EFFECTS OF STRATIFIED CROSS-VALIDATION AND HYPERPARAMETER TUNING ON SENTIMENT CLASSIFICATION WITH THE CHI2-RFE HYBRID FEATURE SELECTION TECHNIQUE IN THE IMDB DATASET

Pankaj Kumar Gautam[1], Dr. Akhilesh A. Waoo[2]

[12] Department of Faculty of Computer Applications & Information Technology and Sciences (AKS University Satna, M.P., INDIA 485001)

Corresponding Author
Akhilesh A. Waoo,
akhileshwaoo@gmail.com

DOI
10.29121/shodhkosh.v5.i5.2024.1889

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

ABSTRACT

Data analysis from social networking sites provides government entities, businesses, and event planners with insights into public sentiments and perceptions. Sentiment analysis (SA) resolves this need by classifying the sentiment of social network users into multiple classes. Despite their usefulness, data from social networking platforms frequently exhibits challenges, including unstructured formats, high volume, and redundant or irrelevant information, which can cause issues like overfitting, underfitting, and the curse of dimensionality. In response to these challenges, this study proposes using the term frequency-inverse document frequency (TF-IDF) for feature extraction along with a hybrid feature selection method that combines Chi2 and recursive feature elimination (RFE), called Chi2-RFE. This approach seeks to identify the optimal feature subset by filtering out irrelevant and redundant features. The proposed method is tested with several classifiers, including KNN, LR, SVC, GNB, DT, and RFC, employing stratified K-fold cross-validation and hyperparameter tuning on an IMDb dataset obtained from Kaggle. By effectively addressing overfitting and underfitting issues, this approach shows that before using StratefiedKfold cross-validation and hyperparameter tuning, LR gives 0.81975 training accuracy and test accuracy 0.815 on training data. After the method mentioned above, overfitting is removed by enhancing accuracy to 0.864833 on test data. KNN also enhanced its test accuracy to 0.891667 from 0.857333. SVC from 0.846666 to 0.883667, and GNB from 0.809666 to 0.829583. Precision is also improved from 0.826 to 0.853 for LR, from 0.848 to 0.897 for KNN, from 0.852 to 0.868 for SVC, and from 0.809666 to 0.799 for GNB. Recall also shows improvement from 0.815 to 0.600 for LR, from 0.857 to 0.894 for KNN, from 0.847 to 0.873 for SVC, and from 0.810 to 0.815 for GNB. F1-score also increased from 0.764 to 0.600 for LR, from 0.843 to 0.883 for KNN, from 0.819 to 0.862 for SVC, and from 0.790 to 0.815 for GNB.

Keywords: Sentiment Analysis, Feature Selection, Hybrid Feature Selection Method, Pre-Processing, Machine Learning, Stratified K-Fold Cross-Validation, Hyperparameter Tuning, Chi2, Recursive Feature Elimination (RFE)

OPEN ACCESS

1. INTRODUCTION

The growth of communication technologies and the spread of high-speed internet have opened up global avenues for engagement on social networks (SN), enabling individuals from diverse cultural backgrounds to express their viewpoints and feelings on a range of topics [9]. IMDb is a popular social networking platform where users discuss their opinions on different movies. Analyzing this data allows organizations to gain a deeper understanding of individuals' sentiments and perceptions about movies. The swift increase in the number of social network users has caused data overflow, resulting in an overwhelming volume of comments and posts that makes it challenging for humans to extract precise and relevant information [7, 23, 25, 26].

The limitations of human analysis in dealing with online-generated text highlight the urgent need for automated approaches to extract valuable insights and hidden information. Sentiment analysis (SA) addresses this issue by using machine learning (ML), natural language processing (NLP), or deep learning (DL) techniques to classify social network (SN) users' sentiments into several distinct categories. [7, 8, 33, 34, 35,39].

Sentiment analysis (SA) has emerged as a prominent research focus because of the NLP challenges it presents and its practical relevance to business and society. These challenges include the integration of structured and unstructured data, the management of irrelevant and redundant features, and the prevention of overfitting and underfitting [1, 9, 11, 19, 21, 25, 27, 28, 29, 36].

When a model is too simplistic and fails to capture the underlying data trends, it leads to underfitting, resulting in poor performance on both training and test datasets. Conversely, overfitting happens when the model is excessively complex, fitting the noise in the training data and demonstrating high performance on the training data but failing to generalize to new, unseen data [3,11].

To achieve a well-generalized model, it is essential to balance performance on both training and unseen data. This is done by splitting the data into two sets: one for training and another for testing the model's effectiveness. The model is first trained using the training set and then evaluated on the test set. K-fold cross-validation refines this process by breaking the data into K-independent subsets. Each subset is used as the test set once, with the remaining K-1 subsets used for training. This approach is repeated K times, ensuring each subset acts as the test set one time [26].

The organization of the article is outlined below: Section 1: Introduction. Section 2: Literature review. Section 3: Methodology. Section 4: Data Collection. Section 5: Preprocessing. Section 6: Feature extraction Section 7: Feature selection. Section 8: Stratified k-fold cross-validation and Hyperparameter tuning. Section 9: Evaluation metrics. Section 10: Classifiers. Section 11: Results and discussion.

