

Feature Selection and Dimensionality Reduction Strategies for Optimizing Machine Learning Classification Models

Dr. Beatriz Almeida

Assistant Professor, Federal Institute of Advanced Computing Studies, Brazil

Abstract

Machine learning classification models frequently encounter computational and statistical difficulties when processing datasets containing large numbers of explanatory variables. High-dimensional datasets may include irrelevant attributes, redundant measurements, correlated features, and complex relationships that influence predictive performance. Although extensive feature representations can provide detailed information about individual observations, they may also increase computational requirements and introduce additional opportunities for model overfitting. Feature selection and dimensionality reduction provide methodological approaches for controlling these difficulties by identifying informative variables or constructing compact representations of the original data. The present paper develops a hypothetical comparative framework for examining feature selection and dimensionality reduction strategies used to optimize machine learning classification models. The proposed investigation considers filter-based selection, wrapper-based selection, embedded feature selection, Principal Component Analysis, and Linear Discriminant Analysis. A fictional dataset containing 3,600 observations and 600 numerical features is introduced to illustrate the experimental design. A linear Support Vector Machine serves as the common classification model for evaluating alternative feature representations.

The proposed methodology incorporates data preprocessing, feature ranking, transformation procedures, model development, and independent performance evaluation. Classification accuracy and macro-averaged F1-score are considered primary performance measures. Two hypothetical analytical tables demonstrate possible differences in classification performance, retained dimensionality, and final classifier fitting time. The methodological discussion examines feature relevance, information preservation, selection stability, computational complexity, and evaluation bias. Particular attention is given to distinguishing feature selection from feature extraction and recognizing that reduced dimensionality does not necessarily guarantee improved classification. The paper concludes that feature optimization offers a valuable methodological direction for developing efficient predictive models when supported by appropriate validation procedures, reproducible analysis, and statistically reliable performance assessment.

Keywords: Feature Selection, Dimensionality Reduction, Machine Learning, Classification Optimization, Principal Component Analysis, Linear Discriminant Analysis, Recursive Feature Elimination, Embedded Learning, Predictive Accuracy, High-Dimensional Data.

1. Introduction

Machine learning classification has become an important component of computational research because it enables automated systems to identify relationships among observations and assign them to predefined categories. Classification algorithms are widely applied in information processing, scientific analysis, digital systems, and other data-intensive environments. Their effectiveness depends substantially on the quality and organization of the explanatory variables used during model development. As datasets become increasingly complex, the number of available features may expand considerably. Although these variables can provide extensive information, excessive dimensionality may introduce challenges associated with computational efficiency, predictive stability, and the generalization of learned classification patterns.

High-dimensional data classification presents difficulties when the available features contain redundant information, irrelevant attributes, or strong correlations. Variables that contribute limited predictive information may increase model complexity without improving classification performance. In some circumstances, the inclusion of unnecessary attributes can cause algorithms to identify relationships that are specific to the training observations rather than representative of independent data. This difficulty becomes particularly significant when the number of explanatory variables is large relative to the available sample size. Consequently, developing reliable classification models requires careful consideration of feature relevance, information redundancy, and the statistical characteristics of the learning problem.

Feature selection provides a methodological approach for reducing unnecessary information by identifying a subset of relevant variables from the original feature space. Filter methods examine statistical relationships between features and target outcomes, while wrapper approaches assess candidate subsets through the performance of a predictive model. Embedded methods integrate feature selection into the model estimation process. These strategies differ in computational requirements, sensitivity to feature interactions, and dependence on the selected classification algorithm. Their comparative evaluation is therefore important because a selection procedure that performs effectively for one dataset or classifier may not provide equivalent benefits under different conditions.

