predicts target class labels based on mode of the predicted individual class labels.

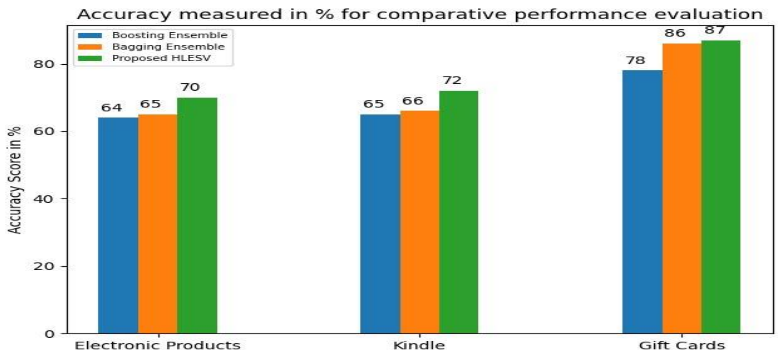

Figure 4: Performance Evaluation Using Accuracy Score (in %)

- **Receiver Operating Characteristic Curve**

The proposed HLESV model is also being evaluated using the ROC curve on kindle and Gift Cards dataset on multi class labels namely Positive, Neutral, And Negative and non-interpretive Sentiment to assess its effectiveness for sentiment classification as shown in the below Figure 5 and Figure 6. ROC curve is a probability-based curve between TPR and NPR at various thresholds. Steeper curve between TPR and FPR explains about effectiveness of the classification system.

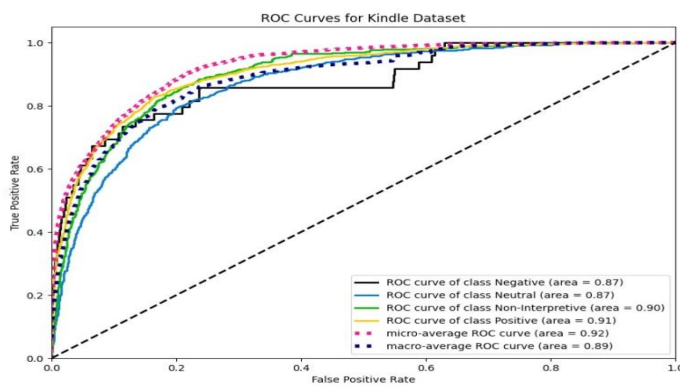

Figure 5: ROC Curve of Proposed HLESV Using Kindle Dataset

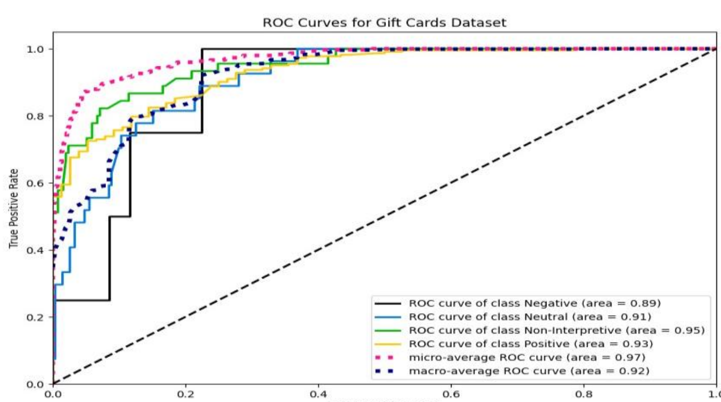

Figure 6: ROC Curve of Proposed HLESV Using Gift Cards Dataset

As shown in the figure, area under curve (AUC) for positive, neutral, negative and Non-Interpretive Sentiment are 0.91, 0.87, 0.87 and 0.90 for kindle dataset and 0.93,0.91,0.89 and 0.95 for Gift Cards dataset respectively. The greater value of AUC explains better performance results.