2. LITERATURE REVIEW

Nguyn has investigated the 5-Fold Cross Validation and Confusion Matrix to overcome underfitting and overfitting. A Vietnamese dataset of customer reviews of the hotel was used with a combination of the BoW technique and TF-IDF is used to construct feature vectors.SGD, Logistic Regression, SVM, Naïve Bayes, Random Forest, Decision Tree, and K-Neighbors classifiers with ensemble methods like Stacking, voting, bagging, and boosting are used to evaluate the model. It gives a maximum F1-score of 96.03 [37].

Sharma and Jain used a hybrid ensemble learning model that used a combination of information gain and CHI-squared feature selection methods and classifiers like Ada Boost with Logistic Regression and SMO-SVM. It works on Twitter data with a low error rate and an accuracy of 88.2% [1].

Parlak and Uysal explored the Extensive Feature Selector (EFS) feature selection method. This method is compared with Chi-Squared (CHI2), Discriminative Power Measure (DPM), Class Discriminating Measure (CDM), Odds Ratio (OR), Distinguishing Feature Selector (DFS), Comprehensively Measure Feature Selection (CMFS), Normalised Difference Measure (NDM), Discriminative Feature Selection (DFSS), and Max-Min Ratio (MMR) using Support-Vector Machines (SVMs), Multinomial Naive Bayes (MNB), and k-Nearest Neighbour (KNN) classifiers on four datasets: Reuters-21578, 20-Newsgroup, Mini 20-Newsgroup, and Polarity. EFS has performed better than the other nine state-of-the-art FS methods in terms of the highest scores. In the future, EFS may be explored in other domains of text mining [38].

Alhussan et al. developed a binary waterwheel plant algorithm (BWWPA) to select features. The BWWPA algorithm performs better than others among the 30 benchmark datasets in the UCI ML repository. Wilcoxon signed-rank and statistical one-way (ANOVA) tests have proved that the proposed algorithm outperforms other feature selection methods [2].

Benarafa et al. proposed WordNet with SVM to minimize overfitting and underfitting problems on product, restaurant, and laptop datasets. Moreover, it has improved SVM kernel computation and accuracy for supporting the IAI task through the use of WordNet semantic relations [6].

Popoola et al. studied one of the most popular feature extraction methods for sentiment analysis: term frequency and inverse document frequency (TF-IDF) from Twitter data. NB, KNN, and RF classifiers were used in this experiment. The result shows that both RF and NB outperform KNN, with an accuracy of 0.81 [7].

3. METHODOLOGY

The study's process, as illustrated in Fig. 1, encompasses data collection and progresses to the evaluation of the classification model.

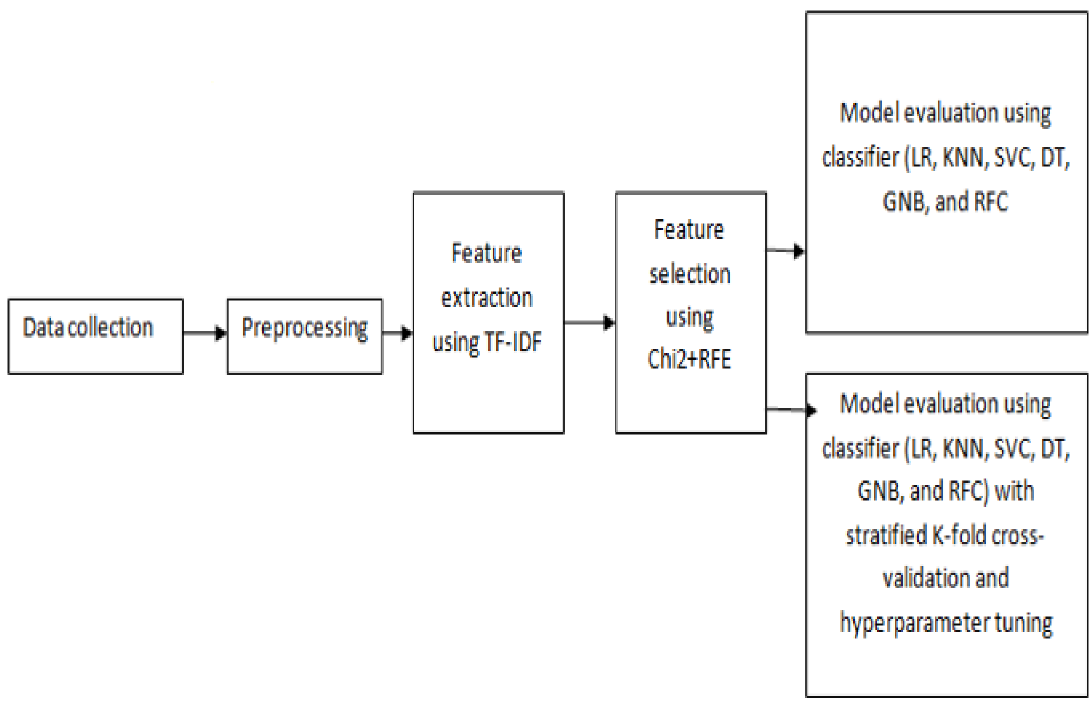

Fig.1 Proposed system framework.

