American Journal of Engineering Research (AJER)
e-ISSN: 2320-0847 p-ISSN : 2320-0936
Volume-9, Issue-4, pp-33-40
www.ajer.org
Research Paper Open Access

Examination of Feature Selection Methods and an Application

Ozer Celik[1]
1(Department of Mathematics-Computer Science/Eskisehir OsmangaziUniversity, TR)
Corresponding Author: Dr. Ozer Celik, email: ozer@ogu.edu.tr

ABSTRACT : *With the improvements in the processing capabilities of machines, the use of artificial intelligence (AI), also called machine intelligence, is becoming widespread for processes that are thought to require "intelligence". Machine learning (ML), one of the sub-topics of artificial intelligence, which is frequently mentioned recently, continues its development. Feature selection in machine learning process is of serious importance for various reasons such as generalization performance, working time requirements, constraints related to the problem, cost and interpretation problems. In addition, feature selection methods have become more important especially for the purposes of creating simpler and more understandable models and increasing the predictive performance for the big data structures that are frequently encountered today. In this study, the most commonly used feature selection methods were examined and it was aimed to provide information to researchers and academicians who want to work on this subject by getting review literature. The developments, advantages and disadvantages of feature selection from past to present have been examined. In addition, an application study was carried out to determine the variables (features) affecting the incontinence status of children by using machine learning techniques on pediatric urology data, estimation of urinary incontinence in children, predictive performance and to investigate the effects of some feature selection methods on predictive performance.*

KEYWORDS *feature selection, machine learning, artificial intelligence.*

Date of Submission: 20-03-2020 Date of Acceptance: 06-04-2020

I. INTRODUCTION

Feature selection, also known as variable subset selection in machine learning and statistical science, is the process of selecting a subset of related properties (variables, predictors) for use in model making.

- Feature selection techniques are used for four reasons:
- Simplify models to facilitate interpretation [1],
- Shorten training times,
- Dimensionality reduction,
- Eliminating overfitting status,
- Reducing variance [1]

Feature selection techniques should be distinguished from feature extraction. Feature extraction creates new features from the functions of the original features, while feature selection returns a subset of the features. Feature selection techniques are used in areas with many features and in relatively few samples (or data points).

1.1. Literature review

As a result of their study, Jain et al. reported that the sequential forward floating selection (SFFS) algorithm proposed by Pudil and others, relatively better than other algorithms tested. Using four different tissue models, they investigated an optimal feature selection problem set for land use classification based on SAR satellite images. They applied feature selection process for pooling features derived from different tissue models [3].

In his study, Hall et al. discussed the problem of feature selection for machine learning with a correlation-based approach. He argued that good feature sets contain features that are highly related to the class but unrelated to each other. Hall stated that Correlation based Feature Selection (CFS) is an algorithm that combines this evaluation formula with an appropriate correlation measure and an intuitive search strategy [4].

Hall et al. stated that filters are more practical than packages when applying the feature selection process to large databases. In these studies, a fast, correlation-based filter algorithm that can be applied to problems with continuous and discrete type data is explained. Naive Bayes reports that the algorithm performs as well as the well-known ReliefF attribute estimator when used as a preprocessing step for sample-based learning, decision trees, locally-weighted regression and model trees. They say that ReliefF chooses more variables because it reduces the data size by fifty percent in most cases. They also noted that decision and model trees created from preprocessed data are usually smaller [5].

Watson et al. introduced feature selection algorithms for SVMs in their articles. They have developed a new method. The resulting algorithms have been shown to be superior to some standard feature selection algorithms, both in toy data and in real-life problems such as facial recognition, pedestrian detection, and analysis of DNA microarray data [6].

In the articles of Lui and Yu, they investigated the topic of feature selection, current feature selection algorithms for classification and clustering. They grouped and compared different algorithms with a framework that categorized them based on search strategies, evaluation criteria, and data mining tasks. An illustrative example is presented to show how existing feature selection algorithms can be integrated into a meta algorithm that can take advantage of individual algorithms [7].

In his study, Uguz used two-stage feature selection and feature extraction to increase text classification performance. In the first stage, each term in the document is listed according to their importance for classification using the Information Gain (IG) method. In the second stage, genetic algorithm (GA) and principal component analysis (PCA) feature selection and feature extraction methods were applied to the terms sorted in descending order and size reduction was performed. Thus, less important terms during text categorization are ignored and feature selection and subtraction methods are applied to terms that have the highest importance [8].

