International Journal of Advances in Engineering Research http://www.ijaer.com

(IJAER) 2020, Vol. No. 20, Issue No. IV, October **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

A SURVEY ON DATA MINING CLASSIFICATION FEATURE SELECTION

Isha Sachdeva

ABSTRACT

This investigation sums up the element determination measure, its significance, various kinds of highlight choice calculations, for example, Filter, Wrapper and Hybrid. Besides, it examinations a portion of the current mainstream highlight choice calculations through a writing overview and furthermore addresses the qualities and difficulties of those calculations. When there are numerous strategies are close by to be acquired, at that point Review of Literature is the best way to deal with find out about existing techniques before going for another model. Highlight choice is a transcendent preprocessing system in Data Mining, which helps in propelling the presentation of mining by choosing just the applicable highlights and staying away from the excess highlights. There are a lot of Feature Selection calculations created and utilized by most scientists. Yet at the same time, it is a rising territory in AI to be engaged for information digging and investigation measure for design acknowledgement. Many component choice calculations stand up to severe difficulties regarding viability and productivity due to the ongoing expansion in information assortment and speed. Various kinds of highlight determination calculations are accessible in writing, for example, Filter-based, Wrapper based and Hybrid measures. Additionally, investigations a portion of the current famous element determination calculations through a writing review likewise addresses the qualities and difficulties of those calculations. There is a requirement for a compelling brought together system, which ought to give highlight choice to any estimate of a dataset without boisterous information, low computational intricacy and most elevated precision.

1. INTRODUCTION

Lately, information put away and gathered for various designs are broad. Such informational index may comprise of millions of records and every one of which might be spoken to by hundreds or thousands of highlights. These days, dataset turned out to be extensive information with amazingly more number of highlights. At the point when information mining and AI calculations are applied to high-dimensional information, dimensionality is the primary issue that ought to be handled1,2. It alludes to the marvel that the story becomes sparser in high-dimensional space, antagonistically influencing calculations intended for low dimensional space.

Additionally, with countless highlights, learning models tend to overfit; this prompts execution debasement on surreptitious information. Information of high dimensionality can altogether expand the memory stockpiling necessities and computational expenses for information examination. Manual administration of these datasets is unrealistic. Accordingly, information mining and AI strategies were created to find information and perceive designs from this information, consequently. Nonetheless, for the most part, more commotion is related to this gathered information. Numerous reasons are causing a stir in

(IJAER) 2020, Vol. No. 20, Issue No. IV, October **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

this information, among which blemish in the advances that gathered the knowledge and the wellspring of the story itself are two significant reasons.

The individual property for the information examination, which is thought of, is the element. A bunch of highlights are utilized for performing order in any AI techniques. Already those applications were employing hundreds or thousands of highlights for the investigation cycle. A large number of the highlights in such informational index contain useful data for understanding the information, pertinent to the issue. All things considered, it additionally includes an enormous number of immaterial highlights and repetitive highlights. This prompts decreasing the learning execution and computational efficiency1,3. The individual ought to have significant learning involved with the problematic field to choose every one of those highlights to be used to build up a classifier from the current colossal number of factors.

The highlights, which are generally relevant to the issue, can be chosen consequently. The valuable data, which is required, ought not to be disappeared during subset choice. This cycle is called highlight choice, which has different names, for example, factor choice and traits choice. This preprocessing step lessens the dimensionality of the dataset before applying the information mining process1. It very well may be helpful for any information mining measure like arrangement, grouping, affiliation rule mining. It very well may be a choice of traits by choosing a subset of applicable highlights for utilizing model development automatically2. Dimensionality decrease is another most mainstream method, which is additionally useful in include determination through commotion evacuation. Clamour implies immaterial or excess information. Highlight choice and dimensionality decrease are diverse in numerous perspectives. All things considered, the two strategies are tending to drop unessential ascribes in a dataset, and highlight choice cycle chooses whether a specific quality is to be utilized or not without alteration.