4. DATA COLLECTION

In this study, we utilized the IMDb Movie Reviews dataset, which was downloaded from https://www.kaggle.com/code/nileshely/imdb-movie-data-genres-descriptions-emotions/ in CSV format. The dataset has 46,173 user reviews and ratings for movies (see Fig. 2). The "Rating" column represents the score given by the user to each film, while the sentiment column divides the review into positive, negative, or neutral. Ratings of 4 and 5 come under the positive sentiment, ratings of 1 and 2 come under the negative sentiment, and a rating of 3 denotes a neutral sentiment. To manage computational resources, we have taken 15,000 reviews from the dataset, which includes 11456 positive, 2221 negative, and 1323 neutral reviews (see Fig. 2 and Table 1).

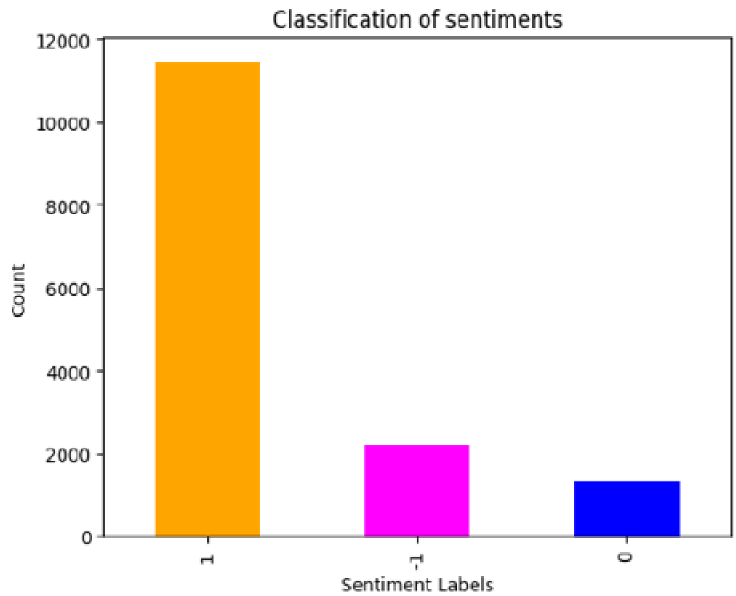

Fig.2 Bar plot of sentiment score

Table 1 Description of the Dataset

Dataset	Reviews	Class	Category	No of Reviews
IMDb Movie Reviews dataset	15000	Positive Negative Neutral	Movie Reviews	Positive: 11456 Negative: 2221 Neutral: 1323

5. PREPROCESSING

Data preprocessing is vital for improving data quality and preparing it for analysis. This process includes reducing noise, standardizing text, correcting spelling errors, normalizing words, and converting text into a numerical format that machines can understand [3, 8, 10].

Punctuation Marks and Number Elimination: Text often contains various punctuation marks such as periods, commas, quotation marks, exclamation marks, question marks, hyphens, and apostrophes, as well as numbers [7]. Since these elements do not influence the polarity of sentences, they must be removed during preprocessing [19, 22, 23].

Case Conversion: In text processing, words in uppercase and lowercase convey the same polarity. Case conversion standardizes text by converting all uppercase words to lowercase, thereby reducing complexity in the analysis process [7, 22].

URL Elimination: URLs are excluded from text processing because they do not impact the polarity of the content, thus helping to focus on the relevant textual information [10].

Lemmatization: This method reduces words to their base form, or lemma, which is the standard dictionary representation of the word. During lemmatization, the part of speech (POS) is considered to find the appropriate lemma, as different POS categories influence the base form [8, 10, 19].

Stop Word Removal: In text processing, stop words are common words like articles, prepositions, and pronouns that do not contribute to the sentiment of the text. However, some stop words, such as “don’t,” are important in phrases like “don’t like,” “don’t try,” and “don’t play” and should not be removed [10, 14, 23].

Tokenization: Tokenization involves dividing text into a set of distinct, meaningful elements known as tokens. This process breaks the text into smaller parts that are used to construct the vocabulary for text analysis [3, 8, 10, 22].

6. FEATURE EXTRACTION

TF-IDF (Term Frequency-Inverse Document Frequency) is a popular feature extraction method. In this approach, TF (Term Frequency) measures how often a word appears in a specific text, while IDF (Inverse Document Frequency) evaluates how common or rare the word is across a collection of texts. The core concept of TF-IDF is that words that occur frequently in a single document but infrequently across others are considered more significant, as they provide better indicators for classification and relevance. The TF-IDF algorithm is employed to quantify the relevance of each word in a document by computing a weight value that reflects both the Term Frequency (TF) of a word within a specific text and the Inverse Document Frequency (IDF) of the word across a collection of documents. TF represents the frequency of a term's occurrence in a given text, while IDF measures the term's significance about its frequency across the entire document corpus [37].

