Published by
NIGERIAN SOCIETY OF PHYSICAL SCIENCES
Available online @ https://journal.nsps.org.ng/index.php/jnsps

J. Nig. Soc. Phys. Sci. 6 (2024) 1823

Journal of the Nigerian Society of Physical Sciences

Ensemble feature selection using weighted concatenated voting for text classification

Oluwaseun Peter Ige[a,b,*], Keng Hoon Gan[a]

[a]School of Computer Sciences, Universiti Sains Malaysia, Gelugor, Pulau Pinang, 11800, Malaysia.
[b]Universal Basic Education Commission, Wuse Zone 4, Abuja, 900284, Nigeria.

Abstract

Amidst the surge in high-dimensional data, filter feature selection techniques have emerged as preferred tools for selecting relevant features, owing to their advantages such as enhanced generalization, faster training times, dimensionality reduction, and improved model performance. However, traditional feature selection methods often exhibit instability, resulting in the selection of varying feature subsets and subsequently different classification accuracies. Addressing this challenge, we propose a novel approach termed Multi-Univariate Hybrid Feature Selection (MUNIFES) to bolster the discriminative power between features and target classes in text classification tasks. MUNIFES integrates the local feature relevance of filter methods, a facet that has been overlooked in prior literature, through a multi-iterative process to select optimal feature subsets from each univariate feature selection method. Leveraging an ensemble of discriminative strength metrics including Chi-Square (Chi2), Analysis of Variance (ANOVA), and Infogain methods, MUNIFES achieves superior performance in selecting optimal feature subsets. Evaluation conducted on the 20newsgroup dataset and its variant (17newsgroup) with 10 classifiers, including ensemble, classification, optimization algorithms, and Artificial Neural Network (ANN), demonstrates the efficacy of MUNIFES compared to state-of-the-art feature selection methods, showcasing improved accuracy in classification tasks.

DOI:10.46481/jnsps.2024.1823

Keywords: Ensemble feature selection, Text classification, Dimensionality reduction, Univariate filter methods

Article History :
Received: 02 October 2023
Received in revised form: 05 January 2024
Accepted for publication: 31 January 2024
Published: 18 February 2024

Communicated by: B. J. Falaye

1. Introduction

With the era of big data arising from the increased usage of the Internet and electronic devices, high-dimensional data has become prevalent. High-dimensional data enriches classification performance because it generally has more features. Nevertheless, some irrelevant and redundant features in high-dimensional data are inimical to classification performance; such features considerably increase computational complexity and memory requirements [1]. The major challenges for data mining [2], discovery of knowledge [3] and pattern recognition [4] are notably the large number of datasets containing noisy, irrelevant, or redundant features. It is necessary to filter out the irrelevant and redundant features by choosing a suitable subset of relevant features to avoid over-fitting and tackle the curse of dimensionality. This could be addressed through feature selection and feature extraction. While feature selection selects a subset of the original features, feature extraction transforms the

*Corresponding Author Tel.: +2348033662954.
Email address: igeoluwaseun@yahoo.com (Oluwaseun Peter Ige)

original features into a new set of features.

There exist four distinct Feature Selection (FS) methodologies as documented in [5]. These are: (i) Filter: This technique relies on ranking strategies that are independent of classifiers, thereby enhancing generality, speed, and scalability while mitigating overfitting. However, it tends to overlook feature interdependency. Noteworthy examples include Information Gain (IG), Document Frequency (DF), minimum Redundancy Maximum Relevance (mRMR), Chi2, and ANOVA. (ii) Wrapper: Unlike filter methods, wrapper techniques utilize a learning and training set to select features tailored to specific classifiers, resulting in higher accuracy albeit at a higher computational cost and an increased risk of overfitting. Sequential algorithms and recursive feature elimination represent common implementations. (iii) Embedded: Embedded approaches leverage algorithms with inherent feature selection mechanisms, offering reduced computational overhead compared to wrapper techniques. However, they face challenges in constructing suitable optimization functions. Popular examples include RIDGE, LASSO, and decision trees. (iv) Ensemble: Ensemble methods amalgamate outcomes from diverse feature selection techniques into a unified framework. Techniques such as voting, stacking, bagging, and boosting are illustrative examples, representing a specialized form of hybrid approach. Hybrid methodologies combine multiple FS approaches to derive an optimal feature subset, with instances including filter-wrapper, embedded, and metaheuristic hybrids.

When selecting subsets of features, the complex requirements of each classification algorithm may lead to a higher runtime. As a result, filter-based FS methods are more commonly used than wrappers and embedded methods. Different filtering techniques determine optimum features in different ways, i.e., the selection process of deterministic factors differs from one method to another. A single algorithm does not assure satisfactory performance in various experiments. The aim of combining several FS methods thus is to increase the maximum precision achieved by a single method since they can overcome mistakes made by other methods in different parts of the input space and increase accuracy through their complementary view of how important features are. This has removed the chance of search space being too large and not composed entirely of relevant features, as envisaged in the previous research [6]. Researchers have discussed the interaction of features with their target classes [1, 6], but feature relevancy to their target classes cannot be overemphasized because it is the basis of all relationships. The importance of a feature or word is measured by its relevance to its target class. This results in a positive influence on its prediction performance.

