Black Sea Journal of Engineering and Science
doi: 10.34248/bsengineering.1541950

Open Access Journal
e-ISSN: 2619 – 8991

Research Article
Volume 7 - Issue 6: 1224-1231 / November 2024

# AN IMPROVED HYBRID MODEL BASED ON ENSEMBLE FEATURES AND REGULARIZATION SELECTION FOR CLASSIFICATION

**Tohid YOUSEFİ[1]*, Özlem VARLIKLAR[1], Mehmet Serhat ODABAS[2]**

[1]*Dokuz Eylül University, Faculty of Engineering, Department of Computer Engineering, 35100, İzmir, Türkiye*
[2]*Onfokuz Mayis University, Bafra Vocational School, 55400, Samsun, Türkiye*

**Abstract:** Feature selection is a pivotal process in machine learning, essential for enhancing model performance by reducing dimensionality, improving generalization, and mitigating overfitting. By eliminating irrelevant or redundant features, simpler and more interpretable models are achieved, which generally perform better. In this study, we introduce an advanced hybrid method combining ensemble feature selection and regularization techniques, designed to optimize model accuracy while significantly reducing the number of features required. Applied to a customer satisfaction dataset, our method was first tested without feature selection, where the model achieved a ROC AUC value of 0.946 on the test set using all 369 features. However, after applying our proposed feature selection method, the model achieved a higher ROC AUC value of 0.954, utilizing only 12 key features and completing the task in approximately 43% less time. These findings demonstrate the effectiveness of our approach in producing a more efficient and superior-performing model.

**Keywords:** Feature selection, Basic filter method, Regularization, Logistic regularization, Tree based feature selection

***Corresponding author:** Dokuz Eylül University, Faculty of Engineering, Department of Computer Engineering, 35100, İzmir, Türkiye
**E mail:** tohid.yousefi@ogr.deu.edu.tr (T. YOUSEFI)
Tohid YOUSEFİ https://orcid.org/0000-0003-4288-8194
Özlem VARLIKLAR https://orcid.org/0000-0001-6415-0698
Mehmet Serhat ODABAS https://orcid.org/0000-0002-1863-7566

**Received:** September 01, 2024
**Accepted:** October 16, 2024
**Published:** November 15, 2024

**Cite as:** Yousefi T, Varlıklar Ö, Odabas MS. 2024. An improved hybrid model based on ensemble features and regularization selection for classification. BSJ Eng Sci, 7(6): 1224-1231.

## 1. Introduction

Feature selection is an essential step in machine learning, aimed at selecting the most relevant features from a dataset to enhance model performance (Miao and Niu, 2016). By reducing the number of features, it simplifies the learning process, speeds up computation, and improves model interpretability. This process also helps prevent overfitting, especially in high-dimensional datasets, by eliminating irrelevant or redundant features that add noise. As a result, models become more efficient and capable of generalizing better to unseen data, striking a balance between complexity and accuracy (Shardlow, 2016). Ultimately, feature selection enables more robust, interpretable, and scalable machine learning models (Li et al., 2017).

When feature selection is not implemented, several problems arise. First, irrelevant and unnecessary features introduce noise into the model, making it difficult for the algorithm to focus on truly important patterns in the data. This can lead to overfitting, where the model becomes too specific to the training data and performs poorly on unseen data. Additionally, adding unnecessary features increases computational costs, slowing down model training and inference (Kalousis et al., 2007; Sugiyama, 2015). In real-world applications, especially with large datasets, this can lead to inefficiency, wasted resources, and difficult-to-interpret models. Without feature selection, it becomes more difficult to understand which features contribute most to the model's predictions, reducing confidence and explainability in critical applications (Remeseiro and Bolon-Canedo, 2019; Luftensteiner et al., 2021).

In this paper, we introduce an enhanced hybrid model method that integrates ensemble feature selection with regularization techniques, designed to optimize feature selection for machine learning models. Our approach effectively balances model complexity and accuracy by leveraging regularization to penalize the inclusion of irrelevant features. This strategy ensures that only the most critical features are retained, thereby improving model performance and reducing computational time. In our experiments with a customer satisfaction dataset, our method achieved superior performance, attaining a ROC AUC of 0.954 using just 12 features, compared to a ROC AUC of 0.946 with the full set of 369 features. Additionally, the process was completed in approximately 43% less time, demonstrating that our approach simplifies the model and accelerates computation without compromising on performance.

