Frontiers in Health Informatics www.healthinformaticsjournal.com
ISSN-Online: 2676-7104
2024; Vol 13: Issue 3 Open Access

Improved Hybrid Feature Selection Approach for Sentiment Classification: Integrating Chi-Square and Recursive Feature Elimination

Pankaj Kumar Gautam[1] (0000-0003-2918-8855), **Akhilesh A. Waoo**[2](0000-0001-6788-710X)

[1,2] Department of Faculty of Computer Applications & Information Technology and Sciences
(AKS University Satna, M.P., INDIA 485001)
Corresponding author's email: [1]pankajgautam82@gmail.com, [2]akhileshwaoo@gmail.com

Cite this paper as: Pankaj Kumar Gautam, Akhilesh A. Waoo (2024) Improved Hybrid Feature Selection Approach for Sentiment Classification: Integrating Chi-Square and Recursive Feature Elimination. *Frontiers in Health Informatics, 13* (3), 10570-10582

Abstract: *Feature selection process select important features that participate in deciding the sentiment of the text and enhance the classification accuracy. Reducing dimensionality, overfitting and underfitting also enhance precision, recall, and F1 score. FS also reduce complexity, storage, and computing time. In this paper, combination of Chi2 and recursive feature elimination is used as a hybrid feature selection method on Amazon review dataset. Three other state of the art feature selection methodsGenetic Algorithm (GA), Mutual Information (MI), and Principal Component Analysis PCA with six classifiers (like Random Forest Classifier (RFC),Logistic Regression (LR),K-Nearest Neighbor (KNN),Linear Support Vector Classifier (Linear SVC), Naïve Bayes (NB),Decision Tree (DT)) are used in this study. Chi2+RFE and MI with 50 percent (10936 features) feature selection methods have given improved accuracy, precision, recall, and f1-score concerning the base condition, where all features (21873 features) are included as well as the above-mentioned classifiers, it demonstrates that Chi2+RFC gives 0.821 maximum accuracy with the LR classifier, 0.821 maximum recall with the LR classifier, and 0.742 maximum f1-score with the DT classifier. Chi2+RFE performs better than other other FS techniques in terms of accuracy, precision, recall, and f1-score.*

Keywords: sentiment analysis; pre-processing; feature selection; hybrid feature selection method; Chi2; recursive feature elimination (RFE).

1. Introduction

Sentiment analysis, or opinion mining, is the process of extracting opinions, reviews, feedback, comments, or emotions from a social networking site or online platform [7, 25, 27]. This data is unstructured, contains stop words, numbers, and URLs, is huge in amount, and must be processed systematically to represent it in positive, neutral, or negative. Manual processing of this data is not possible. SA does this job. Companies and governments use these results to decide policy and consumer behavior [8, 9, 10, 13, 26].Three levels employ sentiment analysis: aspect level, sentence level, and document level.This study used sentence-level SA.

Feature extraction and feature selection are the two initial steps in SA.In word level SA, the smallest unit is the word whose sentiment is found. In sentence level, the smallest unit is a sentence. At the document level, the smallest unit is the document [8].

SA commonly categories into two categories: lexicon-based and supervised. In the lexicon-based method, each text word is mapped to its respective categories to find their sentiment in positive, negative, and neutral. The whole text sentiment is calculated by the sentiment of all these words [10]. In supervised method classification, the model learns from the training data's level, and based on this learning, the model decides the category of the testing data [1].

Feature extraction takes training data as an input and converts it into vector form. The machine understands this form. FE emphasizes the finding and extraction of phrases and emotions. On the other hand, FS selects only the most relevant and significant feature subset from the output subset of FE. FE and FS reduce complexity and storage and enhance accuracy [1].

Machine learning is a kind of supervised method [24, 26]. In ML, three FS methods are filter, wrapper, and hybrid. Filter methods apply feature statistics to decide feature relevance, like Chi2, mutual information, and one-way ANOVA [2]. On the other side, the wrapper method is dependent on the predictive performance of the classifier [9].

The filter method suffers from accuracy while wrapping takes time [9]. A large dataset has a huge amount of redundant and irrelevant features that can decrease the performance and effectiveness of the model. This can create overfitting, underfitting, and the curse of dimensionality problems [4, 8].

The hybrid method is used to overcome the limitations of the above methods. It takes advantage of both methods by combining the simplicity of the filter method and the accuracy of the wrapper method.

In this study, the hybrid feature selection method (Chi2+RFE) of the combination of the Chi2 filter method and the recursive feature elimination (RFE) wrapper method is used to reduce the limitations of both techniques.