Chandrashekar et al. reports that there are many feature selection methods available because there are hundreds of variables with very high characteristics in the literature. They explained that the benefits of feature selection methods are to reduce computation time, improve predictive performance, and better understand data in machine learning or pattern recognition applications. The aim of their work is to provide a general introduction to variable elimination that can be applied to various machine learning problems. Focusing on Filter, Wrapper and Embedded methods, they applied some of the feature selection techniques to standard datasets to demonstrate the applicability of feature selection techniques [9].

Barlaud et al. applied a feature selection process for biological variables in order to increase the accuracy performance of the controlled classification on large data in multi-dimensional space and to decrease the cost of processing [10].

II. FEATURE SELECTION METHODS

Feature selection is defined as the selection of the best subset that can represent the original data se. This process aims to reduce the number of features in the dataset, reducing the size of the data, by selecting the most useful and most important features for the problem of interest. A feature selection algorithm can be seen as an evaluation measure that scores different feature subsets with the combination of a search technique to suggest new feature subsets. The simplest algorithm is to test each subset of its variables that minimize the error rate. The selection of the evaluation metric significantly affects the algorithm. Feature selection methods are defined in three main categories; wrappers, filters and embedded methods.

2.1. Wrapper Method

Unlike filter approaches, Wrapper methods evaluate variable subsets that allow to detect possible interactions between variables [11]. Two main disadvantages of these methods:

- When the number of observations is insufficient, the risk of compatibility increases.
- When the number of variables is high, the processing time is extended.

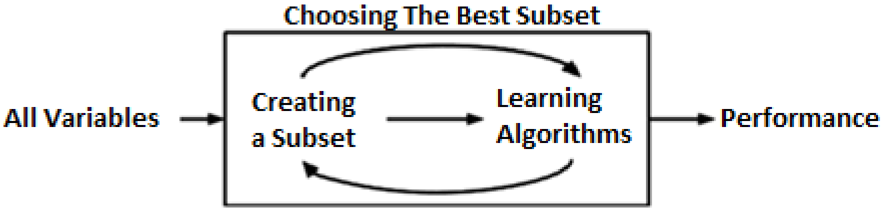

Fig.1. Wrapper method process flow chart

Wrapper methods use a predictive model to score feature subsets. Each new subset is used to train a model tested in a hold-out set. Counting errors (error rate of the model) in this trial set gives a score value for this subset. Since Wrapper methods train a new model for each subset, they are quite affordable but not always performing well.

2.2. Filter Method

Filter-type methods choose variables regardless of model. They are based only on general features, such as correlation with the predicted variable. Filter methods eliminate variables that are least related to the dependent variable (output) [12]. Filter methods tend to choose unnecessary variables, since they do not take into account the relationships between the independent variables.

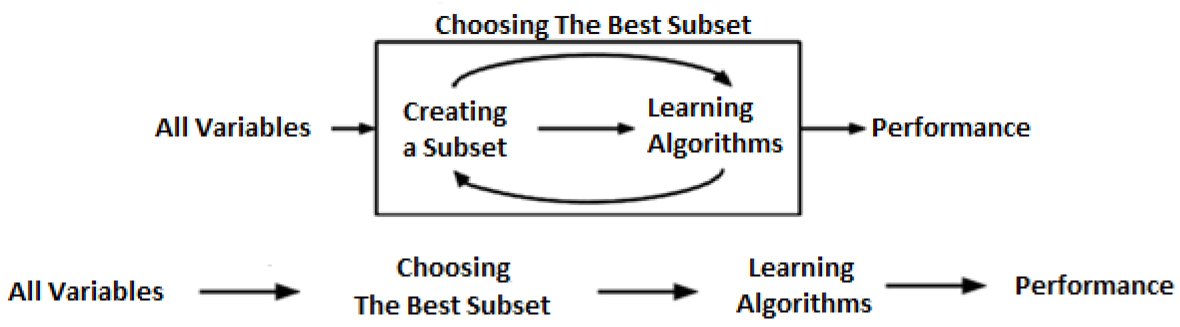

Fig. 2. Filter method process flow chart

Filter methods use a proxy measurement instead of an error rate to score a subset of variables. This measure is used especially to calculate the usefulness of the variable set. Common criteria include mutual information [13], mutual information from a point [14], Pearson correlation coefficient, Relief-based algorithms [15] and in-class distance or significance test scores for each class / variable [16].