This paper makes the following contributions:

- Our proposed method outperformed models using all 369 features by achieving better accuracy with just 12

selected features, resulting in a simpler and more efficient model.

- When applied to the customer satisfaction dataset, our proposed feature selection method reduced the computation time by 43%, demonstrating significantly faster performance compared to methods that do not use feature selection.
- We introduce an enhanced hybrid model method that integrates ensemble feature selection with regularization techniques, effectively reducing dimensionality while maintaining or improving model performance.
- We provide an empirical evaluation of the trade-offs between feature reduction and model accuracy, offering insights into the balance between complexity and computational efficiency.

The paper is structured as follows: Material and Methods outlines the proposed hybrid ensemble and regularization feature selection method, detailing its mathematical foundation, advantages, and the dataset used in the study. This section also explains the preprocessing steps and the machine learning algorithms employed to evaluate the method. Results present a comparative analysis of model performance, both with and without feature selection, highlighting improvements in accuracy and reductions in computational time. Finally, the Conclusion summarizes the key findings, emphasizes the efficiency of the proposed method, and suggests directions for future research.

## 2. Materials and Methods

### 2.1. Feature Selection

Feature selection is an important technique in machine learning that improves model performance by focusing on the most relevant features and discarding irrelevant or redundant ones. This process helps simplify models, improve accuracy, and reduce computational overhead. In high-dimensional datasets where many features may contribute little to model performance, feature selection is vital to reduce overfitting and improve generalization (Ramchandran and Sangaiah, 2018). Figure 1 illustrates this concept by showing how irrelevant features are removed and important features are retained, and highlights the process of refining the feature set to optimize model efficiency. The figure provides a clear visual representation of how feature selection streamlines the data and ultimately leads to more efficient and accurate predictive models (Jimenez-del-Toro et al., 2017; Yousefi and Aktaş, 2024).


**Figure 1.** Feature selection process.

Feature selection commonly employs methods such as wrapper (Kohavi and John, 1997), filter (Chandrashekar and Sahin, 2014), embedded (Zheng and Casari, 2018), ensemble (Opitz and Maclin, 1999), and hybrid (Kabir et al., 2010) approaches to evaluate feature importance and enhance model performance. Before applying these advanced methods, it's crucial to use basic filter techniques to clean the dataset (Moldovan et al., 2017). Methods like constant features, quasi-constant features, duplicated features, and highly correlated feature removal are employed to eliminate low-information or redundant features. These preliminary filtering steps ensure a cleaner, more relevant feature space, setting the stage for more effective and comprehensive feature selection, and ultimately improving model generalization (Yousefi and Aktaş, 2024).

### 2.2. Basic Filter Methods

Feature selection is an essential step in machine learning and data analysis, ensuring that only the most relevant features are utilized in predictive models. The process usually begins with basic filter-based techniques, which help detect and remove irrelevant or redundant features from the dataset. As shown in Figure 2, four commonly used filter-based methods include constant feature removal, quasi-constant feature removal, duplicated feature elimination, and highly correlated feature detection. These methods streamline the feature selection process, leading to more efficient and higher-performing models (Yousefi and Varlıklar, 2024).

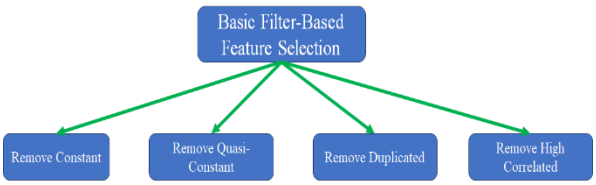


**Figure 2.** Basic filter-based feature selection methods.

- Constant Feature Removal: This approach focuses on identifying features that maintain a constant value across all data instances. Since these features contribute little or no useful information to predictive modelling, they can be safely discarded. Removing constant features helps reduce the dimensionality of the dataset, making the subsequent analysis and model-building steps more efficient and streamlined.
- Quasi-Constant Feature Removal: Quasi-constant features show very little variation across instances, offering limited value for predictive modeling. Although not completely constant, their low variance makes them ineffective in differentiating between classes or outcomes. Detecting and removing quasi-constant features helps to enhance the overall quality of the feature space by eliminating features that do not contribute meaningfully to the model's performance.
- Duplicated Feature Removal: Duplicated features are those that are nearly identical or highly similar to one another, adding unnecessary redundancy to the dataset. These features can increase complexity and potentially lead to bias or overfitting in machine learning models. Removing duplicated features helps streamline the

feature space, making the dataset more manageable and improving the robustness and reliability of further analysis.