The Chi2 filter FS evaluates a feature's influence on a target class, while the RFE will select the significant features on the basis of the rank of the features. This hybrid feature selection method is compared with the other well-known methods.

This article is organized as follows: 1. Introduction; 2. Literature Review; 3. Methodology; 4. Word Cloud; 5. Feature Extraction; 6. Feature Selection; 7. Evaluation; 8. Classifiers; 9. Results and Discussion.

2. Literature Review

Sharma and Jain employed a hybrid ensemble learning approach integrating information gain and CHI-squared feature selection methods alongside classifiers such as Ada Boost with Logistic Regression and SMO-SVM. Their model operates on Twitter data, demonstrating a commendable accuracy of 88.2% and minimal error rate.

Nguyn explored 5-fold cross-validation and confusion matrix methods to mitigate underfitting and overfitting issues. They employed a Vietnamese dataset containing hotel customer reviews, utilizing a blend of Bag-of-Words (BoW) and TF-IDF techniques to construct feature vectors. Various classifiers, including Logistic Regression, Decision Tree, Naïve Bayes, SVM, Random Forest, SGD, and K-Neighbors were employed, alongside ensemble methods such as Stacking, Voting, Boosting, and Bagging to assess the model. Their approach achieved an impressive maximum F1-score of 96.03.

Khare et al. studied the TF-IDF feature extraction method and SVC for handling sarcastic tweets from the LokSabha election obtained from Twitter.

J. and U explored the possibility of Pearson correlation coefficient-based Harris Hawks optimization with the Recurrent Neural Network and Long Short Term Memory (PCCHHO-RNN-LSTM) technique to reduce the dimension of the data. Amazon data on the MATLAB software with an RNN-LSTM classifier gives 95.8% accuracy.

Yuce, Nielsen, and Wargocki studied the possibility of ANOVA, the Taguchi method, and Grey Relational Analysis (GRA) in construction to optimize CFD analyses of ventilation performance.

Chakraborty, Nawar, and Chowdhury investigated AdaBoost, Decision Tree (DT), Random Forest Classifier (RF), Naive Bayes (NB), Support Vector Machine (SVM), deep learning methods Long Short-Term Memory Network (LSTM), and Convolutional Neural Network for finding the polarity of Bengali Facebook posts and comments. LSTM enhances its accuracy with 96.95%, RF with 78.37%, and SVM with 78.23%.
Kausar, Fageeri, and Soosaimanickam proposed a novel bag-of-words feature extraction method based on the reviews obtained from Amazon to extract features. The word cloud method displays words at the frequency with which they appear in the review text. A decision tree and logistic regression classifier were used, and DT got the maximum frequency of 99%.

3. Methodology

The study's process is depicted in Fig. 1, starting with data collection and culminating in the evaluation of the classification model.

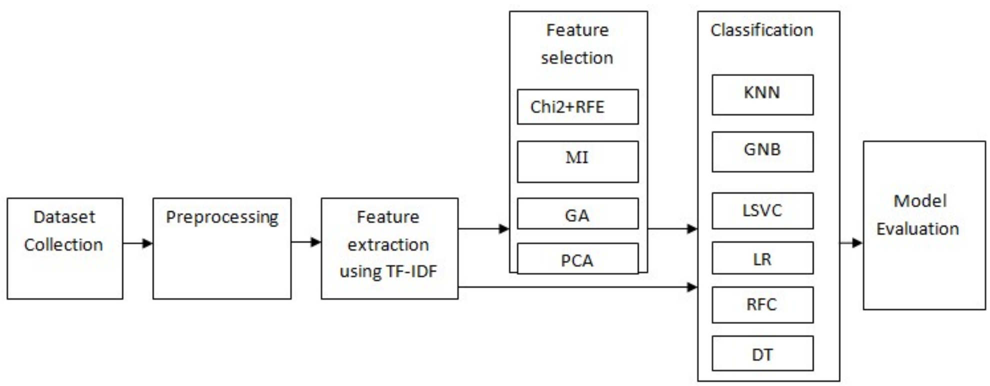

Fig.1 The framework for this study.

3.1 Data Collection

In this study, the Amazon dataset is used in CSV format, which is obtained from https://www.kaggle.com/datasets/snap/amazon-fine-food-reviews/. This dataset is composed of 568,454 user ratings and reviews on food products. One new column sentiment is added to denote the sentiment of reviews. Reviews with ratings of 4 and 5 are categorized as positive, ratings of 1 and 2 as negative and rating 3 as neutral. Due to power constraints, we have selected 12,000 reviews from this dataset, comprising 9,190 positive reviews, 1,811 negative reviews, and 999 neutral reviews (see Fig. 5 and Table 1).

