Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization

by turnitin turnitin

Available online at :
http://ejournal.amikompurwokerto.ac.id/index.php/telematika/
Telematika
Accreditated SINTA "2" Kemenristek/BRIN, No. 85/M/KPT/2020

UNIVERSITAS AMIKOM PURWOKERTO

Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization

Sarjon Defit[1*], Agus Perdana Windarto[2] and Putrama Alkhairi[3]

[1]Faculty of Computer Science, Universitas Putra Indonesia "YPTK" Padang, West Sumatra, Indonesia
[2,3]Information Systems, STIKOM Tunas Bangsa, Pematangsiantar, North Sumatra, Indonesia
E-mail: sarjon_defit@upiyptk.ac.id[1], agus.perdana@amiktunasbangsa.ac.id[2], putrama@amiktunasbangsa.ac.id[3]

ARTICLE INFO

History of the article:
Received January 1, 2019
Revised January 21, 2019
Accepted January 12, 2019

Keywords:
Sentiment analysis,
Classification methods,
Feature selection,
Ensemble techniques,
Text mining

Correspondece:
Telepon: +62 (0751) 12345678
E-mail:
sarjon_defit@upiyptk.ac.id

ABSTRACT

Optimizing classification methods is critical for sentiment analysis accuracy and reliability. This study contrasts enhanced classification models with standard ones, using methods such as Decision Tree, Random Tree, Naive Bayes, Random Forest, K-NN, Neural Network, and Generalized Linear Model, in the context of political sentiment on social media. Analyzing 1200 tweets from December 10-11, 2023, containing "Indonesia" and "capres," the research includes 490 positive, 355 negative, and 353 neutral sentiments. The refined models achieve a high accuracy of 96.37%. Particularly, the backward selection model excels with 100% accuracy for the Negative class, beneficial where false positives incur high costs. The forward selection model ensures balanced precision across all classes. The low misclassification rate solidifies the Naive Bayes classifier's effectiveness with feature selection and ensemble methods. This paper calls for further hybrid feature selection research and recommends the refined classifiers for high-stakes sentiment analysis. Rigorous pre-deployment testing and ongoing post-deployment updates are advised to sustain model effectiveness. The study underscores the significant impact of feature selection and ensemble methods on sentiment analysis classification performance, especially in sensitive political contexts.

INTRODUCTION

Sentiment analysis is a process aimed at determining the content of text-based datasets, which involves understanding and interpreting the implicit emotions and opinions (Hartmann et al., 2023). Currently, public opinion has become a crucial source in an individual's decision-making process regarding a product (Edara et al., 2023; Kaur & Sharma, 2023). Furthermore, sentiment analysis is a popular field of research, as it offers benefits for various aspects, ranging from social media and public opinion (Bhargav, 2022; Keakde, 2022; Talaat, 2023; Uma, 2022), stock markets and finance (Chong, 2022), brand management and marketing (Kumar, 2022; Win, 2022), elections and political analysis (Keakde, 2022; Sutriawan, 2023; Talaat, 2023), customer service and feedback (Bharathi, 2023; Bhargav, 2022), health and medical research (Che, 2023; Chong, 2022), entertainment and film (Zheng, 2019), education (Derisma, 2020; Keakde, 2022), environmental issues and natural disasters (Behl, 2021; Navarro, 2023; Nguyen, 2023; Pappas, 2017; V. Priya, 2016; Ragini, 2018), and travel and tourism (Gholipour, 2020; Luo, 2021; P. S. Priya, 2023; Sontayasara, 2021; Zapata, 2019). In today's rapidly evolving digital era, sentiment analysis has become an essential tool for interpreting public opinion, especially in political contexts like presidential elections. This activity is important as it provides insights into how presidential candidates are perceived by the public, potentially influencing campaign directions and political strategies. On the other hand, sentiment analysis faces the challenge of complexity in interpreting large and diverse text data (Krishna, 2023; Suhaimin, 2023). In this context, classification is at the heart of sentiment analysis, as it allows the sorting of opinions into different categories (positive, negative, neutral), offering a clearer and more structured view of public sentiment (Errami, 2023; Hung, 2023; Lasri, 2023; G. Li, 2023).

Classification methods for sentiment analysis towards political candidates, such as presidents, were suggested by (Ali, 2022) who recommended techniques like Naive Bayes, SVM, or deep learning models such as CNN or LSTM for large-scale sentiment analysis on tweets. However, the extensive volume of Twitter data presents challenges in processing and analysis, as well as difficulties in handling sarcasm and slang. Following this, (Bringula, 2023) proposed NLP methods for text analysis (comments or transcripts) and image or audio analysis techniques for YouTube videos. Yet, sentiment analysis on video content may be limited by transcript quality and the complexity of interpreting visual and audio content. Furthermore, (Budiharto, 2018) suggested Naive Bayes, SVM, or neural networks like LSTM for sentiment analysis of tweets. However, Twitter data often contains slang and abbreviations that can complicate the analysis process. Next, (Buntoro & G A, 2021) proposed using Decision Trees, Random Forest, SVM, or deep learning approaches. However, there are issues with overfitting, especially when using complex algorithms on noisy data. Additionally, (Endsuy, 2021) recommended the VADER method for lexicon and rule-based sentiment analysis. Yet, VADER may not always be effective in capturing more subtle sentiment nuances, especially in political contexts. Following that, (Fagbola, 2019) proposed lexicon-based approaches for sentiment analysis, possibly with additional methods to identify content from bots. However, lexicon approaches may be limited in capturing context and irony in text. Subsequently, (Hananto, 2023) proposed Decision Trees, Random Forest to test and compare several algorithms to find the most effective for sentiment analysis on Twitter, but issues with overfitting remain. Next, (Murfi, 2019) suggested LDA for topic modeling to integrate topic modeling with sentiment analysis, although integrating topic modeling with sentiment analysis may be complex and computationally demanding. Lastly, (Syahriani, 2020) proposed Naive Bayes for sentiment analysis on Facebook comments. However, the Naive Bayes method might oversimplify due to the assumption of feature independence. GPT