Let us assume one document contains N texts. Text s has word w whose weight is calculated by as under

$$tsw = tf\ w \times idf\ w = tf\ w \times \log(N / Nw) \qquad \text{eq (1)}$$

7. FEATURE SELECTION

To boost the effectiveness of classifiers, feature selection is a key pre-processing step aimed at identifying and retaining the most relevant features while removing those that do not contribute [2, 22, 24]. This process also decreases computational complexity [24]. The main techniques for feature selection are wrapper methods and filter methods [2]. The wrapper method relies on learning algorithms to obtain the feature subsets. It gives higher performance than a filter, but it is computationally expensive. On the other side, filters select features regardless of the learning algorithm, like chi2, information gain, mutual information, one-way ANOVA, etc. [2, 22, 24, 28].

Feature selection is commonly approached through three main methods: filter, wrapper, and embedded. The filter method evaluates features based on statistical measures related to the dataset's characteristics, assigning scores to features without considering the classifier's performance. The wrapper method assesses the effectiveness of feature subsets by measuring their impact on improving classification accuracy. The embedded method incorporates feature selection into the model training process, where the search for the best feature subset occurs concurrently with classifier construction [24, 28].

7.1 Chi2

Chi2 is a filter feature selection method. The Chi2 method measures the divergence between the actual frequency of a feature and the frequency that would be expected under the assumption that the feature's occurrence is independent of the class Cj [38].

7.2 Recursive Feature Elimination

Recursive Feature Elimination (RFE) is one of the best wrapper feature selection methods in its category. It works on a brute force technique and searches a subset of features. Start with all features in the training dataset and gradually remove the weakest one until it reaches the specified number of features. Model characteristics play a significant role in determining the feature category in RFE. Using iterative feature removal [30], RFE removes a specified number of features in each cycle and systematically eliminates dependencies and co-linearity from the model.

8. STRATIFIED K-FOLD CROSS-VALIDATION AND HYPERPARAMETER TUNING

Hyperparameters are special parameters that control the learning capability and speed of the algorithm. Each ML algorithm has a specific set of hyperparameters that need to be initialized with a predetermined set of values to enhance the quality of the algorithm. While these are not directly part of the model, they influence the quality of the model. Hyperparameters should be decided before the training phase starts [5, 14, 18, 19, 20, 21, 22].

Cross-validation works as a statistical technique to evaluate the performance of a classifier [22]. It divides the dataset into training and test datasets. The training dataset is used to train the classifier in the training phase, while the test dataset is used to evaluate the performance of the classifier in the evaluation phase. K-fold cross-validation partitions the training dataset into K-independent subsets. In each iteration, one subset will serve as a test set, the remaining k-1 will become a training set, and the test set will change in each iteration.

9. EVALUATION METRICS

The evaluation matrices used in this study are accuracy, precision, recall, and F1-score. These are calculated using the following equations [4, 11, 13, 14, 15, 16, 17, 19, 23].

The Confusion Matrix is a crucial instrument for evaluating classification models, as it requires the actual values for comparison with the predicted values to assess model performance effectively (see Table 2) [5].

Table 2 Laws from the confusion matrix

		Predicted	
		Negative	**Positive**
Actual	**Negative**	True Negative	False Positive
	Positive	False Negative	True Positive

Accuracy=(Σ true prediction)/(Σ test set)*100 eq. (2)

Precision=(true positive)/(true positive + false positive)*100 eq. (3)

Recall=(true positive)/(true positive + false negative)*100 eq. (4)

F1-score=(2x Recall * Precision)/(Recall+Precision) eq. (5)

10. CLASSIFIER

Various classifiers (SVM, GNB, KNN, LR, DT, and RFC) are used in this study.

Support Vector Machine (SVM)

SVM is a widely used supervised ML method that is used in classification, regression, and clustering. SVM uses a plane called a hyperplane to divide the features with the maximum margin in a plane [5, 9, 13]. SVC is capable of handling non-linear data and handling data when conventional ML models are unable to handle these data [13, 19, 23, 25].

Gaussian Naïve Bayes (GNB)

GNB is a classifier that works on the principle of Bayes theorem. Naïve Bayes determines the class of the input data by evaluating a specific set of attributes or parameters. It is mathematically denoted by.

$$P(A/B)=P(B/A)\ P(A)/P(B) \qquad \text{eq. (6)}$$

Where P (B/A) represents the posterior probability and P (A) is the representation of the prior probability [13].

k-nearest Neighbor (KNN)

The k-nearest neighbor classifier uses the theory that closed data points should be in the same class and classifies newly input data using the data that has already been categorized. Learning samples are data that has been previously categorized while test samples are unclassified data. KNN selects all test samples as learning k-samples, which are closed to test samples. This is done after finding the distance between the learning sample and the test sample [9].