Dimensionality reduction represents a related but distinct approach to controlling the complexity of high-dimensional information. Unlike feature selection, which generally retains selected original variables, dimensionality reduction may construct new representations by transforming combinations of existing attributes. Principal Component Analysis identifies directions associated with substantial variance, whereas Linear Discriminant Analysis constructs discriminative projections using known class labels. These methods can reduce the number of variables processed by a classifier, although their effectiveness depends on how well the transformed representation preserves information relevant to the classification task. Understanding the relationship between dimensionality reduction and predictive performance remains an important methodological concern.

The present paper proposes a hypothetical comparative investigation of feature selection and dimensionality reduction strategies for optimizing machine learning classification models. The study considers filter, wrapper, and embedded feature selection alongside Principal Component Analysis and Linear Discriminant Analysis. A fictional multiclass dataset provides the basis for demonstrating the intended analytical framework. A common linear Support Vector Machine classifier is proposed to support consistent performance comparison. Two illustrative tables demonstrate potential classification

and computational outcomes. The investigation does not report completed experiments; instead, it establishes a structured methodology for future empirical research involving feature optimization.

2. Objectives of the Study

The primary objective of this study is to examine the methodological effectiveness of feature selection and dimensionality reduction strategies for optimizing machine learning classification models. The investigation concentrates on predictive performance, feature relevance, computational efficiency, and evaluation reliability. Three specific objectives establish the analytical direction of the proposed comparative research framework.

2.1 Examination of Feature Relevance and Selection Strategies

The first objective is to examine how different feature selection strategies identify relevant explanatory variables within high-dimensional classification datasets. Filter-based, wrapper-based, and embedded methods are selected to represent alternative approaches to feature optimization. The proposed study considers how these methods evaluate relationships between individual attributes and classification outcomes. Particular attention is given to identifying variables that contribute meaningful predictive information while reducing unnecessary redundancy. This objective emphasizes the importance of selecting informative features rather than simply minimizing the number of available variables.

The investigation further aims to distinguish statistical feature relevance from practical classification usefulness. A variable may exhibit a strong individual relationship with the target outcome while contributing limited additional information when combined with other attributes. Conversely, certain features may become useful through interactions that are not evident from individual statistical comparisons. Consequently, the proposed framework considers both feature ranking and model-dependent selection procedures. This objective seeks to identify methodological conditions under which different strategies may provide stable and interpretable feature subsets.

2.2 Comparative Assessment of Dimensionality Reduction Techniques

The second objective focuses on examining dimensionality reduction approaches that transform high-dimensional datasets into compact representations. Principal Component Analysis and Linear Discriminant Analysis are considered because they follow different mathematical principles. PCA identifies directions explaining substantial variation in the explanatory data, whereas LDA uses class information to construct discriminative projections. The proposed comparison examines how these transformations may affect classification performance and computational requirements. Particular attention is given to the relationship between information preservation and the number of retained components.

A further objective involves evaluating whether compact feature representations maintain sufficient information for reliable classification. Reducing dimensionality can remove unnecessary variation, but it can also discard characteristics that contribute to class separation. Consequently, the proposed investigation considers prediction performance alongside dimensionality reduction rather than treating smaller feature spaces as inherently superior. The comparison also recognizes that supervised transformations must be fitted using development data only. This objective emphasizes controlled validation and the careful interpretation of differences among alternative representation strategies.

2.3 Evaluation of Classification Efficiency and Model Reliability

The third objective is to identify how feature optimization procedures may influence the computational efficiency and predictive reliability of classification models. High-dimensional inputs can increase training requirements, particularly when algorithms evaluate numerous variable relationships. Reducing the number of input dimensions may decrease selected computational costs, although feature optimization itself can introduce additional processing requirements. The proposed framework therefore considers both final classifier complexity and the resources necessary to construct the optimized representation.

The investigation also seeks to establish evaluation safeguards for comparing feature optimization methods. Selection procedures may produce misleading results when fitted using information from validation or independent testing observations. Repeated optimization can further increase the risk of overfitting to development data. Consequently, this objective emphasizes clearly separated evaluation partitions, appropriate hyperparameter selection, and transparent computational reporting. The intended comparison focuses on reliable classification performance rather than improvements arising from inappropriate data preparation or evaluation procedures.