5 Conclusion

The aim of this research is to propose a model namely HLESV combining both lexicon and ensemble learning using soft voting technique with proper distribution of weights to the classifier. After initial pre-processing, the features are extracted from review text using TF-IDF vectorizer word embeddings using combination of unigram and bigram text features. In this paper, an innovative approach has been proposed to compute subjectivity score to predict subjective expressions. The HLESV model is to achieve the below three objectives.

- To use subjectivity, score as additional measures for sentiment prediction
- To establish the sentiment classification of consumer reviews based on Hybrid Learning model
- To propose a new sentiment class namely Non-Interpretive Reviews

The proposed HLESV model is efficient and can perform better on large volume of data as tested on CEPR dataset. The main challenges of proposed HLESV model are to identify the efficient mechanism for assigning appropriate weights to the best performing classifier during the ensemble learning process.

6 Future Work

The topic of interest for future research will be to identify a mechanism to automatically assign the weight pattern for best performing classifiers. This paper does not handle complex sarcasm, negation, spam reviews. As future scope of research we will extend our efforts to further improve the design of current model and work on transformer deep learning models for handing spam and fake reviews. We will also focus to implement an efficient product recommendation system with collaborative filtering process by integrating the matrix factorization and tensor factorization along with other similarity metrics.

References

[1] Alfrjani, R., Osman, T., & Cosma, G. (2019). A hybrid semantic knowledgebase-machine learning approach for opinion mining. *Data & Knowledge Engineering*, *121*, 88-108.

[2] Alrehili, A., & Albalawi, K. (2019). Sentiment analysis of customer reviews using ensemble method. *In IEEE International conference on computer and information sciences (ICCIS)*, 1-6.

[3] Al-Saqqa, S., Obeid, N., & Awajan, A. (2018). Sentiment analysis for arabic text using ensemble learning. *In IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA).*

[4] Behera, R.N., Roy, M., & Dash, S. (2016). Ensemble based hybrid machine learning approach for sentiment classification-a review. *International Journal of Computer Applications*, *146*(6), 31-36.

[5] Bhoir, P., & Kolte, S. (2015). Sentiment analysis of movie reviews using lexicon approach. *In IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, 1-6.

[6] Bonta, V., & Janardhan, N.K.N. (2019). A comprehensive study on lexicon-based approaches for sentiment analysis. *Asian Journal of Computer Science and Technology*, *8*(S2), 1-6.

[7] Buitinck, L., Louppe, G., Blondel, M., Pedregosa, F., Mueller, A., Grisel, O., & Varoquaux, G. (2013). API design for machine learning software: experiences from the scikit-learn project.

[8] Cambria, E. (2013). An introduction to concept-level sentiment analysis. In *Advances in Soft Computing and Its Applications: 12th Mexican International Conference on Artificial Intelligence, MICAI, Mexico, Proceedings, Part II 12*, 478-483. Springer Berlin Heidelberg.

[9] Chathuranga, P.D.T., Lorensuhewa, S.A.S., & Kalyani, M.A.L. (2019). Sinhala sentiment analysis using corpus-based sentiment lexicon. *In IEEE 19th international conference on advances in ICT for emerging regions (ICTer)*, *250*, 1-7.

[10] Handhika, T., Fahrurozi, A., Sari, I., Lestari, D.P., & Zen, R.I.M. (2019). Hybrid Method for Sentiment Analysis Using Homogeneous Ensemble Classifier. *In IEEE 2nd international conference of computer and informatics engineering (IC2IE)*, 232-236.

[11] McAuley, J., Targett, C., Shi, Q., & Van Den Hengel, A. (2015). Image-based recommendations on styles and substitutes. *In Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval*, 43-52.

[12] Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, *5*(4), 1093-1113.

[13] Mukwazvure, A., & Supreethi, K.P. (2015). A hybrid approach to sentiment analysis of news comments. *In IEEE 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions)*, 1-6.

[14] Ohana, B., & Tierney, B. (2009). Sentiment classification of reviews using Senti Word Net. *Proceedings of IT&T*, *8*.

[15] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, *12*, 2825-2830.