Logistic regression (LR)

Logistic regression uses two values to determine the probability of outcome. Linear regression becomes ineffective with binary values such as true/false and yes/no because it generates a logistic curve on the value between 0 and 1. Logistic regression works on the logarithm of the target variable's odds. While linear regression builds its model using a probability curve [9].

$$\text{Logit}(p) = \log\left(\frac{p}{1-p}\right) \qquad \text{eq. (7)}$$

3.5 **Decision trees (DTs)**

The decision tree is an example of a hierarchical supervised machine learning model. It performs learning inductively from already known data classes to categorized data. It represents a tree-like structure where leaf nodes represent class labels while internal nodes represent test conditions. Data space is divided recursively until leaf nodes reach a certain number of records required for categorization [9].

11. RESULTS AND DISCUSSION

Table 3 Accuracy, Precision, Recall, and F1-Score comparison of six different classifiers using Chi2 + RFE without and with stratified K-fold cross-validation and hyperparameter tuning on the IMDb dataset

	Chi2+ RFE without stratified K fold cross-validation and hyperparameter tuning					Chi2 + RFE with stratified K fold cross-validation and hyperparameter tuning			
Classifiers	Training Data		Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
	Training Accuracy	Test Accuracy							
LR	0.81975	**0.815**	**0.826**	**0.815**	**0.764**	**0.864833**	**0.853**	**0.600**	**0.853**
KNN	0.89025	**0.857333**	**0.848**	**0.857**	**0.843**	**0.891667**	**0.897**	**0.894**	**0.883**
SVC	0.857666	**0.846666**	**0.852**	**0.847**	**0.819**	**0.883667**	**0.868**	**0.873**	**0.862**
DT	0.980166	0.877	0.871	0.877	0.872	0.799000	0.767	0.795	0.741
GNB	0.832833	**0.809666**	**0.794**	**0.810**	**0.790**	**0.829583**	**0.799**	**0.815**	**0.800**
RFC	0.980166	0.881	0.875	0.881	0.874	0.822667	0.849	0.824	0.780

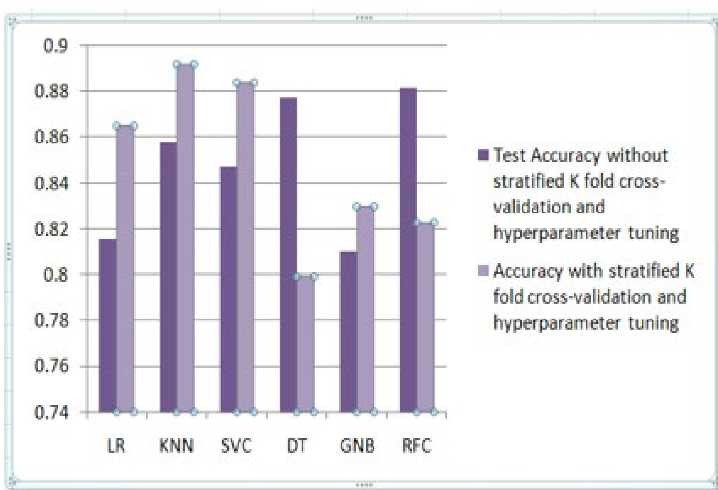

Fig. 3 The graph representation of the accuracy proposed Chi2 + RFE without StratefiedKfold cross-validation and hyperparameter tunning compared to Chi2 + RFE with StratefiedKfold cross-validation and hyperparameter tuning

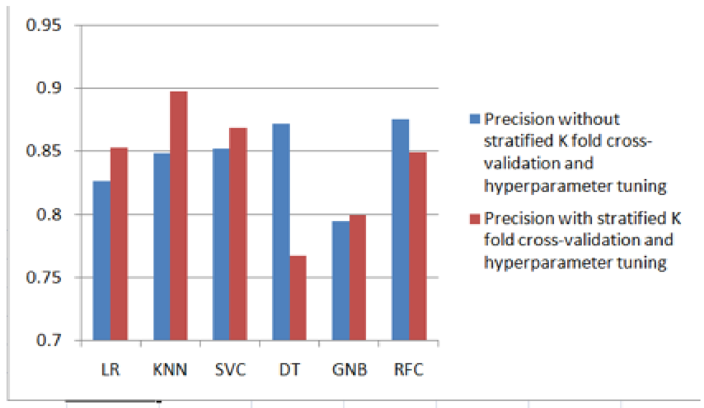

Fig. 4 The graph representation of the Precision proposed Chi2 + RFE without StratefiedKfold cross-validation and hyperparameter tunning compared to Chi2 + RFE with StratefiedKfold cross-validation and hyperparameter tuning.