Table 1. Description of the dataset

Dataset	Reviews	Class	Category	No of Reviews
Amazon	12000	Positive Negative Neutral	Food products	Positive: 9190 Negative: 1811 Neutral: 999

3.2 Data preprocessing

Analyzing the input dataset, we find noise, like misspellings, numbers, incorrect grammar, acronyms, URLs, slang, and stop words. Since these do not contribute to deciding sentiment, they must be removed to enhance classification accuracy [1, 30, 32].

This process will include the following steps:

Number and punctuation mark removal: Punctuation marks and numbers do not play a role in deciding the sentiment, so they must be removed [6, 11].

Case conversion: A conversation of lower and upper case words must be in the same case because they have the same sentiment [6].

URL removal: A URL is an element that does not decide the polarity, so it is removed [3, 6].

Stop word removal: prepositions, conjunctions, pronouns, articles, etc. are the stop word removals that are necessary. Tokenization: Tokens are the words, phrases, keywords, and symbols that make sentences. Before analysis, a sentence must be divided into tokens [6, 12].

Lemmatization: Words in different variants should be converted into base form. Dancing, dancing, and dancing should be converted into dance-based forms [7, 12].

4. Word cloud

The dataset contains various common words. Word cloud is a widely used visualization technique to display the most common words in the word cloud. Positive, negative, and neutral word clouds are used in this study

(see in Fig. 2, Fig. 3, and Fig. 4) [14].

Fig.2 World cloud for positive reviews.

Fig.3 World cloud for negative reviews.

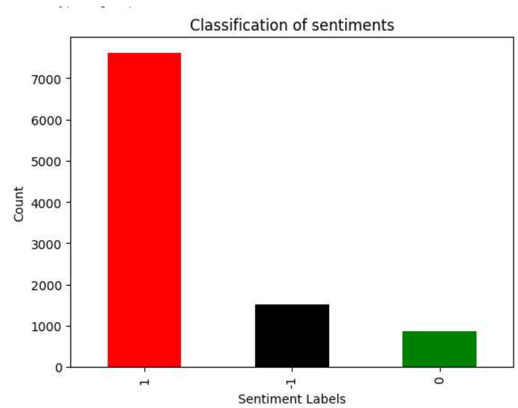

Fig.4 World cloud for neutral reviews.

Fig.5 Bar plot of sentiment counts

5. Feature Extraction

Raw input data must be converted into a machine-understandable format for processing. TF-IDF term frequency (TF) and inverse document frequency (IDF) are some of the best feature extraction methods [5, 15, 30].

Let's assume a document has N texts. In text s, composed of word w, whose weight is calculated by the as under:

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

6. Hybrid Feature Selection

Feature selection is a method to select optimum features that are used to construct an efficient classification model. FS helps in improving the prediction accuracy in sentiment analysis by using features that are necessary for prediction, leaving the unnecessary features. It is a very crucial task.

The hybrid method solves this problem by using the Chi2 filter method in combination with the recursive feature elimination (RFE) wrapper method. The output feature subset of the Chi2-RFE elimination method is passed to the classifier to evaluate the model [28].

6.1 Chi2

The chi2 filter feature selection method calculates the divergence between the actual frequency and the expected frequency of the feature. The occurrence of the feature is not dependent on the Cj class.Chi2 is calculated by the formula [31]:

$$Chi2(t, Cj) = \frac{N.[P(t,Cj).P(\bar{t},\overline{Cj}) - P(t,\overline{Cj}).P(\bar{t} - Cj)]^2}{P(t).P(\bar{t}).P(Cj).P(\overline{Cj})} \quad \text{eq. (2)}$$

6.2 Recursive feature elimination (RFE)

RFE is the best wrapper method in its category. The brute force method is used by RFE to find subsets of features. The training dataset includes all features, and RFE sequentially eliminates the weakest features until it obtains the mentioned number of features [28].
The RFE algorithm is described as understood [30].

