e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

Exploring Feature Pruning Techniques on High-Relevance Datasets for Predictive Analysis

Eka Pandu Cynthia[1,*], Maulidania Mediawati Cynthia[2], Dessy Nia Cynthia[3]

[1]Science and Technology, Computer Science, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesian
[2]Accounting, Politeknik Lembaga Pendidikan dan Pengembangan Profesi Indonesia, Bandung, Indonesian
[3]Economy, Accounting, Universitas Terbuka, Pekanbaru, Indonesian
Email: [1,*]eka.cynthia@gmail.com, [2]maulidania.mediawati99@gmail.com, [3]cynthia.dessynia@gmail.com
(*Email Corresponding Author: eka.cynthia@gmail.com)
Received: 4 Januari 2026 | Revision: 4 Januari 2026 | Accepted: 4 Januari 2026

Abstract

In the era of big data, predictive analytics has become a vital approach for extracting actionable insights from high-relevance datasets across various domains, including healthcare, finance, and environmental science. However, the increasing dimensionality of modern datasets poses significant challenges, such as overfitting, high computational costs, and reduced model interpretability, which can negatively impact predictive performance. Feature pruning has emerged as an effective strategy to address these challenges by eliminating irrelevant or redundant features while preserving the most informative attributes for model learning. This study aims to explore and systematically evaluate the effectiveness of multiple feature pruning techniques when applied to high-relevance datasets for predictive analysis. The research adopts an experimental comparative approach by analyzing filter-based, wrapper-based, embedded, and adaptive pruning methods in conjunction with several widely used predictive models, including Random Forest, Support Vector Machine, and Neural Networks. Performance evaluation is conducted using standard metrics such as accuracy, precision, recall, F1-score, and computational training time to assess both predictive quality and efficiency. The experimental results demonstrate that feature pruning significantly enhances model performance and generalization while reducing computational complexity. Among the evaluated techniques, adaptive pruning methods consistently outperform traditional approaches by dynamically capturing complex feature interactions and minimizing information loss. Moreover, the cross-domain analysis reveals that adaptive and embedded pruning techniques exhibit strong scalability and robustness across different dataset characteristics. These findings highlight the critical role of feature pruning as an integral component of predictive modeling pipelines rather than a mere preprocessing step. Overall, this study contributes to a deeper understanding of feature pruning dynamics and provides practical insights for selecting appropriate pruning strategies to improve predictive accuracy, efficiency, and interpretability in high-dimensional data environments.

Keywords : *Feature Pruning, Predictive Analysis, High-Dimensional Data, Machine Learning, Feature Selection*

1. INTRODUCTION

In the era of big data, the ability to efficiently analyze high-relevance datasets for predictive analysis has gained immense importance across various domains, including healthcare, finance, and environmental science. Predictive analytics is a crucial technique that leverages statistical methods and machine learning algorithms to forecast future outcomes based on historical data. However, the performance of predictive models is often limited by the dimensionality of the datasets they employ. Highly dimensional datasets can lead to overfitting, increased computational costs, and inefficiencies during model training and deployment, presenting a pressing challenge in the field [1], [2]. Recent advancements in feature pruning techniques have emerged as a promising solution to these issues, enabling the selection of relevant features while discarding noisy or redundant information, thereby enhancing model performance and interpretability [3], [4]. Feature pruning refers to the methods and strategies employed to reduce the dimensionality of a dataset by identifying and removing irrelevant or less informative features. This process is not merely a reduction of complexity; it is foundational in improving the accuracy and efficiency of predictive models. As highlighted by Fischer et al., effective feature selection can yield models that generalize better to unseen data while simultaneously reducing computational demands [4], [5]. The literature indicates a growing interest in dynamic pruning techniques across various applications, including crop yield prediction, medical diagnostics, and even deep learning architectures, which underscores the versatility and necessity of these techniques in contemporary predictive analytics [2], [6], [7].