This study proposes a novel multi-univariate hybrid feature selection method (MUNIFES) for enhanced discriminative power between the features and the target class. This is produced by hybridizing three univariate filter feature selection methods (Chi2, Infogain, and ANOVA) through a weighted concatenated voting ensemble to produce the unique distinct discriminative features with the target class. Thus, a multi-univariate filter method is produced. The main contributions of this study are stated as follows:

- A ranked multi-univariate concatenated ensemble FS method with unique features is produced that shows the relevance of features across multiple FS methods.
- Features with less frequency across methods but with discriminative high weights according to threshold are considered.
- The ensembled FS method, namely MUNIFES, is proposed through multiset theory.
- Experimental comparisons with three univariate FS algorithms on public datasets using 10 classifiers show that the proposed MUNIFES improves the accuracy performance of text classification.

The rest of this study is organized as follows: Related works about FS methods are summarized in section **??**. Section **??** presented ensemble FS types and filter FS methods used. Section **??** contained the proposed method. Section **??** described experimental settings consisting of datasets, performance analysis, and statistical significance. Finally, section **??** presented a conclusion.

2. Related Works

In the literature, there are many studies associated with filter-based FS methods. These include correlation-based feature selection-CFS [7], Distinguishing Feature Selector-DFS [8], Comprehensively Measure Feature Selection-CMFS [9], Discriminative Feature Selection-DFSS [10], Relative Discrimination Criterion-RDC [11], minimum Redundancy Maximum Relevancy-mRMR [12], Normalized Difference Measure-NDM [13], Max–Min Ratio-MMR [14], Trigonometric Comparison Measure-TCM [15]. Hybrid FS algorithms exploit the benefits of both filter and wrapper approaches. The execution time gain of filter methods, as a first step, is used to reduce the high dimension of the problem to some extent, and then in the second step, the effectiveness of the wrapper approaches is used to obtain the best subset of features [16]. As FS is inevitable for high dimensional problems such as text analyzing, researchers must evaluate various methods to obtain enhanced classification accuracies since there is no well-defined method to prefer one FS method to another [17]. A solution to this problem can be found in the Ensemble FS method, which relies on an ensemble learning concept. EFS-MI combines the subsets using Mutual Information (MI) to reduce the redundancy among the selected features [6].

Another researcher used a heterogeneous approach of filters to generate multiple top k feature candidate rankings with various clustering-based methods using the mean-shift algorithm [18]. VIKOR method was also used as a Multi-Criteria Decision Making (MCDM) algorithm to rank the features based on evaluating several feature selection methods as different decision-making criteria [19]. The researchers proposed the EFS-MCDM method, a rank features vector, as an output for users to select a desired number of features. Ensemble learning presumes that there are usually better performances from

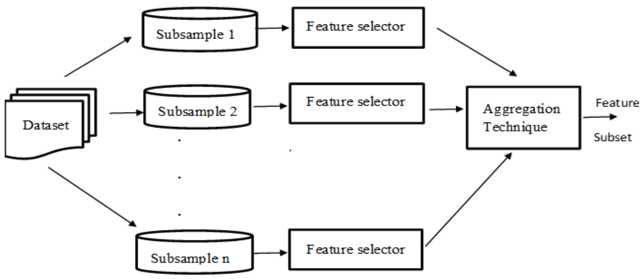

Figure 1. Block diagram of homogeneous ensemble FS.

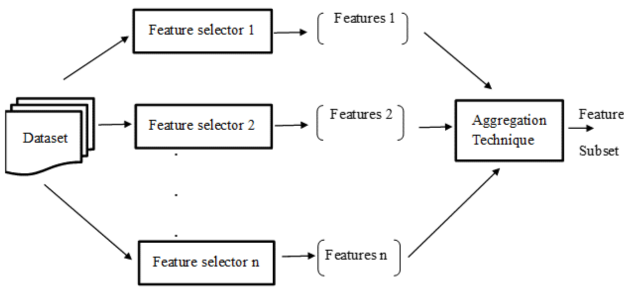

Figure 2. Block diagram of heterogenous ensemble FS.

a combination of more than one predictor. However, harnessing the strength of different univariate filter-FS methods to yield improved classification accuracy through ensemble aggregation cannot be overemphasized. This is lacking in many literatures. A fundamental element of ensemble feature selection is combining several FS methods to improve feature performance [20]. This stimulates this research work.

3. Ensemble types and filter feature selectionb methods

This section describes the ensemble FS methods. It classifies ensemble into different types stating the purpose of each type in feature selection. It also describes the filter feature selection methods that are used for the proposed work.

3.1. Ensemble Types

Ensemble FS is the process of combining multiple models instead of one model. This can be categorized into two main groups based on the type of feature selectors used: Homogeneous if the feature selectors are all the same type, and heterogenous if feature selectors are different. Homogenous ensembles exploit data diversity (data is partitioned into multiple divisions), and a feature selection method is implemented in each division, which results in achieving the final feature subset (Figure 1). Heterogeneous ensembles, which exploit function diversity (multiple feature selection methods), are executed on the same data, and the results of these FS methods aggregate to find the best subset of features (Figure 2) [19].