Interestingly, dimensionality decrease checks the characteristics by surveying various collections of the accessible rundown of properties. Particular Value Decomposition, Principal Component Analysis and Sammon's Mapping are a portion of the instances of Dimensionality decrease. Truth be told, the element determination measure performs screening measure among the accessible highlights through a channel, so the undesirable highlights that are taken out. Change isn't applied during highlight choice; then, the first list of capabilities is kept up without changing its importance. Henceforth highlight choice improves the lucidness. This property has its essentialness in numerous functional applications, for example, finding essential qualities to a particular sickness and building an opinion dictionary for assumption investigation. For the grouping issue, highlight choice plans to choose a subset of profoundly separated highlights. As such, highlights are chosen that are refined to specific examples from classes of different. For the issue of highlight determination for order, because of the accessibility of name data, the pertinence of highlights is surveyed as the ability. The problems accompanying in the grouping are comprehended by Feature Selection measure.

• Predictive models ought to be precise, and highlight determination strategies need to help in this angle. On the off chance that a modest number and just the element which is necessary for forecasts are given to the model, at that point great exactness can be expected.

• Accuracy in the classifier is the main viewpoint. Highlight Selection contributes a crucial job mistake by distinguishing and eliminating immaterial and excess highlights.

• Simple models are anything but difficult to embrace and clarify, and if the quantities of highlights are less, at that point, it tends to be versatile in a straightforward manner.

• Actually, a ton of highlight choice techniques and calculations are accessible in writing. Datasets likewise contain various types of highlights with high or high measurement factors.

• If the quantity of factors is decreased and excess or superfluous factors are taken out, at that point the computational time is diminished, and the forecast exhibition is upgraded.

• In Pattern acknowledgement or AI applications, the element choice calculations can give refined origin on the information. This examination attempts to outline out the scrutinize include choice practices from the organization of a few methodologies. A vast scope of AI applications can be utilized with various sorts of highlight choice calculations, which incorporates Filter based strategies, Wrapper Based Methods and Hybrid or Embedded techniques. The objective of this investigation is to furnish an extensive thought concerning include choice. The motivations to utilize include choice are:

• If the model is furnished with the correct arrangement of factors, it will improve the exactness.

• The exhibition of AI calculations in the classifier can be quicker.

• The model unpredictability is diminished, and the translation is likewise simple.

• Overfitting can be decreased.

2. REVIEW OF LITERATURE

Strategies for dissecting the repetition and significance of highlights as solo and multivariate channel based element choice techniques were proposed. The highlights are assessed utilizing subterranean insect settlement advancement calculation. The precision of the methods is estimated with the novel heuristic data measure by considering the closeness between subsets of features3. A recommender framework for step biometric portrayal utilized Robita Gait framework. Another element choice calculation called Incremental Feature Selection (IFS) with Analysis of Variance (ANOVA) was proposed. Measurable centrality is expanded when applied with a classifier combination model4.

The difficulties of highlight choice for extensive information examination are grasped. As the size of the information develops quickly, the component determination calculation additionally should be actually improved for lessening excess data5. A relative report on four distinct kinds of highlight choice calculations was provided6. Choice trees, entropy measure for positioning highlights, assessment of dispersion calculations, and the bootstrapping analysis were analyzed and discovered every count has its own benefits and bad marks. Additionally demonstrated that the disposal of commotion is the main thought to measure. A fluffy unpleasant reliance is utilized as a rule for highlight determination and presented another soft, harsh set model to ensure that the enrollment level of a part of a similar sort impacts the most elevated amount7. It likewise successfully keeps tests from being misclassified. A covetous forward calculation for highlight determination is also utilized.

A quick consecutive component choice calculation is proposed utilizing fondness engendering clustering8. This calculation partitions the dataset into numerous groups, and afterwards, straight element choice is applied to each bunch independently. All the outcomes are gathered together for the component choice cycle. This calculation works quicker and gives high precision.

A tale cross breed highlight choice calculation is proposed9 utilizing channel based harsh restrictive common data and covering based guileless Bayesian classifier. This decreased the figuring unpredictability and number of unessential highlights likewise diminished. Another element choice calculation called subordinate class thickness based component disposal for paired datasets utilizing highlight positioning methodology called diff-measure is proposed10. It beats in numerous perspectives, for example, dimensionality and computational unpredictability decrease. An element determination calculation called hybridization of Genetic Algorithm and Particle Swarm Optimization through incorporating the speed and update rules, for example, choice principles, hybrid and transformation. It outflanks with restricted example size and naturally chooses the highlights. It is a dataset conveyance independent11. An improved Ant Colony Optimization calculation utilizing covering based technique for characterization is proposed. It beats in exactness when used with a biomedical dataset. The alpha band was discovered that it has a more significant number of highlights than beta band12.