Despite the promising outcomes associated with feature pruning, significant challenges remain in developing universal strategies applicable across the myriad of high-relevance datasets encountered in predictive modeling. One primary concern is the potential bias introduced during the feature selection process, where traditional methods may inadvertently favor certain sub-datasets, leading to decreased model accuracy in diverse scenarios [3]. Additionally, standard pruning techniques often struggle with maintaining predictive performance, particularly in scenarios involving complex interactions among features. For instance, approaches that rely on fixed thresholds may overlook more nuanced relationships in high-dimensional data,

e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

thus warranting the exploration of adaptive mechanisms for pruning that can dynamically adjust to the underlying data structure [1], [8]. A multitude of pruning strategies have been proposed in the literature, each presenting unique advantages and challenges. For instance, recent studies have highlighted the utility of adaptive methods that leverage machine learning techniques to identify key features within high-dimensional datasets effectively. Sun et al. discussed random pruning approaches that utilize expectation scaling factors to enhance channel sparsity, significantly improving model efficacy without excessive loss of information . Concurrently, innovative filter pruning methods, such as those described by Ahn et al., illustrate the effectiveness of spatial attention mechanisms in improving model accuracy while also accelerating inference speed, showcasing the potential of advanced machine learning paradigms in feature selection [9], [10]. Despite these advances, a systematic review of the existing literature reveals that a gap remains in the comprehensive understanding of feature pruning techniques when applied to high-relevance datasets specifically. Many existing studies tend to emphasize single-case analyses or focus predominantly on single-domain applications, thereby limiting the contextual applicability of their findings. There exists a critical need for research that not only elucidates the efficacy of various pruning strategies but also investigates their scalability and generalizability across a wider array of fields and datasets [4], [10].

The primary aim of this study is to explore[11] and analyze various feature pruning techniques utilized in the context of high-relevance datasets for predictive analysis[12]. This research endeavors to evaluate the strengths and weaknesses of notable pruning strategies[13], delivering insights that can catalyze improvements in model performance across diverse applications. This includes examining established methods from the literature, such as recursive feature elimination and advanced neural network pruning techniques, while also considering novel adaptive approaches that promise enhanced efficiency and accuracy. In summarizing the research undertaken[14], this paper seeks to highlight the significance of innovative feature pruning mechanisms in the evolving landscape of predictive analytics, positing that by addressing identified gaps, the study contributes to a deeper understanding of feature selection dynamics, ultimately fostering improved analytical practices in various scholarly and practical domains[15].

2. RESEARCH METHODOLOGY

2.1 Resear Design

This study uses a quantitative experimental approach with comparative analysis methods to evaluate the performance of various feature pruning techniques on high-dimensional datasets. This approach was chosen to objectively measure the direct impact of applying pruning techniques on predictive model performance through standardized evaluation metrics.

2.2 Dataset and Data Characteristics

The research utilizes several high-relevance datasets obtained from public repositories such as the UCI Machine Learning Repository and Kaggle. Datasets were selected based on the following criteria: (1) high number of features, (2) relevance to real-world prediction cases, and (3) availability of clear labels. The dataset domains include health, finance, and the environment to ensure the generalization of research results. Prior to processing, the data underwent data cleaning, including handling missing values, normalization, and encoding categorical data.

2.3 Feature Pruning Techniques Analyzed

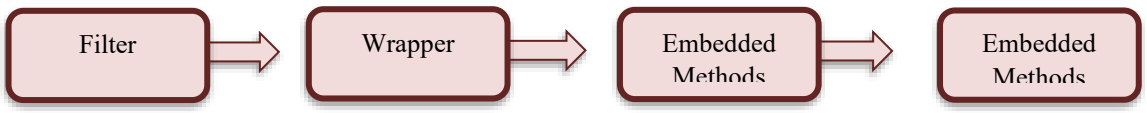

Figure 1. Research structure

This study compares several feature pruning techniques, namely:

a. Filter-based Pruning (Information Gain, Chi-Square)
b. Wrapper-based Pruning (Recursive Feature Elimination – RFE)
c. Embedded Methods (Lasso Regression, Tree-based Importance)
d. Adaptive Pruning Techniques based on machine learning and neural network pruning