- Highly Correlated Feature Removal: Highly correlated features are features that exhibit a strong linear relationship with each other and provide similar information to the model. Including such features can create redundancy and negatively impact the model by increasing its complexity and increasing the risk of multicollinearity. Identifying and removing highly correlated features helps simplify the feature space, reduce overfitting, and improve the overall performance and interpretability of the model.

**2.3. Regularization**

Regularization is a crucial technique in machine learning used to prevent overfitting, especially in models with high complexity. Overfitting occurs when a model captures noise and patterns specific to the training data, resulting in poor generalization to unseen data. Regularization techniques add a penalty to the model's loss function, discouraging it from fitting too closely to the training data and promoting a simpler, more general model that performs well on new data (Tian and Zhang, 2022).

In the context of regression models, regularization modifies the cost function by adding a penalty term that controls the magnitude of the model's parameters (weights). The most commonly used regularization techniques are L1 regularization (Lasso) (Ranstam and Cook, 2018; Tibshirani, 1996) and L2 regularization (Ridge) (McDonald, 2009; Hoerl and Kennard, 1970), both of which are applied in linear and logistic regression models to enhance their robustness.

Logistic regression is a classification algorithm used to estimate the probability of a binary outcome (Kleinbaum et al., 2002; Hosmer et al., 2013). Regularization is important in logistic regression to prevent the model from becoming too complex, especially when dealing with high-dimensional data. In logistic regression, the regularization term is added to the loss function, which is the log-likelihood of the model (Salehi et al., 2019). The regularized cost function for logistic regression can be expressed as in equation 1.

$$j(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)}\log(h_\theta(x^{(i)})) + (1 - y^{(i)})\log(1 - h_\theta(x^{(i)}))\right] + \lambda\sum_{j=1}^{n} R(\theta_j) \quad (1)$$

Where $j(\theta)$ is the regularized cost function, m is the number of training examples, $y^{(i)}$ is the actual class label for the $i^{th}$ training example, $h_\theta(x^{(i)})$ is the predicted probability of the class, $\lambda$ is the regularization parameter, and $R(\theta_j)$ is the regularization term, which could be L1 (Lasso) or L2 (Ridge).

The regularization term $R(\theta_j)$ helps control the magnitude of the model parameters, ensuring that the model does not overfit to the training data. In the L1 configuration, the penalty term is the absolute value of the coefficients and is given by equation 2:

$$R(\theta_j) = \sum_{j=1}^{n}|\theta_j| \quad (2)$$

In the L2 configuration, the penalty is the squared value of the coefficients and is as in equation 3:

$$R(\theta_j) = \sum_{j=1}^{n}\theta_j^2 \quad (3)$$

**2.4. Tree Based Feature Selection**

Tree-based feature selection is a powerful method in machine learning used to identify the most important features within a dataset (Freeman et al., 2013). This approach leverages decision trees or ensembles of trees, such as Random Forest, to rank features based on their contribution to the predictive model. In tree-based methods, the importance of a feature is determined by how often and how effectively it is used to split the data across all trees in the ensemble. Features that result in the greatest reduction in impurity or increase in information gain are ranked higher (Azhagusundari and Thanamani, 2013). The feature importance score for a given feature $f_i$ can be mathematically expressed as in equation 4 (Zhou et al., 2021):

$$\text{Importance}(f_i) = \sum_{t\in\text{trees}}\frac{\Delta\text{Impurity}(f_i, t)}{\text{Number of Trees}} \quad (4)$$

where $\Delta\text{Impurity}(f_i, t)$ represents the decrease in impurity caused by feature $f_i$ in tree t.

In our study, we employed the Random Forest algorithm (Liaw and Wiener, 2002), a widely used ensemble method for tree-based feature selection (Hasan et al., 2016). Random Forest constructs multiple decision trees during training and aggregates their predictions to improve accuracy and reduce overfitting (Biau and Scornet, 2016). The feature importance score provided by Random Forest highlights the contribution of each feature to the model's predictive power. By focusing on these scores, we were able to streamline our model by selecting only the most critical features. This approach demonstrated the effectiveness of tree-based feature selection, particularly with Random Forest, in simplifying models while maintaining or even enhancing their accuracy and robustness. The importance of a feature in the Random Forest is calculated by averaging the reduction in impurity across all trees, as described by the formula above. This process ensures that only the most influential features are retained, leading to a more efficient and interpretable model (Menze et al., 2009).

