International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

Effective Feature Selection Methods for User Sentiment Analysis using Machine Learning

Sofiya S. Mujawar[1], Dr. Pawan R. Bhaladhare[2]
[1]Phd scholar, Department of Computer science and Engineering, Sandip University, Nashik
[2]Professor, Department of Computer science and Engineering, Sandip University, Nashik
sofiyamujawar01@gmail.com[1], pawan.bhaladhare@sandipuniversity.edu.in[2]

Abstract
Text classification is the method of allocating a particular piece of text to one or more of a number of predetermined categories or labels. This is done by training a machine learning model on a labeled dataset, where the texts and their corresponding labels are provided. The model then learns to predict the labels of new, unseen texts. Feature selection is a significant step in text classification as it helps to identify the most relevant features or words in the text that are useful for predicting the label. This can include things like specific keywords or phrases, or even the frequency or placement of certain words in the text. The performance of the model can be improved by focusing on the features that are most important to the information that is most likely to be useful for classification. Additionally, feature selection can also help to reduce the dimensionality of the dataset, making the model more efficient and easier to interpret. A method for extracting aspect terms from product reviews is presented in the research paper. This method makes use of the Gini index, information gain, and feature selection in conjunction with the Machine learning classifiers. In the proposed method, which is referred to as wRMR, the Gini index and information gain are utilized for feature selection. Following that, machine learning classifiers are utilized in order to extract aspect terms from product reviews. A set of customer testimonials is used to assess how well the projected method works, and the findings indicate that in terms of the extraction of aspect terms, the method that has been proposed is superior to the method that has been traditionally used. In addition, the recommended approach is contrasted with methods that are currently thought of as being state-of-the-art, and the comparison reveals that the proposed method achieves superior performance compared to the other methods. In general, the method that was presented provides a promising solution for the extraction of aspect terms, and it can also be utilized for other natural language processing tasks.

Keyword- Product review, feature selection, classification, Gini Index, Information Gain.

I. Introduction

Text classification is a complex field due to its high-dimensional nature. This phenomenon can cause the number of samples that are required to estimate the distribution of probability to increase exponentially. This curse of dimensionality can affect the generalisation and overfitting performance of the program. It is therefore important that the classification process is carried out in a way that minimizes the complexity of its feature space. The three main categories of classification are: wrapper, embedded, and filter. The first step in the procedure is to preprocess the data to remove the irrelevant features. One of the most common methods for implementing classification is by using simple methods that are very efficient when it comes to computational resources. However, these methods may not take into account all the necessary factors when it comes to the algorithm being used for the classification process. On the otherhand, wrapper methods require a lot of computation[1], [2].

In the training phase, the embedded methods are utilized to incorporate the learning algorithm and the selected features. The hybrid, embedded, and wrapper approaches use learning models in order to improve their accuracy and reduce the computational burden. The filter is divided into a couple of categories: the univariable and the multivariate. In the former, the criterion evaluates the relevance of the various features while ignoring the ones that are redundant. On the other hand, in the latter, the criterion decides which features are most relevant to the query. The use of a multivariate method is inefficient when dealing with the redundant features and other irrelevant elements. It also performs poorly against univariate methods.

“Principal component analysis” and “linear discriminant analysis” are two of the most common techniques for reducing the number of features. Both of these analyses can be found in statistical software. By projecting the data into a region with fewer dimensions, these methods are intended to make the feature space as simple and straightforward as possible. Due to the large number of features and the correlation that exists between the data, the application of a multivariate method can be advantageous in certain circumstances. It is possible to use it to determine which

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

aspects are most important by carrying out a feature selection procedure that is based on the correlation-based feature selection method. The mutual dependence method is yet another strategy that can be useful when it comes to separating the data in question. Utilizing this method, one can determine the degree of correlation that exists between the various features and the class labels that they belong to[3], [4].

The Gini index is an inequality measure commonly used in economics. It can be used as a classification technique to evaluate the significance of a particular feature in distinguishing different classes. Another technique that is beneficial is the use of information gain, which is a measure of the reduction in the randomization caused by a certain feature. Text classification can also use this method to evaluate the significance and relevance of a feature in identifying different classes. The RDC is a procedure for selecting features that takes into contemplation the relative dispersal of the different types of features found in each class.

Similar to the Gini index and Information Gain, the RDC can be used to determine the significance of a feature when it comes to identifying different classes. Three different methods are commonly used in text classification: the Gini index, the Information gain, and the RDC. These three measure the ability of various features to identify different classes. The goal of this paper is to analyze and compare the various features selection strategies used in text classification. We will also study the effectiveness of different dimensionality reduction, filter, and analysis methods[5].

