

Variable Selection Methods, Comparison and their Applications in Machine Learning: A Review

Kirti Thakur¹, Harish Kumar² and Snehmani³

^{1,2}University Institute of Engineering (UIET), Panjab University, Chandigarh, India
Email: kirtithakur1991@gmail.com

³Defence Geoinformatics Research Establishment (DGRE), DRDO Chandigarh, India

Abstract—In the past decade, the availability of voluminous and high-dimensional datasets has immensely emerged with continuous technological innovations to foster new ways to gather and analyze data. Thus, feature selection has become a challenging task in areas of application like text classification, data analysis, prediction, information retrieval etc. Knowledge extraction using machine learning models usually depends on the quality and quantity of data that they operate on. Feature selection is one of the core concepts to identify and remove irrelevant as well as redundant information that may impact accuracy or have no impact on the results. Feature selection methods are discussed in this review paper along with practical limitations. Subsequently, the workflow to solve a feature selection problem is also elaborated with feature selection methods. In feature selection, many surveys and empirical assessments were performed in many areas like classification, prediction, regression, and clustering, respectively.

Index Terms— feature selection, feature ranking, filter, hybrid, wrapper, embedded.

I. INTRODUCTION

The abrupt growth in volume as well as dimension of datasets is problematic for algorithms and leads to high computational cost and memory usage. Mobiles, social media, cameras, infrastructure-less wireless networks, weblogs, radio frequency identification (RFID) readers, internet search, web-based data are few names of information gathering sources. According to the IDC Digital Universe study, “By 2020, around 40 zettabytes will be the size of the digital universe” [1]. Apart from structured data in traditional datasets, a large volume of unstructured and semi-structured data is present.

In the late 19th century, feature selection emerged as an important technique for selection of most relevant, significant, and important features. Variable Selection is a part of Feature Engineering, also known as ‘feature selection’ (or attribute selection). It is the process to identify and removal of irrelevant or less vital features to achieve better accuracy and performance enhancement of the model [2]. The identification or extraction of most relevant features also known as dimensionality reduction techniques [1]. The advantage of feature selection is that no information loss of a single feature importance. In case, when original features are very diverse, and a small set of features is required then removal of some features may lead to a chance of loss of information. Whereas drawback of feature extraction is that the linear combinations of original features are not interpretable and the information about the contribution of each feature is often lost. Feature extraction often decreases the feature space size without losing a lot of information. The choice of feature extraction or selection methods totally depends on the data type of application domain [3].

“The objective of feature selection is three-fold: improving the prediction performance of the predictors, providing faster and cost-effective predictors, and providing a better understanding of the underlying process that generated the data.” [2]

II. LITERATURE REVIEW

The feature selection method consists of four critical steps: feature subset creation, subset assessment, stopping criterion, and outcome validation. The objective of feature selection is to select a subset e.g., $b = \{f_2, f_4, f_8, f_9\}$ from the complete set of input features e.g. $a = \{f_1, f_2, f_3, f_4, \dots, f_n\}$ where n is the total number of features in a dataset. The subset ‘b’ can predict the output with improved accuracy and reduced computational cost comparable to the performance of the set ‘a.’ All the features in a dataset are not always relevant and the redundancy in the features may not lead to potential results. Less number of highly relevant features gives a better generalization with less training and testing time. L. Ladha et al and N. Krishnaveni et al in [4] enlisted foremost benefits of performing feature selection before modelling the data:

- Reduces dimensionality of feature space,
- Less storage requirements,
- Removes redundant, irrelevant, and noisy data,
- Overfitting reduction,
- Accuracy improvement of resulting model,
- Training time reduction,
- Improve interpretation with complexity reduction,
- Performance improvement, to gain predictive accuracy.

According to literature review, many novel methods have been implemented based on local and global relations. Genetic Algorithm and Support Vector Machine are the most used methods on the other hand many authors consider high dimensionality to be crucial concepts to address [5]. Feature selection algorithms (FSA) are classified into Filter, Wrapper, and Hybrid feature selection methods [6]. While based on data type, FSA are broadly categorized into four groups: similarity-based, sparse-learning-based, information-theoretical-based, and statistical-based methods [7]. In 2013 using affinity propagation clustering, a fast feature selection algorithm was proposed named ‘Sequential Feature Selection’ (SFS). SFS provides high accuracy and was applied separately to each cluster of a dataset [8]. In 2014 Naïve Bayesian based novel hybrid algorithm was proposed to minimize the feature selection computing complexity using filter and wrapper algorithms [9]. On the other hand, Unsupervised and multivariate filter- based feature selection methods were proposed for analyzing the redundancy and relevance of features [10].