Another calculation is utilizing a covering based component determination strategy by using a mixture search technique and molecule swarm streamlining, and the nearby hunt is proposed. It works with disorder interia weight and nearby pursuit to look among 2d conceivable cases13. Elitism based Multi-Objective Differential Evolution calculation for include determination (FAEMODE) given Filter approach is proposed. A new target recipe is produced by considering the natural and nonlinear conditions for include choice cycle. It gives an incredible outcome when contrasted and seven-channel approaches14.

A covering based component choice calculation utilizing Harmony Search for Holistic Bangla word acknowledgement is proposed. Multi-Layer Perception was being used with HS to improve precision. This calculation is contrasted and Particle Swarm Optimization and Genetic Algorithms to guarantee classifier accuracy15. A mixture channel covering highlight choice calculation for transient burden estimating is proposed. Fractional shared data from the channel approach and a covering based firefly calculation were mutually used to remove pertinent highlights. This calculation is actualized in settled help vector relapse. It performs all the more proficiently contrasted and other algorithms16. A blend of EMD–LDA–PNN–SFAM (Empirical mode Decomposition-Linear Discriminant Analysis-Probabilistic Neural Network-Simplified Fuzzy Adaptive Resonance Theory Map) calculation for include choice cycle

(IJAER) 2020, Vol. No. 20, Issue No. IV, October **e-ISSN: 2231-5152, p-ISSN: 2454-1796**

is proposed, and it outflanks in precision. J-Measure is determined. Highlight decrease is finished utilizing Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). The calculation is tried with online conclusion dataset17. A streamlining procedure for troupe frameworks utilizing channel-based methodology is proposed. It is applied with mono and bi-target variants using Particle Swarm Optimization, Ant Colony Optimization, and Genetic Algorithms. Bi-Objective variant beats in this method18.

3. HIGHLIGHT SELECTION PROCESS

The element choice cycle includes four significant advances, for example, highlight subset age, subset assessment, halting basis and result approval. The element subset age helps in the competitor determination subset for evaluation. As a matter of fact, it follows a heuristic methodology. The looking through methods it follows to deliver subsets are reformist, far-reaching and arbitrary inquiry of highlights. The nature of the subgroup produced is surveyed with an assessment measure. The new subgroup is contrasted and the past subset and found the best one. The first-evaluated subset is additionally utilized for the following examination. This examination cycle is rehashed till the halting measure is reached, and the best subset is created. The last best subset is additionally approved by various tests or with an earlier knowledge19. Figure 1 represents the element choice cycle.

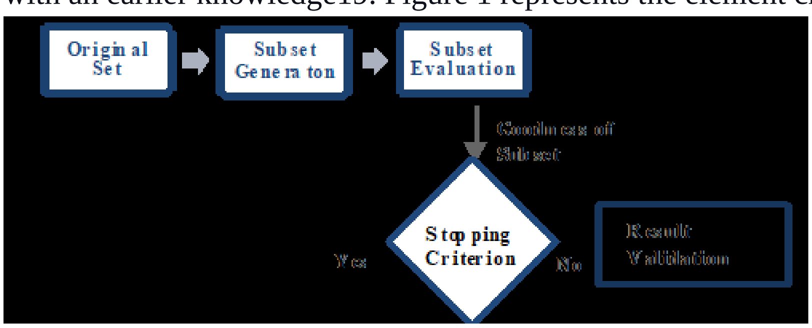

Figure 1. General framework for feature selection.

3.1 Feature Selection Algorithms

Highlight choice calculations are extensively named Filter based Feature Selection, Wrapper Based Feature Selection and Hybrid Feature Selection methods19,20. At any rate, it is additionally classified into four primary gatherings: similitude based, data hypothetical based, scanty learning-based, and factually based techniques while considering the kind of data21.

3.1.1 Filter Methods

By and large, the selection of highlights is sovereign of any AI calculations. Various sorts of objective tests are completed, and the scores are produced. The relationship between's these scores frames a foundation for Filter-based component determination. The connection is a subjective term here. The channel strategies don't eliminate multicollinearity. That is at least two indicators are exceptionally connected, which prompts factual deduction. An accurate measure is applied to allot a scoring to each component. Either the chose highlight should be kept or taken out will be selected through this positioning. The strategies are frequently alluding to a solitary trademark or quality freely regardless of whether the variable is subject to one another.