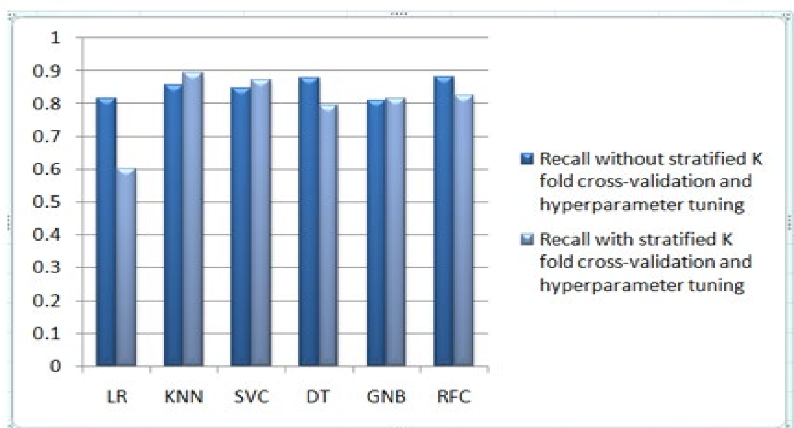

Fig. 5 The graph representation of the recall proposed Chi2 + RFE without StratefiedKfold cross-validation and hyperparameter tunning compared to Chi2 + RFE with StratefiedKfold cross-validation and hyperparameter tuning.

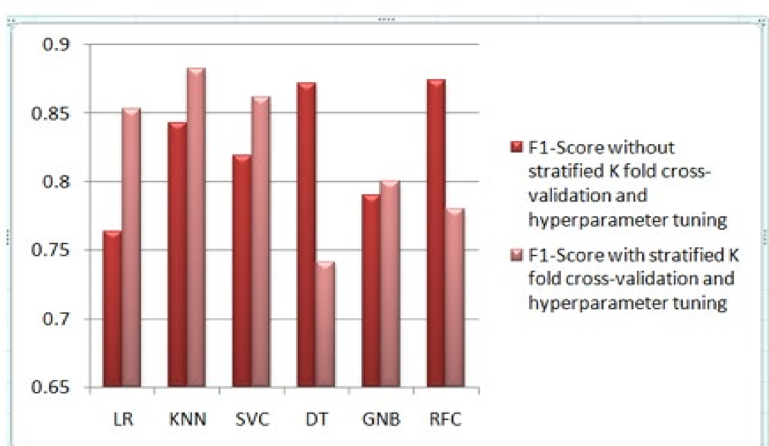

Fig. 5 The graph representation of the f1-score proposed Chi2 + RFE without StratefiedKfold cross-validation and hyperparameter tunning compared to Chi2 + RFE with StratefiedKfold cross-validation and hyperparameter tuning

Table 4 Hyperparameters of six classifiers and the potential values they can obtain during the tuning phase for finding the best results

Classifiers	Chi2 + RFE hybrid feature selection method with hyperparameters	Accuracy
LR	{'C': 100, 'penalty': '12', 'solver': 'libniear'}	0.864833
KNN	{'metric': 'euclidean', 'n_neighbors': 17, 'weights': 'distance'}	0.891667
SVC	{'C': 10, 'gamma': 'scale', 'kernel': 'rbf'}	0.883667
DT	{'criterion': 'entropy', 'max_depth': 15, 'max_features': 'sqrt', 'min_samples_leaf': 3, 'min_samples_split': 20}	0.799000
GNB	{'var_smoothing': 0.01}	0.829583
RFC	{'criterion': 'entropy', 'max_depth': 19, 'max_features': 'sqrt', 'n_estimators': 250}	0.822667

In this study, we proposed a new hybrid feature selection method called Chi2-RFE. Besides, we have used StratefiedKfold cross-validation with 10-fold and hyperparameter tuning on the IMDB dataset with LR, KNN, SVC, DT, GNB, and RFC classifiers. The proposed method with StratefiedKfold cross-validation and hyperparameter tuning enhances the performance of LR, KNN, SVC, and GNB. Before using StratefiedKfold cross-validation and hyperparameter tuning, LR gives 0.81975 training accuracy and test accuracy 0.815 on training data. After the method mentioned above, overfitting is removed by enhancing accuracy to 0.864833 on test data. KNN also enhanced its test accuracy to 0.891667 from 0.857333. SVC from 0.846666 to 0.883667, and GNB from 0.809666 to 0.829583. Precision is also improved from 0.826 to 0.853 for LR, from 0.848 to 0.897 for KNN, from 0.852 to 0.868 for SVC, and from 0.809666 to 0.799 for GNB. Recall also shows improvement from 0.815 to 0.600 for LR, from 0.857 to 0.894 for KNN, from 0.847 to 0.873 for SVC, and from 0.810 to 0.815 for GNB. F1-score also increased from 0.764 to 0.600 for LR, from 0.843 to 0.883 for KNN, from 0.819 to 0.862 for SVC, and from 0.790 to 0.815 for GNB (see in Table 3; Fig 3; Fig. 4; Fig. 5; Fig 6)

As a result, the hyperparameters for LR are {'C': 100, 'penalty': '12',' solver': 'libniear'} that give the highest accuracy of 0.864833, KNN {'metric': 'euclidean', 'n_neighbors': 17, 'weights': 'distance'} with an accuracy of 0.891667, and SVC {'C': 10. 'gamma':'scale', 'kernel': 'rbf'} with accuracy 0.883667, DT {'criterion': 'entropy', 'max_depth': 15, 'max_features':'sqrt', 'min_samples_leaf': 3, 'sqrt', 'min_samples_leaf': 3, 'min_samples_split': 20} with accuracy 0.799000, GNB {'var_smoothing': 0.01} with accuracy 0.829583, RFC {'criterion': 'entropy', 'max_depth': 19, 'max_features':'sqrt', 'n_estimators': 250} with accuracy 0.822667. DT and RFC do not improve performance after using StratefiedKfold cross-validation and hyperparameter tuning. As a future work, other feature selection methods can be used with these hyperparameters and StratefiedKfold cross-validation on other domains (see Table 4).