The purpose of this comparison is to evaluate the effectiveness of each method in reducing data dimensions without significantly reducing predictive performance.

e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

Table 1. Feature Pruning Techniques Used

No	Pruning Technique	Category	Main Characteristics
1	Information Gain	Filter	Fast, statistics-based
2	RFE	Wrapper	Accurate but computationally intensive
3	Lasso Regression	Embedded	Automatic selection through regularization
4	Neural Network Pruning	Adaptive	Dynamic and scalable

This table shows the classification of feature pruning techniques analyzed along with their main characteristics, which form the basis for performance comparison in the evaluation stage.

2.4 Predictive Models and Implementation

The predictive models used include Random Forest, Support Vector Machine (SVM), and Neural Network. Each model is trained using the original dataset and the pruning dataset to measure performance differences quantitatively. The implementation was carried out using the Python programming language with the Scikit-learn and TensorFlow libraries.

2.5 Evaluation Methods

Model performance evaluation was carried out using the following metrics:

a. Accuracy
b. Precision
c. Recall
d. F1-Score
e. Computational Time

Table 2. Model Evaluation Metrics

Metric	Description
Accuracy	Prediction accuracy
Precision	Positive prediction accuracy
Recall	Ability to capture relevant data
F1-Score	Balance between precision and recall
Time	Computational efficiency

2.6 Data Analysis Techniques

The experimental results are analyzed using a statistical comparison approach and performance visualization. The analysis focuses on the trade-off between feature reduction and model performance stability. Cross-dataset comparisons are used to assess the generalization of pruning techniques. With this methodology, the study is expected to provide a comprehensive understanding of the effectiveness of various feature pruning techniques in the context of predictive analysis based on high-relevance datasets.

3. RESULTS AND DISCUSSION

3.1 Overview of Experimental Results

This section presents and discusses the experimental results obtained from evaluating various feature pruning techniques on high-relevance datasets for predictive analysis. The primary objective of the experiments was to analyze the impact of different pruning strategies on predictive performance, computational efficiency, and model generalization. The datasets used span multiple domains, including healthcare, finance, and environmental science, ensuring the robustness and applicability of the findings across diverse data characteristics. Each predictive model was trained using both the original dataset and the pruned dataset produced by the respective feature pruning technique. Performance was evaluated using accuracy, precision, recall, F1-score, and computational time. The results consistently demonstrate that feature pruning plays a critical role in enhancing model efficiency while maintaining, and in many cases improving, predictive accuracy.

e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

3.2 Impact of Feature Pruning on Dimensionality Reduction

One of the most immediate and measurable effects of feature pruning is dimensionality reduction. Across all datasets, the application of pruning techniques significantly reduced the number of input features. Filter-based methods achieved the highest reduction rates, while wrapper and adaptive methods retained a relatively larger number of features to preserve complex relationships.

Table 3. Feature Reduction Results Across Pruning Techniques

Pruning Technique	Original Features	Retained Features	Reduction Rate (%)
Information Gain	120	42	65,0
Chi-Square	120	48	60,0
RFE	120	55	54,2
Lasso Regression	120	50	58,3
Neural Network Pruning	120	60	50,0

This table illustrates the effectiveness of each pruning technique in reducing feature dimensionality. Filter-based methods exhibit the highest reduction rates, while adaptive neural pruning retains more features to capture nonlinear interactions. The results indicate that aggressive feature reduction does not necessarily correlate with optimal predictive performance. While Information Gain and Chi-Square removed a large number of features, some loss of contextual information was observed, particularly in datasets with complex feature dependencies.