In efforts to enhance classification accuracy in sentiment analysis, several crucial aspects need to be considered, especially in light of the various weaknesses present in the methods described. One issue in text sentiment classification is the abundance of attributes used in a dataset (Bordoloi & Biswas, 2023; Choi & Lee, 2017). Generally, the attributes in text sentiment classification are vast, and using all these attributes can diminish the performance of the classifier (Y. Li et al., 2018; Saraswathi et al., 2023; Yu et al., 2019). In the conducted research, optimizing the sentiment analysis model is identified as one of the best solutions to address the various challenges that arise (Wankhade et al., 2022). Optimization in sentiment analysis aims to enhance the accuracy and efficiency of classification methods (Nayak et al., 2023). Feature selection is a critical part of optimizing the performance of classifiers by reducing a large feature space, for example, by eliminating less relevant attributes (Alirezanejad et al., 2020; Urbanowicz et al., 2018). Additionally, the proper use of feature selection can increase accuracy (Khaire & Dhanalakshmi, 2022). Several researchers have compared various classification methods and feature selection techniques to achieve optimal results. Studies conducted by (Pande et al., 2023) employed classification methods such as Support Vector Machine (SVM), Perceptron, K-Nearest Neighbor (KNN), Stochastic Gradient Descent (SGD), and XGBoost. For feature selection, techniques like correlation-based feature selection, principal component analysis (PCA), linear discriminant analysis (LDA), recursive feature elimination (RFE), and univariate feature selection were used. The findings indicated that the best feature selection was correlation-based feature selection, with the highest accuracy achieved being 99.87% when using the XGBoost classifier. Further research by (Rahmadani et al., 2018) applied Naive Bayes and Decision Tree classification methods with genetic algorithm (GA) for feature selection. This study demonstrated that feature selection using GA could enhance accuracy with the Decision Tree method. Subsequently, ensemble methods (AdaBoost and Bagging) can also improve the performance of classifiers (Nti et al., 2020; Teoh et al., 2022). This was evidenced by (Zaini & Awang, 2022) who utilized methods such as logistic regression (LR), support vector classifier (SVC), random forest (RF), extra tree classifier (ETC), naïve bayes (NB), extreme gradient boosting (XGB), decision tree (DT), k-nearest neighbor (KNN), multilayer perceptron (MLP), and stochastic gradient descent (SGD). Moreover, the ensemble method known as stacking was shown to yield the best results when logistic regression was used for classification, achieving an accuracy of 90.16%. From all the research findings presented, there is potential to further improve classification performance through appropriate optimization. Therefore, this study aims to explore the effectiveness of various classification methods (K-NN, Naive Bayes, Random Forest, Decision Tree, Neural Network, Support Vector Machine, Linear Regression, Generalized Linear Model) in analyzing sentiment towards presidential candidates, integrated with several feature selection techniques (Forward Selection, Backward Elimination, Optimize Selection) and ensemble methods (AdaBoost and Bagging). The comparative results will determine the best classification in the context of presidential elections. These findings also contribute new insights to the research conducted.

RESEARCH METHODS

In the study titled "Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization", the research methods involved processing data using a laptop with specifications of an Intel Core i9 CPU at 1.9 GHz, 32GB RAM, and Microsoft Windows 11 Professional 64-bit operating system. The application utilized for this purpose was RapidMiner 9.1. The research data were collected through crawling Twitter data using keywords “capres” or “presidential candidate”, and “Indonesia”, with a total of 1200 tweets gathered between December 10 and 11, 2023. The data comprised 490 positive tweets, 355 negative tweets, and 353 neutral tweets.

The flowchart outlined below presents a detailed process for research methodologies that are widely utilized in machine learning and data analysis, with a particular focus on natural language processing (NLP) and classification tasks. The research commenced with the acquisition of data, during which data was extracted and harvested from Twitter using a specialized scraper. This gathered data was subsequently auto-saved in an Excel spreadsheet format (.xls) to facilitate further processing and data labeling. Presented here is the complete research framework pertaining to the study as depicted in Figure 1 below.

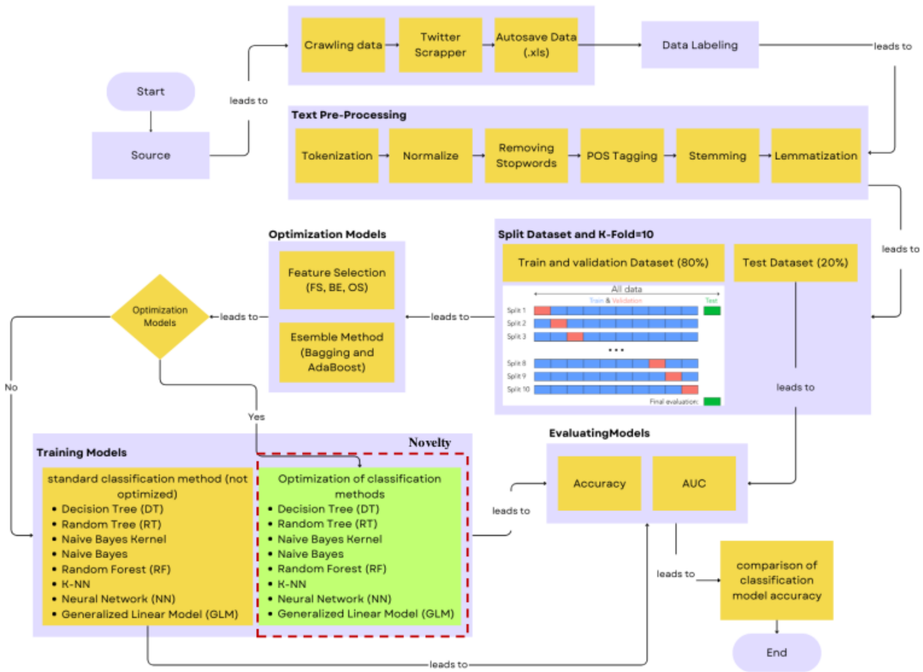

Figure 1. Research framework

1. Text Pre-Processing

Once the data is acquired, it undergoes a crucial preprocessing phase to ensure its quality and usability for machine learning tasks. This phase involves several steps:

a. **Tokenization**: The text is broken down into individual words or tokens.

b. **Normalization**: The tokens are normalized to ensure consistency, which may include converting to lowercase.

c. **Removing Stopwords**: Commonly used words that do not contribute to the meaning of the text, such as "the" and "is," are removed.

d. **POS Tagging**: Part-of-speech tags are assigned to each word to identify their grammatical role.

e. **Stemming**: Words are reduced to their root form, which may sometimes result in non-actual words.

f. **Lemmatization**: Similar to stemming, but ensures that the root word belongs to the language.