Filters generally need to calculate less than wrapper methods, but they produce a set of variables that are not adjusted for a particular prediction model [17]. This lack of adjustment means that a variable adjusted from a filter is more general than that set with a wrapper, and usually gives lower predictive performance than a wrapper. Also the variable set does not contain the assumptions of a prediction model therefore more useful for revealing the relationships between the variables. Many filters provide a variable order instead of a clear best variable subset, and the breakpoint in the order is selected through cross-validation. Filter methods are also used as a pre-processing step for wrapper methods, which allows a wrapper to be used for larger problems. Another popular approach is the Recursive Feature Elimination (RFE) algorithm, which is often used with the Support Vector Machine (SVM) to create a pattern and remove low-weight variables [18].

2.3. Embedded Method

Recently embedded methods have been proposed that try to combine the advantages of both methods. A learning algorithm takes advantage of its own feature selection process and simultaneously performs feature selection and classification.

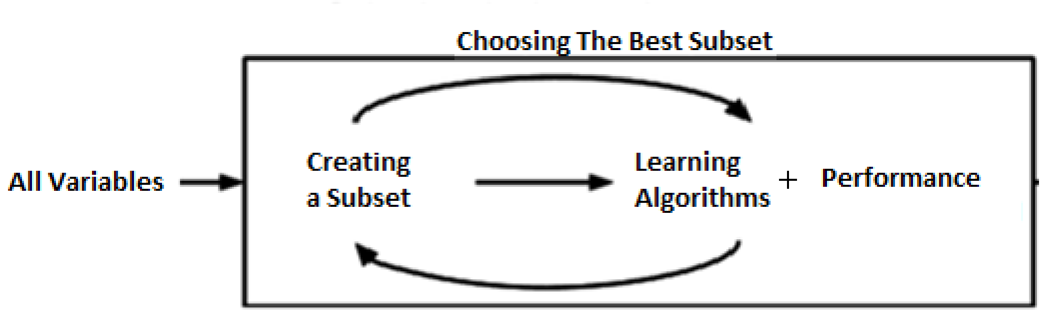

Fig. 3.Process Flow Chart of the Embedded Method

Embedded methods are a group of techniques that cover everything that makes the choice of variables as part of the modeling process. An example of this approach is the LASSO method used to create a linear model that punishes regression coefficients with an L1 penalty and draws most of them to zero. Any variable with non-zero regression coefficient is selected by the LASSO algorithm. Bolasso [19], the Lasso prediction model associated with Bootstrap, which is among the improvements of LASSO, features FeaLect [20], which scored all the features based on the integrative analysis of LASSO's L1 penalty and the L2 penalty of the Ridge Regression, and the integrative analysis of the regression coefficients. In terms of complexity, it is among the filters and the wrapper method.

In traditional regression analysis, the most popular feature selection method is stepwise regression, which is the winding technique. It is an algorithm that adds (or deletes the worst variable) the best variable in each round. The important thing is to decide when to stop the algorithm. In machine learning, this is usually done through cross-validation.

2.4. Subset Selection

Subset selection evaluates a subset of variables as a group for suitability. Wrappers use a search algorithm to search the field of possible variables and evaluate each subset by running a model in the subset. Wrappers can be expensive in calculation and have a risk of memorization. Filters are similar to wrappers in the search approach, but rather than evaluating against a model, a simpler filter is evaluated.

Many popular search approaches use the greedy hill climbing feature, which repeatedly evaluates the candidate subset of features, then replaces the subset and evaluates whether the new subset is an old improvement. Evaluation of subsets requires a scoring measurement that ranks a subset of variables. Comprehensive search is often impractical, so at the stop point set by some practitioner (or operator), the variable subset with the highest score discovered up to that point is selected as the satisfactory subset. Stop criterion varies by algorithm. Alternative search-based techniques are based on the search for targeted projections that find low-dimensional projections of high-scoring data, then variables with the largest projections in low-dimensional space are selected.