In paper [11], four feature selection methods are compared as Decision trees, entropy measure for ranking features, estimation of distribution algorithms, and the bootstrapping algorithm on the other side also proved that the elimination of noise is vital to consider in the classification process. Bi-Objective version outperforms Particle Swarm Optimization, Ant Colony Optimization, and Genetic Algorithms as an optimization technique for ensemble systems using filter-based feature selection approach [12]. Whereas Wrapper-based approach was used by many researchers for feature selection using different optimization techniques such as Ant Colony Optimization algorithm [13], a hybrid search method and particle swarm optimization [14], Harmony Search [15].

The output of feature selection algorithms is based on various dimensions such as ranking or subset, tasks involved as supervised or unsupervised, and principle as filter/ wrapper/ embedded. The feature selection method must be enhanced to reduce redundant data for large data analytics [4]. Various studies demonstrate that various algorithms are available for feature selection and each algorithm behaves differently for different types of datasets. Therefore, analysis is required to find out the suitable algorithm for feature selection.

This section provides introduction to feature selection and brief review on feature selection algorithms. In the following section 3, different feature selection algorithms are discussed to check situation-based suitability of feature selection techniques for classification/prediction and clustering algorithms; In section 4, feature selection algorithms are compared based on different merits such as efficiency, computation cost, feature dependencies, etc. Various application and technical domains for feature selection are provided in Section 5. Finally, conclude the paper with a discussion on advanced issues section 6.

III. FEATURE SELECTION METHODS AND LIMITATIONS:

The feature selection problem has been studied form many years by statistics, researcher, and machine learning communities. With emerging data mining research, more attention has been given to feature selection techniques. Feature selection is also known as subset selection and a pre-processing technique used in machine learning to increase learning accuracy by removing irrelevant features [16]. Taxonomy of feature selection techniques for different datasets is shown in “Fig. 1”.

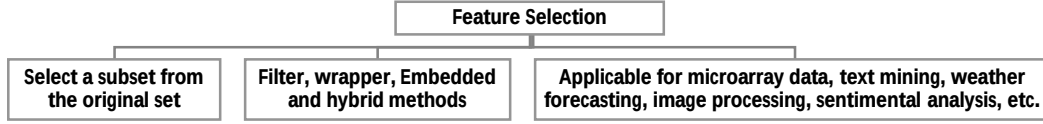


Fig 1. Taxonomy of feature selection techniques for different datasets

Complete (Exhaustive) search, Sequential Search, Exponential Search and Randomized Search are most common search Strategies used for feature selection. A large range of application areas (text analysis, microarray data analysis, climate change prediction, digital image processing, sentimental analysis, etc.) uses different types of feature selection algorithms (Filter, wrapper, Embedded and hybrid methods).

A. Filter Methods

Filter methods is the most generic approach among all the four and works irrespective of data modelling algorithm. Optimal feature set is selected by analysing general features of the dataset. In literature, methods with respect to filter class are as univariate (evaluate a single feature using ranking) and multivariate (evaluate an entire feature subset using search strategy). Filter methods cannot be applied universally for different knowledge discovery operations thus has been classified as regression, classification, or clustering [6], [10], [17]–[32] . Filter approach to feature subset is shown in Fig. 2.

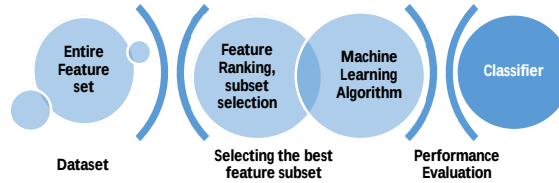


Fig 2. Filter method

According to [33], filter methods are faster and have low computational cost in contrast to wrapper and embedded methods but inefficient reliability in classification problems. As a result, this method is more popular to both academicians and industry practitioner. According to survey analysis, in Table I filter methods are summarised into 2 categories, along with application utility and references.