Algorithm. Recursive Feature Elimination

Tune/Train the model on the training set using all predictors
Calculate model performance
Calculate variable importance or rankings
for each subset size S_i, $i=1,2,..$ Sdo
    Keep the S_i's most important variables
    Tune/Train the model on the training set using S_i, predictors
    Calculate model performance
end for
Calculate the performance profile over the S_i
Determine the appropriate number of predictors
Use the model corresponding to the optimal S_i

6.3 Mutual Information (MI)

Mutual Information is useful to find features that give the most information of the output class. Assume we have two random variables M and N, where M= {m1, m2,…,mk} and N={n1,n2,…,nk}, and k is the total number of samples. M and N share the quantity of knowledge that is known as MI. MI is calculated by the following formula [16]:

$$I(M, N) = \Sigma_{m\in \mathrm{M}}\Sigma_{\mathrm{n}\in \mathrm{N}} pd(m.n)log\frac{pd(m,n)}{pd(m)p(n)} \quad \text{eq. (3)}$$

6.4 Principal Component Analysis (PCA)

A higher-dimensional dataset is reduced to a lower-dimensional dataset, which is possible with the use of principal component analysis. PCA does this by obtaining principal components (PCs), which are uncorrelated feature data sets obtained, by conversion of interrelated feature data sets. The first few PCs represent most of the variations of the entire data set. PCs combine features linearly, where the first PC has the highest variations and the second PC is the highest orthogonal to the PC in terms of variance among all subsequent PCs. Further PCs follow the same pattern [18].

6.5 Genetic Algorithm (GA)

Natural evaluation of concepts is developed to obtain a probabilistic optimization technique formally known as genetic algorithms. In a genetic algorithm, the solution or response to the problem is the point in the search space that denotes chromosomes. Chromosomes are a set of genes. The set of chromosomes is the population. The solution to the problem can be found by developing fitness functions, which must be applied before using genetic algorithms.

7. Evaluation

Evaluation of the model's performance is essential. Four standard evaluation metrics, precision, recall, accuracy, and f1-score, are used for this purpose. Tp and Tn are positive and negative reviews for the correct model classification, respectively. At the same time, Fp and Fn are false positive and negative reviews for incorrect model classification [3, 7, 22, 23, 28, 32].

$$\text{precision} = \frac{\text{Tp}}{(\text{Tp+Fp})} \qquad \text{eq. (4)}$$

$$\text{recall} = \frac{\text{Tp}}{(\text{Tp+F}\ \ )} \qquad \text{eq. (5)}$$

$$\text{accuracy} = \frac{\text{Tp+Tn}}{(\text{Tp+Tn+Fn+Fp})} \qquad \text{eq. (6)}$$

$$\text{f1} - \text{score} = \frac{2\ (\text{recall} * \text{precision})}{\text{recall+ precision}} \qquad \text{eq. (7)}$$

8. Classifiers

In this study, KNN, GNB, LSVC, LR, RFC, and DT are six classifiers used:

8.1 K-Nearest Neighbor (KNN)

The K-Nearest Neighbour (KNN) algorithm is a highly effective but simple machine learning algorithm.KNN is used in both the classification and regression analysis.The input is assigned to the most suitable category based on its proximity to the nearest neighbors belonging to each class.A KNN classifier is used to find the optimal value of the number of neighbours parameter k by using grid search and cross-validation so that the performance of the model can be enhanced [17, 30].

8.2 Random Forest Classifier (RFC)

Random Forest Classifier is one of the most effective classifier for classification as well as regression. It is a tree of a variety of algorithms that can be utilized with decision trees. In this algorithm, more number of trees gives more performance and efficiency [17, 20].

8.3 Linear Support Vector Machines (LSVMs)

Linear support vector machine is a supervised machine learning algorithm that can be used in both classification and regression. A hyperplane is used to classify classes. As the number of dimensions increases, the efficiency of the SVM also increases. When the number of dimensions is higher than the samples, then also SVM works well. The cross-validation technique is used in SVM to enhance computational efficiency [21, 30].

8.4 Logistic regression (LR)

In classification problem when the target variable is categorical logistic regression supervised classification algorithm is used. Logistic regression models a function from dataset attributes to targets, predicting the probability that a new example belongs to specific classes. It is also referred to as Maximum Entropy in some contexts [17, 19, 20].

8.5 Gaussian Naive Bayes (GNB)

Gaussian Naive Bayes is a flavor of Naive Bayes. When the attribute data continues and consists of a numeric value, GNB calculates the probability. In the case of continuous data, the common assumption is that continuous values are correlated with each class, and Gaussian distributions are used to access these values. In GNB, training data is categorized based on classes, and the standard deviation and mean for each class is calculated. It is mathematically calculated by the formula [18, 29]:

$$P\left(\frac{x_i}{y}\right) = \frac{1}{\sqrt{2\pi\sigma^2 y}} \exp\left(-\frac{(x_i-\mu_y)}{2\sigma^2 y}\right) \qquad \text{eq. (8)}$$

8.6 Decision Tree (DT)

Decision tree is one of the well-known supervised machine learning classifier used to solve classification as well as regression problem. Training data is situated in the root of DT [5]. DT plays a significant role in representing choices and their respective results in the graph forms. They used predictive models to find the elevation of an item in branches and predict the target value of an item in leaves [30].