3. Review of Literature

3.1 Statistical Foundations of Feature Selection

Guyon et al. (2002) investigated gene selection using Support Vector Machines and demonstrated the methodological relevance of recursive feature elimination for high-dimensional classification. Guyon and Elisseeff (2003) subsequently examined variable and feature selection objectives, including predictive performance, computational efficiency, and improved understanding of the underlying data. Peng et al. (2005) developed a mutual-information-based approach emphasizing maximum relevance and minimum redundancy. These contributions established important foundations for selecting informative variables while addressing unnecessary feature overlap in classification problems.

Zou and Hastie (2005) introduced the Elastic Net as a regularization approach combining variable selection with the handling of correlated predictors. Saeys et al. (2007) reviewed feature selection techniques in bioinformatics and emphasized their importance for high-dimensional data analysis. Fan and Lv (2008) investigated screening methods for ultrahigh-dimensional feature spaces. Together, these studies indicate that effective feature selection depends on statistical assumptions, variable relationships, and the characteristics of the prediction problem. Their contributions support systematic comparison rather than reliance on one universally preferred selection strategy.

3.2 Development of Dimensionality Reduction Approaches

Roweis and Saul (2000) introduced Locally Linear Embedding as a nonlinear dimensionality reduction method designed to preserve selected neighborhood relationships. Tenenbaum et al. (2000) developed a geometric framework for representing nonlinear low-dimensional structure within high-dimensional observations. Jolliffe (2002) provided a comprehensive statistical treatment of Principal Component Analysis and its applications. These studies established important foundations for transforming extensive data representations into compact forms while emphasizing the mathematical relationships underlying alternative dimensionality reduction procedures.

Belkin and Niyogi (2003) investigated Laplacian Eigenmaps for dimensionality reduction and data representation. Zou et al. (2006) examined sparse Principal Component Analysis as an approach to

improving interpretability within component-based representations. Van der Maaten and Hinton (2008) developed t-distributed Stochastic Neighbor Embedding for visualizing high-dimensional information. These contributions demonstrate the diversity of available representation-learning techniques. However, visualization-oriented methods and transformations designed for prediction serve different purposes, making task-specific evaluation essential when optimizing classification models.

3.3 Feature Stability and Comparative Model Evaluation

Haury et al. (2011) examined the relationship between feature selection, predictive accuracy, stability, and interpretability in high-dimensional molecular data. Their investigation demonstrated that more complex selection procedures do not necessarily outperform simpler methods. Kursu and Rudnicki (2010) introduced the Boruta feature selection approach, while Meinshausen and Bühlmann (2010) investigated stability selection as a means of improving the reliability of variable identification. These contributions emphasize that classification accuracy alone does not provide a complete assessment of feature selection quality.

Brown et al. (2012) examined information-theoretic feature selection criteria through a unified statistical framework. Varma and Simon (2006) investigated evaluation bias associated with model selection, while Cawley and Talbot (2010) examined overfitting during model optimization. Demšar (2006) discussed statistical procedures for comparing classification algorithms across datasets. Collectively, these studies support reproducible feature selection, independent validation, and careful interpretation of performance differences. They also establish the importance of accounting for uncertainty when comparing alternative classification optimization strategies.

4. Research Methodology

4.1 Research Design and Classification Framework

The present paper adopts a hypothetical comparative quantitative research design to examine feature selection and dimensionality reduction strategies within a multiclass classification environment. Six feature configurations are considered: the complete feature set, mutual-information-based selection, recursive feature elimination, Elastic Net selection, Principal Component Analysis, and Linear Discriminant Analysis. A linear Support Vector Machine is proposed as the common final classifier. This arrangement allows the investigation to concentrate on differences in feature representation rather than changes in the underlying classification algorithm.