TABLE I. MOST COMMON FILTER METHODS

Filter feature selection	Univariate Filter class	Information Gain (IG) [17]; Gain Ratio [18]; Chi-square [18]; Fisher Score (F-Score) [19]; Symmetrical uncertainty [20]	Classification
		ReliefC [21]	Clustering
		Laplacian Score (L-Score) and Spectral feature selection (SPEC) [22]	Classification, Clustering
		Correlation [22]	Regression
		Relief and ReliefF [23]; ANOVA / Term Variance/ Variance Threshold [24]; Count/ frequency based or Count Vectorizer [25]; Gini Index (GI) [26]	Classification, Regression
	Multivariate Filter class	Fast Correlation based filter (FCBF) [22]; RRVSACO: Relevancy-Redundancy Variable Selection Ant Colony [10]; RSM: Random subspace (an ensemble) method [27] ; Relevance-redundancy feature selection (RRFS) [28]	Classification
		Correlation based feature selection (CFS) [20]; Minimal redundancy-maximal-relevance (mRMR) [6]	Classification, Regression
		Variable selection using sparse cluster analysis [29]; LFSBSS: Localized Feature Selection Based on Scatter Separability [30]; MCVS: Multi-Cluster Variable Selection [22]; Variable weighting K-means [31]; Graph Clustering Ant Colony (GCACO) [32]; Graph Clustering with Node Centrality (GNCC) [31]	Clustering
		Unsupervised Feature Selection Ant Colony (UFSACO) [10]	Classification, Clustering

B. Wrapper Methods

In 1996, Ron Kohavi and George H. John proposed the wrapper procedure to decrease irrelevant features [34]. Black-box is used as a predictor to evaluate the variable subset using wrapper methods. Wrapper methods performs better in case of a smaller number of features in contrast it is expensive to be used for huge feature set due to high computational cost and feature selection process got slower if each feature is evaluated with trained classifier [16]. Another drawback is higher risk of overfitting as if using classifier model data learning rate is too high and provides poor generalization [4], [6], [9]. The induction technique (also known as the ‘black box’) has been used to represent the supervised problems. Each training instance is characterised by feature vectors and a class label as shown in Fig. 3. Further, ranking is used for irrelevant features elimination [34].

In 2015, Diao R and Shen Q. suggested that to form a hybrid approach with intellectual properties for feature selection, development of a meta-framework may be beneficial for dynamic identification of suitable algorithms [35]. According to [14], Genetic algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) are widely used among existing heuristic methods for the variable selection problem. It has been observed that due to its cogent structure GA is preferable even though PSO and ACO provide higher accuracy for similar tasks. According to literature survey, Wrappers algorithms can be categorized into Sequential selection and Heuristic search. In Table II Wrapper methods are summarized, along with application utility and references.

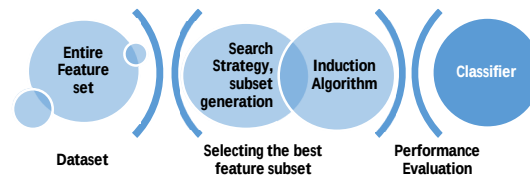


Fig 3. Wrapper method

Wrappers perform feature selection based on performance of modelling algorithm (black box). E.g., for classification tasks, features subsets will be evaluated based on classifier performance, whereas, for clustering, it will be based on performance of clustering algorithm [13], [36]–[42]. While implementing Wrappers, the subset generation is dependent on search strategy same as filter methods. The model evaluation steps are repeated for each subset until all features are ranked. According to literature, wrappers performance is better than filters because it evaluates subsets using real modelling algorithms. Generally, any combination of search strategy and modelling algorithms can be implemented for variable selection. While modelling methods like Naïve Bayes, linear support vector and Extreme Learning Machines are best with wrappers for greedy search problems.

TABLE II. MOST COMMON WRAPPER METHODS

Wrapper feature selection	Sequential/ Greedy selection algorithms	Sequential backward selection and Sequential forward [38]	Classification
	Global/ Heuristic algorithms	Random/ selection	Classification, Clustering
		Ant Colony Optimization (ACO) and Genetic Algorithm (GA) [39]; Particle Swarm Optimization (PSO-SVM) [42]; Artificial Bee Colony (ABC) [40]; Random mutation hill-climbing [36]; Simulated annealing (SA) [41]	

C. Embedded Methods

The filter methods have a major drawback that is independent of classifiers which results worse performance than wrapper methods. However, wrapper has high computational cost. A midway solution is the use of embedded or hybrid methods and both use the principal of the classifier to generate criteria to rank most optimal features [43]. Embedded methods are robust and effective while dealing with high dimensional dataset and have lower risk of overfitting in comparison to wrapper [44]. First, these methods train a machine learning model then drive feature importance. Finally, remove non-important features using the feature subset. Fig. 4 illustrates the embedded variable selection methodology.