Figure 1 shows that the same feature selection method is applied to subsets of the same dataset (subsample 1 to subsample n) for feature selection. It is homogenous because the same feature selection method is applied to the subsamples of the same dataset while it is data diversified because different subsets of the dataset are extracted. Figure 2 on the other hand depicts different feature selectors on the same dataset hence it is heterogenous, and function diversified.

3.2. Filter feature selection methods

Filters are FS methods that use a statistical rating metric to measure the relevance of features. For a given set $X = \{X_1, X_2, \ldots, X_n\}$ of feature size n, the filter methods calculate a score function (X_i) according to the contribution of feature $X_i \in X$ to solve the text classification task. The feature weights are ranked according to their estimated scores, and the features with a score above the threshold are retained while the others are discarded. Filter techniques can be cost-effective and easy to use because the FS task does not involve any learning model. Consequently, the researchers use a wide range of filtering-based FS algorithms. Filters, in which they improve classification accuracies while reducing processing time, are applied for the purpose of obtaining more discriminative terms [17]. For text classification purposes, several filter methods have been used. Some of the most used filtering algorithms from the literature are ANOVA, Term Variance (TV), Distinguishing Feature Selector (DFS), Information Gain (Infogain), Document Frequency (DF), Relief, Symmetric Uncertainty (SU), Chi-square (Chi2), and Correlation-based Feature Selection (CFS). These are based on different metrics, univariate with reduced processing time and less computational complexity than the wrapper and have been proven effective in text classification applications [21–24]. Hence, they offer a unique perspective on feature importance. Since filter methods assume feature independence and cannot remove redundant features, this ensemble's main motivation is that the computational processing power of univariate filters has been combined with improved voting mechanisms to enhance discriminative filtering performance to obtain the most important features. This will make features more pertinent to the label classes while reducing redundancies. The filter univariate algorithms are briefly explained below:

1. Chi-square (Chi2): This method utilizes the test of independence to assess whether the feature f is independent of the target variable. It evaluates the association between the presence or absence of a feature and the target variable. It calculates the chi-squared statistic for each feature and the target variable. The higher the value, the more relevant the feature with respect to the class C (target).

$$\chi^2 = \sum \frac{(OV_i - EV_i)^2}{EV_i}, \tag{1}$$

where: χ^2 = chi-square, OV_i = observed value, and EV_i = expected value.

2. Infogain: It measures the reduction in entropy (uncertainty) about the target variable after observing a fea-

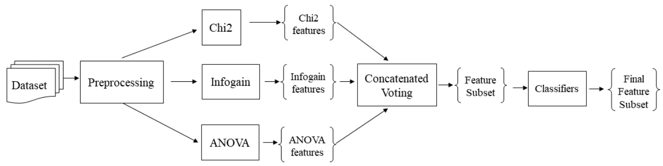

Figure 3. The proposed framework-MUNIFES.

ture. It measures how much information (infogain) a feature adds to the target variable. Features with higher infogain are considered informative for distinguishing between classes.

$$\text{Infogain}(X, y) = H(y) - H(y|X), \tag{2}$$

where: X = feature matrix, y = target vector,

$H(y)$ = target variable entropy, and

$H(y|X)$ = conditional entropy.

3. ANOVA: It shows the difference between two or more means through significant tests. It measures the ratio of between-group variance to within-group variance. It evaluates the significance of feature variations with respect to different classes. Features with higher scores are considered more relevant. It computes the F-statistic for each feature with respect to the target variable.

$$\text{ANOVA}(X, y) = \frac{\frac{SSB}{k}}{\frac{SSE}{(n-k-1)}}, \tag{3}$$

where: X = feature matrix, y = target vector,

SSB = sum of squares between groups,

SSE = sum of squares within groups,

k = number of groups, n = total number of samples.

4. Proposed method

This section discusses the proposed method as a multi-univariate heterogenous filter-based approach for ensemble learning, as shown in Figure 3. MUNIFES framework consists of three stages: (i) preprocessing, (ii) concatenated voting ensemble with weighting condition (iii) classification methods - evaluation of the selected feature subset for improved accuracy. These are explained in the following subsections.

4.1. Preprocessing

Text-based preprocessing was carried out that involved the removal of headers, punctuations, special characters, stop-words. The text passed through Tokenization and Lemmatization and is represented by the Term Frequency-Inverse Document Frequency (TF-IDF). Despite the above preprocessing steps, the feature dimension of the documents was still very high, and the computation complexity was undesirable, resulting in minimal accuracy. To solve this problem, this paper applies a trimming process to remove rare terms whose DF is lower than 3 [20]. Removing terms with very low DF reduces noise, which could lead to overfitting in the data. Features with varied numbers from 100 to 1000 were selected for comparison across feature selection methods.