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCES

Sharma S, Jain A. Hybrid ensemble learning with feature selection for sentiment classification in social media. In Research Anthology on Applying Social Networking Strategies to Classrooms and Libraries 2023 (pp. 1183-1203). IGI Global.

Alhussan AA, Abdelhamid AA, El-Kenawy ES, Ibrahim A, Eid MM, Khafaga DS, Ahmed AE. A binary waterwheel plant optimization algorithm for feature selection. IEEE Access. 2023 Sep 4.

Ahmad S, Asghar MZ, Alotaibi FM, Awan I. Detection and classification of social media-based extremist affiliations using sentiment analysis techniques. Human-centric Computing and Information Sciences. 2019 Dec;9:1-23.

Razaque A, Frej MB, Bektemyssova G, Almi'ani M, Amsaad F, Alotaibi A, Jhanjhi NZ, Ali M, Amanzholova S, Alshammari M. Quality of Service Generalization using Parallel Turing Integration Paradigm to Support Machine Learning. Electronics. 2023 Feb 25;12(5):1129.

Sandeep MS, Tiprak K, Kaewunruen S, Pheinsusom P, Pansuk W. Shear strength prediction of reinforced concrete beams using machine learning. InStructures 2023 Jan 1 (Vol. 47, pp. 1196-1211). Elsevier.

Benarafa H, Benkhalifa M, Akhloufi M. An enhanced SVM model for implicit aspect identification in sentiment analysis. International Journal of Advanced Computer Science and Applications. 2023;14(5).

Popoola G, Abdullah KK, Fuhnwi GS, Agbaje J. Sentiment Analysis of Financial News Data using TF-IDF and Machine Learning Algorithms. In2024 IEEE 3rd International Conference on AI in Cybersecurity (ICAIC) 2024 Feb 7 (pp. 1-6). IEEE.

Safira W, Prabaswara B, Stevens Karnyoto A, Pardamean B. Leveraging ALBERT for Sentiment Classification of Long-Form ChatGPT Reviews on Twitter. International Journal of Computing and Digital Systems. 2024 Feb 24;16(1):1-1.

Cam H, Cam AV, Demirel U, Ahmed S. Sentiment analysis of financial Twitter posts on Twitter with the machine learning classifiers. Heliyon. 2024 Jan 15;10(1).

Thomas R, Jeba JR. A novel framework for an intelligent deep learning-based product recommendation system using sentiment analysis (SA). Automatika. 2024 Apr 2;65(2):410-24.

Bhargava R, Arora S, Sharma Y. Neural network-based architecture for sentiment analysis in Indian languages. Journal of Intelligent Systems. 2019 Jul 26;28(3):361-75.

Chen C, Guo X, Zhang W, Zhao Y, Wang B, Ma B, Wei D. Application of GA-WELM Model Based on Stratified Cross-Validation in Intrusion Detection. Symmetry. 2023 Sep 7;15(9):1719.

Kaliappan J, Bagepalli AR, Almal S, Mishra R, Hu YC, Srinivasan K. Impact of Cross-validation on Machine Learning models for early detection of intrauterine fetal demise. Diagnostics. 2023 May 10;13(10):1692.

Elgeldawi E, Sayed A, Galal AR, Zaki AM. Hyperparameter tuning for machine learning algorithms used for Arabic sentiment analysis. InInformatics 2021 Nov 17 (Vol. 8, No. 4, p. 79). MDPI.

Safira W, Prabaswara B, Stevens Karnyoto A, Pardamean B. Leveraging ALBERT for Sentiment Classification of Long-Form ChatGPT Reviews on Twitter. International Journal of Computing and Digital Systems. 2024 Feb 24;16(1):1-1.

Yu Y, Qiu D, Wan H. Sentiment Analysis Method of Epidemic-related Microblog Based on Hesitation Theory. ACM Transactions on Asian and Low-Resource Language Information Processing. 2024 Apr 15;23(4):1-25.

Omuya EO, Okeyo G, Kimwele M. Sentiment analysis on social media tweets using dimensionality reduction and natural language processing. Engineering Reports. 2023 Mar;5(3):e12579.

Rao NN. Iterative Ensemble Learning over High Dimensional Data for Sentiment Analysis. Scalable Computing: Practice and Experience. 2024 Feb 24;25(2):1219-34.