2. Model Training

Following preprocessing, the data can follow two distinct paths depending on whether optimization models are applied.

a. **Without Optimization**: The data is used to train standard classification models without any optimization. These models include:

   1. Decision Tree (DT)
   2. Random Tree (RT)
   3. Naive Bayes Kernel
   4. Naive Bayes

5. Random Forest (RF)
6. K-NN
7. Neural Network (NN)
8. Generalized Linear Model (GLM)

b. **With Optimization**: Before training, the data is processed through optimization models, which involve feature selection methods like Filter Selection (FS), Backward Elimination (BE), or other techniques (OS). Additionally, ensemble methods like Bagging and AdaBoost are employed to enhance the performance of the classifiers. The optimized classifiers include the same list as the standard methods but are expected to perform better due to the optimization.

1) Forward Selection (FS)
Start with no variables in the model, test the addition of each variable using a chosen model fit criterion (like R-squared, AIC, BIC, etc.), add the variable that improves the model the most, and repeat until no significant improvement is made.

2) Backward Elimination (BE)
Start with all variables in the model, remove the variable that has the least statistical significance (like the one with the highest p-value), and repeat until all variables in the model are significant.

3) Optimized Selection (OS)
This can be a combination of both forward selection and backward elimination, or any other optimization algorithm that evaluates the importance of each feature based on model performance metrics. The mathematical representation of the model fit criterion might be:

for R-squared: $R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$ (1)

for Akaike Information Criterion (AIC): AIC= 2k – 2 ln(L) (2)

for Bayesian Information Criterion (BIC): BIC = ln(n)k – 2 ln(L) (3)

where SS_{res} is the sum of squres of residuals, SS_{tot} is the total sum of squares, k is the number of parameters in the model, L is the maximized value of the likelihood function of the models, and n is the number of observations.

4) Bagging
Bagging involves creating multiple models (usually of the same type) from different subsets of the training dataset. The final model's output is the average of all the models' outputs for regression or the majority vote for classification. Mathematical representation for regression could be:

F(x)= $\frac{1}{B}\sum_{b=1}^{B} F_b(x)$ (4)

for classification, it's a majority vote among the B classifiers

5) AdaBoost
AdaBoost combines weak classifiers to form a strong classifier. Each weak classifier's vote is weighted based on its accuracy, and after each iteration, the weights of the training instances are updated to focus on the more difficult cases.Split Dataset and K-Fold=10: Dividing the dataset into training and testing

data with an 80:20 ratio and using K-Fold Cross Validation with K=10 to validate the model. The final strong classifier is:

$$F(x)=\sum_{t=1}^{T} \alpha_t f_t(x) \quad (5)$$

where $f_t(x)$ is the output of the weak classifier, α_t is weight assigned to that classifier, and T is the totals number of weak classifiers. The weights α_t are calculated using the formula:

$$\alpha_t = \frac{1}{2} \ln(\frac{1-E_t}{E_t}) \quad (6)$$

where E_t is the errore rate of the weak classifier.

3. Dataset Splitting and Validation

Independently of the optimization, the dataset is split into training and validation sets using a K-Fold (with K=10) cross-validation method. Typically, 80% of the data is used for training and the remaining 20% for testing.

4. Model Evaluation

Models are evaluated based on metrics such as Accuracy and Area Under the Receiver Operating Characteristic Curve (AUC). These metrics provide insight into the performance of the classifiers, taking into account both the true positive rate and the false positive rate.

5. Comparison and Conclusion

Finally, the classification models, whether optimized or not, are compared based on their accuracy. This comparison allows for a critical assessment of the impact of optimization techniques on model performance. The research concludes with the selection of the best-performing model, marking the end of the machine learning pipeline.

RESULTS AND DISCUSSION

This research involved the development of an optimized classification model and a thorough analysis comparing it to basic classifications. The classification methods utilized encompass Decision Tree (DT), Random Tree (RT), Naive Bayes Kernel, Naive Bayes, Random Forest (RF), K-NN, Neural Network (NN), and Generalized Linear Model (GLM). We utilized various optimization techniques, including feature selection (Forward Selection, Backward Elimination, Optimize Selection) and ensemble methods. Afterwards, the results were examined using RapidMiner Studio software, where they were assessed using a confusion matrix that included metrics such as accuracy, classification error, weighted mean recall, weighted mean precision, root mean squared error, and correlation. Here are the model measurement results, providing a concise summary of the comparative analysis of the classification methods.

Table 1. Matrix confusion results in standard classification

Classifiers 80:20 and k-fold= 10	Standard					
	accuracy	classification error	weighted mean recall	weighted mean precision	root MSE	correlation
Decision Tree	67,36	32,64	55,54	51,53	0,5240	0,180
Random Tree	40,23	59,77	33,33	13,41	0,642	0,000
Naive Bayes Kernel	55,27	44,73	44,59	52,54	0,5910	0,070
Naive Bayes	70,38	29,62	63,88	63,07	0,544	0,068
Random Fores	67,76	32,24	55,87	49,85	0,636	0,156

Classifiers 80:20 and k-fold= 10	Standard					
	accuracy	classification error	weighted mean recall	weighted mean precision	root MSE	correlation
K-NN	73,34	26,66	60,88	62,32	0,492	0,199
Neural Network	40,43	59,57	33,33	13,48	0,642	0,000
Generalized Linear Model (GLM)	72,47	27,53	60,42	58,56	0,487	0,106

After examining Table 1, which presents a comparison of different classification models, it becomes evident that the K-Nearest Neighbors (K-NN) model stands out as the most efficient option. It boasts an impressive accuracy rate of 73.34%, along with remarkable recall, precision, and the lowest RMSE. These findings highlight the model's exceptional classification precision. The Generalized Linear Model (GLM) also showed strong performance, closely matching K-NN in terms of accuracy. On the other hand, the effectiveness of Random Tree and Neural Network models was found to be lower. The Decision Tree and Random Forest models, although they achieved an accuracy rate of over 67%, did not perform as efficiently as the K-NN and GLM models. The performance of Naive Bayes and Naive Bayes Kernel varied, with standard Naive Bayes showing slightly higher accuracy. In this case, the best model choice depends on the specific data application. K-NN and GLM are considered to be the top choices.

