

Adaptive Method for Feature Selection in the Machine Learning Context

Yamen El Touati

Department of Computer Science, Faculty of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia
yamen.touati@nbu.edu.sa (corresponding author)

Jihane Ben Slimane

Department of Computer Science, Faculty of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia
jehan.saleh@nbu.edu.sa

Taoufik Saidani

Department of Computer Science, Faculty of Computing and Information Technology, Northern Border University, Kingdom of Saudi Arabia
taoufik.saidan@nbu.edu.sa

Received: 2 April 2024 | Revised: 11 April 2024 | Accepted: 12 April 2024

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: <https://doi.org/10.48084/etasr.7401>

ABSTRACT

Feature selection is a fundamental aspect of machine learning that is crucial for improving the accuracy and efficiency of models. It carefully analyzes the abundance of data to identify the most significant characteristics, hence improving the accuracy of predictions and minimizing the likelihood of model overfitting. This technique not only optimizes model training by reducing computational requirements, but also enhances the model's interpretability, resulting in more transparent and reliable predictions. The deliberate omission of unnecessary variables is a process of improving the model and also constitutes a crucial measure toward achieving more flexible and comprehensible results in machine learning. An analysis to assess the effectiveness of feature selection on regression models was conducted. The impact was measured using Mean Squared Error (MSE) metrics. A variety of regression algorithms were evaluated, and then feature selection techniques, including statistical and algorithmic methods, such as SelectKBest, PCA, and RFE with Linear Regression and Random Forest, were applied. After selecting the features, linear models demonstrated improvements in mean squared error (MSE), highlighting the value of removing unnecessary data. This study emphasizes the subtle impact of feature selection on model performance, calling for a tailored strategy to maximize prediction accuracy.

Keywords-cloud computing; cyber security; preventive approach; prediction techniques; artificial intelligence

I. INTRODUCTION

Machine Learning (ML) models face certain challenges as they handle increasingly large datasets, resulting in a rise in model complexity and a significant demand on processing resources [1]. The increasing number of features not only augments the reaction time of regressors, but also enhances the risk of overfitting, by introducing redundant or irrelevant data that obscure the models [2]. Within this multi-dimensional realm, the pursuit of rapid and precise forecasts becomes progressively more demanding. Feature selection is a suitable solution to this problem, as it reduces the complexity of the data, improves the focus of the model, and saves computational resources. Feature selection is a crucial technique in modern ML pipelines. It improves the efficiency of models by

extracting the most useful variables from datasets, without affecting the prediction accuracy [3].

The field of feature selection is characterized by a diverse range of strategies, each with particular benefits and drawbacks. Essential techniques comprise filter methods such as Chi-squared and correlation coefficients, wrapper methods like Recursive Feature Elimination (RFE), and embedded methods established by Least Absolute Shrinkage and Selection Operator (LASSO) as well as Decision Trees (DTs) [4]. Although these strategies strive to improve model performance by identifying the most informative features, they nevertheless face challenges. While filter techniques are scalable, they may fail to consider feature dependencies. On the other hand, wrapper methods can be computationally demanding and often necessitate extensive search tactics.

Embedded approaches provide a harmonious compromise but are limited to particular model structures [2]. A major problem is finding the right balance between the effectiveness of feature selection and the complexity of the model. This requires a careful approach to guarantee that important data are not discarded, which is crucial for maintaining the model's ability to make accurate predictions. This work aims to compare feature selection approaches across several regression models and identify the most appropriate method based on the specific context of the case study.

II. BACKGROUND ON FEATURE SELECTION TECHNIQUES

Eliminating irrelevant or superfluous data enhances the performance of the model, lowering complexity, and increasing interpretability. Subsequently, a concise summary of each category of feature selection is provided.

A. Filter Methods

Filter methods select features based on the latter's statistical properties and are independent of any ML algorithms [5]. They are usually fast and scalable to high-dimensional datasets. Common techniques include the following.