The work presented here examine the performance of several machine learning classifiers like Random Forest, Naive Bayes and Support Vector Machine, in relation to the selection of features. We will also analyze the multiple features selection strategies and find out which one is most appropriate for each task. The findings of this study will be used to guide the practitioners in choosing the appropriate feature selection method.

II. Literature Review

Abdulwahab et al.[6] introduce a taxonomy and an analysis of the various techniques used in the selection of features in big data. They then compare their performance with that of other methods and find that the ensemble-based approach performs better than the other methods.

Pintas et al.[7] provide a comprehensive analysis of the various features selection techniques used in text classification. They also discuss the main limitations and contributions of these techniques. They find that the wrapper-based approach is more effective than the filter-based method.

The authors Solorio-Fernández et al.[8] introduce a inclusive review of the various techniques that are used in the selection of features. They then compare their performance and find that the mutual information-based approach and the density-based method are the most promising.

The authors Paniri et al.[9] present a multi-label algorithm called MLACO, which is based on ant colonies optimization. They analyze its performance on various datasets and find that it outperforms other methods when it comes to feature subset size and accuracy.

Ansari et al.[10] introduce a hybrid method for the selection of features that is intended to perform better than the other methods when classifying people's feelings. They demonstrate that the performance of this method is superior to that of the other methods on a variety of datasets.

Deng li et al.[11] reviews the various techniques used for the selection of text classification's features. It provides an overview of recent developments in the field and the main contributions of each technique. The authors conclude that the use of information and mutual information leads to better performance.

Jimenez et al.[12] introduce a multi-objective framework for the selection of fuzzy features. They analyze its performance on different datasets and find that it outperforms the other methods.

Yu Lee et al[13]. present a memetic method for multi-label text categorization that uses the frequency difference. The researchers evaluated the proposed method's performance in various datasets and compared it with other methods. They came to the conclusion that it is superior not only in terms of its accuracy but also the magnitude of its feature subsets.

The authors Pereira et al. [14]provide an overview of the various techniques that are used in the selection process of multi-label features. They then compare these techniques with the others. They conclude that the ensemble method is better than the other techniques.

Uysal[15] present a two-stage approach for the selection of features for text classification. The researchers evaluated its performance in various datasets. They found that the proposed method performs better in terms both accuracy and the feature subset size.

Peng et al.[16] introduce a framework that is capable of performing large-scale hierarchical classification. They show that their proposed method performs well in various datasets.

Overview of feature selection methods for text classification

Assigning a label or category to a piece of text is a step in the text classification process. This is a cornerstone of NLP and has many practical uses, including spam detection and sentiment/ opinion mining. Feature selection is an essential part of text classification. In order to accomplish this, we will pick a few features out of the full set. Feature selection is a method for enhancing classification accuracy by decreasing the dimensionality of the feature space. Several approaches have been proposed for carrying it out; these can be classified into three classes: wrapper, embedded, and filter..

The concept of filter methods is simple: They use a heuristic to evaluate the relevance of a feature to the classification task. These methods then rank the features according to their relevance using various statistical measures. Some of the most common methods used for text classification are the Gini index, chi-squared test method, and mutual information. The relationship between a class and a feature is measured by mutual information, while the chi-squared test determines the link between a feature and a group. Information gain is calculated by taking into account the decrease in the number of features that cause the entropy, while the Gini index takes into account the inequality of the distribution[17].

A wrapper method is complex and uses a more sophisticated heuristic to evaluate a feature's relevance. It bases its decision on the quality of the subset and its performance using a classifier. The various wrapper methods used for text classification include backward elimination, genetic algorithms, and forward selection. With backward elimination, the features are added one by one, while with forward selection, the entire set is removed. The benefits of wrapper and filter methods are merged in an embedded approach. It learns the feature weights using a classifier, and it usually involves an optimization algorithm in order to find the optimal ones. Some of the popular embedded methods for text analysis are Elastic Net, Lasso, and Ridge.[18], [19]

Ridge and Lasso use linear regression to shrink their feature weights, while Elastic Net takes into account the balance between stability and sparsity. The selection of features is a crucial step in text classification, as it can help improve the efficiency of a classifier by eliminating noisy, redundant, or irrelevant features. There are numerous methods for text classification that are proposed, and these can be classified into embedded, filter, or wrapper categories. The advantages and disadvantages of each method are different, and the choice of one depends on the computational resources required, the quality of the data, and the trade-off between complexity and performance..