The proposed analytical procedure consists of dataset preparation, variable standardization, feature optimization, classifier training, and independent performance evaluation. Feature selection and transformation procedures would be fitted exclusively using development observations. Hyperparameters and retained dimensions would be determined through appropriate validation procedures. The final classifier would then be trained using the selected representation. The proposed design has not been implemented experimentally, and no actual classifier predictions were generated while preparing this illustrative manuscript.

4.2 Hypothetical Dataset and Feature Optimization

For illustrative purposes, the study assumes a fictional dataset containing 3,600 observations and 600 numerical explanatory variables. The target outcome consists of three equally represented classification categories. The features are assumed to include relevant, redundant, correlated, and weakly informative

attributes. The proposed data partition contains 2,520 training observations, 540 validation observations, and 540 independent test observations. These values are manually specified and do not describe collected or computationally simulated observations.

The proposed feature optimization stage would retain 60 original variables through each feature selection strategy. Mutual-information-based minimum-redundancy selection would serve as the filter approach, while Support Vector Machine Recursive Feature Elimination would represent the wrapper method. Elastic Net-based coefficient selection would provide an embedded strategy. PCA would retain 60 transformed components, whereas LDA would produce two discriminant components because the proposed dataset contains three classes. All procedures would be fitted without accessing independent test labels.

4.3 Model Training and Performance Assessment

The proposed classification model would use a linear Support Vector Machine trained separately on each optimized feature representation. Input scaling would be estimated using training observations, with identical transformations subsequently applied to validation and test data. Classification parameters would be selected through development-stage validation. For wrapper-based and embedded approaches, feature selection would be repeated within the relevant validation procedures to prevent leakage. The proposed framework would preserve equivalent class distributions across evaluation partitions through stratified sampling.

The intended evaluation would report classification accuracy, macro-averaged F1-score, retained dimensionality, and computational requirements. Accuracy would measure the proportion of correctly classified observations, while macro-F1 would assign equal importance to each of the three categories. Computational evaluation would distinguish the cost of final classifier fitting from the additional expense of feature selection or transformation. Repeated experiments and uncertainty estimates would be necessary before interpreting performance differences as meaningful. The present manuscript uses hypothetical values solely for academic illustration.

5. Analysis

5.1 Illustrative Comparison of Classification Performance

Classification performance provides an important basis for examining whether feature optimization procedures preserve useful information while reducing unnecessary variables. Table 1 presents hypothetical classification outcomes for six feature configurations evaluated using a common linear Support Vector Machine. The fictional test dataset contains 540 observations equally distributed among three classes. Correct classification counts are included to demonstrate how overall accuracy could be calculated. The reported macro-F1 values are illustrative rather than derived from independently generated class-level predictions.

The hypothetical comparison assigns the highest classification accuracy to recursive feature elimination, followed by Elastic Net selection and mutual-information-based selection. PCA and LDA receive different illustrative performance values, reflecting a possible effect of transforming the original variables. However, these differences cannot establish actual methodological superiority because the procedures were not executed. A genuine investigation would require class-specific predictions, repeated experimental trials, confidence intervals, and appropriate statistical testing before determining whether feature optimization improves classification reliability.

Table 1: Hypothetical Classification Performance Across Feature Optimization Strategies

Feature Optimization Strategy	Correct Predictions (out of 540)	Accuracy (%)	Macro-F1
Full Feature Set	424	78.52	0.780
Mutual Information / mRMR	453	83.89	0.837
SVM Recursive Feature Elimination	459	85.00	0.848
Elastic Net Feature Selection	456	84.44	0.842
Principal Component Analysis	443	82.04	0.817
Linear Discriminant Analysis	435	80.56	0.803

The illustrative pattern demonstrates how retaining a smaller number of informative features might be associated with changes in predictive performance. However, dimensionality reduction may improve, preserve, or reduce classification accuracy depending on the dataset and selected learning procedure.