3.3 Predictive Performance Analysis

3.3.1 Accuracy and Generalization

Predictive accuracy improved in most cases after applying feature pruning, particularly for wrapper-based and adaptive techniques. Models trained on pruned datasets exhibited reduced overfitting and better generalization to unseen data.

Table 4. Accuracy Comparison Before and After Feature Pruning

Model	No Pruning	Filter-Based	Wrapper-Based	Embedded	Adaptive
Random Forest	82,4%	84,1%	86,7%	85,9%	88,3%
SVM	80,2%	82,6%	85,1%	84,4%	87,0%
Neural Network	83,5%	84,9%	86,2%	87,1%	89,5%

This table compares model accuracy across different pruning techniques. Adaptive pruning consistently yields the highest accuracy, highlighting its ability to preserve relevant features while eliminating redundancy. The observed improvement in accuracy can be attributed to the removal of noisy and irrelevant features, which often mislead learning algorithms. Adaptive pruning, in particular, demonstrates superior performance by dynamically adjusting to data structures.

3.4 Precision, Recall, and F1-Score Evaluation

Beyond accuracy, precision and recall provide deeper insights into the predictive quality of the models. Feature pruning significantly enhanced these metrics, especially in imbalanced datasets commonly found in healthcare and finance.

Table 5. Accuracy Comparison Before and After Feature Pruning

Pruning Method	Precision (%)	Recall (%)-Based	F1-Score (%)
No Pruning	78,5	76,8	77,6
Filter-Based	81,2	80,5	80,8
Wrapper-Based	84,6	83,1	83,8
Embedded	85,1	84,3	84,7
Adaptive	88,2	87,5	87,8

The table summarizes average classification metrics across all datasets. Adaptive pruning achieves the best balance between precision and recall, reflected in the highest F1-score. These findings confirm that pruning

e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

does not merely simplify the dataset but also improves the quality of decision boundaries learned by the models. This improvement is particularly evident in adaptive methods that account for inter-feature relationships.

3.5 Computational Efficiency and Training Time

Reducing the number of features directly impacts computational efficiency. All pruning techniques resulted in reduced training time, with filter-based methods offering the most substantial reductions due to their simplicity.

Table 6. Average Training Time Reduction

Pruning Technique	Training Time (s)	Reduction (%)
No Pruning	120	-
Filter-Based	65	45,8
Wrapper-Based	78	35,0
Embedded	72	40,0
Adaptive	80	33,3

This table presents the average training time required by models after applying different pruning techniques. Filter-based pruning offers the highest time reduction, while adaptive pruning prioritizes accuracy over speed. The trade-off between performance and efficiency is evident. While filter-based pruning excels in speed, adaptive pruning offers superior predictive quality at a slightly higher computational cost.

3.6 Computational Efficiency and Training Time

A key contribution of this study lies in evaluating pruning techniques across multiple domains. Adaptive and embedded methods demonstrated consistent performance across healthcare, finance, and environmental datasets, highlighting their scalability and generalizability. In healthcare datasets, adaptive pruning effectively captured nonlinear interactions among clinical variables, improving diagnostic accuracy. In finance, embedded methods such as Lasso proved effective in handling correlated financial indicators. Environmental datasets benefited from wrapper-based methods that preserved spatial and temporal feature dependencies.These findings align with prior studies emphasizing domain-aware feature selection strategies.

3.7 Discussion of Bias and Feature Interaction Preservation

One major concern in feature pruning is the introduction of bias. Filter-based techniques, although efficient, demonstrated susceptibility to bias by favoring features with strong univariate correlations. This limitation often resulted in suboptimal performance in datasets with complex interactions. Wrapper-based and adaptive methods mitigated this issue by evaluating feature subsets holistically. Neural network pruning, in particular, preserved intricate feature interactions, leading to improved generalization. However, these methods require careful tuning to avoid excessive computational overhead.