Table 2. Matrix confusion results in optimized classification

Classifiers 80:20 and k-fold= 10	Feature Selection (Forward Selection) + esemble method					
	accuracy	classification error	weighted mean recall	weighted mean precision	root MSE	correlation
Decision Tree	68,910	31,090	56,810	52,110	0,515	0,183
Random Tree	40,430	59,570	33,330	13,480	0,642	0,000
Naive Bayes Kernel	88,780	11,220	87,740	91,840	0,278	0,869
Naive Bayes	96,370	3,630	96,520	95,230	0,188	0,904
Random Fores	70,990	29,010	58,530	52,380	0,629	0,193
K-NN	79,720	20,280	66,770	82,680	0,367	0,290
Neural Network	40,430	59,570	33,330	13,480	0,642	0,000
Generalized Linear Model (GLM)	85,160	14,840	75,750	90,850	0,414	0,510

Table 3. Matrix confusion results in optimized classification

Classifiers 80:20 and k-fold= 10	Feature Selection (Backward Selection) + esemble method					
	accuracy	classification error	weighted mean recall	weighted mean precision	root MSE	correlation
Decision Tree	68,3	31,7	56,31	51,95	0,517	0,181
Random Tree	40,43	59,57	33,33	13,48	0,642	0,000
Naive Bayes Kernel	89,05	10,95	88,08	92,06	0,268	0,876
Naive Bayes	96,37	3,63	96,52	95,24	0,19	0,906
Random Fores	41,84	58,16	34,56	56,04	0,631	0,000
K-NN	94,43	5,57	92,32	95,32	0,362	0,878
Neural Network	40,43	59,57	33,33	13,48	0,647	0,000
Generalized Linear Model (GLM)	85,56	14,44	76,39	90,42	0,415	0,516

Table 4. Matrix confusion results in optimized classification

Classifiers 80:20 and k-fold= 10	Feature Selection (Optimize Selection Burte-Force) + esemble method					
	accuracy	classification error	weighted mean recall	weighted mean precision	root MSE	correlation
Decision Tree	58,81	20,99	46,71	42,01	0,503	0,171
Random Tree	30,33	49,47	23,23	12,38	0,630	0,000
Naive Bayes Kernel	78,68	10,12	77,64	81,74	0,266	0,857
Naive Bayes	86,27	2,53	86,42	85,13	0,176	0,892
Random Fores	60,89	18,91	48,43	42,28	0,617	0,181
K-NN	69,62	10,18	56,67	72,58	0,355	0,278
Neural Network	30,33	49,47	23,23	12,38	0,630	0,000
Generalized Linear Model (GLM)	75,06	13,74	65,65	80,75	0,402	0,498

Tables 2, 3, and 4 present the confusion matrices' outcomes for different classification models that were optimized through the implementation of feature selection and ensemble methods. A number of noteworthy patterns and discovered insights are presented in these tables. **Table 2** presents the results of forward feature selection. Notably, the Naive Bayes and Naive Bayes Kernel models exhibit exceptional accuracy enhancements, surpassing 88% and 96% correspondingly. In contrast, K-NN and GLM demonstrate substantial advancements as well, whereas Random Tree and Neural Network continue to exhibit subpar performance. **Table 3** demonstrates that when backward feature selection is implemented, the accuracy of K-NN nearly soars to 94%. With exceptionally high accuracy, Naive Bayes and Naive Bayes Kernel continue to exhibit their superiority, whereas GLM sustains its performance enhancement. Nevertheless, Random Tree and Neural Network fail to demonstrate substantial advancements once more. Finally, certain models, including the Random Tree and Decision Tree, exhibit a performance decrease when optimal feature selection is implemented, as shown in **Table 4.** In contrast, Naive Bayes maintains its strength despite a marginal decline from prior outcomes, while GLM and K-NN exhibit comparatively moderate performance in comparison to the aforementioned tables. The findings suggest that employing a brute-force strategy for feature selection does not consistently result in enhanced performance, and its impact is significantly contingent upon the specific model being utilized. In general, the outcomes presented in these three tables indicate that ensemble methods and optimization via feature selection can significantly affect the performance of classification models.

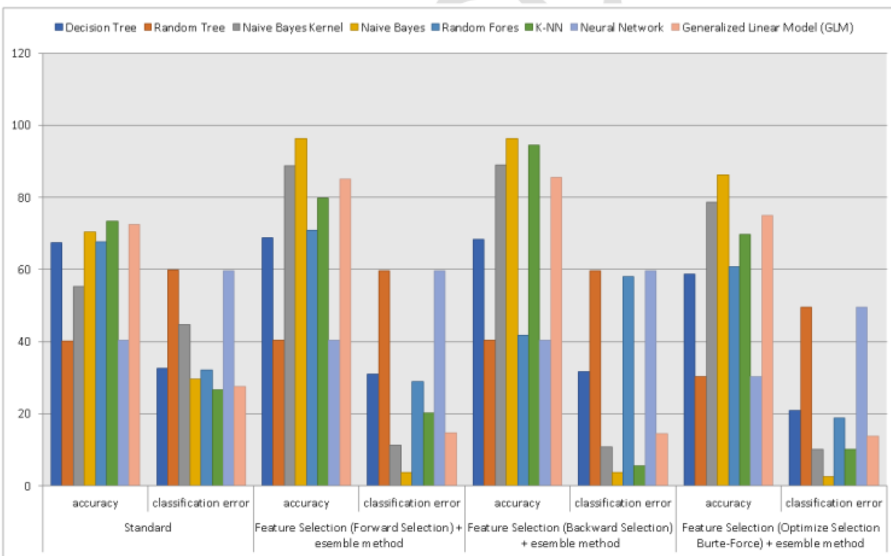

Figure 1. The graphic results summarize the standard classification method and the optimized classification method (accuracy and classification error)

In **Figure 1** (source table 1 to 4), there is a comparison between basic classification and three different approaches in feature selection combined with ensemble methods for the purpose of classification in sentiment analysis. **The first approach**, using Forward Selection, shows significant variation in accuracy rates, ranging from 40.43% to 96.37%, with classification errors ranging from 3.63% to 59.57%. This indicates that this approach might be sensitive to the dataset used and the initial feature selection. **The second approach**, using Backward Selection, appears to produce similar levels of accuracy as the first approach, but with slightly better consistency in reducing classification errors, the lowest still at 3.63% and the highest at 59.57%. This similarity suggests that both methods may have comparable effectiveness, but Backward Selection might have an advantage in handling overfitting. **The third approach**, using the Brute-Force method in feature selection optimization, tends to have lower accuracy, with the highest value only reaching 86.27%, and classification errors ranging from 2.53% to 49.47%. Although it does not always yield higher results in terms of accuracy, this approach may offer a better balance between adapting to training data and generalizing to unknown data, indicating a potentially higher reliability in practical applications. **Overall**, the Backward Selection approach, along with ensemble methods, emerges as a more promising strategy, offering better consistency in performance.