5.2 Illustrative Assessment of Dimensionality and Computational Efficiency

Dimensionality reduction and feature selection can influence computational requirements by changing the number of variables processed during final classifier development. Table 2 presents hypothetical dimensionality and fitting-time values for the proposed classification strategies. The feature selection approaches retain 60 original variables, representing a 90% reduction from the full feature space. PCA also produces 60 components, while LDA produces two discriminant components. These differences demonstrate the distinction between selecting existing attributes and constructing transformed representations.

The hypothetical fitting times illustrate how a smaller classification input may require fewer computational resources during the final model fitting stage. However, these values exclude the time required for feature ranking, recursive elimination, embedded selection, and transformation estimation. Wrapper-based procedures may require substantial additional computation because the classifier is repeatedly trained during feature evaluation. Consequently, the table should not be interpreted as evidence that every dimensionality reduction strategy improves end-to-end computational efficiency.

Table 2: Hypothetical Dimensionality Reduction and Classifier Fitting Characteristics

Feature Strategy	Final Input Dimensions	Dimensionality Reduction (%)	Final Classifier Fitting Time (Seconds)
Full Feature Set	600	0.00	21.0
Mutual Information / mRMR	60	90.00	9.1

SVM Recursive Feature Elimination	60	90.00	10.6
Elastic Net Feature Selection	60	90.00	9.7
Principal Component Analysis	60	90.00	7.8
Linear Discriminant Analysis	2	99.67	2.1

The illustrative comparison emphasizes that computational efficiency must be assessed across the complete modeling pipeline. A procedure that produces a compact representation may still require substantial computational effort during its initial construction.

6. Discussion

6.1 Feature Relevance and Classification Improvement

Feature selection provides a methodological approach for controlling unnecessary information within machine learning classification models. When datasets contain redundant or weakly informative variables, selecting a smaller subset may simplify the learning process and reduce opportunities for overfitting. However, the relationship between feature count and classification performance is not necessarily linear. Removing attributes can improve predictive stability under certain conditions, while excessive reduction may eliminate information required to distinguish observations belonging to different classes.

The hypothetical comparison assigns favorable accuracy and macro-F1 values to recursive feature elimination and Elastic Net selection. These examples illustrate how model-dependent procedures might be evaluated against statistical filters and complete feature representations. Nevertheless, the numerical differences do not demonstrate actual performance improvements. A valid empirical conclusion would require consistent feature selection procedures, repeated training, independent testing, and statistical uncertainty assessment. The reliability of selected feature subsets should also be examined across different development samples.

6.2 Dimensionality Reduction and Information Preservation

Dimensionality reduction methods transform extensive explanatory data into representations containing fewer variables. Principal Component Analysis emphasizes directions that capture substantial variance, whereas Linear Discriminant Analysis seeks directions relevant to separating known classes. These objectives are mathematically different, and their suitability depends on the characteristics of the classification problem. A high-variance direction may contribute limited information about class membership, while a low-variance feature can remain useful for distinguishing categories.

The proposed framework illustrates how PCA and LDA could be compared with methods that preserve original variables. However, transformed components may be more difficult to interpret because they represent combinations of the original attributes. LDA also has a structural dimensionality limitation associated with the number of classes. Consequently, fewer dimensions do not automatically

imply superior classification. The practical value of dimensionality reduction depends on predictive performance, representational quality, interpretability, and computational requirements.

6.3 Evaluation Reliability and Model Optimization

Reliable comparison of feature optimization strategies requires careful separation between model development and independent performance assessment. Feature selection conducted before data partitioning can introduce information leakage because the chosen variables may reflect relationships present within future test observations. Supervised dimensionality reduction procedures are particularly sensitive to this problem because class labels influence their transformations. Consequently, feature selection and transformation fitting should occur within the development process rather than using the complete dataset.

Repeated model optimization also introduces the possibility of overfitting to validation information. Researchers may examine numerous feature counts, hyperparameters, and classification configurations before selecting a preferred model. This extensive search can produce favorable development results that do not generalize. The proposed methodology therefore emphasizes predefined evaluation procedures and independent testing. Statistical comparisons should consider uncertainty and performance variability rather than treating the highest observed metric as definitive evidence of methodological superiority.