Embedded methods consider not only feature dependencies via relationship between input and output feature, but also search features locally that allow local discrimination. In other words, it uses independent criteria to find optimal subset for known cardinality. The final optimal subset is selected among optimal subsets by using learning algorithm having best accuracy level. Various types of decision tree algorithms are used by different embedded methods as CART, C4.5, random forest, multinomial logistic regression, and its variants [45]. Among

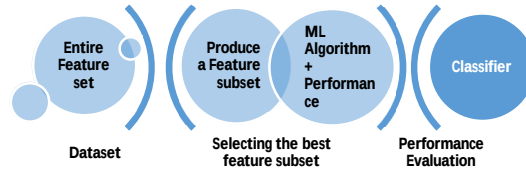


Fig 4. Embedded method

different embedded methods, Support Vectors based on Recursive Feature Elimination (SVM-RFE) is widely used [46] Whereas, regularization method also known as penalization, is one of the common embedded types of feature selection. Among regression based embedded methods, LASSO and RIDGE regression are mostly used for hyperspectral data to reduce overfitting through inherent correction [47], [48]. According to literature survey, some mostly used Embedded algorithms are mentioned in Table III. Embedded methods are summarised, along with application utility and references [45], [47]–[52].

TABLE III. MOST COMMON EMBEDDED METHODS

Embedded feature selection	Tree based methods / Decision tree algorithms	Classification and Regression Trees (CART) [45]	Classification, Regression, Clustering
		ID3 [50]; Random Forest [51]	Classification
		C4.5 Decision Trees [49]	Clustering
	Regularization / Regression methods	LASSO regression (L1 regularization) [47]	Regression
		RIDGE regression (L2 regularization) [48]	
		Elastic nets (L1/L2 regularization) [52]	

D. Hybrid Methods

Hybrid methods are developed to utilize the advantages of filters as well as wrappers methods [44]. These methods are sequential based as the first step is commonly based on filter methods to reduce the number of features that help to remove redundant features then a wrapper method is applied to select the desired number of features from a reduced set for optimal feature selection [53]. It can be any combination of classical methods such as filter-filter, filter-wrapper, and filter-filter-wrapper where the output of one method as feature subset is provided as input to another. Independent test and performance evaluation functions are used by this approach for feature subset selection. Thus, it helps to improve the efficiency and accuracy for high dimensional dataset with better computational cost [54]. Hybrid feature selection method layout is shown in Fig. 5. Several methodologies are developed using hybrid methods, such as: hybrid ant colony optimization; fuzzy random forest-based feature selection mixed gravitational search algorithm or hybrid genetic algorithms.

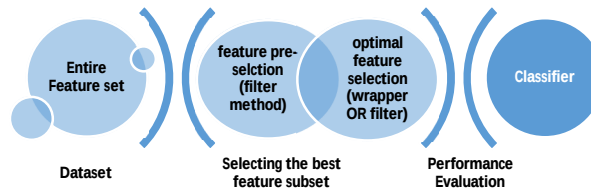


Fig 5. Hybrid method

IV. COMPARISON

All above mentioned methods have been widely used by many researchers for different applications. The performance of feature selection methods totally depends upon the dimensionality of the dataset. Thus, new methods for feature selection are always needed. Table IV, summaries and compared the feature selection methods along with merits and demerits [9], [13], [14], [37], [43], [44].

V. APPLICATIONS DOMAINS FOR FEATURE SELECTION

In this section, application and technical domains for feature selection are briefly discussed. Feature selection problems aim at selecting a subset of variables to describe the data with maximum accuracy. The chosen subset must be short in size and contain only the information required for the given application. In the literature, feature selection issues are often tackled using search strategies, where the assessment of a specified subset is performed

TABLE IV. COMPARISON OF FILTER, WRAPPER, EMBEDDED AND HYBRID METHODS

Filter methods	Wrapper methods	Embedded methods	Hybrid methods
High computational efficiency.	Better results than filter, but computational expensive for huge dataset.	The performance degrades if number of irrelevant features is high in target set.	Better computational complexity than wrapper methods.
No interaction with modelling algorithm.	Dependent on modelling algorithm performance.	Dependent on modelling algorithm performance.	Dependent on modelling algorithm performance.
Low computational cost for large dataset.	High computational cost for large dataset thus works better for small dataset.	Computational cost is less in comparison to wrapper methods.	More flexible and robust against high dimensional data
Fast execution than wrapper methods	Slow execution than filter methods.	Faster than wrapper methods.	High performance than filter methods.
Less prone to overfitting	High risk of overfitting as if data learning rate is too high.	Generally used to reduce over fitting; Least prone to overfitting	Overcomes the demerits of wrapper, enhanced search algorithm thus least prone to overfitting
Mostly ignores feature dependencies and consider each feature separately.	Consider feature dependencies.	Identification of small set of features may be problematic.	Depends on the combination of different feature selection methods.
The features subsets are evaluated by using statistical tests	Cross validation methods are used for method assessment.	Cross validation methods are used for valuation of methods.	Cross validation methods are used for assessment.

by an appropriate function (filter methods) or directly by the execution of a Data Mining tool (wrapper methods) [6], [22], [53].