**2.5. Proposed Method**

Feature selection plays a crucial role in machine learning and data analysis by enhancing model performance

through the identification and selection of the most relevant features from a dataset. This process focuses on removing redundant, irrelevant, or noisy features, which helps reduce model complexity, mitigate overfitting, and improve computational efficiency. The primary objective is to concentrate on the most informative features that significantly boost the model's predictive accuracy (Li et al., 2017).

In this study, we propose an advanced hybrid approach that combines ensemble feature methods with regularization techniques to optimize both feature selection and model performance. Our proposed method demonstrates superior efficiency, achieving better results in less time with fewer features compared to traditional methods.

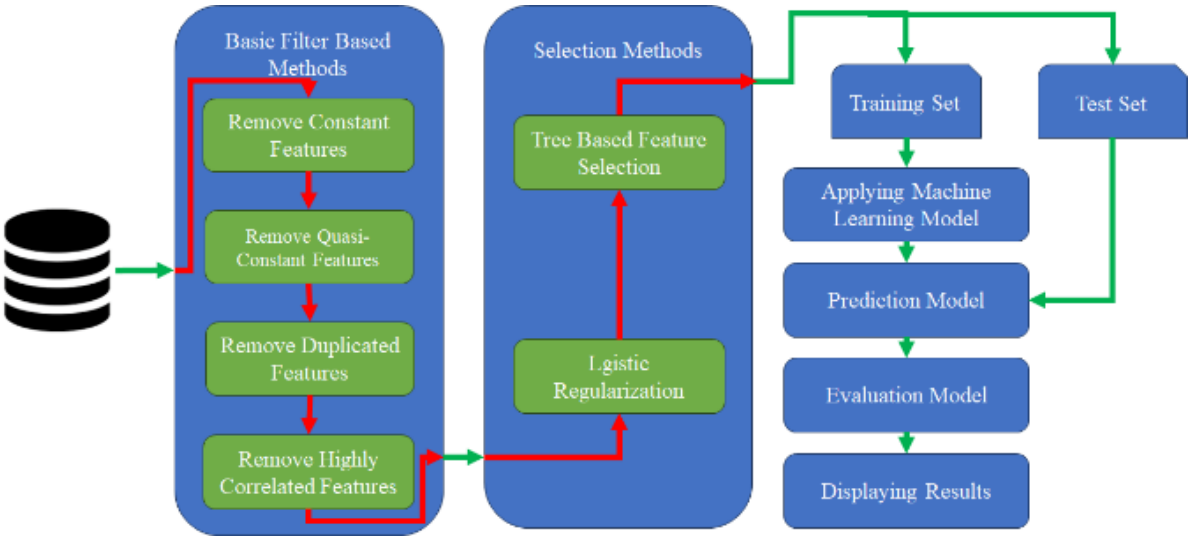


**Figure 3.** Proposed method.

In our proposed approach, the process begins with the application of basic filter-based feature selection methods, as depicted in Figure 3. The initial phase involves removing features with zero variance, which are features where all values are identical. These features are discarded because they lack informative value for the model. This initial removal can be mathematically described by equation 5:

$$X_{clean} = X - \{x_j | \, Var(x_j) = 0\} \quad (5)$$

Here, X denotes the original dataset, and $X_{clean}$ represents the dataset after removing constant features. In the second stage, features with low variance, which could potentially impact model performance, are removed based on a predefined variance threshold, as shown in equation 6. For our study, this threshold was set to 0.998.

$$X_{clean} = X_{clean} - \{x_j | \, Var(x_j) < \tau\} \quad (6)$$

where τ is a specified variance threshold.

In the third stage, features carrying the same or nearly the same information are removed to reduce redundancy. This is done by comparing the features for exact copies as in equation 7:

$$X_{clean} = X_{clean} - \{x_j | \, \exists x_i \text{ suhc that } x_i = x_j \text{ and } j \neq i\} \quad (7)$$

Here, $X_{clean}$ represents the dataset after removing duplicate features, and $x_i$=$x_j$ denotes features that are identical.

The final stage of feature selection involves using a simple filter to eliminate highly correlated variables. In this stage, only one feature is retained from among those with strong correlations to minimize model complexity and enhance generalization. This process is mathematically represented by equation 8:

$$X_{selected} = \{x_j | \, Corr(x_j, x_i) < p_{max})\} \quad (8)$$

Here $p_{max}$ represents the maximum correlation threshold, which was determined as 80% in our study.