4.2. Concatenated-weighted voting ensemble

The multiplicity concept of a multiset is utilized in this paper [25]. A multiset is a collection of objects (called elements) in which elements may occur more than once. The number of times an element occurs in a multiset (multiplicity) is relevant, and each occurrence contributes to the multiset's cardinality (the sum of the multiplicities of its elements). The word "multiset" (often shortened to mset) abbreviates the term "multiple-membership set."

$$x \in^n y, \tag{4}$$

where: x is an element of y with multiplicity n.

$$x \cup \varphi = x. \tag{5}$$

$$x \cap \varphi = \varphi. \tag{6}$$

$$x \cup x = x. \tag{7}$$

$$x \cap x = x. \tag{8}$$

$$x \cup y = y \cup x. \tag{9}$$

$$x \cap y = y \cap x. \tag{10}$$

$$x \cup (y \cup z) = (x \cup y) \cup z. \tag{11}$$

$$x \cap (y \cup z) = (x \cap y) \cup z. \tag{12}$$

$$x \cap (y \cup z) = (x \cap y) \cup (x \cap z). \tag{13}$$

$$x \cup (y \cap z) = (x \cup y) \cap (x \cup z). \tag{14}$$

For example A = {1,2,3,4,5,6,7,8}, B = {1,2,4,5,7,8,9,11}, C = {1,2,4,6,7,8,9,10}

$$D = A \cup B \cup C$$
$$= \{|1, 1, 1, 2, 2, 2, 3, 4, 4, 4, 5, 5, 6, 6, 7, 7, 7, 8, 8, 8, 9, 9, 10, 11|\}$$

Observe the different symbols adopted to stress the distinction between a traditional set: {}, and a multiset: {||}. This could be represented as "1 is an element in D with multiplicity 3", "2 is an element in D with multiplicity 3", "3 is an element in D with multiplicity 1". It could also be written as

$$A \cup (B \cup C) = (A \cup B) \cup C = (A \cup B \cup C) \in^n D. \tag{15}$$

That is D(1) = 3, D(2) = 3, D(3) = 1, D(4) = 3, D(5) = 2, D(6) = 2, D(7) = 3, D(8) = (3), D(9) = 2, D(10) = 1, D(11) = 1. From here, it could be noted that the relevancy of an element (feature) is its multiplicity across the sets (feature selection methods). An element of frequency greater or equal to 2 is said to be relevant across sets. Elements 1, 2, 4, 5, 6, 7, 8, and 9 are more relevant in the above example. Also, though elements 3, 10, and 11 are

Table 1. Recap of experimental atasets

Dataset	Documents	Number of Classes
$20 - Newsgroup$	18,846	20
$17 - Newsgroup$	16,075	17

less popular, either may contain weight that could be discriminative enough to improve accuracy.

$$E = \{|1, 2, 4, 5, 6, 7, 8, 9|\}, \quad F = \{|3, 10, 11|\}, \quad G = \{|E \cap F|\}$$

Fx represents elements in F with a weight greater or equal to a given threshold.

4.3. Classification methods

The proposed features were tested on four different ensemble methods, four linear classification and regression algorithms, one optimization algorithm (Stochastic Gradient Descent-SGD) and one Artificial Neural Network (Multi-Layer Perceptron-MLP). The ensemble methods are to combine multiple weak classifiers to create a strong classifier [26]. They are Random Forest (RF), Gradient Boost (GB), AdaBoost (ADA), and Ensemble (It combines the projections of 8 base classifiers to make a final prediction – Voting Classifier). The linear classification and regression algorithms are Logistic Regression (LR), Support Vector Machine (SVM), K-Nearest Neighbours (KNN), and Multinomial Naïve Bayes (NB). They are to evaluate the selected features' relevancy for improved classification accuracy.

5. Experiments and performance evaluation

To evaluate MUNIFES, 20-Newsgroups and its variant 17-Newsgroup datasets were chosen. The threshold feature selection k is set to range from 100 to 1000 at an interval of 100. The varied numbers of feature selection from 100 to 1000 is justified from the previous work [20]. The MUNIFES and other methods were coded in Python 3.11 with numpy, pandas, and scikit-learn packages (latest versions). All experiments were conducted with an Intel® Core ™ i7-1165G7 2.80GHz CPU, 1TB 16GB RAM. The experimental results compared with benchmark feature selection methods are analyzed, and their averages are calculated via 5-fold cross-validation.

5.1. Datasets

The 20-Newsgroups dataset, a popular benchmark collection [20], comprises about 20,000 documents collected from 20 different newsgroups, and the 17-Newsgroups dataset is a variant of the 20-Newsgroup with 16,075 documents (It excepts 3 classes of soc.religion.christian, misc.forsale and alt.atheism from 20-Newsgroup). After the preprocessing stage (removing stop-words, lowercasing, and stemming), documents are represented by the TF-IDF model with a corresponding DF greater than 3. The final selected features have sizes ranging from 100 to 1000 with an interval of 100. Table 1 recaps the experimental datasets.

The dataset is divided into training, validation, and test sets. A validation set is used to tune hyperparameters and make decisions during training. To mitigate the effect of redundancy, unique features were selected in the experiment to improve interpretability, reduce the risk of overfitting, and enhance the model's generalization ability to new data.

5.2. Measurement criteria

This study evaluates the performance of MUNIFES through accuracy (balanced dataset). Accuracy displays the percentage of the correct classification of negative and positive samples.

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN}, \tag{16}$$

where: TP = True Positive, TN = True Negative,

FP = False Positive, FN = False Negative.