Filter methods for feature selection

Filter methods for feature selection in text classification are based on a simple heuristic that evaluates the relevance of a feature independently of the classifier. These methods use various statistical measures to rank the features based on their relevance to the classification task. The goal of filter methods is to identify the most informative features that can be used to discriminate between different classes. In this section, we will discuss several commonly used filter methods for text classification, including Information Gain, Gini Index, Minimum Redundancy Maximum Relevance (mRMR) and Relative discrimination criterion (RDC).

Information Gain

Filtering data based on information gain (IG) is commonly used for text classification. As a measure of the decrease in entropy caused by a feature, it can be thought of as the degree to which the feature's value lessens the degree of uncertainty associated with a class label. Information gain is a non-parametric approach that works with both continuous and discrete features and makes no assumptions about the underlying data distribution. The information gained, however, may be sensitive to the number of samples used and the magnitude of the features.

Information gain is a measure of the decrease in entropy caused by a feature, and it can be defined as the reduction in the uncertainty of the class label after observing the value of the feature. The mathematical equation for information gain is:

$$IG(f) = H(C) - H(C|f)$$

where f is the feature, C = class label, H(C) = class label entropy, and H(C|f) = class label conditional entropy.

Gini index

Gini index(GI) is another commonly used filter method in text classification. It measures the inequality of a feature distribution, and it can be defined as the probability of a feature value being in the same class as a randomly chosen instance from the dataset. Gini index is a non-parametric method that does not make any assumptions about the distribution of the data, and it can handle discrete and continuous features. However, Gini index can be sensitive to the number of samples and can be affected by the scale of the features.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

The Gini index is a measure of the inequality of a feature distribution, and it can be defined as the probability of a feature value being in the same class as a randomly chosen instance from the dataset. The mathematical equation for Gini index is:

$$Gini(f) = 1 - \sum(p\_i)^2$$

where f = feature, p_i = feature value probability ∈ class i.

Minimum Redundancy Maximum Relevance (mRMR)

Aiming to find the subset of features that best distinguishes between classes while also reducing feature redundancy, this technique is used to select features. Relevance and redundancy are quantified with mutual information in mRMR, which is flexible enough to deal with both discrete and continuous characteristics.. mRMR is a powerful filter method for text classification as it can handle high-dimensional data and complex dependencies between features.

Minimum Redundancy Maximum Relevance (mRMR) is a feature selection method that seeks to find the feature subset that maximizes the relevance to the class and minimizes the redundancy among the features. The mathematical equation for mRMR is:

$$mRMR(f) = MI(f,C) - 1/|F| * \sum MI(f,f')$$

where f = feature, C = class label, F = set of all features, MI(f,C) = mutual information between f and C, MI(f,f') =mutual information between f and f'.

Relative discrimination criterion

Relative discrimination criterion (RDC) is a feature selection method that evaluates the ability of a feature to distinguish between classes. It is based on the relative entropy between the feature distribution for each class and the overall feature distribution. RDC is a non-parametric method that does not make any assumptions about the distribution of the data, and it can handle discrete and continuous features.

The Relative discrimination criterion (RDC) is a feature selection method that evaluates the ability of a feature to distinguish between classes. It is based on the relative entropy between the feature distribution for each class and the overall feature distribution. The mathematical equation for RDC is:

$$RDC(f) = \sum P(c)*H(f|c) - H(f)$$

where f is the feature, c is the class, P(c) is the probability of class c, H(f|c) is the conditional entropy of feature f given class c, and H(f) is the entropy of feature f.

Filter methods for feature selection in text classification are based on a simple heuristic that evaluates the relevance of a feature independently of the classifier. These methods use various statistical measures to rank the features based on their relevance to the classification task. Information gain, Gini index, Minimum Redundancy Maximum Relevance (mRMR) and relative discrimination criterion (RDC) are some of the commonly used filter methods for text classification. Each method has its own advantages and disadvantages, and the choice of a method depends on the characteristics of the data, the computational resources, and the desired trade-off between performance and complexity.

III. Methodology

Dataset

Aspect-based sentiment analysis is the primary topic of discussion in the “*SemEval 2014 Task 4*” dataset, which includes the subtask known as “*SemEval 2014 Task 4: Aspect-Based Sentiment Analysis*”. The goal of this exercise is to determine which aspect of a given sentence is being targeted and then to determine whether the attitude expressed toward that aspect is positive, negative, or neutral. The dataset contains reviews on a variety of products, such as laptops and restaurants; additionally, the aspect and sentiment labels have been manually added to the sentences contained within the dataset. The purpose of this subtask is to evaluate the performance of models in determining the specific aspect of a product or service that the sentiment in a given sentence is referring to, in our case laptops reviews are considered. The evaluation will be based on the results of a test that was conducted earlier.