In this section, as illustrated in Figure 3, we first apply logistic regularization within the selection methods. Techniques such as L1 and L2 regularization are employed in logistic regression models to decrease model complexity and mitigate overfitting. The mathematical representation of the regularization term is given by equation 9:

$$j(\theta) = -\frac{1}{m}\sum_{i=1}^{m}\left[y^{(i)} \log(h_\theta(x^{(i)})) + (1 - y^{(i)}) \log(1 - h_\theta(x^{(i)}))\right] + \lambda \sum_{j=1}^{n} w_j^2 \quad (9)$$

Where $j(\theta)$ denotes the regularized cost function, m represents the number of training examples, $y^{(i)}$ is the

actual class label for the $i^{th}$ training example, $h_\theta(x^{(i)})$ is the predicted probability of the class, $\lambda$ is the regularization coefficient, and the L2 penalty term $\lambda \sum_{j=1}^{n} w_j^2$ plays a crucial role in managing the magnitude of the coefficients, thereby helping to prevent overfitting.

In the last stage of the selection method, we applied tree-based feature selection. Features were ranked based on their importance using methods like decision trees. A key metric for assessing feature importance in decision trees is Information Gain. This measure evaluates the reduction in entropy (uncertainty) achieved by a feature, helping to identify those that offer the greatest amount of information about the target variable. The Information Gain for feature $x_j$ can be calculated as equation 10:

$$\text{Information Gain}\left(x_j\right) = \text{Entropy}\,(S) - \sum_{v \in \text{Values}(x_j)} \frac{|S_v|}{|S|} \text{Entropy}(S_v) \quad (10)$$

Where Entropy(S) is the entropy of the original dataset S, $\text{Values}\left(x_j\right)$ are the distinct values of the feature $x_j$, $|S_v|$ is the number of instances in subset $S_v$ corresponding to each value v of $x_j$, and $\text{Entropy}(S_v)$ is the entropy of the subset $S_v$.

The entropy of a dataset S can be calculated as equation 11:

$$\text{Entropy}\,(S) = -\sum_{k=1}^{K} p_k \log_2(p_k) \quad (11)$$

where $p_k$ is the probability of each class k in the dataset and K is the number of classes.

Finally, after splitting the data into training and test sets, we build the machine learning model using the Random Forest algorithm. We then proceed with the estimation process and evaluate the model to measure its performance. The results are presented to show how well the model has performed.

### 2.6. Experiments

#### 2.6.1. Dataset

The dataset, accessible via the provided link, provides detailed insights into the satisfaction levels of Santander Bank customers. It is an important resource for analyzing customer experiences and preferences in the banking sector. The dataset contains 369 features, providing a comprehensive view of various aspects that affect customer satisfaction.

In this study, we did not initially select any features and then applied the proposed method to this dataset. The results and implications of this approach are discussed in the results section of our study. For those who want to explore the dataset further, it is freely available on Kaggle via the following link:

https://www.kaggle.com/competitions/santander-customer-satisfaction/data

#### 2.6.2. Evaluation metrics

Evaluation metrics are essential for assessing machine learning model performance, offering quantitative measures to gauge how well a model is performing and to compare different models. In our study, Accuracy, Precision, Recall, F1 Score, ROC AUC Score, and $R^2$—were employed to evaluate and compare the performance of our model, ensuring a comprehensive assessment of its effectiveness (Hossin and Sulaiman, 2015).

Accuracy measures the proportion of correctly classified instances out of the total. It is given by $\frac{TP+TN}{TP+TN+FP+FN}$ where TP, TN, FP, and FN stand for True Positives, True Negatives, False Positives, and False Negatives, respectively.

Precision assesses the proportion of true positive predictions among all positive predictions, calculated as $\frac{TP}{TP+FP}$.

Recall evaluates the proportion of true positive predictions among all actual positive instances, represented as $\frac{TP}{TP+FN}$.

F1 Score provides a balanced measure between Precision and Recall, computed as $2 * \frac{Precision*Recall}{Precision+Recall}$.

ROC AUC Score measures the area under the Receiver Operating Characteristic curve, which plots the true positive rate against the false positive rate at various thresholds.

$R^2$ used in regression, indicates the proportion of variance in the dependent variable explained by the independent variables, calculated as $1 - \frac{SS_{res}}{SS_{tot}}$, $SS_{res}$ is Residual Sum of Squares, and $SS_{tot}$ is Total Sum of Squares.