1) Correlation Coefficients

The core of this approach involves identifying and eliminating characteristics that exhibit strong correlation with either the target variable or with one another. This method evaluates the magnitude and orientation of the linear correlation between each predictor variable and the response variable [6]. The coefficient values vary from -1 to 1, with values approaching 1 indicating a strong positive connection, and values approaching -1 indicating a strong negative correlation. A value close to 0 indicates a lack of linear association. When it comes to feature selection, features that have a larger absolute value of the correlation coefficient with the target variable are often perceived as more important and are therefore chosen.

2) Chi-Squared Test

This method is employed in feature selection to assess the degree of independence between categorical variables [8]. It functions by comparing the observed frequencies of occurrences in various categories with the frequencies that would be anticipated if there was no connection between the variables. The essence of this approach is centered around the computation of the Chi-Squared statistic, which measures the discrepancy between the actual and anticipated frequencies across all the categories of the variables under examination.

3) Information Gain

This method quantifies the decrease in entropy or uncertainty regarding the target variable when the value of a feature is known [4]. Derived from information theory, this technique assesses the amount of information that a characteristic offers regarding the distribution of classes in the data. It is especially valuable in decision-making procedures, such as constructing DTs for classification problems. Entropy, a fundamental notion in Information Gain (IG), quantifies the degree of disorder or unpredictability present in the dataset. A

greater entropy signifies a more diverse distribution of classes, while a lesser entropy implies a more distinct segregation of classes. IG quantifies the difference in entropy before and after witnessing a feature. A larger IG suggests that the feature greatly reduces uncertainty regarding the outcome of the target variable [9]. IG is particularly effective in situations when the objective is to comprehend the influence of attributes on the predictability of categorical outcomes. Its primary application lies in the development of DTs, where it aids in the selection of features that effectively divide the dataset into subsets with more consistent and less ambiguous class distributions.

4) Variance Threshold

The Variance Threshold (VT) method [9] is a straightforward and efficient technique for selecting features. It operates on the principle that features with low variance are less likely to provide useful information. This strategy functions by establishing a specific threshold value for variance. Any features that fail to satisfy this threshold are eliminated from the dataset. The underlying principle of this approach is simple: if a characteristic remains relatively constant across various examples, it is improbable that it will have a substantial impact on the model's ability to make accurate predictions. To effectively implement the VT technique, it is crucial to select a suitable threshold. Setting a low threshold may preserve noisy or unnecessary characteristics, whilst setting a high threshold could lead to the omission of potentially valuable information. The selection of the threshold frequently relies on domain expertise and the distinct attributes of the dataset.

5) SelectKBest

This approach chooses features based on the top k highest scores obtained from a specified scoring function [9], such as ANOVA F-test, chi-squared test, or mutual information. This strategy is especially valuable for preserving the most pertinent features based on their statistical significance in relation to the target variable.

B. Wrapper Methods

Wrapper methods are a type of feature selection strategy that evaluates subsets of features by training models and measuring their prediction power [10]. Contrary to filter techniques that assess features apart from models, wrapper methods employ a search algorithm to examine all the possible feature subsets and choose the ones that produce the highest performance based on a predetermined evaluation criterion, usually involving the correctness of the model. Wrapper approaches can greatly improve the performance of models. However, their computational requirements and the possibility of overfitting necessitate their cautious usage.

1) Recursive Feature Elimination (RFE)

This flexible and customizable technique involves the iterative construction of models and the removal of the least significant feature at each phase [10]. This procedure is iterated until the target quantity of features is attained. RFE is founded on the idea that by systematically removing the least important features, as determined by the model's own assessment of

feature relevance, it can identify the most influential predictors for the model's performance.

2) Sequential Feature Selection (SFS)

The process incrementally chooses features according to a specific criterion, assessing the impact of adding or removing features on the model's performance [11]. The technique can be classified into two primary strategies: Sequential Forward Selection (SFoS) and Sequential Backward Selection (SBaS).