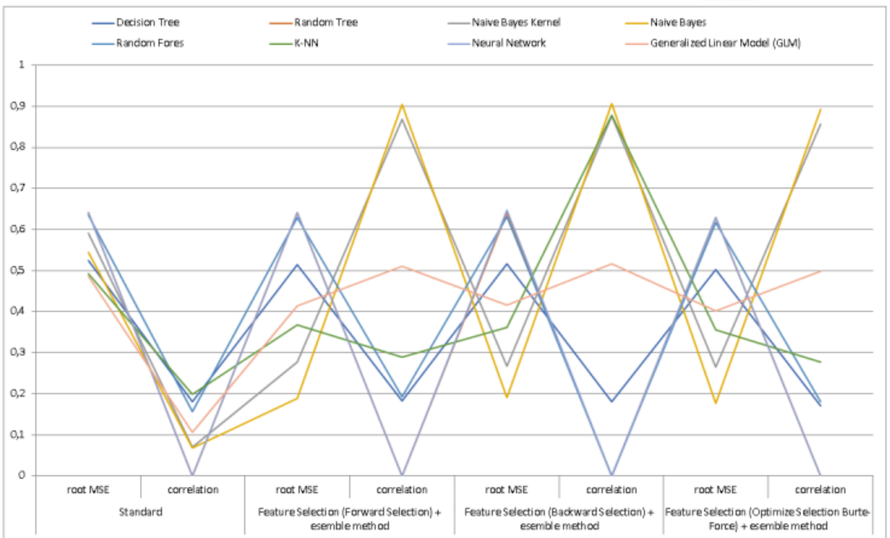

Figure 2. The graphic results summarize the standard classification method and the optimized classification method (root MSE and correlation)

In **Figure 2** (source table 1 to 4), there are standard approaches and three feature selection techniques combined with ensemble methods. The two main metrics used for evaluation are the square root of Mean Square Error (root MSE) and correlation coefficient, applied in an 80:20 data split scheme and 10-fold cross-validation. On closer examination, the Decision Tree algorithm shows a slight improvement in both metrics after the application of feature selection techniques, with the greatest improvement seen in the use of Optimize Selection Brute-Force. For Random Tree, there is almost no change in root MSE or correlation, except for a minimal decrease in root MSE with the application of Optimize Selection Brute-Force. Then, both the Naive Bayes Kernel and the standard Naive Bayes experience significant spikes in correlation and substantial decreases in root MSE when feature selection techniques are applied. This suggests that feature selection is highly beneficial for these two Naive Bayes variants. K-NN also benefits dramatically, particularly in correlation improvement when feature selection techniques are used, with the greatest increase coming from Backward Selection.

Meanwhile, Random Forest shows only marginal improvement with the application of feature selection techniques. For Neural Network, there is no significant change with Forward and Backward Selection, but there is a decrease in root MSE with Optimize Selection Brute-Force. Lastly, the Generalized Linear Model (GLM) shows a consistent and beneficial increase in correlation and decrease in root MSE with all the tested feature selection techniques. **Overall**, the Optimize Selection Brute-Force technique appears to be the most effective overall in enhancing the performance of various classification algorithms.

Based on two figures comparing various classification algorithms (figure 1 and 2), it is evident that the Naive Bayes method integrated with Feature Selection techniques, both Forward and Backward, combined with ensemble methods, stands out as the front runner. This method not only scores an impressive accuracy rate, peaking at 96.37%, but also records the lowest classification error, at 3.63%. The very high correlation and low root MSE displayed in the second table affirm the superiority of this method, with correlation nearing 0.9 and root MSE around 0.19, indicating predictions that are highly consistent with the reality of the observed data. Therefore, the **Naive Bayes method** with Forward or Backward Feature Selection plus ensemble methods is the optimal choice. This conclusion is drawn considering the balance between accuracy, classification error, prediction consistency, and alignment with actual data, meaning the Naive Bayes method has a high predictive capability and a high level of reliability in correlating prediction outcomes with actual values. Here are the complete results of the analysis of the Naive Bayes method optimized with feature selection (both Forward or Backward and ensemble method) as shown in the following figure.

accuracy: 96.37%

	true Negatif	true Positif	true Netral	class precision
pred. Negatif	556	0	1	99.82%
pred. Positif	11	602	7	97.10%
pred. Netral	35	0	277	88.78%
class recall	92.36%	100.00%	97.19%	

Figure 3. Analysis results of the Naive Bayes method with Feature Selection (Forward Selection) and ensemble method

In **Figure 3**, the model demonstrates very high precision and recall in identifying the Positive class, with a perfect recall of 100%. The Negative class also shows high precision at 99.82%, but with a slightly lower recall at 92.36%. The Neutral class, while having the lowest precision and recall at 88.78% and 97.19% respectively, still exhibits strong performance. Classification errors are minimal, with only 11 instances of Negative predicted as Positive and 35 instances of Neutral predicted as Negative.

accuracy: 96.37%

	true Negatif	true Positif	true Netral	class precision
pred. Negatif	556	0	0	100.00%
pred. Positif	11	602	8	96.94%
pred. Netral	35	0	277	88.78%
class recall	92.36%	100.00%	97.19%	

Figure 4. Analysis results of the Naive Bayes method with Feature Selection (Backward Selection) and ensemble method

In **Figure 4,** the precision for the Negative class has increased to 100%, indicating that every instance predicted as Negative is indeed Negative. However, the precision for the Positive class has slightly decreased compared to the first model, now being 96.94%. The precision for the Neutral class remains unchanged. The recall rates are identical to the first model. The classification errors are similar, with 11 instances of Negative predicted as Positive, but there is an increase in the misclassification of Neutral instances as Positive, with 8 occurrences.