Confusion Matrix summarizes the number of correct and incorrect predictions across classes, showing True Positives, True Negatives, False Positives, and False Negatives.

## 3. Results and Discussion

In this study, we highlight the importance of feature selection in machine learning, emphasizing its role in improving model performance by reducing complexity, preventing overfitting, and increasing computational efficiency. Feature selection allows us to focus on the most relevant and informative features, which is crucial for building effective models.

Our proposed method has shown that the model can achieve better performance in less time by using fewer features. This is especially important in practical applications where computational resources and time are limited. The results clearly show that when the model is trained with selected features, it not only reduces the overall training time but also improves the prediction accuracy.

To illustrate the impact of feature selection, we present the results in two ways. First, we show the model's performance using all 369 features without any feature selection. The results indicate that while the model performs adequately, it is not optimized in terms of time

and complexity. Second, we apply our proposed feature selection method, and the results reveal a significant improvement in performance, with the model achieving higher accuracy and efficiency using a reduced set of features.

In the first stage, we performed the prediction process using the Random Forest algorithm without applying any feature selection method, utilizing all 369 features. As shown in Table 1, this approach yielded an accuracy of 0.946, and the results were obtained in approximately 546 seconds.

**Table 1.** Results of the model without using feature selection

| Accuracy | $R^2$ | F1 | Recall |
|---|---|---|---|
| 0.946 | 0.946 | 0.947 | 0.952 |
| Precision | ROC_AUC | Time | Feature |
| 0.944 | 0.946 | 564s | 369 |

Note: All results are from the test dataset.

As illustrated in Figure 4, the confusion matrix was generated, which provides a detailed breakdown of the model's performance across different classes. The matrix highlights the number of true positives, true negatives, false positives, and false negatives, giving us insight into the model's strengths and areas for improvement.

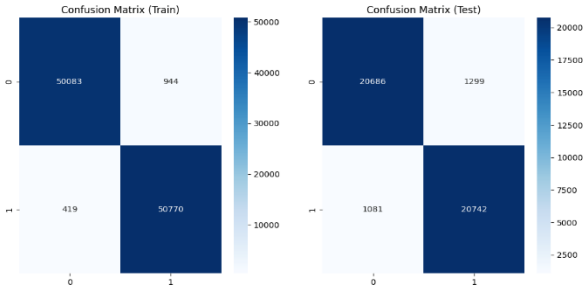


**Figure 4.** Confusion matrix of random forest model without using the proposed method.

In Figure 5, we can observe the ROC AUC curve, which demonstrates the model's performance when all features are used. The ROC AUC score achieved was 0.946, indicating a strong ability of the model to discriminate between the classes.

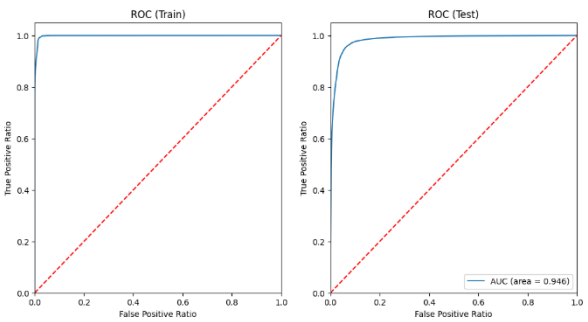


**Figure 5**. ROC_AUC graph of random forest model without using the proposed method.

In the second stage, following the application of the simple filter feature selection method outlined in this article, we applied logistic regularization and subsequently selected the most relevant features using ensemble methods. As shown in Table 2, this approach allowed us to achieve an accuracy of 0.954 using just 12 features instead of the original 369. This resulted in not only improved performance but also a significant reduction in processing time. Our proposed method reached the outcome in 327 seconds, which is 43% faster than the previous approach.

**Table 2.** Results of the model using the proposed method

| Accuracy | $R^2$ | F1 | Recall |
|---|---|---|---|
| 0.954 | 0.954 | 0.954 | 0.952 |
| Precision | ROC_AUC | Time | Feature |
| 0.958 | 0.954 | 327s | 12 |

Not: All results are from the test dataset.

Figure 6 presents a confusion matrix that details the model's performance across various classes. It highlights the counts of true positives, true negatives, false positives, and false negatives, demonstrating that the model shows improved performance compared to the previous one.