C. Embedded Methods

Embedded techniques [12] incorporate feature selection within the learning algorithm itself, providing a balanced approach between the simplicity of filter methods and the accuracy-oriented focus of wrapper methods. These methods share the capability to conduct feature selection during training the model, by employing the intrinsic characteristics of the learning algorithms to assess the significance of each feature. This integration enables embedded approaches to take into account the interplay between the features and the target variable, resulting in models that are more efficient and tend to be more accurate.

1) Least Absolute Shrinkage and Selection Operator (LASSO)

The LASSO is a regression analysis technique that combines variable selection and regularization to improve the accuracy and interpretability of the resulting statistical model [7, 13]. This method addresses the constraints of conventional regression techniques by applying a constraint on the sum of the absolute values of the model parameters, known as the L1 penalty. This penalty effectively reduces the coefficients of less significant features to zero and effectively selects only the variables that make a substantial contribution to the predictive capability of the model. The LASSO method may be utilized in a wide range of data types and regression models, such as linear regression, logistic regression, and survival models. This makes it a versatile tool that can be deployed for both prediction and inference purposes.

2) Decision Trees

DTs are an efficient non-linear predictive modeling technique extensively employed in ML for applications including classification and regression [14]. These systems function by dividing the data into subsets according to the values of their features, leading to a tree-like model that represents decisions and their potential outcomes. Every node in the tree corresponds to a feature in the dataset, each branch corresponds to a decision rule, and each leaf node corresponds to an outcome. The root of the tree is the primary feature that effectively divides the data according to a purity measure, such as Gini impurity or entropy in classification tasks, and variance reduction in regression [15, 16]. The interpretability of DTs is highly appealing due to its clear nature. The tree structure emulates the cognitive processes of humans, facilitating the comprehension and elucidation of the rationale behind forecasts. DTs have the ability to process both numerical and categorical data and are eligible to represent intricate non-linear connections without requiring data manipulation or assuming linearity.

To summarize, embedded approaches are notable for their efficiency and model-specific optimization as they integrate feature selection directly into the model training process. Nevertheless, these methods are not exempt from limitations. Their main constraint emerges from their reliance on the particular model employed, which can impede the applicability of the chosen features across various models. In addition, they demand substantial resources in comparison to filter approaches, especially when dealing with extensive datasets. Additionally, the inherent bias in the feature selection process might occasionally hide important features that may become critical when using different modeling approaches.

D. Dimensionality Reduction Techniques: Principal Component Analysis (PCA)

PCA is a technique that converts the original features into a new collection of orthogonal features, known as principle components, which effectively capture the highest amount of variance present in the data [12]. The principle components are formed by combining the original features in a linear manner and are chosen depending on the extent to which they capture the variation in the data. PCA is commonly employed to decrease the number of dimensions in data while preserving as much information as feasible. This indirectly achieves the goal of feature selection by reducing the feature space to a smaller collection of significant components.

III. CASE STUDY: HAPPINESS INDICATOR

This study perceives a dataset regarding the happiness indicator, derived from UN's annual happiness report, as a case study [17]. This dataset aggregates 26 features to describe happiness according to region. These features encompass a range of features reflecting economic, social, and health-related factors that are believed to influence a nation's or population's overall happiness levels. It is important to recognize that while all these features can contribute to understanding the factors that influence happiness, not all of them may effectively participate in predicting happiness levels in every context. The relevance of each feature can vary depending on cultural, economic, and social dynamics. ML models applied to such datasets often reveal interesting insights into which factors are most predictive of happiness, highlighting the complex interplay of various elements contributing to well-being. This necessitates a comprehensive approach to feature selection, ensuring that the most influential factors are included in predictive modeling efforts while acknowledging that some features might not significantly impact the prediction of happiness. Figure 1 presents the structured approach undertaken in this research. Initially, a feature selection process is employed on the dataset, which encompasses various indicators of happiness, to distill the data to their most informative features. This refined dataset is subsequently input into one of the 10 regression models outlined in Figure 1. The performance of each model is quantitatively assessed using the Mean Square Error (MSE) metric. This assessment is crucial as it allows for a comparative analysis between the outcomes of the regressors' post-feature selection and their performance when utilizing the entire, unaltered dataset. Such comparison seeks to highlight the efficacy of feature selection in enhancing model accuracy. The selection of specific regressors and