In Table V, feature selection methods are summarized with respect to application domains and evaluation metrics. Although there is no ideal solution for a particular statement of problem and difficult to conclude until a significant number of test situations are efficiently addressed.

TABLE V. SUMMARY OF FEATURE SELECTION METHODS BY VARIOUS APPLICATION DOMAINS

Application domain	Feature Selection methods	Evaluation metrics	Best performing
Bioinformatics [13], [55]	Information Gain, Chi-square, t-Statistics, Gain Ratio, Symmetrical Uncertainty, ReliefF, Gini Index, Max Minority, Sum of Variance	Accuracy, Stability, AUC	Information Gain, Symmetrical Uncertainty, Chi-square and ReliefF
Text Classification [56], [57]	Probability ratio, Bi-normal Separation, Accuracy Balanced, F1 Measure, Information Gain, Power, Random, Correlation, Chi-square,	Recall, Accuracy, Precision, and F-measure	Information Gain and Bi-normal Separation
Clustering[21], [58]	Information Gain, Iterative Feature Selection, Document Frequency, Chi-square, Entropy based ranking	Entropy and Precision	Iterative feature selection
Rule Induction[15], [59]	RIPPER, Induction algorithms, Pruning, LEM1, LEM2, AQ, LERS, Multi-strategy approach	Accuracy, Recall, and Precision	RIPPER and Multi-strategy approach
System Monitoring[60], [61]	Distance, entropy, Wrapper (SVM/NN), Global geometric similarity scheme, Correlation, Principal component analysis, Frequency-domain analysis	Accuracy	Overall, all methods
Image Recognition [40], [62]	Genetic Search, K-mean, Relief, Sequential Floating Forward or Backward Selection (SFFS/SFBS), different combination of all, and Random Search	Accuracy and MSE	Relief, SFFS or SFBS and their combinations

VI. CONCLUSION

In this paper, various strategies are reviewed in the context of feature selection. Each algorithm selects the variables without computing redundancy out of which some algorithms also do not consider performance and accuracy. On the other hand, some algorithms do not consider the existence of noisy data when picking features. According to literature, if the computational time is extended, then the learning process becomes negligible. Filter methods can be used on huge datasets with many features because they are faster. However, it has no effect on accuracy. Wrapper approaches pick the best characteristics with high precision. However, the computational cost is high. Some hybrid solutions attempted to address the shortcomings of both methodologies. The objective of this study is to provide an in-depth comprehension of feature selection. If the high dimensional dataset contains irrelevant, insignificant, and unimportant features which in turns prevents effective modelling and may produce less accurate as well as less understandable results or may fail to achieve desirable results. Based on the research, an efficient unified framework for any size of dataset with minimal computing cost, and highest accuracy is required for variable selection.