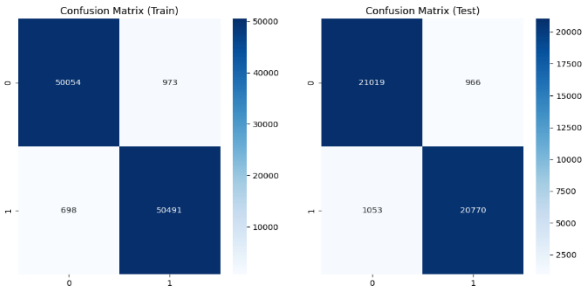


**Figure 6.** Confusion matrix of the proposed method.

Figure 7 illustrates the ROC AUC curve for the model that was trained using just 12 features. The ROC AUC score of 0.954 reflects the model's strong performance with the reduced feature set. This high score confirms that our proposed method is effective in maintaining predictive accuracy while simplifying the feature set.

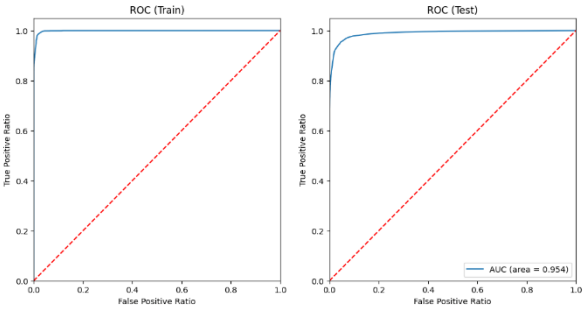


**Figure 7.** ROC_AUC graph of the proposed method.

## 4. Conclusion

Feature selection plays a crucial role in the development of efficient and accurate machine learning models. By identifying and retaining the most relevant features, feature selection helps reduce model complexity, prevent overfitting, and improve computational efficiency. It ensures that the model focuses on the most informative aspects of the data, leading to better generalization and more reliable predictions.

In this study, we proposed a hybrid feature selection method that combines simple filter techniques with logistic regularization and ensemble methods. This approach was designed to optimize both the performance and efficiency of the model. When we applied the Random Forest algorithm to the customer satisfaction dataset without any feature selection, using all 369 features, we achieved an ROC AUC score of 0.946. However, by implementing our proposed feature selection method, we improved this score to 0.954 while drastically reducing the number of features to just 12. Additionally, the processing time was reduced by 43%, highlighting the efficiency gains of our method.

The significant improvements in both accuracy and processing time underscore the advantages of our proposed feature selection method. By eliminating redundant and less informative features, the model not only becomes more efficient but also delivers superior performance. This demonstrates the effectiveness of the method in handling large datasets, making it a valuable tool for similar applications in other domains.

Looking ahead, we believe that further optimization of the regularization parameters through the use of metaheuristic algorithms could yield even better results. Future work will focus on exploring this approach to enhance the model's performance. By fine-tuning these parameters more precisely, we aim to achieve even greater accuracy and efficiency in subsequent studies.

**Author Contributions**

The percentages of the authors' contributions are presented below. The authors reviewed and approved the final version of the manuscript.

| | T.Y. | Ö.V. | M.S.O. |
|---|---|---|---|
| C | 45 | 35 | 20 |
| D | 45 | 35 | 20 |
| S | 45 | 35 | 20 |
| DCP | 45 | 35 | 20 |
| DAI | 45 | 35 | 20 |
| L | 45 | 35 | 20 |
| W | 45 | 35 | 20 |
| CR | 45 | 35 | 20 |
| SR | 45 | 35 | 20 |
| PM | 45 | 35 | 20 |
| FA | 45 | 35 | 20 |

C=Concept, D= design, S= supervision, DCP= data collection and/or processing, DAI= data analysis and/or interpretation, L= literature search, W= writing, CR= critical review, SR= submission and revision, PM= project management, FA= funding acquisition.

**Conflict of Interest**

The authors declared that there is no conflict of interest.

**Ethical Consideration**

Ethics committee approval was not required for this study because of there was no study on animals or humans.

## References

Azhagusundari B, Thanamani AS. 2013. Feature selection based on information gain. Inter J Innov Technol Explor Engin (IJITEE), 2(2): 18-21.

Biau G, Scornet E. 2016. A random forest guided tour. Test, 25: 197-227.

Chandrashekar G, Sahin F. 2014. A survey on feature selection methods. Comput Electr Engin, 40(1): 16-28.

Freeman C, Kulić D, Basir O. 2013. Feature-selected tree-based classification. IEEE Transact Cybernet, 43(6): 1990-2004.

Hasan MAM, Nasser M, Ahmad S, Molla KI. 2016. Feature selection for intrusion detection using random forest. J Inform Sec, 7(3): 129-140.