7. Challenges

7.1 Redundant Features and Selection Instability

High-dimensional datasets frequently contain variables that represent overlapping information or exhibit substantial statistical correlation. Such relationships complicate feature selection because several alternative subsets may provide similar predictive capabilities. A selection method may identify one variable from a correlated group while another method retains a different variable. Consequently, differences among selected subsets do not necessarily indicate differences in predictive usefulness. The proposed framework recognizes the importance of evaluating feature relevance in relation to both individual variables and their interactions.

Selection instability introduces further difficulties when small changes in training observations produce substantially different feature subsets. This problem can limit the interpretation of selected variables and complicate reproducibility. Feature selection procedures may therefore require repeated sampling and appropriate stability assessment. However, high stability does not independently establish strong predictive performance. A method can repeatedly select variables that provide limited classification information. Reliable evaluation should consider feature stability, predictive accuracy, and the substantive meaning of retained attributes together.

7.2 Information Loss and Representation Interpretability

Dimensionality reduction introduces the possibility of losing predictive information when the selected representation excludes relevant relationships. PCA retains directions associated with substantial variance, but these directions may not correspond to the features most useful for separating classes. Similarly, supervised transformations can become unstable when available observations are limited or class distributions are difficult to estimate. Consequently, the number of retained components should be selected through suitable validation rather than arbitrary dimensionality reduction targets.

Interpretability presents an additional challenge because transformed components generally represent combinations of original explanatory variables. This can complicate attempts to understand why particular observations receive certain classifications. Feature selection may preserve more direct connections between variables and predictions, although those connections are not automatically causal. The proposed framework therefore emphasizes a distinction between predictive usefulness and explanatory interpretation. Researchers should consider whether classification improvement justifies potential reductions in the transparency of the resulting feature representation.

7.3 Computational Expense and Validation Bias

Wrapper-based feature selection can require substantial computational resources because multiple candidate feature subsets must be evaluated through repeated classifier training. Embedded methods may reduce some search requirements but still introduce optimization complexity. Dimensionality reduction techniques also require additional computation to estimate projection parameters. Consequently, reduced classifier fitting time does not necessarily indicate lower total computational cost. A fair comparison must account for preprocessing, selection, transformation, hyperparameter optimization, and final model training.

Validation bias presents another significant challenge when researchers repeatedly evaluate feature configurations using the same development observations. Such practices can produce optimistic performance estimates, particularly when the number of tested alternatives is large. Information leakage may also arise if feature selection uses independent test labels. The proposed methodology therefore requires clear separation of evaluation stages and appropriately nested optimization procedures. These safeguards are essential for distinguishing genuine classification improvements from artifacts introduced by model selection and repeated testing.

8. Conclusion

Feature selection and dimensionality reduction provide important methodological strategies for controlling the complexity of machine learning classification models. The present paper developed a hypothetical comparative framework involving filter-based selection, recursive feature elimination, embedded regularization, Principal Component Analysis, and Linear Discriminant Analysis. Its primary focus was the relationship among feature relevance, compact data representation, classification performance, and computational efficiency. The proposed framework demonstrates how alternative optimization strategies could be evaluated under consistent classification conditions.

The illustrative tables demonstrate possible procedures for reporting classification accuracy, macro-F1, retained dimensionality, and final classifier fitting time. However, all numerical values were manually constructed and do not represent executed machine learning experiments. Meaningful conclusions would require authentic or reproducibly generated datasets, implemented feature optimization methods, independent testing, and appropriate statistical analysis. Consequently, the hypothetical comparisons cannot establish the superiority of any particular selection or dimensionality reduction approach.