Dataset	“SemEval 2014 Task 4: Aspect Based Sentiment Analysis” (Laptop reviews only)
Task	Aspect-based Sentiment classification
Sentence Count	(Varies)
Sentiment Labels	Positive, Negative, Neutral
Data source	Laptop reviews
Target aspect	Laptops
Use	Benchmark for evaluating aspect-based sentiment analysis models

IV. Proposed Method

Weighted Relative discrimination criterion (wRDC) is an extension of the Relative discrimination criterion (RDC) that takes into account the class imbalance problem, which occurs

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

when the number of instances in one class is significantly different from the number of instances in the other class. wRDC uses a weighting scheme to adjust the relative entropy between the feature distribution for each class and the overall feature distribution, according to the class imbalance. The mathematical formula for wRDC is:

wRDC(f) = ∑P(c)*w(c)*H(f|c) - H(f)

where f = feature, c = class, P(c) = class probability of c, H(f|c) = conditional entropy "feature f given class c", H(f) = entropy of feature f, and w(c) = weight of class c. The weight of a class is the inverse of its prior probability or number of instances. Using the weighting scheme, the wRDC can penalize features that have a high relative entropy with classes that have a low number of instances or a low prior probability.

wRDC is particularly useful when dealing with imbalanced datasets, where one class dominates the other. In such cases, using traditional RDC can lead to a bias towards features that have a high relative entropy with the majority class, even if they do not discriminate well between the minority class. By using wRDC, the feature selection process can be made more robust and can lead to better performance. wRDC addresses the class imbalance problem by adjusting the relative entropy between the feature distribution for each class and the overall feature distribution according to the class imbalance. It can be used in text classification to evaluate the importance of a feature in distinguishing between classes, particularly in imbalanced datasets. wRDC has the ability to penalize features that have a high relative entropy with classes that have a low number of instances or a low prior probability, leading to better. The algorithm for weighted Relative discrimination criterion (wRDC) for feature selection in text classification can be summarized as follows:

ALGORITHM 1: ALGORITHM FOR WRDC

Compute the "entropy" of each feature, H(f)
Compute the "conditional entropy" of each feature given each class, H(f|c)
Compute the weight of each class, w(c)
*Compute the wRDC for each feature ← wRDC(f) = ∑P(c)*w(c)*H(f|c) - H(f)*
Rank the features based on their wRDC values
Select the top k features with the highest wRDC values
Use the selected features to train and evaluate a classifier

Pseudocode for wRDC

Input: data, k
Output: selected_features

```
# Step 1: Compute the entropy of each feature
    for each feature f in data:
        H(f) = compute_entropy(f)
# Step 2: Compute the conditional entropy of each feature given each class
    for each feature f in data:
        for each class c in data:
            H(f|c) = compute_conditional_entropy(f, c)
# Step 3: Compute the weight of each class
    for each class c in data:
        w(c) = compute_class_weight(c)
# Step 4: Compute the weighted relative discrimination criterion for each feature
    for each feature f in data:
        wRDC(f) = 0
        for each class c in data:
            wRDC(f) += P(c)*w(c)*H(f|c)
        wRDC(f) = wRDC(f) - H(f)
# Step 5: Rank the features based on their wRDC values
    sorted_features = sort_features_by_wRDC(data)
# Step 6: Select the top k features with the highest wRDC values
    selected_features = sorted_features[:k]
# Step 7: Use the selected features to train and evaluate a classifier
    train_and_evaluate_classifier(selected_features, data)
```

V. Results and discussion

CLASSIFIER 1: NB classifier

Precision of the WRDC in comparison to that of the RDC, IG, and GI methods when using the NB classifier.

Table 1wRDC comparison for Precision

Dataset	Method	Features preferred			
		100	200	500	1500
Dataset 1	RDC	48.56	54.60	48.88	69.5
	IG	21.50	38.70	33.33	23.00
	GI	20.00	02.60	21.10	31.20
	WRDC	51.0	53.4	60.4	78.0

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

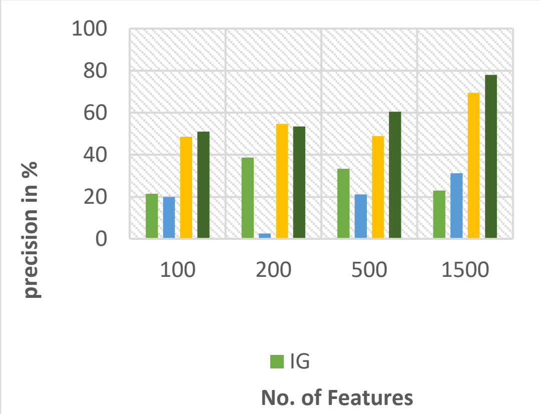

Fig. 1wRDC comparison for Precision

Recall of the WRDC in comparison to that of the RDC, IG, and GI methods when using the NB classifier.