9. Result and discussion

Table 2. Precision comparison of six different classifiers.

Feature Selection Methods	KNN (%)	GNB (%)	LSVC (%)	LR (%)	RFC (%)	DT (%)
BASE (without feature selection)	0.713	0.660	0.816	0.783	0.824	0.728
Chi2+RFE	**0.681**	**0.670**	**0.811**	**0.796**	**0.900**	**0.740**
Mutual Informationwith 10% feature	0.670	0.620	0.798	0.796	0.820	0.739
Mutual Information with 20% feature	0.638	0.640	0.797	0.793	0.824	0.739
Mutual Information with 30% feature	0.679	0.669	0.809	0.794	0.825	0.740
Mutual Informationwith 50% feature	**0.724**	0.651	0.809	**0.797**	0.825	**0.749**
Genetic Algorithm	0.664	0.024	0.575	0.575	0.575	0.575
Principal Component Analysis	0.629	0.586	0.575	0.575	0.663	0.656

Table 3. Recall comparison of six different classifiers.

Feature Selection Methods	KNN (%)	GNB (%)	LSVC (%)	LR (%)	RFC (%)	DT (%)
BASE (without feature selection)	0.763	0.561	0.836	0.815	0.802	0.738
Chi2+RFE	**0.760**	**0.460**	**0.833**	**0.821**	**0.810**	**0.750**
Mutual_infornationwith 10% feature	0.758	0.175	0.824	0.823	0.811	0.742
Mutual_infornation with 20% feature	0.757	0.232	0.823	0.821	0.810	0.748
Mutual_infornation with 30% feature	0.759	0.305	0.832	0.822	0.806	0.750
Mutual_infornation with 50% feature	0.758	0.384	0.832	**0.823**	0.804	**0.759**
Genetic Algorithm	**0.760**	0.155	0.758	0.758	0.758	0.758
Principal component Analysis	0.734	0.733	0.758	0.758	0.723	0.712

Table 4. Accuracy comparison of six different classifiers.

Feature Selection Methods	KNN (%)	GNB (%)	LSVC (%)	LR (%)	RFC (%)	DT (%)
BASE (without feature selection)	0.763	0.561	0.836	0.815	0.802	0.738
Chi2+RFE	**0.760**	**0.460**	**0.833**	**0.821**	**0.811**	**0.750**
Mutual_infornationwith 10% feature	0.758	0.175	0.824	**0.823**	**0.811**	0.742
Mutual_infornation with 20% feature	0.757	0.232	0.823	0.821	0.810	0.748
Mutual_infornation with 30% feature	0.759	0.305	0.832	0.822	0.806	0.750
Mutual_infornation with 50% feature	0.758	0.384	0.832	**0.823**	0.804	**0.759**
Genetic Algorithm	**0.760**	0.155	0.758	0.758	0.758	0.758
Principal component Analysis	0.734	0.733	0.758	0.758	0.723	0.712

Table 5. f1-score comparison of six different classifiers.

Feature Selection Methods	KNN (%)	GNB (%)	LSVC (%)	LR (%)	RFC (%)	DT (%)
BASE (without feature selection)	0.763	0.561	0.836	0.815	0.802	0.738
Chi2+RFE	0.660	**0.522**	**0.820**	0.790	0.760	**0.742**
Mutual_infornationwith 10% feature	0.758	0.175	0.824	0.823	**0.811**	0.742
Mutual_infornation with 20% feature	0.757	0.232	0.823	0.821	0.810	0.748
Mutual_infornation with 30% feature	0.759	0.305	0.832	0.822	0.806	0.750
Mutual_infornation with 50% feature	0.758	0.384	**0.832**	**0.823**	0.804	**0.759**
Genetic Algorithm	**0.760**	0.155	0.758	0.758	0.758	0.758
Principal component Analysis	0.734	0.733	0.758	0.758	0.723	0.712