The above Figure 2 portrays the channel-based component choice calculation steps. The coefficients, for example, Pearson's Correlation, Mutual Information, Kendall Correlation, Spearman Correlation, Linear Discriminant Analysis, Chi-Square test, Fisher Score, Count based and ANOVA (Analysis of Variance) are a portion of the strategies utilized in the channel based methodology. Pearson Correlation: Pearson's connection is a measurement or coefficient used to discover the quality of the relationship between's two factors.

Shared Information: It assists with decreasing vulnerability about the estimation of another variable. Various elements of dataset the proportional information in datasets are expanded between the focused on factors and joint dispersion.

Kendall Correlation: It is an evaluating strategy used to discover the affiliation. The positioning for ordinal factors are determined, for example, various rankings and positioning of multiple factors are considered for finding connections.

Spearman Correlation: The pace of the consistent relationship among two factors is spoken to utilizing the Spearman Correlation coefficient.

Straight Discriminant Analysis (LDA): Closely identified with ANOVA and Regression Analysis. It works in a Linear model and more appropriate for the arrangement classes more than two.

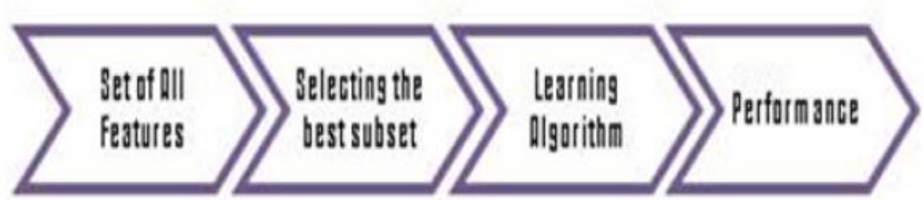

Figure 2. **Filter based feature selection process.**

Chi-Square Test: The separation between the real outcomes and expected outcomes are contrasted, and a factual method called the chi-square test.

Fisher Score: The contrasts between the average and watched values are found through fisher score. The data is amplified when the thing that matters is limited.

Tally Based: The primary data isn't introduced in all segments of information. The heaviness of the qualities from every section is checked to get a thought regarding the information.

ANOVA: Analysis of fluctuation (ANOVA) is a gathering of factual models to test the hugeness between implies.

3.1.2 Wrapper Methods

In covering techniques, there is a subset of highlights with various blends are created and utilized in a model for preparing. Given the surmising's drawn from the past model, it will be concluded whether to add or eliminate highlights from the chose subset through assessment. The covering based component determination strategies are exorbitant, computationally. The essential thought resembles a looking through the system. Model precision is assessed with the blend of highlights with a score appointed through a proactive model. Figure 3 clarifies about the covering based component choice cycle.

There are numerous coverings based element choice strategies utilized broadly. Forward component determination calculation, Backward Feature Selection Algorithm and Recursive element choice calculation are a portion of the standard models.

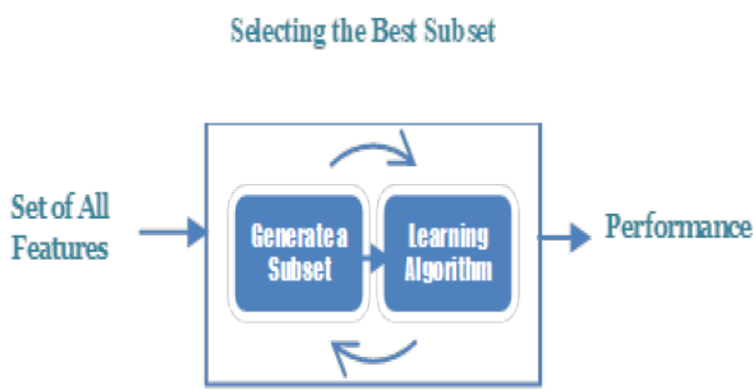

Figure 3. Wrapper based feature selection process.

Forward Feature Selection Algorithm: It is fundamentally an iterative strategy began with zero highlights. In each cycle, another element is added to the model and validated to check whether it gives an improvement in the presentation of the model. In reverse Feature Elimination Algorithm: It is a converse model of sending Feature Selection. This model beginnings with all the highlights. It is likewise an iterative model and eliminates the most un-huge element in every cycle. The presentation of the model is estimated and includes are taken out until no improvement is watched. Recursive Feature end: It is a sort of eager streamlining procedure. The most valuable element subset is found in this strategy. At every emphasis, another model with various component subset is made and discovered the best and most noticeably awful performing highlights. The model development measure proceeds until all highlights are gone through. The end of highlights is done dependent on its positioning.