REFERENCES

- [1] Gantz John and Reinse David, "THE DIGITAL UNIVERSE IN 2020: Big Data, Bigger Digital Shadow s, and Biggest Growth in the Far East," Dec. 2012. <https://www.cs.princeton.edu/courses/archive/spring13/cos598C/idc-the-digital-universe-in-2020.pdf> (accessed Nov. 09, 2022).
- [2] I. Guyon and A. Elisseeff, "An Introduction to Variable and Feature Selection," *Journal of Machine Learning Research*, vol. 3, pp. 1157–1182, 2003, Accessed: Nov. 09, 2022.
- [3] A. Janecek, W. Gansterer, M. Demel, and G. Ecker, "On the Relationship Between Feature Selection and Classification Accuracy," in *Proceedings of the Workshop on New Challenges for Feature Selection in Data Mining and Knowledge Discovery at ECML/PKDD*, Sep. 2008, pp. 90–105.
- [4] Radha V and N Krishnaveni N, "Feature Selection Algorithms for Data Mining Classification: A Survey," *Indian J Sci Technol*, vol. 12, no. 6, pp. 1–11, Feb. 2019,
- [5] T. Li, Y. S. Ho, and C. Y. Li, "Bibliometric analysis on global Parkinson's disease research trends during 1991–2006," *Neurosci Lett*, vol. 441, no. 3, pp. 248–252, Aug. 2008, Accessed: Nov. 10, 2022.
- [6] J. Tang, S. Alelyani, and H. Liu, "Feature selection for classification: A review," *Data Classification: Algorithms and Applications*, pp. 37–64, Jan. 2014,
- [7] J. Li *et al.*, "Feature selection: A data perspective," *ACM Comput Surv*, vol. 50, no. 6, Dec. 2017,
- [8] K. Zhu and J. Yang, "A cluster-based sequential feature selection algorithm," *Proceedings - International Conference on Natural Computation*, pp. 848–852, 2013, Accessed: Nov. 10, 2022.
- [9] Z. Zeng, H. Zhang, R. Zhang, and Y. Zhang, "A Hybrid Feature Selection Method Based on Rough Conditional Mutual Information and Naive Bayesian Classifier," *ISRN Applied Mathematics*, vol. 2014, pp. 1–11, Mar. 2014, Accessed: Nov. 14, 2022.
- [10] S. Tabakhi and P. Moradi, "Relevance–redundancy feature selection based on ant colony optimization," *Pattern Recognit*, vol. 48, no. 9, pp. 2798–2811, Sep. 2015, Accessed: Nov. 10, 2022.
- [11] V. H. Medina Garcia, J. Rodriguez Rodriguez, and M. A. Ospina Usaquén, "A comparative study between feature selection algorithms," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 10943 LNCS, pp. 65–76, 2018,
- [12] L. E. A. Laura Emmanuella and A. M. de Paula Canuto, "Filter-based optimization techniques for selection of feature subsets in ensemble systems," *Expert Syst Appl*, vol. 41, no. 4, pp. 1622–1631, Mar. 2014, Accessed: Nov. 10, 2022.
- [13] T. Tekin Erguzel, C. Tas, and M. Cebi, "A wrapper-based approach for feature selection and classification of major depressive disorder–bipolar disorders," *Comput Biol Med*, vol. 64, pp. 127–137, Sep. 2015, Accessed: Nov. 10, 2022.
- [14] M. M. Javidi and N. Emami, "A hybrid search method of wrapper feature selection by chaos particle swarm optimization and local search," *Turkish Journal of Electrical Engineering and Computer Sciences*, vol. 24, no. 5, pp. 3852–3861, Jan. 2016,
- [15] S. Das, P. K. Singh, S. Bhowmik, R. Sarkar, and M. Nasipuri, "A Harmony Search Based Wrapper Feature Selection Method for Holistic Bangla Word Recognition," *Procedia Comput Sci*, vol. 89, pp. 395–403, Jan. 2016, Accessed: Nov. 10, 2022.
- [16] Chandrashekar Girish and Sahin Ferat, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, Jan. 2014, Accessed: Nov. 10, 2022.
- [17] Hoque N., Bhattacharyya D. K., and Kalita J. K., "MIFS-ND: A mutual information-based feature selection method," *Expert Syst Appl*, vol. 41, no. 14, pp. 6371–6385, Oct. 2014, Accessed: Nov. 10, 2022.
- [18] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, "Data Mining: Practical Machine Learning Tools and Techniques," *Data Mining: Practical Machine Learning Tools and Techniques*, pp. 1–621, Nov. 2016, Accessed: Nov. 10, 2022.
- [19] Duda R.O., Hart P.E., and Stork D.G., "Pattern Classification," *Journal of Classification* 2007 24:2, vol. 24, no. 2, pp. 305–307, Sep. 2007,
- [20] L. Yu and H. Liu, "Feature Selection for High-Dimensional Data: A Fast Correlation-Based Filter Solution," in : *Proc. 20th International Conference on Machine Learning (ICML-2003)*, 2003, pp. 856–863.
- [21] M. Dash and Y. S. Ong, "RELIEF-C: Efficient feature selection for clustering over noisy data," *Proceedings - International Conference on Tools with Artificial Intelligence, ICTAI*, pp. 869–872, 2011, Accessed: Nov. 10, 2022.
- [22] S. Alelyani, J. Tang, and H. Liu, *Chapter: Feature Selection for Clustering: A Review*. Chapman and Hall/CRC, 2018.
- [23] M. Robnik-Šikonja and I. Kononenko, "Theoretical and Empirical Analysis of ReliefF and RReliefF," *Machine Learning* 2003 53:1, vol. 53, no. 1, pp. 23–69, Oct. 2003,
- [24] F. Ahmed and M. L. Gavrilova, "Two-Layer Feature Selection Algorithm for Recognizing Human Emotions from 3D Motion Analysis," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11542 LNCS, pp. 53–67, 2019,
- [25] H. Nematzadeh, R. Enayatifar, M. Mahmud, and E. Akbari, "Frequency based feature selection method using whale algorithm," *Genomics*, vol. 111, no. 6, pp. 1946–1955, Dec. 2019, Accessed: Nov. 10, 2022.
- [26] H. Liu, M. Zhou, X. S. Lu, and C. Yao, "Weighted Gini index feature selection method for imbalanced data," *ICNSC 2018 - 15th IEEE International Conference on Networking, Sensing and Control*, pp. 1–6, May 2018,
- [27] C. Lai, M. J. T. Reinders, and L. Wessels, "Random subspace method for multivariate feature selection," *Pattern Recognit Lett*, vol. 27, no. 10, pp. 1067–1076, Jul. 2006, Accessed: Nov. 10, 2022.