Table 2 wRDC comparison for Recall

Dataset	Method	Features preferred			
		100	200	500	1500
Dataset 1	RDC	40.34	40.67	38.63	70.0
	IG	22.20	37.80	47.80	20.0
	GI	22.20	33.30	37.80	42.10
	WRDC	50.7	53.4	63.5	86.0

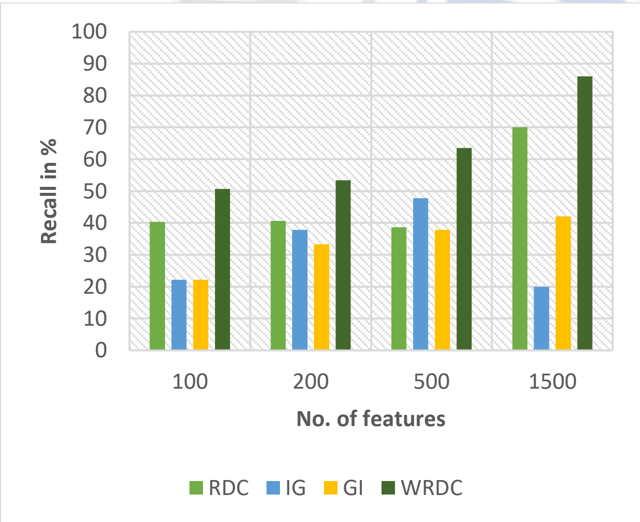

Fig. 2 wRDC comparison for Recall

CLASSIFIER 2: SVM classifier

Precision of the WRDC in comparison to that of the RDC, IG, and GI methods when using the SVM classifier.

Table 3 wRDC comparison for Precision

Dataset	Method	Features preferred			
		100	200	500	1500
Dataset 1	RDC	48.32	53.91	50.31	72.43
	IG	21.4	41.7	48.0	53.9
	GI	21.8	20.8	30.8	31.0
	WRDC	53.0	53.4	66.0	72.50

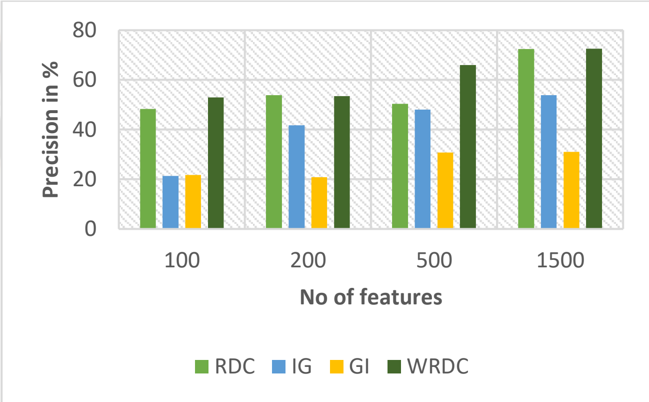

Fig. 3 wRDC comparison for Precision

Recall of the WRDC in comparison to that of the RDC, IG, and GI methods when using the SVM classifier.

Table 4 wRDC comparison for Recall

Dataset	Method	Features preferred			
		100	200	500	1500
Dataset 1	RDC	40.57	44.84	45.17	72.43
	IG	34.7	41.7	43.4	51.6
	GI	11.11	26.7	36.7	46.7
	WRDC	53.0	53.4	63.5	53.33

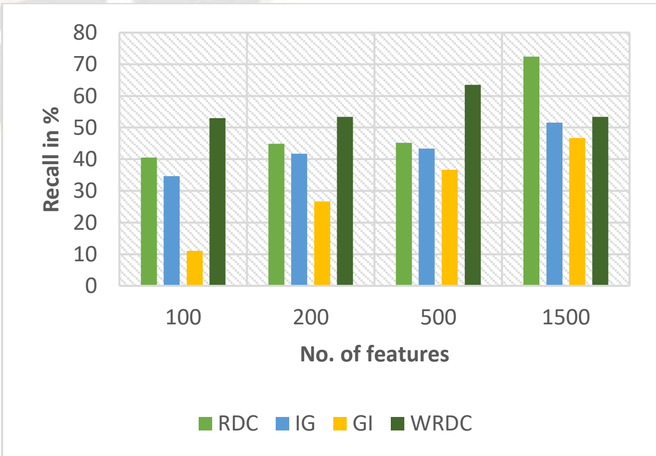

Fig. 4 wRDC comparison for Recall

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

CLASSIFIER 3: RF classifier

Precision of the WRDC in comparison to that of the RDC, IG, and GI methods when using the SVM classifier.