TABLE I: FEATURE SELECTION ALGORITHMS

Search Methods	Other Available Filter Measurements
Exhaustive Search	Feature selection Base on Consistency
Best First	Feature selection Based on Correlation
Simulated Annealing	Class Separability • Posibility of Error • Distance Between Classes • Probability Distance • Entropy
Genetic Algorithm [21]	
Forward Selection [22]	
Backward Elimination	
Particle Swarm Optimization [23]	
Targeted Projection Pursuit	
Scatter Search [24]	
Variable Neighborhood Search [25, 26]	

TABLE II: SOME FEATURE SELECTION METHODS AND ALGORITHMS USED IN HE LITERATURE

Subject	Algorithm	Method	Classifier	Source
SNPs	Genetic Algorithm	Wrapper	Decision Tree	[27]
SNPs	Hill Climbing	Filtre + Wrapper	Naive Bayesian	[28]
SNPs	Simulated Annealing	Wrapper	Naive bayesian	[29]
Marketing	Simulated Annealing	Wrapper	Regression	[30]
Economy	Simulated Annealing, Genetic Algorithm	Wrapper	Regression	[31]
Spectral Mass	Genetic Algorithm	Wrapper	Multiple Linear Regression, Partial Least Squares	[32]
Spam	Binary PSO + Mutation	Wrapper	Decision tree	[23]
Microarray	Tabu Search + PSO	Wrapper	Support Vector Machine, K Nearest Neighbors	[33]
Microarray	PSO + Genetic Algorithm	Wrapper	Support Vector Machine	[34]
Microarray	Genetic Algorithm + Iterated Local Search	Embedded	Support Vector Machine	[35]
Microarray	Iterated Local Search	Wrapper	Regression	[36]

Microarray	Genetic Algorithm	Wrapper	K Nearest Neighbors	[37]
Microarray	Hybrid Genetic Algorithm	Wrapper	K Nearest Neighbors	[38]
Microarray	Genetic Algorithm	Wrapper	Support Vector Machine	[39]
Microarray	Genetic Algorithm	Wrapper	All paired Support Vector Machine	[40]
Microarray	Genetic Algorithm	Embedded	Support Vector Machine	[41]
Microarray	Genetic Algorithm	Hybrid	Support Vector Machine	[42]
Microarray	Genetic Algorithm	-	Support Vector Machine	[43]
Microarray	Genetic Algorithm	Wrapper	Support Vector Machine	[44]
Alzheimer's disease	Welch's t-test	Filtre	Kernel Support Vector Machine	[45]
Computer vision	Infinite Feature Selection	Filtre	Independent	[46]
Microarrays	Eigenvector Centrality FS	Filtre	Independent	[47]

2.5. Performance Evaulation

The Confusion Matrix contains information about the actual and predicted classifications made by a classification system. The performance of such systems is usually evaluated using the data in the matrix [48]. Accuracy, ACC, and F1 Score, which are widely used in our study, are a method of success assessment.

TABLE III: CONFUSION MATRIX

		Real Result		
		1	0	Success (%)
Predict	1	TP	FP	Precision Score
	0	FN	TN	Negative Predictive Value (NPV)
		Recall Score, Sensitivity	Specificity	Accuracy (ACC)

ACC = (TP + TN) / (TP + TN + FP + FN)
Precision = TP / (TP + FP)
NPV = TN / (TP + FP)
Recall = TP / (TP + FN)
Specificity = TN / (FP + TN)
F1 Score = 2(Precision*Recall)+ (Precision+Recall)

The ROC (receiver operating characteristic) curve is a graphical representation of the diagnostic capability of the binary classification system. AUC (Area Under Curve) shows the classification performance of the installed model and takes a value between 0 and 1. AUC value close to 1 means that the classification performance of the model is high [49].

III. MATERIAL AND METHOD

Machine learning algorithms were used to determine the factors affecting the sales of vehicles using the 82 variable data set of 2549 used cars between 01/06 / 2015-01 / 06/2019 on a famous website for used car sales. During this process, the performances of the prediction models, which are established by using Backward Elimination, Stepwise Selection, Forward Selection and Feature Selector methods, which are among the feature selection methods, were compared. The data set has undergone data preprocessing steps, such as SMOTE, which is a stratified sampling and data replication technique to prevent missing observation analysis, correlation analysis and data imbalance. The data ready for machine learning was modeled with 4 different feature selection methods and success rates were compared.

3.1. Backward Elimination Method

While determining whether the variables affect the dependent variable, it performs elimination according to the calculated p values (threshold = 0.05). In each iteration, the variable with the highest p value is eliminated and this iteration process continues until all the ones with p value greater than 0.05 are eliminated.