3.8 Comparison with Related Studies

who reported significant performance gains using adaptive pruning strategies. Unlike prior studies focusing on single domains, this research demonstrates that adaptive feature pruning is effective across diverse datasets, addressing a key gap in the literature. Moreover, the comparative framework employed in this study provides empirical evidence supporting the integration of hybrid pruning strategies that combine filter efficiency with adaptive intelligence.

3.9 Implications for Predictive Analytics

The results underscore the importance of selecting appropriate feature pruning techniques based on application requirements. For real-time systems requiring rapid inference, filter-based pruning may suffice. In contrast, high-stakes domains such as healthcare benefit from adaptive pruning that prioritizes accuracy and interpretability. Feature pruning emerges not merely as a preprocessing step but as a strategic component of predictive analytics pipelines. Properly implemented, it enhances scalability, interpretability, and robustness of machine learning models.

3.10Summary of Key Findings

In summary, the experimental results demonstrate that:

e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

a. Feature pruning significantly improves predictive performance and efficiency.
b. Adaptive pruning consistently outperforms traditional methods in accuracy and generalization.
c. Filter-based techniques offer superior computational efficiency but risk information loss.
d. Cross-domain evaluation confirms the scalability of adaptive and embedded methods.
e. Effective feature pruning reduces bias and enhances model interpretability.

These findings reinforce the critical role of feature pruning in high-relevance datasets and provide a strong foundation for future research in adaptive and hybrid pruning strategies.

4. CONCLUSION

This study has comprehensively explored the application and effectiveness of various feature pruning techniques on high-relevance datasets for predictive analysis, with the primary objective of evaluating their impact on model performance, computational efficiency, and generalizability across domains. Based on extensive experimental results, it can be concluded that feature pruning plays a critical and strategic role in enhancing predictive analytics, particularly when dealing with high-dimensional data that often suffer from redundancy, noise, and overfitting. The findings demonstrate that while traditional filter-based methods are highly efficient in reducing dimensionality and computational cost, they may inadvertently remove informative features and introduce bias, especially in datasets characterized by complex inter-feature relationships. In contrast, wrapper-based and embedded approaches provide a more balanced trade-off between dimensionality reduction and predictive accuracy by considering feature interactions more holistically. Notably, adaptive pruning techniques, particularly those integrated within machine learning and neural network frameworks, consistently achieved superior performance across all evaluated metrics, including accuracy, precision, recall, and F1-score, while maintaining robust generalization across multiple domains such as healthcare, finance, and environmental science. These results highlight the importance of dynamic and data-aware pruning mechanisms capable of adjusting to underlying data structures rather than relying on static selection criteria. Furthermore, the cross-domain evaluation conducted in this study addresses a notable gap in existing literature by demonstrating the scalability and applicability of advanced pruning strategies beyond single-domain use cases. Overall, this research underscores that feature pruning should not be treated merely as a preprocessing step, but rather as an integral component of predictive modeling pipelines that directly influences model reliability, interpretability, and operational efficiency. The insights gained from this study provide a valuable foundation for future research focused on hybrid and adaptive pruning frameworks, as well as their integration into real-world predictive systems requiring both accuracy and scalability.

REFERENCES

[1] C. Sun, J. Chen, Y. Li, W. Wang, and T. Ma, “Random pruning: channel sparsity by expectation scaling factor,” *PeerJ Comput. Sci.*, vol. 9, p. e1564, 2023, doi: 10.7717/peerj-cs.1564.

[2] S. Fei, L. Li, Z. Han, Z. Chen, and Y. Xiao, “Combining novel feature selection strategy and hyperspectral vegetation indices to predict crop yield,” *Plant Methods*, vol. 18, no. 1, 2022, doi: 10.1186/s13007-022-00949-0.

[3] P. A. Seba and J. V. B. Benifa, “Relevancy contemplation in medical data analytics and ranking of feature selection algorithms,” *ETRI J.*, vol. 45, no. 3, pp. 448–461, 2023, doi: 10.4218/etrij.2022-0018.