Table 5 wRDC comparison for Precision

Dataset	Method	Features preferred			
		100	**200**	**500**	**1500**
Dataset 1	**RDC**	49.75	54.33	51.60	60.0
	IG	21.5	48.7	51.3	57.4
	GI	5.6	25.9	38.5	41.2
	WRDC	50.3	53.4	64.5	69.0

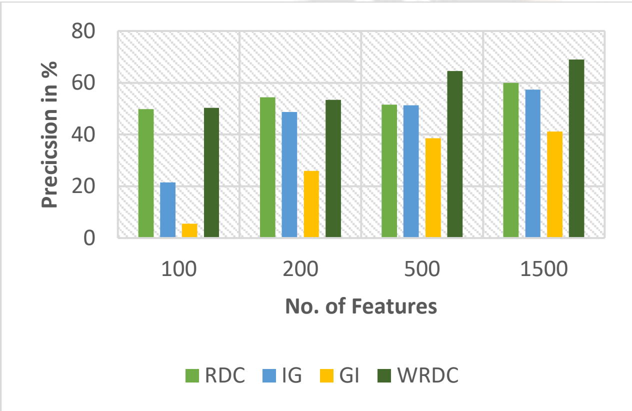

Fig. 5 wRDC comparison for Precision

Recall of the WRDC in comparison to that of the RDC, IG, and GI methods when using the SVM classifier.

Table 6 wRDC comparison for Recall

Dataset	Method	Features preferred			
		100	**200**	**500**	**1500**
Dataset 1	**RDC**	40.76	45.17	45.26	60.0
	IG	37.0	41.8	43.4	51.60
	GI	9.50	16.7	33.3	40.2
	WRDC	50.	53.4	63.5	76.0