5.3. Performance analysis of MUNIFES

To evaluate the strength and stability of MUNIFES, feature selection methods, an ensemble of all the classifiers, and ANN (MLP) are integrated to compare the two datasets. This can be seen in Figures 4-7, evaluated by 8 classifiers, ANN and an ensemble according to the size variation of the selected feature subsets.

5.4. Performance of MUNIFES with 20-Newsgroup

The performance comparison between MUNIFES and base FS methods on 8 classifiers is in Figure 4. MUNIFES outperforms other methods in all cases in terms of accuracy except the KNN classifier in the chi2 method. More specifically, Chi-square outperformed MUNIFES by 100 features on GB with a mild difference of 0.003. For the KNN classifier, MUNIFES recorded an increase of 0.006 on 100 features while it outperformed our method slightly on other feature intervals. As for the performance comparison between MUNIFES and ensemble FS methods in Figure 5, MUNIFES recorded the best performance among the ensemble methods for accuracy. Also, MUNIFES on ANN classifier recorded the best performance among the base FS methods.

5.5. Performance of MUNIFES with 17-Newsgroup

The performance comparison between MUNIFES and base FS methods is in Figure 6. MUNIFES outperforms other methods in all cases in terms of accuracy except the KNN classifier in the chi2 method. More specifically, Chi-square outperformed MUNIFES in the case of 400 features on GB with a mild difference of 0.0038 and 300 and 400 features on RF with a slight difference. For the KNN classifier, MUNIFES recorded a higher number of 0.0010 on 100 features while it outperformed our method slightly on other feature numbers. As for the performance comparison between MUNIFES and ensemble FS methods in Figure 7, MUNIFES recorded the best performance among the ensemble methods for accuracy. Also, MUNIFES on ANN classifier recorded the best performance among the base FS methods.

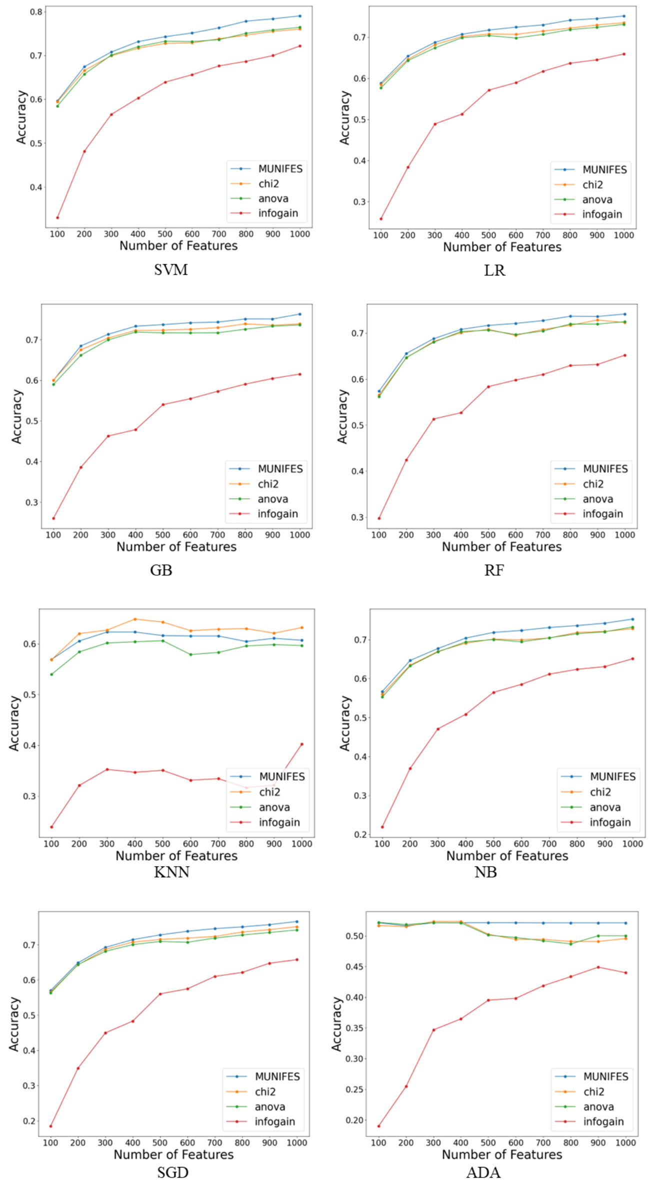

Figure 4. Accuracy comparison between MUNIFES and base FS methods for 20-Newsgroup on 8 classifiers.

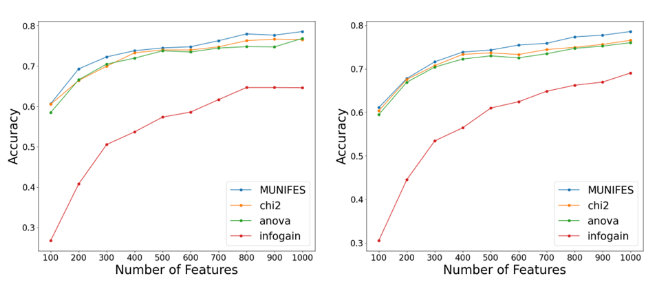

Figure 5. Accuracy comparison between MUNIFES and base FS methods for 20-Newsgroup on ANN and Ensemble classifiers.