Gharaibeh H, Al Mamlook RE, Samara G, Nasayreh A, Smadi S, Nahar KM, Aljaidi M, Al-Daoud E, Gharaibeh M, Abualigah L. Arabic sentiment analysis of Monkeypox using deep neural network and optimized hyperparameters of machine learning algorithms. Social Network Analysis and Mining. 2024 Jan 24;14(1):30.

Zhao X, Peng H, Dai Q, Bai X, Peng H, Liu Y, Guo Q, Yu PS. Rdgcn: Reinforced dependency graph convolutional network for aspect-based sentiment analysis. In Proceedings of the 17th ACM International Conference on Web Search and Data Mining 2024 Mar 4 (pp. 976-984).

Amin MS, Ayon EH, Ghosh BP, MD MS, Bhuiyan MS, Jewel RM, Linkon AA. Harmonizing Macro-Financial Factors and Twitter Sentiment Analysis in Forecasting Stock Market Trends. Journal of Computer Science and Technology Studies. 2024 Jan 7;6(1):58-67.

Aghware FO, Yoro RE, Ejeh PO, Odiakaose CC, Emordi FU, Ojugo AA. Sentiment analysis in detecting sophistication and degradation cues in malicious web content. Kongzhi yu Juece/Control Decis. 2023;38(01):653-65.

Haque R, Islam N, Tasneem M, Das AK. Multi-class sentiment classification on Bengali social media comments using machine learning. International journal of cognitive computing in engineering. 2023 Jun 1;4:21-35.

Ihianle IK, Machado P, Owa K, Adama DA, Otuka R, Lotfi A. Minimising redundancy, maximising relevance: HRV feature selection for stress classification. Expert Systems with Applications. 2024 Apr 1;239:122490.

Khan TA, Sadiq R, Shahid Z, Alam MM, Su'ud MB. Sentiment Analysis using Support Vector Machine and Random Forest. Journal of Informatics and Web Engineering. 2024 Feb 14;3(1):67-75.

Asri AM, Ahmad SR, Yusop NM. Feature Selection using Particle Swarm Optimization for Sentiment Analysis of Drug Reviews. International Journal of Advanced Computer Science and Applications. 2023;14(5).

Nguyen-Thanh T, Tran GT. Vietnamese sentiment analysis for hotel review based on overfitting training and ensemble learning. In Proceedings of the 10th International Symposium on Information and Communication Technology 2019 Dec 4 (pp. 147-153).

Li X, Zhang J, Safara F. Improving the accuracy of diabetes diagnosis applications through a hybrid feature selection algorithm. Neural processing letters. 2023 Feb;55(1):153-69.

Alassaf M, Qamar AM. Improving sentiment analysis of Arabic Tweets by One-way ANOVA. Journal of King Saud University-Computer and Information Sciences. 2022 Jun 1;34(6):2849-59.

Yin Y, Jang-Jaccard J, Xu W, Singh A, Zhu J, Sabrina F, Kwak J. IGRF-RFE: a hybrid feature selection method for MLP-based network intrusion detection on UNSW-NB15 dataset. Journal of Big Data. 2023 Feb 5;10(1):15.

AbdulGani AF, Al Ahmad M. Autoregressive parametric modeling combined ANOVA approach for label-free-based cancerous and normal cells discrimination. Heliyon. 2021 May 1;7(5).

ACCENTURE GLOBAL SOLUTIONS LIMITED. 2024. TARGET IDENTIFICATION USING BIG DATA AND MACHINE LEARNING.INDIA Patent 512287, filed Feb. 6, 2018, and issued Feb. 19, 2024.

Mohammed Hadi, Michal Koblas, Saeed Shoaraee. 2021. SENTIMENT ANALYSIS. US 2020/0065383 A1, filed Aug 24, 2018, and issued Mar 23, 2021.

Lotfi A Zadeh, Saied Tadayon, BijanTadayon. 2024. System and method for extremely efficient image and pattern recognition and artificial intelligence platform. US 11914674 B2, filed Dec 6, 2021, and issued Feb 27, 2024.

Alsayat A. Improving sentiment analysis for social media applications using an ensemble deep learning language model. Arabian Journal for Science and Engineering. 2022 Feb;47(2):2499-511.

Rani S, Kumar P. Deep learning based sentiment analysis using convolution neural network. Arabian Journal for Science and Engineering. 2019 Apr 1;44:3305-14.

Liu, Hao, Xi Chen, and Xiaoxiao Liu. A study of the application of weight distributing method combining sentiment dictionary and TF-IDF for text sentiment analysis.IEEE Access 10 (2022): 32280-32289.

Parlak B, Uysal A. K. (2023). A novel filter feature selection method for text classification: Extensive Feature Selector. Journal of Information Science, 49(1), 59-78. https://doi.org/10.1177/0165551521991037

Sharma, S , Waoo, AA, (2023) Decision Tree Machine Learning Approach for Customer Behavior Analysis on Online Product Review, Journal of Em.erging Technologies and Innovative Research (JETIR), Volume 10, Issue 3 www.jetir.org (ISSN-2349-5162)