feature selection approaches in this study was driven by the goal of including a broad spectrum of predictive modeling techniques and examining how they interact with different feature selection strategies. The selected predictors encompass a range of models, varying from straightforward to intricate, guaranteeing a thorough assessment of algorithmic difficulties. Similarly, the feature selection methods were chosen to investigate various aspects of feature importance, ranging from statistical significance and PCA to recursive elimination and regularization procedures. This deliberate choice enables an in-depth examination of the influence of each feature selection method on the performance of the model, offering a significant understanding of the improvement of predictive models using various regression techniques.

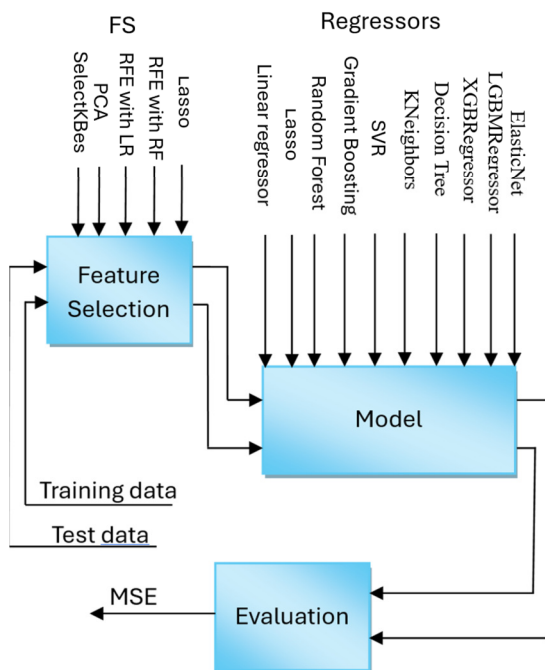


Fig. 1. The adopted framework.

TABLE I. MSE FOR UNPROCESSED DATA.

Classifier	MSE (before FS)
Linear Regression	1.052277
LASSO	1.38633
RF	0.528901
Gradient Boosting	0.60818
SVR	0.592134
KNeighbors	0.618426
DT	1.409413
XGBRegressor	0.543812
LGBMRegressor	0.521754
ElasticNet	1.363167

The efficacy of various regression scenarios when applied to the unprocessed dataset is systematically detailed in Table I through the lens of MSE metrics. This quantitative analysis aims to clarify the relative performance of each model, providing a distinct viewpoint on their predictive accuracy in their most basic form. Table I demonstrates that the MSE values clearly contrast the performance of the regression

models used on the unprocessed dataset. Ensemble regressors such as Random Forest (RF) and LGBMRegressor have superior performance compared to other methods, as indicated by their lowest MSE values. This highlights their ability to effectively capture intricate patterns in the data. The XGBRegressor performs favorably, capitalizing on its advanced boosting algorithms. On the other hand, less complex models such as Linear Regression and DT have larger MSEs, indicating difficulties in dealing with the complexities of the dataset. Regularization-based regression models such as LASSO and ElasticNet have limitations in their original form, but they have the potential to improve when used in conjunction with customized feature selection techniques. The findings emphasize the importance of model selection in achieving high predicted accuracy, with ensemble methods outperforming other approaches in this particular situation.

Table II presents the results of deploying Recursive Feature Elimination (RFE) with Linear Regression (LR) and Random Forest (RF) for feature selection. It manifests that there are varying gains in MSE for numerous regression models. Significantly, RFE combined with RF exhibits substantial MSE reductions for models such as Linear Regression and Support Vector Regression (SVR), indicating a robust correlation between the RF-selected features and these models. Nevertheless, LASSO and ElasticNet, which have built-in regularization, do not show any gain in MSE, indicating their diminished sensitivity to the reduced feature space provided by RFE. The varied outcomes underscore the complex interaction among the selected feature selection technique, the regression model, and the distinctive structure of the dataset. More precisely, the advantage of the RFE combined with the RF in identifying important features for specific models, resulting in improved performance, is not applicable to all cases.