- [28] A. J. Ferreira and M. A. T. Figueiredo, "An unsupervised approach to feature discretization and selection," *Pattern Recognit*, vol. 45, no. 9, pp. 3048–3060, Sep. 2012, Accessed: Nov. 10, 2022.
- [29] D. M. Witten and R. Tibshirani, "A framework for feature selection in clustering," *J Am Stat Assoc*, vol. 105, no. 490, pp. 713–726, Jun. 2010,
- [30] Y. Li, M. Dong, and J. Hua, "Localized feature selection for clustering," *Pattern Recognit Lett*, vol. 29, no. 1, pp. 10–18, Jan. 2008,
- [31] D. S. Modha and W. S. Spangler, "Feature Weighting in k-Means Clustering," *Machine Learning 2003* 52:3, vol. 52, no. 3, pp. 217–237, Sep. 2003,
- [32] P. Moradi and M. Rostami, "Integration of graph clustering with ant colony optimization for feature selection," *Knowl Based Syst*, vol. 84, pp. 144–161, Aug. 2015, Accessed: Nov. 10, 2022.
- [33] Saptarsi Goswami and Amlan Chakrabarti, "Feature Selection: A Practitioner View," *International Journal of Information Technology and Computer Science(IJITCS)*, vol. 6, no. 11, pp. 66–77, 2014,
- [34] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artif Intell*, vol. 97, no. 1–2, pp. 273–324, Dec. 1997, Accessed: Nov. 10, 2022.
- [35] R. Diao and Q. Shen, "Nature inspired feature selection meta-heuristics," *Artificial Intelligence Review* 2015 44:3, vol. 44, no. 3, pp. 311–340, Jan. 2015,
- [36] D. B. Skalak, "Prototype and Feature Selection by Sampling and Random Mutation Hill Climbing Algorithms," *Machine Learning Proceedings 1994*, pp. 293–301, Jan. 1994, Accessed: Nov. 10, 2022.
- [37] J. C. Cortizo and I. Giraldez, "Multi criteria wrapper improvements to Naive Bayes learning," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4224 LNCS, pp. 419–427, 2006,
- [38] K. Z. Mao, "Orthogonal forward selection and backward elimination algorithms for feature subset selection.," *IEEE Trans Syst Man Cybern B Cybern*, vol. 34, no. 1, pp. 629–34, Feb. 2004,
- [39] H. R. Kanan, K. Faez, and S. M. Taheri, "Feature selection using Ant Colony Optimization (ACO): A new method and comparative study in the application of face recognition system," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 4597 LNCS, pp. 63–76, 2007,
- [40] M. Schiezero and H. Pedrini, "Data feature selection based on Artificial Bee Colony algorithm," *EURASIP Journal on Image and Video Processing* 2013 2013:1, vol. 2013, no. 1, pp. 1–8, Aug. 2013,
- [41] S. W. Lin, Z. J. Lee, S. C. Chen, and T. Y. Tseng, "Parameter determination of support vector machine and feature selection using simulated annealing approach," *Appl Soft Comput*, vol. 8, no. 4, pp. 1505–1512, Sep. 2008, Accessed: Nov. 10, 2022.
- [42] Chung-Jui Tu, Li-Yeh Chuang, Jun-Yang Chang, and Cheng-Hong Yang, "Feature selection using PSO-SVM," *IAENG Int J Comput Sci*, vol. 33, no. 1, 2007,
- [43] I. S. Oh, J. S. Lee, and B. R. Moon, "Hybrid genetic algorithms for feature selection," *IEEE Trans Pattern Anal Mach Intell*, vol. 26, no. 11, pp. 1424–1437, Nov. 2004, Accessed: Nov. 10, 2022.
- [44] das Sanmay, "Filters, Wrappers and a Boosting-Based Hybrid for Feature Selection," in *Proceedings of the Eighteenth International Conference on Machine Learning (ICML'01)*, 2001, pp. 74–81.
- [45] M. Sandri and P. Zuccolotto, "Variable Selection Using Random Forests," *Data Analysis, Classification and the Forward Search*, pp. 263–270, Aug. 2006,
- [46] Maciej Serda *et al.*, "Sparse multinomial logistic regression via Bayesian L1 regularisation," *Uniwersytet śląski*, vol. 7, no. 1, pp. 209–216, 2007,
- [47] R. Muthukrishnan and R. Rohini, "LASSO: A feature selection technique in predictive modeling for machine learning," *2016 IEEE International Conference on Advances in Computer Applications, ICACA 2016*, pp. 18–20, Mar. 2017, Accessed: Nov. 10, 2022.
- [48] M. Imani and H. Ghassemian, "Ridge regression-based feature extraction for hyperspectral data," <https://doi.org/10.1080/01431161.2015.1024894>, vol. 36, no. 6, pp. 1728–1742, Mar. 2015,
- [49] S. W. Lin and S. C. Chen, "Parameter determination and feature selection for C4.5 algorithm using scatter search approach," *Soft Computing* 2011 16:1, vol. 16, no. 1, pp. 63–75, May 2011,
- [50] S. Nizamani, N. Memon, U. K. Wiil, and P. Karampelas, "Modeling Suspicious Email Detection using Enhanced Feature Selection," Dec. 2013,
- [51] B. H. Menze *et al.*, "A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC Bioinformatics*, vol. 10, Jul. 2009, Accessed: Nov. 10, 2022.
- [52] A. Destrero, S. Mosci, C. de Mol, A. Verri, and F. Odone, "Feature selection for high-dimensional data," *Computational Management Science* 2008 6:1, vol. 6, no. 1, pp. 25–40, Apr. 2008,
- [53] Muhammad Shakil Pervez and Dewan Md. Farid, "Literature Review of Feature Selection for Mining Tasks," *International Journal of Computer Application*, vol. 116, no. 21, pp. 30–33, 2015.
- [54] Veerabhadrapa and L. Rangarajan, "Bi-level dimensionality reduction methods using feature selection and feature extraction," *Int J Comput Appl*, vol. 4, no. 2, pp. 33–38, 2010.
- [55] M. Yousef, A. Kumar, and B. Bakir-Gungor, "Application of biological domain knowledge based feature selection on gene expression data," *Entropy*, vol. 23, no. 1. 2021.