In this experiment, we introduce a novel hybrid feature selection method called Chi2-RFE. Classifiers experimented without FS (base) conditions and with FS methods. Chi2+RFE, Mutual Information with 10% feature (MI with 10), Mutual Information with 20% feature (MI with 20), Mutual Information with 50% feature (MI with 50), Genetic Algorithm (GA), and Principal Component Analysis (PCA) are the methods. Chi2+RFE and MI with 50 percent give maximum accuracy of 0.821, 0.823 on LR; Chi2+RFE gives 0.811 on RFC; Chi2+RFE and MI with 50 percent give maximum accuracy of 0.750, 0.759 on DT, while base conditions give accuracy of 0.815, 0.802, and 0.738 with LR, RFC, and DT, respectively.Chi2+RFE gives maximum precision of 0.670 on GNB; Chi2+RFE gives 0.900 precision on RFC; Chi2+RFE and MI with 50 percent give maximum precision of 0.796, 0.797 on LR; Chi2+RFE and MI with 50 percent give maximum precision of 0.740, 0.749 on DT, while base conditions give precision of 0.660, 0.783, 0.824, and 0.728 with GNB, LR, RFC, and DT, respectively. Chi2+RFE and MI with 50 percent give maximum recall of 0.821, 0.823 on LR; Chi2+RFE gives maximum recall of 0.810 on RFC; Chi2+RFE and MI with 50 percent give maximum recall of 0.750, 0.759 on DT; while base conditions give recall of 0.815, 0.802, and 0.738 with LR, RFC, and DT, respectively. MI with 50 percent gives a maximum f1-score of 0.823 on LR; MI with 10 percent gives a maximum f1-score of 0.811 on RFC; Chi2+RFE and MI with 50 percent give a maximum f1-score of 0.742 and 0.759 on DT; while base conditions give f1-scores of 0.815, 0.802, and 0.738 with LR, RFC, and DT, respectively (see in Fig 6, Fig 7, Fig 8, Fig 9).

As a result, Chi2+RFE and MI with 50 percent feature selection methods have given improved accuracy, precision, recall, and f1-score concerning the base condition, where all features are included as well as the above-mentioned classifiers, it demonstrates that Chi2+RFC gives 0.821 maximum accuracy with the LR classifier, 0.821 maximum recall with the LR classifier, and 0.742 maximum f1-score with the DT classifier. Chi2+RFE perform better than other FS techniques in terms of accuracy, precision, recall, and f1-score. As a future work, Chi2+RFE can be compared with other feature selection methods and domains (see Table 2,

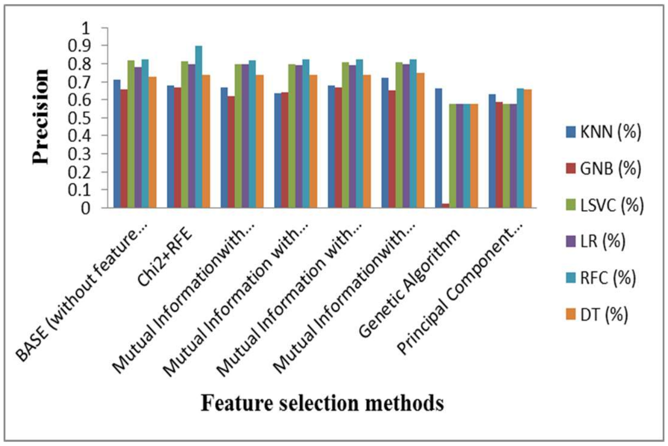

Table 3, Table 4, and Table 5).

Fig. 6 Precision comparison of six different classifiers.

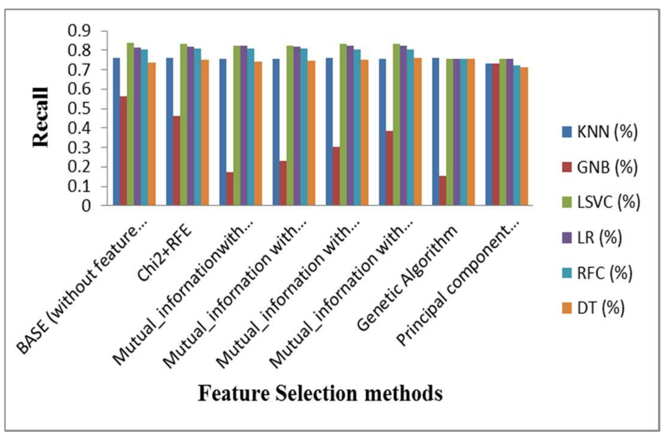

Fig. 7 Recall comparison of six different classifiers.

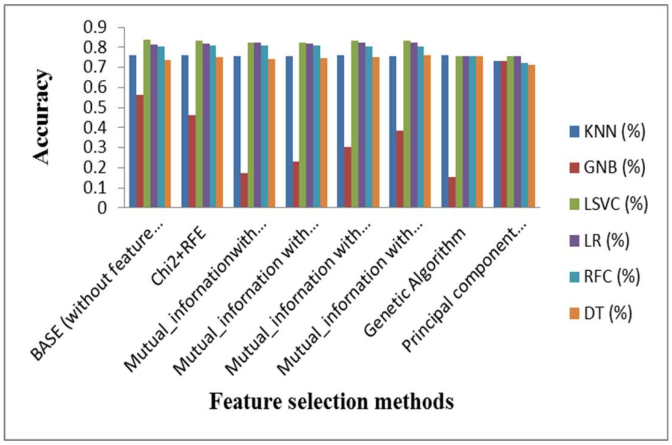

Fig. 8 Accuracy comparison of six different classifiers.