Adding and eliminating highlights is done through sending or in reverse cycle. The heuristic looking through strategies are received for looking. The contrast among Filter and Wrapper strategies Following is a portion of the differences between the channel and covering highlight choice techniques.

• The hugeness of highlights is estimated by the relationship variable in channel techniques. Covering strategies measure the noteworthiness of an element through preparing a model.
• Wrapper strategies are slower than channel techniques since it performs model preparing for include estimation. With the goal that the computational expense of the covering process is high contrasted with Filter strategies.
• The subsets of highlights are weighed by measurable apparatuses in channel techniques. Cross-approval strategies are utilized in covering techniques.
• The covering put together strategies depend concerning discovering the best subsets than channel techniques.
• The covering strategies are furnishing the models with more overfitting contrasted with channel models.

3.1.3 Embedded or Hybrid Methods

The best attributes of both the channel and covering based strategies are joined to frame the implanted or mixture models. Its own underlying component choice strategies are utilized for the usage of calculations. Figure 4 portrays the cycle of crossbreed highlight determination measure. The learning cycle in installed models empowers us to locate the best exactness level during highlight choice. The regularization strategy is one of the standard inserted types include determination. The other name of regularization techniques is punishment strategies. Different imperatives like relapse calculation are brought into the enhancement of a prescient analysis to make a model with fewer coefficients to accomplish lower multifaceted nature. Tether relapse and RIDGE relapse are a portion of the popular relapse strategies which decrease overfitting through natural remedy. Regularized trees, Random multinomial logit and Memetic calculation are a portion of different models.

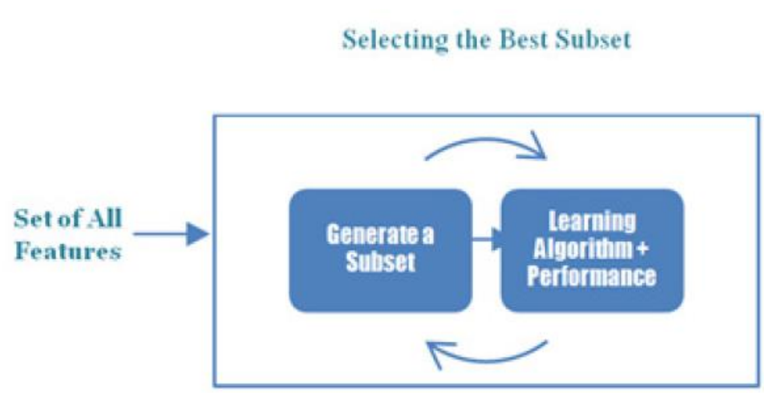

Figure 4. Hybrid feature selection process.

Table 1. Comparison of feature selection algorithms

Algorithm	**Type**	**Factors/ Approaches Used**	**Result/Inference**	**Limitation/s**
Unsupervised and multivariate filter-based feature selection method[3]	Filter Based	Ant Colony Optimization	The performance of the algorithm is improved.	New State Transmission Rule to control the randomness can be developed.
Incremental Feature Selection(IFS) with Analysis of Variance(ANOVA)	Filter	ANOVA	Statistical Significance is increased	Other Validations can be done
Affinity Propagation-Sequential Feature Selection Algorithm[8]	Wrapper Based	Cluster Based	Faster for high dimensional data	Accuracy is comparable
Fuzzy Rough Set Feature selection algorithm	Filter	Fuzzy Based\ Greedy Forward Algorithm	Works better in large degree of overlapping datasets	Does not work for small stack datasets
Novel Hybrid Feature Selection Algorithm	Hybrid	Rough Conditional Mutual Information. Bayesian Classifier	Computational complexity is reduced Irrelevant Features are reduced. Improves prediction accuracy	Accuracy can be improved
Class dependent density based feature elimination	Filter	Feature Ranking Feature Elimination Selection	Works better for High dimensional binary data. Works along with the classifier.	Other data types can be verified
Hybridization of Genetic Algorithm and Particle Swarm Optimization	Hybrid	Genetic Algorithm Particle Swarm Optimization	Automatic Feature Selection with High Accuracy with small number of Samples in High Dimensional Dataset.	SVM can be improved Verified with parameter initialization
Improved Ant Colony Optimization-SVM[12]	Wrapper	Ant Colony Optimization. Support VectorMachines	Accuracy of FS is improved. Found that more relevant features are in alpha band	Data Scalability is not verified. Might consider Beta band also.