[16] Permatasari, R.I., Fauzi, M.A., Adikara, P.P., & Sari, E.D.L. (2018). Twitter sentiment analysis of movie reviews using ensemble features based Naïve Bayes. *In IEEE International Conference on Sustainable Information Engineering and Technology (SIET)*, 92-95.

[17] Rajeswari, A.M., Mahalakshmi, M., Nithyashree, R., & Nalini, G. (2020). Sentiment analysis for predicting customer reviews using a hybrid approach. *In IEEE Advanced Computing and Communication Technologies for High Performance Applications (ACCTHPA)*, 200-205.

[18] Ravi, K., & Ravi, V. (2015). A survey on opinion mining and sentiment analysis: tasks, approaches and applications. *Knowledge-based systems*, *89*, 14-46.

[19] Rehioui, H., & Idrissi, A. (2019). New clustering algorithms for twitter sentiment analysis. *IEEE Systems Journal*, *14*(1), 530-537.

[20] Sadhasivam, J., & Kalivaradhan, R.B. (2019). Sentiment analysis of Amazon products using ensemble machine learning algorithm. *International Journal of Mathematical, Engineering and Management Sciences*, *4*(2), 508-520.

[21] Shelke, N.M., Deshpande, S., & Thakre, V. (2017). Exploiting expectation maximization algorithm for sentiment analysis of product reviews. *In IEEE International Conference on Inventive Communication and Computational Technologies (ICICCT)*, 390-396.

[22] Sridhar, S., & Sanagavarapu, S. (2020). Sentiment Analysis Using Ensemble-Hybrid Model with Hypernym Based Feature Engineering. *In IEEE FORTEI-International Conference on Electrical Engineering (FORTEI-ICEE)*, 13-18.

[23] Srinadi, N.L.P., Hermawan, D., & Jaya, A.A.N.A. (2023). Advancement of Banking and Financial Services Employing Artificial Intelligence and the Internet of Things. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications, 14*(1), 106-117.

[24] Thara, S., & Sidharth, S. (2017). Aspect based sentiment classication: Svd features. *In IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2370-2374.

[25] Wang, J., Sun, C., Li, S., Wang, J., Si, L., Zhang, M., & Zhou, G. (2019). Human-like decision making: Document-level aspect sentiment classification via hierarchical reinforcement learning.

[26] Whitehead, M., & Yaeger, L. (2010). Sentiment mining using ensemble classification models. *Innovations and advances in computer sciences and engineering*, *1*, 509-514.

[27] Yuan, M., Ouyang, Y., Xiong, Z., & Sheng, H. (2013). Sentiment classification of web review using association rules. *In Online Communities and Social Computing: 5th International conference, OCSC, Held as Part of HCI International, Las Vegas, NV, USA. Proceedings 5*, 442-450. Springer Berlin Heidelberg.

Dataset

[28] Electronic Products Dataset from http://archive.ics.uci.edu/ml/machine-learning-databases/00331/sentiment%20labelled%20sentences.zip

[29] Gift Cards Dataset from *https://nijianmo.github. io/amazon/index.html*

[30] Kindle Dataset from *https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products*

Authors Biography

Sourav Sinha received his B.E degree in Information Technology and M.E degree with distinction in Software Engineering from Birla Institute of Technology, Mesra, Ranchi, India in 2009 and 2011 respectively. He is currently pursuing towards his PhD degree at the B.S. Abdur Rahman Crescent Institute of Science and Technology Chennai, India. His areas of interest are Machine Learning, Artificial Intelligence, Information Retrieval and Text Mining.

Dr.S. Revathi obtained her B.E.(CSE) in 1994, M.E.(CSE) in 2000 and PhD in 2013. She is currently working as Professor in the Department of Computer Science and Engineering at B.S. Abdur Rahman Crescent Institute of Science and Technology, Chennai, India. She has published more than 50 peer reviewed journal and conference papers in the domain of Mobile Ad hoc Network, Internet of Things, Network Security, Data mining etc.