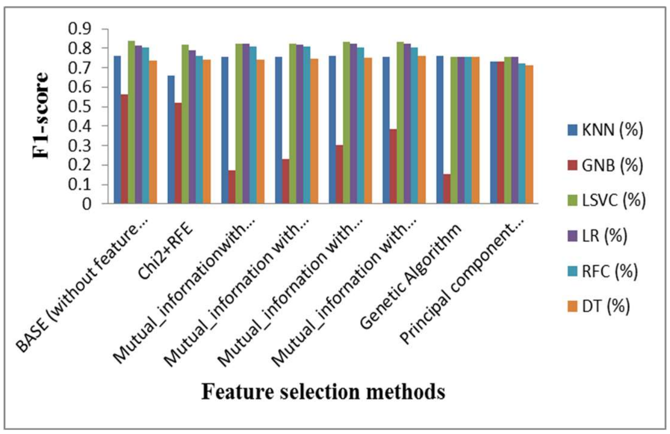

Fig. 9 f1-score comparison of six different classifiers.

References

1. Alassaf, M.,& Qamar, A. M. (2022). Improving sentiment analysis of Arabic tweets by One-Way ANOVA. *Journal of King Saud University-Computer and Information Sciences*, *34*(6), 2849-2859.
2. Yuce, B. E., Nielsen, P. V., & Wargocki, P. (2022). The use of Taguchi, ANOVA, and GRA methods to optimize CFD analyses of ventilation performance in buildings. *Building and Environment*, *225*, 109587.
3. Kausar, M. A., Fageeri, S. O., & Soosaimanickam, A. (2023). Sentiment Classification based on Machine Learning Approaches in Amazon Product Reviews. *Engineering, Technology & Applied Science Research*, *13*(3), 10849-10855.
4. Sangeetha, J., & Kumaran, U. (2023). Sentiment analysis of amazon user reviews using a hybrid approach. *Measurement: Sensors*, *27*, 100790.
5. Khare, A., Gangwar, A., Singh, S., & Prakash, S. (2023). Sentiment Analysis and Sarcasm Detection in Indian General Election Tweets. In *Research Advances in Intelligent Computing* (pp. 253-268). CRC Press.
6. Chakraborty, P., Nawar, F., & Chowdhury, H. A. (2022). Sentiment analysis of Bengali facebook data using classical and deep learning approaches. In *Innovation in Electrical Power Engineering, Communication, and Computing Technology: Proceedings of Second IEPCCT 2021* (pp. 209-218). Springer Singapore.