Hoerl AE, Kennard RW. 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 12(1): 55-67.

Hosmer Jr DW, Lemeshow S, Sturdivant RX. 2013. Applied logistic regression, John Wiley & Sons, London, UK, pp: 254.

Hossin M, Sulaiman MN. 2015. A review on evaluation metrics for data classification evaluations. Inter J Data Dining Knowledge Manage Process, 5(2): 1-8.

Jimenez-del-Toro O, Otálora S, Andersson M, Eurén K, Hedlund M, Rousson M, Müller H, Atzori M. 2017. Analysis of histopathology images: From traditional machine learning to deep learning, Elsevier, New York, USA, pp: 135.

Kabir MM, Islam MM, Murase K. 2010. A new wrapper feature selection approach using neural network. Neurocomputing, 73(16-18): 3273-3283.

Kalousis A, Prados J, Hilario M. 2007. Stability of feature selection algorithms: a study on high-dimensional spaces. Knowledge Inform Systems, 12: 95-116.

Kleinbaum DG, Dietz K, Gail M, Klein M, Klein M. 2002. Logistic regression, Springer, USA, pp: 142.

Kohavi R, John GH. 1997. Wrappers for feature subset selection. Artificial Intel, 97(1-2): 273-324.

Li J, Cheng K, Wang S, Morstatter F, Trevino RP, Tang J, Liu H. 2017. Feature selection: A data perspective. ACM Comput Surveys, 50(6):1-45.

Liaw A, Wiener M. 2002. Classification and regression by random Forest. R news, 2(3): 18-22.

Luftensteiner S, Mayr M, Chasparis G. 2021. Filter-based feature selection methods for industrial sensor data: a review. International Conference on Big Data Analytics and Knowledge Discovery, Virtual Event, September 27–30, pp: 242-249.

McDonald GC. 2009. Ridge regression. Computa Stat, 1(1): 93-100.

Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, Hamprecht FA. 2009. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. BMC Bioinform, 10: 1-16.

Miao J, Niu L. 2016. A survey on feature selection. Procedia Comput Sci, 91: 919-926.

Moldovan D, Cioara T, Anghel I, Salomie I. 2017. Machine learning for sensor-based manufacturing processes. 13th IEEE international conference on intelligent computer communication and processing (ICCP), September 7-9, Cluj-Napoca, Romania, pp: 147-154.

Opitz D, Maclin R. 1999. Popular ensemble methods: An empirical study. J Artificial Intel Res, 11: 169-198.

Ramchandran A, Sangaiah AK. 2018. Unsupervised anomaly detection for high dimensional data—An exploratory analysis, Elsevier, New York, USA, pp: 254.

Ranstam J, Cook JA. 2018. LASSO regression. J British Surg, 105(10): 1348-1358.

Remeseiro B, Bolon-Canedo V. 2019. A review of feature selection methods in medical applications. Comput Biol Med, 112: 103375.

Salehi F, Abbasi E, Hassibi B. 2019. The impact of regularization on high-dimensional logistic regression. Adv Neural Inform Proces Systems, 32: 1-11.

Shardlow M. 2016. An analysis of feature selection techniques. Univ Manchester, 1: 1-7.

Sugiyama M. 2015. Introduction to statistical machine learning. Morgan Kaufmann, New York, USA, pp: 425.

Tian Y, Zhang Y. 2022. A comprehensive survey on regularization strategies in machine learning. Informn Fusion, 80: 146-166.

Tibshirani R. 1996. Regression shrinkage and selection via the lasso. J Royal Stat Soc Series B: Stat Method, 58(1): 267-288.

Yousefi T, Varlıklar Ö. 2024. Breast Cancer Prediction with Hybrid Filter-Wrapper Feature Selection. Inter J Adv Nat Sci Engin Res, 8: 411-419.

Yousefi T, Varlılar Aktaş Ö. 2024. Predicting customer satisfaction with hybrid basic filter-based feature selection method. 4th International Artificial Intelligence and Data Science Congress, 14-15 March, Izmir, Türkiye, pp: 1-10.

Zheng A, Casari A. 2018. Feature engineering for machine learning: principles and techniques for data scientists. O'Reilly Media, Inc., London, UK, pp: 358.

Zhou H, Zhang J, Zhou Y, Guo X, Ma Y. 2021. A feature selection algorithm of decision tree based on feature weight. Expert Syst Applicat, 164: 113842.