Both models exhibit high accuracy, precision, and recall across all classes. The model in **Figure 3** demonstrates a more balanced performance among the classes in terms of precision, whereas the model in **Figure 4** maximizes precision for the Negative class with a slight compromise on the Positive class. Misclassifications are minimal in both models, indicating their effectiveness. Thus, the Naive Bayes method is the best classification method in the sentiment analysis.

PERFORMANCE VECTOR

```
PerformanceVector:
accuracy: 96.37%
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         1
Positif:  11        602       7
Netral:   35        0         277
classification_error: 3.63%
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         1
Positif:  11        602       7
Netral:   35        0         277
weighted_mean_recall: 96.52%, weights: 1,
1, 1
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         1
Positif:  11        602       7
Netral:   35        0         277
weighted_mean_precision: 95.23%, weights:
1, 1, 1
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         1
Positif:  11        602       7
Netral:   35        0         277
root_mean_squared_error: 0.188 +/- 0.000
correlation: 0.904
```

(a) The naive bayes method with feature selection (forward selection) and ensemble method

PERFORMANCE VECTOR

```
PerformanceVector:
accuracy: 96.37%
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         0
Positif:  11        602       8
Netral:   35        0         277
classification_error: 3.63%
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         0
Positif:  11        602       8
Netral:   35        0         277
weighted_mean_recall: 96.52%, weights: 1,
1, 1
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         0
Positif:  11        602       8
Netral:   35        0         277
weighted_mean_precision: 95.24%, weights:
1, 1, 1
ConfusionMatrix:
True:     Negatif   Positif   Netral
Negatif:  556       0         0
Positif:  11        602       8
Netral:   35        0         277
root_mean_squared_error: 0.190 +/- 0.000
correlation: 0.906
```

(b) The naive bayes method with feature selection (backward selection) and ensemble method

Figure 5. Performance Vectors of the Naive Bayes method

In comparing two sentiment analysis methods (Figure 5), both demonstrate high accuracy (96.37%) with minimal differences. The first (a), combining Naive Bayes with forward feature selection and an ensemble approach, excels in classifying negative sentiments and shows good accuracy for positive and neutral ones. The second (b), using backward feature selection with an ensemble, is similarly accurate but slightly better in precision (95.24% vs. 95.23%), indicating a minor advantage in reducing false positives. Both have a classification error of 3.63% and a weighted mean recall of 96.52%, showcasing their effectiveness across sentiment classes.

CONCLUSIONS AND RECOMMENDATIONS

Overall, the evaluation of Naive Bayes classification models that have been improved with feature selection and ensemble approaches reveals a distinct advantage over traditional classification techniques.

The updated models have demonstrated exceptional accuracy, precision, and recall, rendering them very dependable for delicate analytical tasks. The combination of the backward feature selection strategy with ensemble approaches has demonstrated exceptional precision in the Negative class, highlighting the advantages of optimization in classification models. Forward selection, however, has ensured an equitable precision across classes, which is crucial for sustaining a complete predictive performance.

Based on these findings, it is advisable to implement the optimized Naive Bayes technique for tasks that prioritize precision and accuracy. The decision to use forward or backward selection should be based on the unique requirements for achieving balanced class precision or emphasizing specific classes. It is recommended to incorporate ensemble approaches to enhance the models' ability to generalize and reduce the risk of overfitting. Additional exploration of hybrid feature selection techniques may provide even more powerful categorization methodologies..

REFERENCES/BIBLIOGRAPHY

Ali, R. H. (2022). A large-scale sentiment analysis of tweets pertaining to the 2020 US presidential election. *Journal of Big Data*, *9*(1). https://doi.org/10.1186/s40537-022-00633-z

Alirezanejad, M., Enayatifar, R., Motameni, H., & Nematzadeh, H. (2020). Heuristic filter feature selection methods for medical datasets. *Genomics*, *112*(2), 1173–1181. https://doi.org/10.1016/j.ygeno.2019.07.002

Behl, S. (2021). Twitter for disaster relief through sentiment analysis for COVID-19 and natural hazard crises. *International Journal of Disaster Risk Reduction*, *55*. https://doi.org/10.1016/j.ijdrr.2021.102101

Bharathi, R. (2023). Leveraging Deep Learning Models for Automated Aspect Based Sentiment Analysis and Classification. *SSRG International Journal of Electrical and Electronics Engineering*, *10*(5), 120–130. https://doi.org/10.14445/23488379/IJEEE-V10I5P111

Bhargav, M. (2022). Comparative Analysis and Design of Different Approaches for Twitter Sentiment Analysis and classification using SVM. *International Journal on Recent and Innovation Trends in Computing and Communication*, *10*(9), 60–66. https://doi.org/10.17762/ijritcc.v10i9.5706

Bordoloi, M., & Biswas, S. K. (2023). Sentiment analysis: A survey on design framework, applications and future scopes. In *Artificial Intelligence Review* (Vol. 56, Issue 11). Springer Netherlands. https://doi.org/10.1007/s10462-023-10442-2

Bringula, R. (2023). YouTube Videos on the Achievements of Presidential Candidates: Sentiment and Content Analysis. *Journal of Political Marketing*. https://doi.org/10.1080/15377857.2023.2202617

Budiharto, W. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. *Journal of Big Data*, *5*(1). https://doi.org/10.1186/s40537-018-0164-1

Buntoro, & G A. (2021). Implementation of a Machine Learning Algorithm for Sentiment Analysis of Indonesia's 2019 Presidential Election. *IIUM Engineering Journal*, *22*(1), 78–92. https://doi.org/10.31436/IIUMEJ.V22I1.1532

Che, S. P. (2023). Effect of daily new cases of COVID-19 on public sentiment and concern: Deep learning-based sentiment classification and semantic network analysis. *Journal of Public Health (Germany)*. https://doi.org/10.1007/s10389-023-01833-4

Choi, Y., & Lee, H. (2017). Data properties and the performance of sentiment classification for electronic commerce applications. *Information Systems Frontiers*, *19*(5), 993–1012. https://doi.org/10.1007/s10796-017-9741-7

Chong, K. S. (2022). Comparison of Naive Bayes and SVM Classification in Grid-Search Hyperparameter Tuned and Non-Hyperparameter Tuned Healthcare Stock Market Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*, *13*(12), 90–94. https://doi.org/10.14569/IJACSA.2022.0131213

Derisma. (2020). Comparing the classification methods of sentiment analysis on a public figure on indonesian-language social media. *Journal of Theoretical and Applied Information Technology*, *98*(8), 1214–1220.