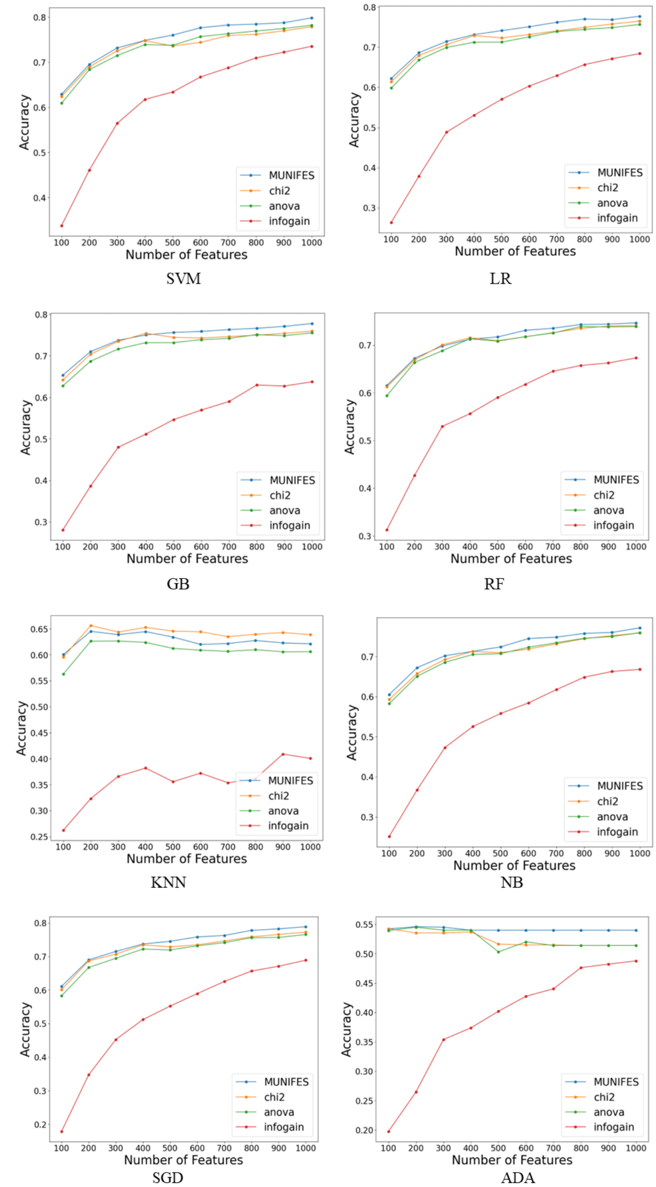

Figure 6. Accuracy comparison between MUNIFES and base FS methods for 17-Newsgroup on 8 classifiers.

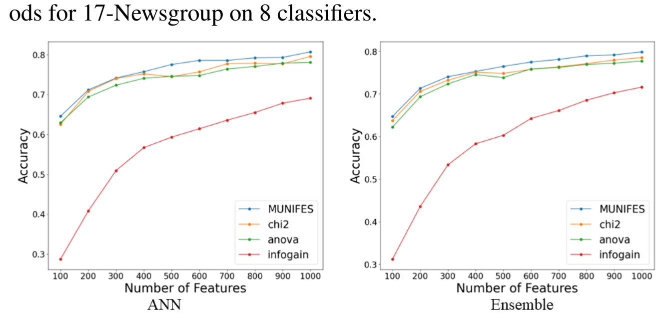

Figure 7. Accuracy comparison between MUNIFES and base FS methods for 17-Newsgroup on ANN and Ensemble classifiers.

5.6. *Statistical significance*

To measure the statistical significance of the accuracy results between MUNIFES and other FS methods, a t-test was

Table 2. Statistical significance of MUNIFES with infogain method - 20-Newsgroup

Intervals	t-statistic	p-value	Significant
100	17.391679	1.06×10^{-12}	+
200	10.791724	2.73×10^{-9}	+
300	6.864304	2.02×10^{-6}	+
400	5.276900	5.12×10^{-5}	+
500	4.323017	4.10×10^{-4}	+
600	3.530491	2.39×10^{-3}	+
700	2.893272	9.69×10^{-3}	+
800	2.374246	2.89×10^{-2}	+
900	2.172257	4.34×10^{-2}	+
1000	1.990538	6.19×10^{-2}	+

Table 3. Statistical significance of MUNIFES with infogain method - 17-Newsgroup

Intervals	t-statistic	p-value	Significant
100	18.53653	3.56×10^{-13}	+
200	12.16363	4.06×10^{-10}	+
300	7.591159	5.13×10^{-7}	+
400	5.816223	1.65×10^{-5}	+
500	4.779981	1.50×10^{-4}	+
600	3.976414	8.85×10^{-4}	+
700	3.300163	3.98×10^{-3}	+
800	2.804303	1.17×10^{-2}	+
900	2.562125	1.96×10^{-2}	+
1000	2.36447	2.95×10^{-2}	+

employed. Tables 2 and 3 show that MUNIFES accuracy results are statistically significant across all classifiers in all intervals of feature selections for both datasets with a 90% confidence based on the P-values for the Infogain method. Chi2 and ANOVA experienced significance at a lower percentage level of confidence.