3.2. Forward Selection Method

While determining whether the variables affect the dependent variable, it performs the selection process according to the calculated p values (threshold = 0.05). In each iteration, the variable with the smallest p value is selected, and this iteration process continues until all of those with a p value less than 0.05 are selected.

3.3. Stepwise Selection Method

Forward-Backward mixture method. While determining whether the variables affect the dependent variable, it performs selection by Backward and selection by Forward according to the calculated p values (threshold = 0.05). In each iteration, the variable with the largest p value is eliminated, the variable with the smallest p value is selected, and this iteration process continues until all the ones with p value greater than 0.05 are eliminated.

IV. RESULTS

30 of the 82 variables in the raw data correlated with each other. Feature Selection methods were applied to the remaining 52 variables.

The number of variables after applying the Backward Elimination Method is 28. The following variables have been found to affect the sales status of the vehicle:

Starting price, Model year, Gear Type, Fuel Type, 2 Vehicle Keys Available (Total), Hill Start Assist, Electric Mirror, Fog lights, Cruise control, Functional Steering, Headlight Washer, Glass Ceiling - Sky Dome, Floor, Radio with CD Player, Mapped Navigation Device, Radio - Cassette Player, Central locking, Bonnet, Left Front Fender, Left Front Door, Left Rear Door, Left Rear Fender, Luggage, Right Front Fender, Right Front Door, Right Rear Door, Right Rear Fender, Ceiling.

The number of variables after applying the Forward Selection Method is 24. The following variables have been found to affect the sales status of the used vehicle:

Model year, Starting price, 2 Vehicle Keys Available (Total), Hill Start Assist, Parking Assistant, Bonnet, Fuel Type, Gear Type, Fog lights, Cruise control, Heated Seat, Left Front Fender, Mapped Navigation Device, Radio with CD Player, Right Front Fender, Headlight Washer, Left Rear Fender, Electric Mirror, Glass Ceiling - Sky Dome, Floor, Functional Steering, Arrow Display Navigation Device, Right Front Door, Left Front Door.

The number of variables after applying the Stepwise Selection Method is 21. The following variables have been found to affect the sales status of the used vehicle:

Model year, Starting price, 2 Vehicle Keys Available (Total), Hill Start Assist, Bonnet, Fuel Type, Gear Type, Fog lights, Cruise control, Left Front Fender, Mapped Navigation Device, Radio with CD Player, Right Front Fender, Headlight Washer, Left Rear Fender, Electric Mirror, Glass Ceiling - Sky Dome, Functional Steering, Floor, Right Front Door, Left Front Door.

In the data set where 3 different feature selection methods are applied, the algorithm that gives the highest success is given in the table below. The data set is divided into two as 70-30% education and test data. The number of units in the training data is 985 and 423 in the test data.

TABLE IV: FEATURESELECTION APPLICATION RESULTS

	Backward Elimination Method		Forward Selection Method		Stepwise Selection Method	
Algorithm	XGBClassifier		XGBClassifier		RandomForestClassifier	
Separation Rate	80/20					
Education Data	2039					
Test Data	510					
Accuracy	0,817647059		0,837254902		0,833333333	
Confusion Matrix	Real Result					
	1	0	1	0	1	0
Prediction Status 1	102	55	116	41	110	47
Prediction Status 0	38	315	42	311	38	315
Result 1 (Recall)	0.8924		0.8810		0.8924	
Result 0	0.6500		0.7400		0.7890	
Precision	0.8514		0.8835		0.8702	
F1 score	0.8714		0.8823		0.8811	
Roc Auc Score	0.7710		0.8099		0.7965	

Note: 0; “no sale”, 1; “sold”.

V. CONCLUSION

While testing the feature selection methods in the application, after using all relevant machine learning algorithms, the result of the algorithm with the highest results was taken into consideration. According to the results of the application, the success rate of the Stepwise Selection method was slightly higher than the Forward Selection and Feature Selector methods, and a significantly higher success rate than the Backward Elimination method. In addition, the highest success rate has been achieved with the Support Vector Classifier algorithm, but this rate is affected not only by the algorithm but also by the method of selecting variables. Therefore, it should not be overlooked that feature selection methods can give different results with different data sets and algorithms. While conducting such studies, the methods with the highest success should be preferred by trying more than one feature selection method.