Edara, D. C., Vanukuri, L. P., Sistla, V., & Kolli, V. K. K. (2023). Sentiment analysis and text categorization of cancer medical records with LSTM. *Journal of Ambient Intelligence and Humanized Computing*, *14*(5), 5309–5325. https://doi.org/10.1007/s12652-019-01399-8

Endsuy, A. R. D. (2021). Sentiment Analysis between VADER and EDA for the US Presidential Election 2020 on Twitter Datasets. *Journal of Applied Data Sciences*, *2*(1), 8–18. https://doi.org/10.47738/jads.v2i1.17

Errami, M. (2023). Sentiment Analysis on Moroccan Dialect based on ML and Social Media Content Detection. *International Journal of Advanced Computer Science and Applications*, *14*(3), 415–425. https://doi.org/10.14569/IJACSA.2023.0140347

Fagbola, T. M. (2019). Lexicon-based bot-aware public emotion mining and sentiment analysis of the Nigerian 2019 presidential election on Twitter. *International Journal of Advanced Computer Science and Applications*, *10*(10), 329–336. https://doi.org/10.14569/ijacsa.2019.0101047

Gholipour, H. F. (2024). Business Sentiment and International Business Travels: A Cross-country Analysis. *Journal of Travel Research*, *59*(6), 1061–1072. https://doi.org/10.1177/0047287519872828

Hananto, A. L. (2023). Best Algorithm in Sentiment Analysis of Presidential Election in Indonesia on Twitter. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(6), 473–481.

Hartmann, J., Heitmann, M., Siebert, C., & Schamp, C. (2023). More than a Feeling: Accuracy and Application of Sentiment Analysis. *International Journal of Research in Marketing*, *40*(1), 75–87. https://doi.org/10.1016/j.ijresmar.2022.05.005

Hung, L. P. (2023). Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection. *Journal of Advanced Computational Intelligence and Intelligent Informatics*, *27*(1), 84–95. https://doi.org/10.20965/jaciii.2023.p0084

Kaur, G., & Sharma, A. (2023). A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-022-00680-6

Keakde, M. K. (2022). Study and analysis of various sentiment classification strategies: A challenging overview. *International Journal of Modeling, Simulation, and Scientific Computing*, *13*(1). https://doi.org/10.1142/S1793962322500015

Khaire, U. M., & Dhanalakshmi, R. (2022). Stability of feature selection algorithm: A review. *Journal of King Saud University - Computer and Information Sciences*, *34*(4), 1060–1073. https://doi.org/10.1016/j.jksuci.2019.06.012

Krishna, R. (2023). Machine Learning Based Twitter Sentiment Analysis and User Influence. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*, 215–221. https://doi.org/10.17762/ijritcc.v11i8s.7192

Kumar, V. V. (2022). Aspect based sentiment analysis and smart classification in uncertain feedback pool. *International Journal of System Assurance Engineering and Management*, *13*, 252–262. https://doi.org/10.1007/s13198-021-01379-2

Lasri, I. (2023). Real-time Twitter Sentiment Analysis for Moroccan Universities using Machine Learning and Big Data Technologies. *International Journal of Emerging Technologies in Learning*, *18*(5), 42–61. https://doi.org/10.3991/ijet.v18i05.35959

Li, G. (2023). Data augmentation for aspect-based sentiment analysis. *International Journal of Machine Learning and Cybernetics*, *14*(1), 125–133. https://doi.org/10.1007/s13042-022-01535-5

Li, Y., Guo, H., Zhang, Q., Gu, M., & Yang, J. (2018). Imbalanced text sentiment classification using universal and domain-specific knowledge. *Knowledge-Based Systems*, *160*, 1–15. https://doi.org/10.1016/j.knosys.2018.06.019

Luo, Y. (2021). Tourism Attraction Selection with Sentiment Analysis of Online Reviews Based on Probabilistic Linguistic Term Sets and the IDOCRIW-COCOSO Model. *International Journal of Fuzzy Systems*, *23*(1), 295–308. https://doi.org/10.1007/s40815-020-00969-9

Murfi, H. (2019). Topic features for machine learning-based sentiment analysis in Indonesian tweets. *International Journal of Intelligent Computing and Cybernetics*, *12*(1), 70–81. https://doi.org/10.1108/IJICC-04-2018-0057

Navarro, J. (2023). Press media impact of the Cumbre Vieja volcano activity in the island of La Palma (Canary Islands): A machine learning and sentiment analysis of the news published during the volcanic eruption of 2021. *International Journal of Disaster Risk Reduction*, *91*. https://doi.org/10.1016/j.ijdrr.2023.103694

Nayak, S., Savita, & Sharma, Y. K. (2023). A modified Bayesian boosting algorithm with weight-guided optimal feature selection for sentiment analysis. *Decision Analytics Journal*, *8*(July), 100289. https://doi.org/10.1016/j.dajour.2023.100289

Nguyen, A. (2023). Managing demand volatility of pharmaceutical products in times of disruption through news sentiment analysis. *International Journal of Production Research*, *61*(9), 2828–2839. https://doi.org/10.1080/00207543.2022.2070044

Nti, I. K., Adekoya, A. F., & Weyori, B. A. (2020). A comprehensive evaluation of ensemble learning for stock-market prediction. *Journal of Big Data*, *7*(1). https://doi.org/10.1186/s40537-020-00299-5

Pande, S., Khamparia, A., & Gupta, D. (2023). Feature selection and comparison of classification algorithms for wireless sensor networks. *Journal of Ambient Intelligence and Humanized Computing*, *14*(3), 1977–1989. https://doi.org/10.1007/s12652-021-03411-6

Pappas, N. (2017). Multilingual visual sentiment concept clustering and analysis. *International Journal of Multimedia Information Retrieval*, *6*(1), 51–70. https://doi.org/10.1007/s13735-017-0120-4

Priya, P. S. (2023). An Aspect based Sentiment Analysis of Tour and Travel Recommendation Approach using Machine Learning. *International Journal of Intelligent Systems and Applications in Engineering*, *11*(10), 754–762.

Priya, V. (2016). Chennai rains sentiment-an analysis of opinion about youngsters reflected in tweets using hadoop. *International Journal of Pharmacy and Technology*, *8*(3), 16172–16180.