Choas Binary Particle Swarm Optimization with Local Search[13]	Wrapper	Particle Swarm Optimization. Local Search	works with chaos interia weight. Searches among 2nd possible cases with local search	Filter based approach Using Global Searches
Filter Approach using Elitism based Multi-objective Differential Evolution algorithm for feature selection (FAEM ODE) [14]	Filter Based	Differential Evolution. Multi objective Optimization	Linear and Nonlinear dependency were considered	Other parameters also to be considered
Harmony Search(HS) for Word Recognition[15]	Wrapper Based	Harmony Search. Multi Layer Perception Classifier	Classifier accuracy is good compared with PSO and GA. Both local and global search were used	Automatic feature selection using stopping criteria. Scheme can be formulated to dynamically identify HS
Hybrid Filter-Wrapper feature selection for short term load forecasting[16]	Hybrid	Filter based Partial Mutual Information. Wrapper based firefly algorithm	Reduced the redundant features without degrading the forecasting accuracy.	Invest some of the exogenious variables and lagged variables. More extensive comparison
Combination of EMD–LDA–PNN–SFAM[17]	Filter	Empirical mode Decomposition Linear Discriminate Analysis Probabilistic Neural Network Simplified Fuzzy Adaptive Resonance Theory Map	J-Measure is improved. Real data set is used and high classification accuracy is attained. Optimal separation of features in different classes. Better categorization	Variable operating conditions are of speed and charge can be considered.
Optimization techniques for ensemble systems[18]	Filter	Particle Swarm Optimization(PSO) Ant Colony Optimization Genetic Algorithms	Compared with Mono and Bi-Objective versions PSO provides better accuracy in both. Found that Bi-Objective works better.	Other Optimization techniques Evaluation criteria can be improved.

4. EXAMINATION OF FEATURE SELECTION ALGORITHMS

The high calculation viability and consensus are the advantages of Filter-based strategies. Covering based approach ensures better outcomes. However, it is computationally costly for the colossal dataset.

The geniuses of the two techniques are gotten through inserted or crossbreed strategies. At any rate, every one of these strategies has been broadly utilized by numerous analysts for the order issues. On the off chance that the dimensionality of a dataset is unique, a similar component choice calculation may not be fit. Thus, new methodologies of Feature Selection Algorithms are consistently out of luck. Table 1 sums up a portion of the element determination calculations with all the three sorts, for example, Filter-based, Wrapper based and Hybrid. Every measure has its own benefits and faults.

5. CONCLUSION

There are many element determination calculations. Every calculation chooses just the highlights without thinking about computational excess. The presentation and exactness are not considered in specific calculations. The presence of boisterous information isn't considered when choosing highlights in particular measures. The computational time is expanded, and the learning cycle will get inconsequential. Channel-based technique rehearses the whole, preparing information while making a subset. Channel techniques can be applied to enormous datasets with voluminous highlights as it works quicker. Yet, it doesn't reflect in better exactness. Covering strategies select the best highlights with high precision. However, the computational expense is enormous. Some cross breed strategies attempted to illuminate the issues that both the techniques have. From this review, obviously, there is a requirement for a powerful brought together system, which ought to give highlight choice to any measure of the dataset without boisterous information, low computational multifaceted nature and most noteworthy exactness.