These results validate the success of the MUNIFES method against other FS methods used for the comparison.

5.7. Discussion

The proposed MUNIFES method is centred on multi-univariate aggregation containing ranking, concatenation, weighting, and voting. This utilized the computational processing power of univariate filters with an enhanced weighted voting mechanism to improve discriminative filtering performance. From the experiment outcomes, SVM, GB, and SGD lead in classification accuracies across the 8 classifiers. The GB showed a higher value from 100 to 400 features ahead of both SVM and SGD, while SVM took over from 500 to 1000 features, leaving GB and SGD in the second and third positions. The two classifiers were swapped between GB and SGD on the second and third positions from 700 to 1000 features. It could be seen that SVM achieved better performance with a higher number of features. This is in line with previous research [20, 27]. A comparison between MUNIFES and traditional base FS methods showed that the former performed better than univariate feature selection methods. Chi2, anova, and infogain methods consider the relevancy between features with their target classes individually with varied considerable results, while MUNIFES concatenated their results through a weighted voting ensemble to pick the best among the equals. Comparison between MUNIFES and the ensemble of classifiers showed on aggregate that the selected features through MUNIFES have better accuracy than others. This is also evident in the ANN classifier results from MUNIFES, which outperformed other FS methods from all the classifiers. This shows that MUNIFES is more robust discriminative and performs better classification than FS univariate base methods.

6. Conclusion

This study aimed to enhance text classification performance by employing a multi-univariate Feature Selection (FS) aggregation strategy encompassing ranking, concatenation, weighting, and voting mechanisms. Leveraging the computational prowess of univariate filters coupled with an advanced weighted voting mechanism, this approach aimed to elevate discriminative filtering efficacy. The novel method, termed MUNIFES, was introduced to realize this objective. Three prominent univariate filter FS techniques, namely Chi-square, ANOVA, and Infogain, were utilized to identify the k best features. Each feature's weight was computed, and scores were documented accordingly. Subsequently, features selected from each FS method were concatenated, preserving their unique attributes and frequencies. Features garnering a minimum majority vote of 2 were chosen from the unique feature pool, with additional consideration given to features with a vote below 2 but exhibiting discriminative weight above a specified threshold. To validate the efficacy of the proposed method, a series of experiments were conducted, assessing accuracy using 8 classifiers (SVM, LR, GB, RF, KNN, NB, SGD, ADA), ensemble classifiers, and ANN classifiers on both the 20-Newsgroups and its variant 17-Newsgroup datasets. Overall, the results demonstrated that MUNIFES surpassed other FS methods in performance. However, it is noted that the predefined threshold in filter-based feature selection methods, as adopted in MUNIFES, may impact both weight selection and feature size. Additionally, the issue of redundancy warrants further investigation, particularly concerning variants of the 20-Newsgroup dataset with fewer classes. Future research endeavors should aim to identify an optimal threshold to stabilize outcomes across diverse text datasets and mitigate redundancy in variant datasets.

Acknowledgment

The authors are grateful to the Editor, and anonymous reviewers for their comments and suggestions to enhance the quality of the paper.

References

[1] Z. Chu, J. He, X. Zhang, X. Zhang & N. Zhu, "Differential privacy high-dimensional data publishing based on feature selection and clustering", Electronics **12** (2023) 1959. https://doi.org/10.3390/electronics12091959.

[2] J. Wan, H. Chen, Z. Yuan, T. Li, X. Yang & B. Bin Sang, "A novel hybrid feature selection method considering feature interaction in neighborhood rough set", Knowledge-Based Systems **227** (2021) 107167. https://doi.org/10.1016/j.knosys.2021.107167.

[3] R. Cekik & A. K. Uysal, "A novel filter feature selection method using rough set for short text data", Expert Systems with Applications **160** (2020) 113691. https://doi.org/10.1016/j.eswa.2020.113691.

[4] G. Taskin, H. Kaya & L. Bruzzone, "Feature selection based on high dimensional model representation for hyperspectral images", IEEE Transactions on Image Processing **26** (2017) 29182928. https://doi.org/10.1109/TIP.2017.2687128.

[5] S. Abasabadi, H. Nematzadeh, H. Motameni & E. Akbari,"Automatic ensemble feature selection using fast non-dominated sorting", Information Systems **100** (2021) 101760. https://doi.org/10.1016/j.is.2021.101760.

[6] N. Hoque, M. Singh & D. K. Bhattacharyya, "EFS-MI: an ensemble feature selection method for classification", Complex & Intelligent Systems **4** (2018) 105. https://doi.org/10.1007/s40747-017-0060-x.

[7] N. Gopika & A. M. Kowshalaya, "Correlation-Based Feature Selection Algorithm for Machine Learning", IEEE, Coimbatore, India, 2018, pp. 692-695. https://doi.org/10.1109/cesys.2018.8723980.

[8] A. K. Uysal & S. Gunal,"A novel probabilistic feature selection method for text classification", Knowledge-Based Systems **36** (2012) 226. https://doi.org/10.1016/j.knosys.2012.06.005.