- [56] R. K. Palacharla and V. K. Vatsavayi, "A novel filter based multivariate feature Selection technique for text classification," *J Theor Appl Inf Technol*, vol. 99, no. 18, 2021.
- [57] O. M. Alyasiri, Y. N. Cheah, and A. K. Abasi, "Hybrid Filter-Wrapper Text Feature Selection Technique for Text Classification," in *International Conference on Communication and Information Technology, ICICT 2021*, 2021.
- [58] K. Gopalipour, E. Akbari, S. S. Hamidi, M. Lee, and R. Enayatifar, "From clustering to clustering ensemble selection: A review," *Engineering Applications of Artificial Intelligence*, vol. 104, 2021.
- [59] A. Adla and S. T. Zouggar, "Performance Assessment of Random Forest Induction Methods," in *2021 International Conference on Decision Aid Sciences and Application, DASA 2021*, 2021.
- [60] M. Tiboni, C. Remino, R. Bussola, and C. Amici, "A Review on Vibration-Based Condition Monitoring of Rotating Machinery," *Applied Sciences (Switzerland)*, vol. 12, no. 3, 2022.
- [61] A. N. Anggraeni, K. Mustofa, and S. Priyanta, "Comparison of Filter and Wrapper Based Feature Selection Methods on Spam Comment Classification," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 15, no. 3, 2021,
- [62] P. Jayapriya and K. Umamaheswari, "Performance analysis of two-stage optimal feature-selection techniques for finger knuckle recognition," *Intelligent Automation and Soft Computing*, vol. 32, no. 2, 2022.