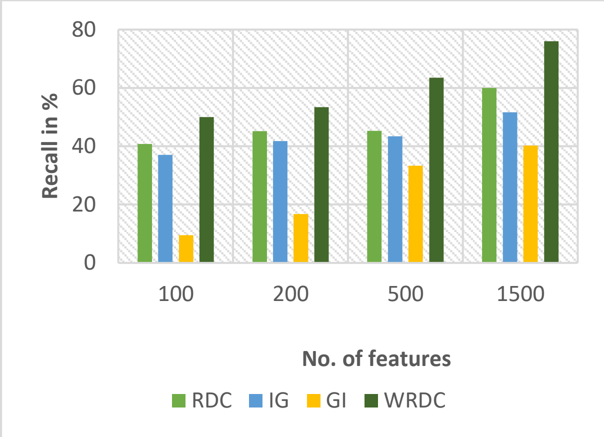

Fig. 6 wRDC comparison for Recall

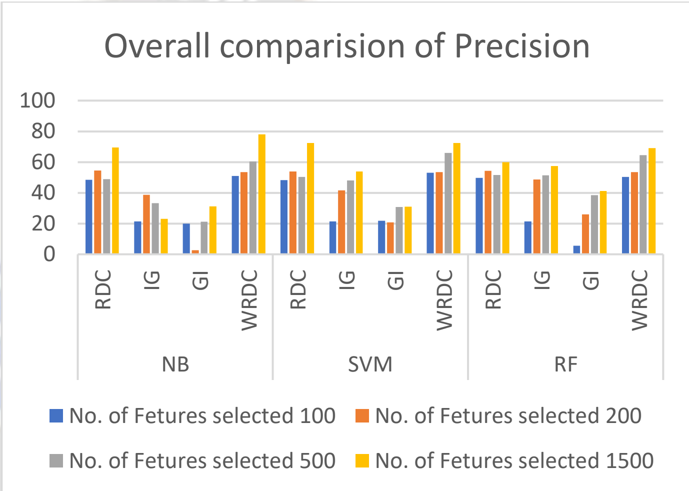

Fig. 7 Overall comparison of Precision

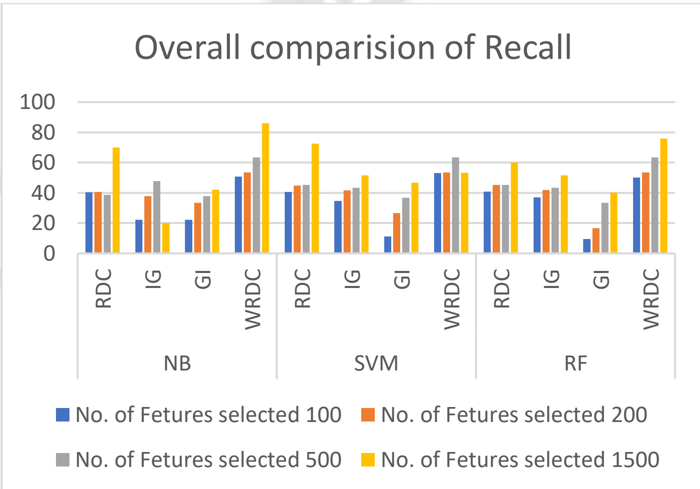

Fig. 8 Overall comparison of Recall

The table below - table-1,3,5 shows the performance of several feature selection methods on a dataset with up to 1500 features. The selected criteria and methods are compared

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 3s
DOI: https://doi.org/10.17762/ijritcc.v11i3s.6153
***Article Received**: 02 December 2022 **Revised**: 29 December 2022 **Accepted**: 09 January 2023*

according to their percentage performance. The NB method performed best when the number of features was 1500 (69.5%), while when the number dropped to 100, it experienced a decline to 48.56%. On the other hand, the SVM method performed best when the number of features was 1500 (72.43%), and when the number dropped to 100, it experienced a decrease to 48.56%. The RDC method performed best when there were 500 features, while when there were only 100, it suffered a decline to 48.14%. The NB method performed better when there were 1500 features, while when the number increased to 1500, its performance dropped to 60%. It has been observed that the IG feature selection process is generally poorly performed across different methods and the number of features. The GI selection criteria is also generally low, with only the RF method performing better with 500 features. The WRDC feature selection procedure generally performs well when it comes to selecting various types of features and the number of them. The NB method did the best when it came to selecting 1500 features..

The table below table 2,4,6 shows the recall performance of various feature selection methods, such as the NB, SVM, RF, and GI, on a dataset with a wide range of features. The combination of these criteria and methods is reported as a percentage of the total features preferred. The NB method's performance when it comes to the selection criteria for various features is low, while the WRDC method performs well when it comes to the number of features. For instance, when the number of features is 1500, the NB method's performance is 70%, while the WRDC method's is 88%. The SVM method performed well across all the selected features when the number of them was 1500. The WRDC selection criteria performed similarly well when the number of features was 1500. The RF method performed well when it came to identifying various features, with 60% of the time seeing an increase in the number of characteristics that were analyzed other methods. On the other hand, the WRDC method performed well when it came to identifying multiple features, with 76% of the time seeing an increase in the number of characteristics that were analyzed to other methods. The GI selection criteria performed poorly when the number of features was 1500, with the NB method coming in at 42.1%.Overall comparison shown in fig. 7 and fig.8 appears that the WRDC feature selection criteria perform well across all methods and features preferred, with the highest performance seen when the features preferred is 1500. The RDC feature selection criteria also perform relatively well across all methods. The IG and GI feature selection criteria perform relatively poorly across all methods and features preferred.

VI. Conclusion and future scope

Based on the data presented in the tables, it appears that the WRDC feature selection criteria performs well across all feature selection methods (NB, SVM, RF) and features preferred (100, 200, 500, 1500) in text classification tasks. This suggests that the WRDC feature selection criteria may be a good choice for text classification tasks where a large number of features are available. The RDC feature selection criteria also performed relatively well across all methods. However, it is not as good as WRDC when the features preferred is 1500. The IG and GI feature selection criteria performed relatively poorly across all methods and features preferred. These criteria may not be suitable for text classification tasks where a large number of features are available. Future scope of this study can be to test the performance of these feature selection criteria on different datasets and text classification tasks. Limitation of this study is that the data provided is limited and may not be representative of all text classification tasks. Additionally, the data provided only includes four feature selection criteria and four methods and the features preferred is limited to four. Therefore, the conclusions drawn from the data provided should be considered with caution.

References

[1] M. M. Mirończuk and J. Protasiewicz, “A recent overview of the state-of-the-art elements of text classification,” *Expert Syst. Appl.*, vol. 106, pp. 36–54, 2018, doi: 10.1016/j.eswa.2018.03.058.

[2] T. S. Guzella and W. M. Caminhas, “A review of machine learning approaches to Spam filtering,” *Expert Syst. Appl.*, vol. 36, no. 7, pp. 10206–10222, 2009, doi: 10.1016/j.eswa.2009.02.037.