TABLE II. MEAN SQUARE ERROR WITH RFE

FS Method	RFE with LR	RFE with RF
Linear Regression	1.171770442	0.509977941
LASSO	1.386330297	1.386330297
RF	0.824556968	0.545563374
Gradient Boosting	0.846468436	0.599220767
SVR	0.766967273	0.508328972
KNeighbors	0.795142402	0.552138526
DT	1.270863684	0.822534526
XGBRegressor	0.733299151	0.604312746
LGBMRegressor	0.867051929	0.588737869
ElasticNet	1.363166521	1.363166521

TABLE III. MSE WITH SELECTKBEST, PCA, AND LASSO

FS Method	SelectKBest	PCA	LASSO
Linear Regression	0.690713043	0.78452869	0.492283778
LASSO	1.386330297	1.3863303	1.386330297
RF	0.695029081	0.67395536	0.583572458
Gradient Boosting	0.93889274	0.63761269	0.530825627
SVR	0.553119458	0.8034124	0.546840816
KNeighbors	0.572774158	0.95055255	0.536134101
DT	1.154119895	1.49901768	1.224861263
XGBRegressor	0.665255965	0.52361983	0.413781451
LGBMRegressor	0.730696933	0.58751475	0.549061012
ElasticNet	1.363166521	1.3863303	1.363166521

Table III illustrates the MSE results obtained by applying three different feature selection techniques: SelectKBest, PCA, and LASSO. Typically, these approaches provide better MSE scores in most regression models, suggesting that they are more effective in selecting features compared to the RFE. It is worth mentioning that using LASSO-based feature selection leads to the largest reductions in MSE for models such as Linear Regression and XGBRegressor. This suggests that LASSO-based feature selection is effective in identifying and keeping the features that have a major impact on the performance of the model. SelectKBest and PCA exhibit notable enhancements in performance, especially when applied to the RF, Gradient Boosting, and LGBMRegressor models. The improved outcomes can be ascribed to the innate qualities of these feature selection techniques. The SelectKBest method efficiently identifies features that have the greatest statistical significance with respect to the target variable, hence directly influencing the correctness of the model. PCA preserves the variance and underlying data structure while translating the feature space into principle components, making it more compatible with specific models. LASSO inherently conducts feature selection by utilizing regularization and reducing feature coefficients, resulting in the elimination of the non-contributory variables and the creation of a more model-relevant subset of features. The relevance of selecting the appropriate approach for feature selection becomes evident when considering the properties of the dataset and the needs of the regression models. This alignment is crucial for enhancing the performance of predictive modeling.

The results of this study, presented in Figures 2 and 3, disclose a significant difference in the efficacy of feature selection techniques among various regression models. The XGBRegressor and LGBMRegressor models, which are gradient boosting models, initially exhibited the lowest MSE, suggesting strong performance when using all the available features. After implementing feature selection, a consistent pattern of decreased MSE for linear models was noticed. This emphasizes the effectiveness of feature selection in improving model accuracy by eliminating irrelevant variables and reducing noise. The utilization of PCA and RFE in conjunction with a RF estimator resulted in substantial enhancements for multiple models, indicating that these feature selection techniques successfully identified and preserved the most influential features. In contrast, specific models demonstrated an increase in MSE when feature selection was included, suggesting the potential elimination of significant predictive characteristics.

IV. CONCLUSION

Feature selection is a crucial step in machine learning, with the goal of selecting and keeping just the input variables that have the highest predictive power for a specific outcome. Extracting the most informative features from the dataset, not only improves the performance of the model by reducing overfitting and increasing accuracy, but also reduces computing complexity and promotes model interpretability. This strategy is essential when working with high-dimensional data, as the curse of dimensionality can otherwise negatively impact the performance of the model. Feature selection acts as a crucial

link between raw data and effective predictive modeling, guaranteeing that the models created are both robust and useful.