6. REFERENCES

1. Chandrashekar G, Sahin F. A survey on feature selection methods. Computers and Electrical Engineering. 2014; 40(1):16–28. https://doi.org/10.1016/j.compeleceng.2013.11.024.
2. Jantawan B, Tsai CF. A comparison of filter and wrapper approaches with data mining techniques for categorical variables selection. International Journal of Innovative Research in Computer and Communication Engineering. 2014; 2(6):4501–8.
3. Tabakhi S, Moradi P. Relevance–redundancy feature selection based on ant colony optimization. Pattern Recognition. 2015; 48(9):2798–811. https://doi.org/10.1016/j.patcog.2015.03.020.
4. Semwal VB, Singha J, Sharma PK, Chauhan A, Behera B. An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification. Multimedia tools and applications. 2017; 76(22):24457–75. https://doi.org/10.1007/s11042-016-4110-y.
5. Li J, Liu H. Challenges of feature selection for big data analytics. IEEE Intelligent Systems. 2017; 32(2):9–15. https://doi. org/10.1109/MIS.2017.38.
6. Garcia VH, Rodriguez JR, Usaquén MA. A comparative study between feature selection algorithms. International Conference on Data Mining and Big Data; 2018. p. 65–76. https://doi.org/10.1007/978-3-319-93803-5_7.
7. Wang C, Qi Y, Shao M, Hu Q, Chen D, Qian Y, Lin Y. A fitting model for feature selection with fuzzy rough sets. IEEE Transactions on Fuzzy Systems. 2017; 25(4):741–53. https://doi.org/10.1109/TFUZZ.2016.2574918.

8. Zhu K, Yang J. A cluster-based sequential feature selection algorithm. Ninth International Conference on Natural Computation (ICNC); 2013. p. 848–52. https://doi. org/10.1109/ICNC.2013.6818094. PMid:23763950 .
9. Zeng Z, Zhang H, Zhang R, Zhang Y. A hybrid feature selection method based on rough conditional mutual information and Naive Bayesian classifier. ISRN Applied Mathematics; 2014. p. 1–11.
10. Javed K, Babri HA, Saeed M. Feature selection based on class-dependent densities for high-dimensional binary data. IEEE Transactions on Knowledge and Data Engineering. 2012; 24(3):465–77. https://doi.org/10.1109/TKDE.2010.263.
11. Ghamisi P, Benediktsson JA. Feature selection based on hybridization of genetic algorithm and particle swarm optimization. IEEE Geoscience and Remote Sensing Letters. 2015; 2(2):309–13. https://doi.org/10.1109/ LGRS.2014.2337320.
12. Erguzel TT, Tas C, Cebi M. A wrapper-based approach for feature selection and classification of major depressive disorder–bipolar disorders. Computers in biology and medicine. 2015; 1(64):127–37. https://doi.org/10.1016/j. compbiomed.2015.06.021. PMid:26164033 .
13. Javidi MM, Emami N. A hybrid search method of wrapper feature selection by chaos particle swarm optimization and local search. Turkish Journal of Electrical Engineering and Computer Sciences. 2016; 24(5):3852–61. https://doi. org/10.3906/elk-1404-220.
14. Elitism based multi-objective differential evolution for feature selection: A filter approach with an efficient redundancy measure [Internet]. [cited 2017 Aug 15]. Available from: https://www.sciencedirect.com/science/article/pii/ S1319157817301003.
15. Das S, Singh PK, Bhowmik S, Sarkar R, Nasipuri M. A harmony search based wrapper feature selection method for holistic bangla word recognition. Procedia Computer Science. 2016; 89:395–403. https://doi.org/10.1016/j. procs.2016.06.087.
16. Hu Z, Bao Y, Xiong T, Chiong R. Hybrid filter–wrapper feature selection for short-term load forecasting. Engineering Applications of Artificial Intelligence. 2015; 40:17–27. https://doi.org/10.1016/j.engappai.2014.12.014.
17. Ali JB, Saidi L, Mouelhi A, Chebel-Morello B, Fnaiech F. Linear feature selection and classification using PNN and SFAM neural networks for a nearly online diagnosis of bearing naturally progressing degradations. Engineering Applications of Artificial Intelligence. 2015; 42:67–81. https://doi.org/10.1016/j.engappai.2015.03.013.
18. Santana LE, de Paula Canuto AM. Filter-based optimization techniques for selection of feature subsets in ensemble systems. Expert Systems with Applications. 2014; 41(4):1622–31. https://doi.org/10.1016/j.eswa.2013.08.059.
19. Tang J, Alelyani S, Liu H. Feature selection for classification: A review. Data classification: Algorithms and applications; 2014. p. 1–33.
20. A survey of feature selection and feature extraction techniques in machine learning [Internet]. [cited 2014 Oct 09]. Available from: https://ieeexplore.ieee.org/document/6918213.
21. Feature selection: A data perspective [Internet]. [cited 2018 Aug 26]. Available from: https://arxiv.org/ abs/1601.07996.