7. Duma, R. A., Niu, Z., Nyamawe, A. S., Tchaye-Kondi, J., & Yusuf, A. A. (2023). A Deep Hybrid Model for fake review detection by jointly leveraging review text, overall ratings, and aspect ratings. *Soft Computing*, *27*(10), 6281-6296.
8. Khan, J., Alam, A., & Lee, Y. (2021). Intelligent hybrid feature selection for textual sentiment classification. *IEEE Access*, *9*, 140590-140608.
9. Nafis, N. S. M., & Awang, S. (2021). An enhanced hybrid feature selection technique using term frequency-inverse document frequency and support vector machine-recursive feature elimination for sentiment classification. *Ieee Access*, *9*, 52177-52192.
10. Rijal, S., Cakranegara, P. A., Ciptaningsih, E. M. S., Pebriana, P. H., Andiyan, A., & Rahim, R. (2023). Integrating Information Gain methods for Feature Selection in Distance Education Sentiment Analysis during Covid-19. *TEM Journal*, *12*(1).
11. Abdulkhaliq, S. S., &Darwesh, A. M. (2020). Sentiment Analysis Using Hybrid Feature Selection Techniques. *UHD Journal of Science and Technology*, *4*(1), 29-40.
12. Qaisar, S. M. (2020, October). Sentiment analysis of IMDb movie reviews using long short-term memory. In *2020 2nd International Conference on Computer and Information Sciences (ICCIS)* (pp. 1-4). IEEE.
13. Gharaibeh, H., Mamlook, A., Emhamed, R., Samara, G., Nasayreh, A., Smadi, S., ...& Abualigah, L. (2024). Arabic sentiment analysis of Monkeypox using deep neural network and optimized hyperparameters of machine learning algorithms. *Social Network Analysis and Mining*, *14*(1), 1-18.
14. Gupta, K., Jiwani, N., &Afreen, N. (2023). A combined approach of sentimental analysis using machine learning techniques. *Revue d'IntelligenceArtificielle*, *37*(1), 1.
(Gupta, Jiwani&Afreen, 2023)
15. Kamarudin, M. H., Maple, C., & Watson, T. (2019). Hybrid feature selection technique for intrusion detection system. *International Journal of High Performance Computing and Networking*, *13*(2), 232-240.
16. Sutoyo, E., Rifai, A. P., Risnumawan, A., & Saputra, M. (2022). A comparison of text weighting schemes on sentiment analysis of government policies: a case study of replacement of national examinations. *Multimedia Tools and Applications*, *81*(5), 6413-6431.
17. Ampomah, E. K., Nyame, G., Qin, Z., Addo, P. C., Gyamfi, E. O., &Gyan, M. (2021). Stock market prediction with gaussian naïve bayes machine learning algorithm. *Informatica*, *45*(2).
18. Hanbal, I. F., Ingosan, J. S., Oyam, N. A. A., & Hu, Y. (2020, April). Classifying wastes using random forests, gaussian naïve bayes, support vector machine and multilayer perceptron. In *IOP conference series: Materials science and engineering* (Vol. 803, No. 1, p. 012017). IOP Publishing.
19. Khavandi, H., Moghadam, B. N., Abdollahi, J., & Branch, A. (2023). Maximizing the Impact on Social Networks using the Combination of PSO and GA Algorithms. *Future Generation in Distributed Systems*, *5*, 1-13.
20. Danyal, M. M., Khan, S. S., Khan, M., Ghaffar, M. B., Khan, B., & Arshad, M. (2023). Sentiment Analysis Based on Performance of Linear Support Vector Machine and Multinomial Naïve Bayes Using Movie Reviews with Baseline Techniques. *Journal on Big Data*, *5*.
21. Khare, A., Gangwar, A., Singh, S., & Prakash, S. (2023). Sentiment Analysis and Sarcasm Detection in Indian General Election Tweets. In *Research Advances in Intelligent Computing* (pp. 253-268). CRC Press.
22. Sharma, S., & Waoo, A. A. (2023). Customer Behavior Analysis in E-Commerce using Machine Learning Approach: A Survey. *IJSRCSEIT*, *9*, 163-170.
23. Sharma, S., & Waoo, A. A. Decision Tree Machine Learning Approach for Customer Behavior Analysis on Online Product.
24. Warner, E., Lee, J., Hsu, W., Syeda-Mahmood, T., Kahn Jr, C. E., Gevaert, O., & Rao, A. (2024). Multimodal Machine Learning in Image-Based and Clinical Biomedicine: Survey and Prospects. *International Journal of Computer Vision*, 1-17.
25. ACCENTURE GLOBAL SOLUTIONS LIMITED. (2024). TARGET IDENTIFICATION USING BIG DATA AND MACHINE LEARNING.*INDIA Patent* 512287, filed Feb. 6, 2018, and issuedFeb. 19, 2024.

26. Hadi Mohammed, Koblas Michal, & Shoaraee Saeed. (2021). SENTIMENT ANALYSIS. *US* 2020/0065383 A1, filed Aug 24, 2018, and issued Mar 23, 2021.
27. Zadeh A Lotfi, Tadayon Saied, Tadayon Bijan. (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.
28. Adewole, K. S., Balogun, A. O., Raheem, M. O., Jimoh, M. K., Jimoh, R. G., Mabayoje, M. A., ... & Asaju-Gbolagade, A. W. (2021). Hybrid feature selection framework for sentiment analysis on large corpora. *Jordanian Journal of Computers and Information Technology*, *7*(2).
29. Andrian, B. W., Tobing, F. A. T., Pane, I. Z., & Kusnadi, A. (2023). Implementation Of Naïve Bayes Algorithm In Sentiment Analysis Of Twitter Social Media Users Regarding Their Interest To Pay The Tax. *International Journal of Science, Technology & Management*, *4*(6), 1733-1742.
30. Alsemaree, O., Alam, A. S., Gill, S. S., & Uhlig, S. (2024). Sentiment analysis of Arabic social media texts: A machine learning approach to deciphering customer perceptions. *Heliyon*, *10*(9).
31. 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
32. Nguyen-Thanh, T., & Tran, G. T. (2019, December). 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* (pp. 147-153).