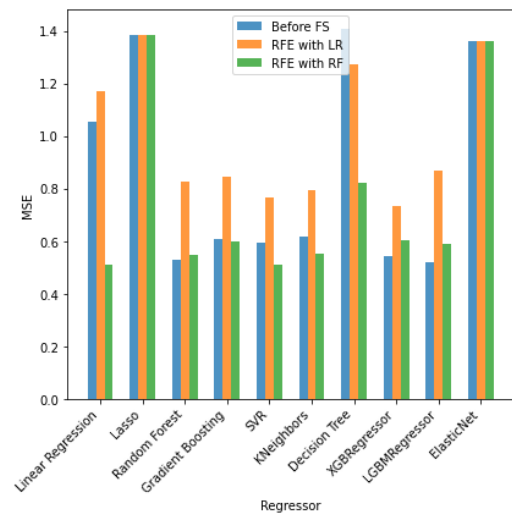


Fig. 2. MSE Comparison for RFE with LR and RFE with RF.

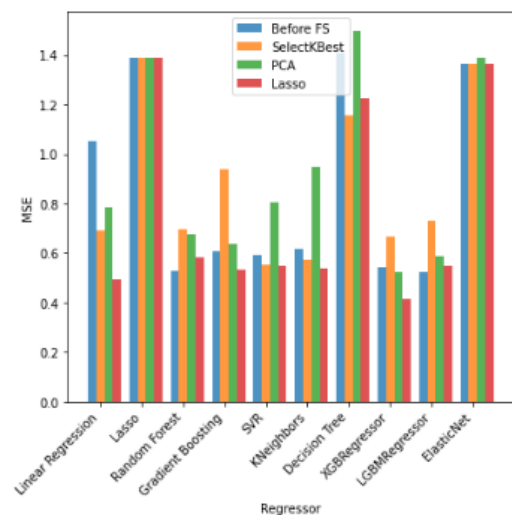


Fig. 3. MSE Comparison for SelectKBest, PCA, and LASSO.

The empirical research highlights the significance of feature selection in the field of predictive modeling. Through careful reduction of the feature space, models frequently achieve improved predicted accuracy by eliminating noise and unnecessary variables that could obscure the underlying signal. This study clearly demonstrates the varying impact of different feature selection strategies on the performance of multiple regression models, ranging from simple linear methods to more complicated ensemble methods. Specifically, techniques such as LASSO and Recursive Feature Elimination (RFE) with Random Forest have demonstrated the ability to greatly enhance model accuracy. This highlights the importance of choosing an appropriate feature selection method, which is just as crucial as choosing the model itself. Therefore, this strengthens the idea that efficient feature selection is not just an

initial step, but a crucial element in constructing reliable predictive models, requiring thoughtful analysis and customized strategies to fully utilize the predictive capabilities of the data.

ACKNOWLEDGMENT

The authors gratefully acknowledge the approval and the support of this research from the Deanship of Scientific Research study by the grant CSCR-2023-12-2177, Northern Border University, Arar, Kingdom of Saudi Arabia.