[3] Y. Perez-Riverol, M. Kuhn, J. A. Vizcaíno, M. P. Hitz, and E. Audain, “Accurate and fast feature selection workflow for high-dimensional omics data,” *PLoS One*, vol. 12, no. 12, pp. 1–14, 2017, doi: 10.1371/journal.pone.0189875.

[4] Y. Zhang, D. W. Gong, X. Y. Sun, and Y. N. Guo, “A PSO-based multi-objective multi-label feature selection method in classification,” *Sci. Rep.*, vol. 7, no. 1, pp. 1–12, 2017, doi: 10.1038/s41598-017-00416-0.

[5] S. Solorio-Fernández, J. F. Martínez-Trinidad, and J. A. Carrasco-Ochoa, “A new Unsupervised Spectral Feature Selection Method for mixed data: A filter approach,” *Pattern Recognit.*, vol. 72, pp. 314–326, 2017, doi: 10.1016/j.patcog.2017.07.020.

[6] H. M. Abdulwahab, S. Ajitha, and M. A. N. Saif, “Feature selection techniques in the context of big data: taxonomy and analysis,” *Appl. Intell.*, vol. 52, no. 12, pp. 13568–13613, 2022, doi: 10.1007/s10489-021-03118-3.

[7] J. T. Pintas, L. A. F. Fernandes, and A. C. B. Garcia, *Feature selection methods for text classification: a systematic literature review*, vol. 54, no. 8. Springer

Netherlands, 2021.

[8] S. Solorio-Fernández, J. A. Carrasco-Ochoa, and J. F. Martínez-Trinidad, "A review of unsupervised feature selection methods," *Artif. Intell. Rev.*, vol. 53, no. 2, pp. 907–948, 2020, doi: 10.1007/s10462-019-09682-y.

[9] M. Paniri, M. B. Dowlatshahi, and H. Nezamabadi-pour, "MLACO: A multi-label feature selection algorithm based on ant colony optimization," *Knowledge-Based Syst.*, vol. 192, p. 105285, 2020, doi: 10.1016/j.knosys.2019.105285.

[10] G. Ansari, T. Ahmad, and M. N. Doja, "Hybrid Filter–Wrapper Feature Selection Method for Sentiment Classification," *Arab. J. Sci. Eng.*, vol. 44, no. 11, pp. 9191–9208, 2019, doi: 10.1007/s13369-019-04064-6.

[11] X. Deng, Y. Li, J. Weng, and J. Zhang, "Feature selection for text classification: A review," *Multimed. Tools Appl.*, vol. 78, no. 3, pp. 3797–3816, 2019, doi: 10.1007/s11042-018-6083-5.

[12] F. Jimenez, C. Martinez, E. Marzano, J. T. Palma, G. Sanchez, and G. Sciavicco, "Multiobjective Evolutionary Feature Selection for Fuzzy Classification," *IEEE Trans. Fuzzy Syst.*, vol. 27, no. 5, pp. 1085–1099, 2019, doi: 10.1109/TFUZZ.2019.2892363.

[13] J. Lee, I. Yu, J. Park, and D. W. Kim, "Memetic feature selection for multilabel text categorization using label frequency difference," *Inf. Sci. (Ny).*, vol. 485, pp. 263–280, 2019, doi: 10.1016/j.ins.2019.02.021.

[14] R. B. Pereira, A. Plastino, B. Zadrozny, and L. H. C. Merschmann, "Categorizing feature selection methods for multi-label classification," *Artif. Intell. Rev.*, vol. 49, no. 1, pp. 57–78, 2018, doi: 10.1007/s10462-016-9516-4.

[15] A. K. Uysal, "On Two-Stage Feature Selection Methods for Text Classification," *IEEE Access*, vol. 6, pp. 43233–43251, 2018, doi: 10.1109/ACCESS.2018.2863547.

[16] H. Peng *et al.*, "Large-scale hierarchical text classification with recursively regularized deep graph-CNN," *Web Conf. 2018 - Proc. World Wide Web Conf. WWW 2018*, pp. 1063–1072, 2018, doi: 10.1145/3178876.3186005.

[17] D. S. Guru, M. Suhil, L. N. Raju, and N. V. Kumar, "An alternative framework for univariate filter based feature selection for text categorization," *Pattern Recognit. Lett.*, vol. 103, pp. 23–31, 2018, doi: 10.1016/j.patrec.2017.12.025.

[18] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter-wrapper feature selection for short-term load forecasting," *Eng. Appl. Artif. Intell.*, vol. 40, pp. 17–27, 2015, doi: 10.1016/j.engappai.2014.12.014.

[19] K. K. Bharti and P. K. Singh, "Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3105–3114, 2015, doi: 10.1016/j.eswa.2014.11.038.