Overall, the study identifies feature redundancy, information preservation, computational complexity, selection stability, and evaluation bias as important considerations in classification optimization. Future empirical investigations should compare these factors across datasets with different sample sizes, feature distributions, and classification structures. Feature optimization may contribute to

reliable predictive modeling when supported by appropriate preprocessing and validation. Nevertheless, genuine performance improvements must be established through independently verified computational evidence rather than illustrative numerical comparisons.

References

1. Al-shalif, S. A., Senan, N., & Saeed, F. (2024). A systematic literature review on meta-heuristic based feature selection techniques for text classification. *PeerJ Computer Science*, 10, Article e2084. <https://doi.org/10.7717/peerj-cs.2084>
2. Alirezapour, H., Mansouri, N., & Mohammad Hasani Zade, B. (2024). A comprehensive survey on feature selection with Grasshopper Optimization Algorithm. *Neural Processing Letters*, 56, Article 28. <https://doi.org/10.1007/s11063-024-11514-2>
3. Barbieri, M. C., Grisci, B. I., & Dorn, M. (2024). Analysis and comparison of feature selection methods towards performance and stability. *Expert Systems with Applications*, 249, Article 123667. <https://doi.org/10.1016/j.eswa.2024.123667>
4. Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2), 245–271. [https://doi.org/10.1016/S0004-3702\(97\)00063-5](https://doi.org/10.1016/S0004-3702(97)00063-5)
5. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, 1157–1182. <https://www.jmlr.org/papers/v3/guyon03a.html>
6. Hall, M. A. (1999). Correlation-based feature selection for machine learning [Doctoral dissertation, University of Waikato]. University of Waikato Research Commons.
7. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer. <https://doi.org/10.1007/978-0-387-84858-7>
8. Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Article 20150202. <https://doi.org/10.1098/rsta.2015.0202>
9. Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. *Artificial Intelligence*, 97(1–2), 273–324. [https://doi.org/10.1016/S0004-3702\(97\)00043-X](https://doi.org/10.1016/S0004-3702(97)00043-X)
10. Liu, H., & Motoda, H. (Eds.). (1998). *Feature selection for knowledge discovery and data mining*. Springer. <https://doi.org/10.1007/978-1-4615-5689-3>
11. McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform manifold approximation and projection for dimension reduction. *arXiv*. <https://arxiv.org/abs/1802.03426>
12. Moslemi, A. (2023). A tutorial-based survey on feature selection: Recent advancements on feature selection. *Engineering Applications of Artificial Intelligence*, 126, Article 107136. <https://doi.org/10.1016/j.engappai.2023.107136>
13. Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), 1226–1238. <https://doi.org/10.1109/TPAMI.2005.159>
14. Roweis, S. T., & Saul, L. K. (2000). Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500), 2323–2326. <https://doi.org/10.1126/science.290.5500.2323>
15. Song, Q., Ni, J., & Wang, G. (2010). A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Transactions on Knowledge and Data Engineering*, 22(1), 1–14. <https://doi.org/10.1109/TKDE.2009.156>

16. Tenenbaum, J. B., de Silva, V., & Langford, J. C. (2000). A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500), 2319–2323. <https://doi.org/10.1126/science.290.5500.2319>
17. Theng, D., & Bhoyar, K. K. (2024). Feature selection techniques for machine learning: A survey of more than two decades of research. *Knowledge and Information Systems*, 66(3), 1575–1637. <https://doi.org/10.1007/s10115-023-02010-5>
18. van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9, 2579–2605. <https://www.jmlr.org/papers/v9/vandermaaten08a.html>
19. Song, L., Smola, A., Gretton, A., Bedo, J., & Borgwardt, K. (2012). Feature selection via dependence maximization. *Journal of Machine Learning Research*, 13, 1393–1434. <https://www.jmlr.org/papers/v13/song12a.html>
20. Villa-Blanco, C., Bielza, C., & Larrañaga, P. (2023). Feature subset selection for data and feature streams: A review. *Artificial Intelligence Review*, 56, 1011–1062. <https://doi.org/10.1007/s10462-023-10546-9>