[9] J. Yang, Y. Liu, X. Zhu, Z. Liu & X. Zhang, "A new feature selection based on comprehensive measurement both in inter-category and intra-category for text categorization", Information Processing & Management **48** (2012) 741. https://doi.org/10.1016/j.ipm.2011.12.005.

[10] W. Zong, F. Wu, L. K. Chu & D. Sculli,"A discriminative and semantic feature selection method for text categorization", International Journal of Production Economics **165** (2015) 215. https://doi.org/10.1016/j.ijpe.2014.12.035.

[11] A. Rehman, K. Javed, H. A. Babri & M. Saeed, "Relative discrimination criterion - A novel feature ranking method for text data", Expert Systems with Applications **42** (2015) 3670. https://doi.org/10.1016/j.eswa.2014.12.013.

[12] N. S. Mohamed, S. Zainudin & Z. Ali Othman, "Metaheuristic approach for an enhanced mRMR filter method for classification using drug response microarray data", Expert Systems with Applications **90** (2017) 224. https://doi.org/10.1016/j.eswa.2017.08.026.

[13] K. Javed & H. A. Babri, "Feature selection based on a normalized difference measure for text classification", Information Processing & Management **53** (2017) 473. https://doi.org/10.1016/j.ipm.2016.12.004.

[14] A. Rehman, K. Javed, H. A. Babri & N. Asim,"Selection of the most relevant terms based on a max-min ratio metric for text classification", Expert Systems with Applications **114** (2018) 78. https://doi.org/10.1016/j.eswa.2018.07.028.

[15] K. Kim & S. Y. Zzang, "Trigonometric comparison measure: A feature selection method for text categorization", Data & Knowledge Engineering **119** (2018) 1. https://doi.org/10.1016/j.datak.2018.10.003.

[16] Z. Manbari, F. AkhlaghianTab & C. Salavati,"Hybrid fast unsupervised feature selection for high-dimensional data", Expert Systems with Applications **124** (2019) 97. https://doi.org/10.1016/j.eswa.2019.01.016.

[17] E. Borandağ, A. Özçift & Y. Kaygusuz, "Development of majority vote ensemble feature selection algorithm augmented with rank allocation to enhance Turkish text categorization", Turkish Journal of Electrical Engineering and Computer Sciences **29** (2021) 514. https://doi.org/10.3906/ELK-1911-116.

[18] P. Drotár, M. Gazda & L. Vokorokos, "Ensemble feature selection using election methods and ranker clustering", Information Sciences **480** (2019) 365. https://doi.org/10.1016/j.ins.2018.12.033.

[19] A. Hashemi, M. B. Dowlatshahi & H. Nezamabadi-pour,"Ensemble of feature selection algorithms: a multi-criteria decision-making approach", International Journal of Machine Learning and Cybernetics **13** (2022) 49. https://doi.org/10.1007/s13042-021-01347-z.

[20] G. Fu, B. Li, Y. Yang & C. Li, "Re-ranking and TOPSIS-based ensemble feature selection with multi-stage aggregation for text categorization", Pattern Recognition Letters **168** (2023) 47. https://doi.org/10.1016/j.patrec.2023.02.027.

[21] T. Sabbah, A. Selamat, M. H. Selamat, R. Ibrahim, & H. Fujita,"Hybridized term-weighting method for Dark Web classification", Neurocomputing **173** (2016) 1908. https://doi.org/10.1016/j.neucom.2015.09.063.

[22] B. Seijo-Pardo, I. Porto-Díaz, V. Bolon-Canedo & A. Alonso-Betanzos,"Ensemble feature selection: Homogeneous and heterogeneous approaches", Knowledge-Based Systems **118** (2017) 124. https://doi.org/10.1016/j.knosys.2016.11.017.

[23] D. S. Guru, M. Suhil, L. N. Raju & N. V. Kumar, "An alternative framework for univariate filter based feature selection for text categorization", Pattern Recognition Letters **103** (2018) 23. https://doi.org/10.1016/j.patrec.2017.12.025.

[24] A. Hashemi, M. Bagher Dowlatshahi & H. Nezamabadi-pour,"A pareto-based ensemble of feature selection algorithms", Expert Systems with Applications **180** (2021) 115130. https://doi.org/10.1016/j.eswa.2021.115130.

[25] L. da F. Costa,"An Introduction to Multisets", Sao Carlos Institute of Physics, Oct. 2021. [Online]. https://doi.org/10.48550/arXiv.2110.12902.

[26] O. J. Ibidoja, F.P. Shan, Mukhtar, J. Sulaiman & M. K. M. Ali, "Robust M-estimators and Machine Learning Algorithms for Improving the Predictive Accuracy of Seaweed Contaminated Big Data", Journal of the Nigerian Society of Physical Sciences **5** (2023) 1137. https://doi.org/10.46481/jnsps.2023.1137.

[27] S. A. Abdulraheem, S. Aliyu & F. B. Abdullahi,"Hyper-parameter tuning for support vector machine using an improved cat swarm optimization algorithm", Journal of the Nigerian Society of Physical Sciences **5** (2023) 1007. https://doi.org/10.46481/jnsps.2023.1007.