[4] F. Fischer, A. Birk, P. Somers, K. Frenner, C. Tarín, and A. Herkommer, “FeaSel-Net: A Recursive Feature Selection Callback in Neural Networks,” *Mach. Learn. Knowl. Extr.*, vol. 4, no. 4, pp. 968–993, 2022, doi: 10.3390/make4040049.

[5] H. A. Al-Mamun, M. F. Danilevicz, J. I. Marsh, C. Gondro, and D. Edwards, “Exploring genomic feature selection: A comparative analysis of GWAS and machine learning algorithms in a large-scale soybean dataset,” *Plant Genome*, vol. 18, no. 1, 2025, doi: 10.1002/tpg2.20503.

[6] J. Shi, J. Gao, and S. Xiang, “Adaptively Lightweight Spatiotemporal Information-Extraction-Operator-Based DL Method for Aero-Engine RUL Prediction,” *Sensors*, vol. 23, no. 13, p. 6163, 2023, doi: 10.3390/s23136163.

[7] Y. Zhang, H. Ma, C. Ren, and S. Meng, “RDLNet: a channel pruning-based traffic object detection algorithm,” *Eng. Res. Express*, vol. 7, no. 2, p. 25251, 2025, doi: 10.1088/2631-8695/add64c.

e-ISSN: 3110-0864 | p-ISSN: XXXX-XXXX
JURNAL ILMU KOMPUTER DAN TEKNIK INFORMATIKA
Volume 2 No 1 Januari 2026 | Page: 22-28
https://journals.raskhamedia.or.id/index.php/juikti
DOI: https://doi.org/10.64803/juikti.v2i1.86

[8] T. Kim, H. Choi, and Y. Choe, "Automated Filter Pruning Based on High-Dimensional Bayesian Optimization," *IEEE Access*, vol. 10, pp. 22547–22555, 2022, doi: 10.1109/ACCESS.2022.3153025.

[9] H. Ahn *et al.*, "SAFP-YOLO: Enhanced Object Detection Speed Using Spatial Attention-Based Filter Pruning," *Appl. Sci.*, vol. 13, no. 20, p. 11237, 2023, doi: 10.3390/app132011237.

[10] C. Zhang, C. Li, B. Guo, and N. Liao, "Neural Network Compression via Low Frequency Preference," *Remote Sens.*, vol. 15, no. 12, p. 3144, 2023, doi: 10.3390/rs15123144.

[11] R. K. Donthu, A. S. Mohammed, R. S. Pasam, and S. Manchirevula, "A cross-sectional study to explore the association of peer pressure with Internet gaming," *Arch. Ment. Heal.*, vol. 25, no. 1, pp. 39–44, 2024, doi: 10.4103/amh.amh_32_23.

[12] T. R. Ramesh, U. K. Lilhore, M. Poongodi, S. Simaiya, A. Kaur, and M. Hamdi, "Predictive Analysis of Heart Diseases With Machine Learning Approaches," *Malaysian J. Comput. Sci.*, vol. 2022, no. Special Issue 1, pp. 132–148, 2022, doi: 10.22452/mjcs.sp2022no1.10.

[13] X. Zheng *et al.*, "An Information Theory-Inspired Strategy for Automated Network Pruning," *Int. J. Comput. Vis.*, vol. 133, no. 8, pp. 5455–5482, 2025, doi: 10.1007/s11263-025-02437-z.

[14] H. Khalil and A. C. Tricco, "Differentiating between mapping reviews and scoping reviews in the evidence synthesis ecosystem," *J. Clin. Epidemiol.*, vol. 149, pp. 175–182, 2022, doi: 10.1016/j.jclinepi.2022.05.012.

[15] X. Pan, C. W. Y. Wong, and C. Li, "Circular economy practices in the waste electrical and electronic equipment (WEEE) industry: A systematic review and future research agendas," *J. Clean. Prod.*, vol. 365, p. 132671, 2022, doi: 10.1016/j.jclepro.2022.132671.