Ragini, J. R. (2018). Big data analytics for disaster response and recovery through sentiment analysis. *International Journal of Information Management*, *42*, 13–24. https://doi.org/10.1016/j.ijinfomgt.2018.05.004

Rahmadani, S., Dongoran, A., Zarlis, M., & Zakarias. (2018). Comparison of Naive Bayes and Decision Tree on Feature Selection Using Genetic Algorithm for Classification Problem. *Journal of Physics: Conference Series*, *978*(1). https://doi.org/10.1088/1742-6596/978/1/012087

Saraswathi, N., Sasi Rooba, T., & Chakaravarthi, S. (2023). Improving the accuracy of sentiment analysis using a linguistic rule-based feature selection method in tourism reviews. *Measurement: Sensors*, *29*(May), 100888. https://doi.org/10.1016/j.measen.2023.100888

Sontayasara, T. (2021). Twitter sentiment analysis of bangkok tourism during covid-19 pandemic using support vector machine algorithm. *Journal of Disaster Research*, *16*(1), 24–30. https://doi.org/10.20965/jdr.2021.p0024

Suhaimin, M. S. M. (2023). Social media sentiment analysis and opinion mining in public security: Taxonomy, trend analysis, issues and future directions. *Journal of King Saud University - Computer and Information Sciences*, *35*(9). https://doi.org/10.1016/j.jksuci.2023.101776

Sutriawan. (2023). Performance Evaluation of Classification Algorithm for Movie Review Sentiment Analysis. *International Journal of Computing*, *22*(1), 7–14. https://doi.org/10.47839/IJC.22.1.2873

Syahriani. (2020). Sentiment analysis of facebook comments on indonesian presidential candidates using the naïve bayes method. *Journal of Physics: Conference Series*, *1641*(1). https://doi.org/10.1088/1742-6596/1641/1/012012

Talaat, A. S. (2023). Sentiment analysis classification system using hybrid BERT models. *Journal of Big Data*, *10*(1). https://doi.org/10.1186/s40537-023-00781-w

Teoh, C. W., Ho, S. B., Dollmat, K. S., & Tan, C. H. (2022). Ensemble-Learning Techniques for Predicting Student Performance on Video-Based Learning. *International Journal of Information and Education Technology*, *12*(8), 741–745. https://doi.org/10.18178/ijiet.2022.12.8.1679

Uma, M. (2022). Analysis of Ensemble Classification of Twitter Sentiments Using New Dependency Tree Based Approach. *International Journal on Artificial Intelligence Tools*, *31*(5). https://doi.org/10.1142/S0218213022500324

Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: Introduction and review. *Journal of Biomedical Informatics*, *85*(June), 189–203. https://doi.org/10.1016/j.jbi.2018.07.014

Wankhade, M., Rao, A. C. S., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. In *Artificial Intelligence Review* (Vol. 55, Issue 7). Springer Netherlands. https://doi.org/10.1007/s10462-022-10144-1

Win, M. N. (2022). Sentiment Attribution Analysis With Hierarchical Classification And Automatic Aspect Categorization On Online User Reviews. *Malaysian Journal of Computer Science*, *35*(2), 89–110. https://doi.org/10.22452/mjcs.vol35no2.1

Yu, C., Zhu, X., Feng, B., Cai, L., & An, L. (2019). Sentiment analysis of Japanese tourism online reviews. *Journal of Data and Information Science*, *4*(1), 89–113. https://doi.org/10.2478/jdis-2019-0005

Zaini, N. A. M., & Awang, M. K. (2022). Performance Comparison between Meta-classifier Algorithms for Heart Disease Classification. *International Journal of Advanced Computer Science and Applications*, *13*(10), 323–328. https://doi.org/10.14569/IJACSA.2022.0131039

Zapata, G. (2019). Business information architecture for successful project implementation based on sentiment analysis in the tourist sector. *Journal of Intelligent Information Systems*, *53*(3), 563–585. https://doi.org/10.1007/s10844-019-00564-x

Zheng, J. (2019). Research and Analysis in Fine-grained Sentiment of Film Reviews Based on Deep Learning. *Journal of Physics: Conference Series*, *1237*(2). https://doi.org/10.1088/1742-6596/1237/2/022152.

Comparative Analysis of Classification Methods in Sentiment Analysis: The Impact of Feature Selection and Ensemble Techniques Optimization

ORIGINALITY REPORT

10%	9%	6%	6%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Tidar Student Paper	2%
2	"Proceedings of the Third International Conference on Computing, Mathematics and Statistics (iCMS2017)", Springer Science and Business Media LLC, 2019 Publication	1%
3	ejournal.amikompurwokerto.ac.id Internet Source	1%
4	link.springer.com Internet Source	1%
5	thesai.org Internet Source	1%
6	www.researchgate.net Internet Source	<1%
7	smartech.gatech.edu Internet Source	<1%

8	journals.vgtu.lt Internet Source	<1 %
9	Submitted to Tarumanagara University Student Paper	<1 %
10	www.joig.net Internet Source	<1 %
11	Submitted to Universitas Raharja Student Paper	<1 %
12	Submitted to Loughborough University Student Paper	<1 %
13	www.escholar.manchester.ac.uk Internet Source	<1 %
14	Enow Takang Achuo Albert, Ngalle Hermine Bille, Bell Joseph Martin, Ngonkeu Mangaptche Eddy Leonard. "Integrating genetic markers and adiabatic quantum machine learning to improve disease resistance-based marker assisted plant selection", Journal of Scientific Agriculture, 2023 Publication	<1 %
15	Submitted to Brunel University Student Paper	<1 %
16	europub.co.uk Internet Source	<1 %

17	nemertes.library.upatras.gr Internet Source	<1%
18	peerj.com Internet Source	<1%
19	www.frontiersin.org Internet Source	<1%
20	atrium.lib.uoguelph.ca Internet Source	<1%
21	insightsociety.org Internet Source	<1%
22	jutif.if.unsoed.ac.id Internet Source	<1%
23	ojs.amhinternational.com Internet Source	<1%
24	www.hightechjournal.org Internet Source	<1%
25	www.mdpi.com Internet Source	<1%
26	www.wootware.co.za Internet Source	<1%
27	Rhodessa J. Cascaro, Bobby D. Gerardo, Ruji P. Medina. "Filter Selection Methods for Multiclass Classification", Proceedings of the	<1%