REFERENCES

- [1] A. L'Heureux, K. Grolinger, H. F. Elyamany, and M. A. M. Capretz, "Machine Learning With Big Data: Challenges and Approaches," *IEEE Access*, vol. 5, pp. 7776–7797, 2017, <https://doi.org/10.1109/ACCESS.2017.2696365>.
- [2] Y. Peng and M. H. Nagata, "An empirical overview of nonlinearity and overfitting in machine learning using COVID-19 data," *Chaos, Solitons, and Fractals*, vol. 139, Oct. 2020, Art. no. 110055, <https://doi.org/10.1016/j.chaos.2020.110055>.
- [3] S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," in *2014 Science and Information Conference*, London, UK, Dec. 2014, pp. 372–378, <https://doi.org/10.1109/SAI.2014.6918213>.
- [4] S. Nuanmeesri and W. Sriurai, "Multi-Layer Perceptron Neural Network Model Development for Chili Pepper Disease Diagnosis Using Filter and Wrapper Feature Selection Methods," *Engineering, Technology & Applied Science Research*, vol. 11, no. 5, pp. 7714–7719, Oct. 2021, <https://doi.org/10.48084/etasr.4383>.
- [5] K. Kira and L. A. Rendell, "A Practical Approach to Feature Selection," in *Machine Learning Proceedings 1992*, D. Sleeman and P. Edwards, Eds. San Francisco, CA, USA: Morgan Kaufmann, 1992, pp. 249–256.
- [6] H.-H. Hsu and C.-W. Hsieh, "Feature Selection via Correlation Coefficient Clustering," *Journal of Software*, vol. 5, no. 12, pp. 1371–1377, Dec. 2010, <https://doi.org/10.4304/jsw.5.12.1371-1377>.
- [7] N. V. Kimmatkar and B. V. Babu, "Human Emotion Detection with Electroencephalography Signals and Accuracy Analysis Using Feature Fusion Techniques and a Multimodal Approach for Multiclass Classification," *Engineering, Technology & Applied Science Research*, vol. 12, no. 4, pp. 9012–9017, Aug. 2022, <https://doi.org/10.48084/etasr.5073>.
- [8] F. L. da Silva, M. L. Grassi Sella, T. M. Franco, and A. H. R. Costa, "Evaluating classification and feature selection techniques for honeybee subspecies identification using wing images," *Computers and Electronics in Agriculture*, vol. 114, pp. 68–77, Jun. 2015, <https://doi.org/10.1016/j.compag.2015.03.012>.
- [9] P. More and P. Mishra, "Enhanced-PCA based Dimensionality Reduction and Feature Selection for Real-Time Network Threat Detection," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6270–6275, Oct. 2020, <https://doi.org/10.48084/etasr.3801>.
- [10] Z. Liu, J. Yang, L. Wang, and Y. Chang, "A novel relation aware wrapper method for feature selection," *Pattern Recognition*, vol. 140, Aug. 2023, Art. no. 109566, <https://doi.org/10.1016/j.patcog.2023.109566>.
- [11] D. K. Singh and M. Shrivastava, "Evolutionary Algorithm-based Feature Selection for an Intrusion Detection System," *Engineering, Technology & Applied Science Research*, vol. 11, no. 3, pp. 7130–7134, Jun. 2021, <https://doi.org/10.48084/etasr.4149>.
- [12] C.-W. Chen, Y.-H. Tsai, F.-R. Chang, and W.-C. Lin, "Ensemble feature selection in medical datasets: Combining filter, wrapper, and embedded feature selection results," *Expert Systems*, vol. 37, no. 5, 2020, Art. no. e12553, <https://doi.org/10.1111/exsy.12553>.
- [13] R. Tibshirani, "The lasso method for variable selection in the Cox model," *Statistics in Medicine*, vol. 16, no. 4, pp. 385–395, Feb. 1997, [https://doi.org/10.1002/\(sici\)1097-0258\(19970228\)16:4<385::aid-sim380>3.0.co;2-3](https://doi.org/10.1002/(sici)1097-0258(19970228)16:4<385::aid-sim380>3.0.co;2-3).
- [14] C. "Ann" Ratanamahatana and D. Gunopulos, "Feature selection for the naive bayesian classifier using decision trees," *Applied Artificial Intelligence*, vol. 17, no. 5–6, pp. 475–487, May 2003, <https://doi.org/10.1080/713827175>.
- [15] T. Suryakanthi, "Evaluating the Impact of GINI Index and Information Gain on Classification using Decision Tree Classifier Algorithm*," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 2, pp. 612–619, Jan. 2020, <https://doi.org/10.14569/IJACSA.2020.0110277>.
- [16] Y. Yuan, L. Wu, and X. Zhang, "Gini-Impurity Index Analysis," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3154–3169, 2021, <https://doi.org/10.1109/TIFS.2021.3076932>.
- [17] E. Shalev, "Countries_Happiness/country_statsd.csv at master · Elaishalev/Countries_Happiness," *GitHub*, https://github.com/Elaishalev/Countries_Happiness/blob/master/country_statsd.csv.