REFERENCES

[1]. James, G., Witten, D., Hastie, T., &Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.

[2]. Bermingham, M. L., Pong-Wong, R., Spiliopoulou, A., Hayward, C., Rudan, I., Campbell, H., ...& Haley, C. S. (2015). Application of high-dimensional feature selection: evaluation for genomic prediction in man. Scientific reports, 5, 10312.

[3]. Jain, A., &Zongker, D. (1997). Feature selection: Evaluation, application, and small sample performance. IEEE transactions on pattern analysis and machine intelligence, 19(2), 153-158.

[4]. Hall, M. A. (1999). Correlation-based feature selection for machine learning.

[5]. Hall, M. A. (2000). Correlation-based feature selection of discrete and numeric class machine learning.

[6]. Weston, J., Mukherjee, S., Chapelle, O., Pontil, M., Poggio, T., &Vapnik, V. (2001). Feature selection for SVMs. In Advances in neural information processing systems (pp. 668-674).

[7]. Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on Knowledge & Data Engineering, (4), 491-502.

[8]. Uğuz, H. (2011). A two-stage feature selection method for text categorization by using information gain, principal component analysis and genetic algorithm. Knowledge-Based Systems, 24(7), 1024-1032.

[9]. Chandrashekar, G., &Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.

[10]. Barlaud, M., Chambolle, A., &Caillau, J. B. (2019). Robust supervised classification and feature selection using a primal-dual method. arXiv preprint arXiv:1902.01600.

[11]. Phuong, T. M., Lin, Z., & Altman, R. B. (2005, August). Choosing SNPs using feature selection. In 2005 IEEE Computational Systems Bioinformatics Conference (CSB'05) (pp. 301-309). IEEE.

[12]. Hamon, J. (2013). Optimisationcombinatoire pour la sélection de variables enrégressionengrande dimension: Application engénétiqueanimale (Doctoral dissertation, Université des Sciences etTechnologie de Lille-Lille I).

[13]. Guyon, I., &Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3(Mar), 1157-1182.

[14]. Yang, Y., & Pedersen, J. O. (1997, July). A comparative study on feature selection in text categorization. In Icml (Vol. 97, No. 412-420, p. 35).

[15]. Urbanowicz, R. J., Meeker, M., La Cava, W., Olson, R. S., & Moore, J. H. (2018). Relief-based feature selection: introduction and review. Journal of biomedical informatics.

[16]. Forman, G. (2003). An extensive empirical study of feature selection metrics for text classification. Journal of machine learning research, 3(Mar), 1289-1305.

[17]. Zhang, Y., Li, S., Wang, T., & Zhang, Z. (2013). Divergence-based feature selection for separate classes. Neurocomputing, 101, 32-42.

[18]. Guyon, I., Weston, J., Barnhill, S., &Vapnik, V. (2002). Gene selection for cancer classification using support vector machines. Machine learning, 46(1-3), 389-422.

[19]. Bach, F. R. (2008, July). Bolasso: model consistent lasso estimation through the bootstrap. In Proceedings of the 25th international conference on Machine learning (pp. 33-40). ACM.

[20]. Zare, H., Haffari, G., Gupta, A., & Brinkman, R. R. (2013, January). Scoring relevancy of features based on combinatorial analysis of lasso with application to lymphoma diagnosis. In BMC genomics (Vol. 14, No. 1, p. S14). BioMed Central

[21]. Soufan, O., Kleftogiannis, D., Kalnis, P., &Bajic, V. B. (2015). DWFS: a wrapper feature selection tool based on a parallel genetic algorithm. PloS one, 10(2), e0117988.

[22]. Figueroa, A. (2015). Exploring effective features for recognizing the user intent behind web queries. Computers in Industry, 68, 162-169.

[23]. Zhang, Y., Wang, S., Phillips, P., & Ji, G. (2014). Binary PSO with mutation operator for feature selection using decision tree applied to spam detection. Knowledge-Based Systems, 64, 22-31.

[24]. López, F. G., Torres, M. G., Batista, B. M., Pérez, J. A. M., & Moreno-Vega, J. M. (2006). Solving feature subset selection problem by a parallel scatter search. European Journal of Operational Research, 169(2), 477-489.

[25]. García-Torres, M., Garcıa-López, F., Melián-Batista, B., Moreno-Pérez, J. A., & Moreno-Vega, J. M. (2004). Solving feature subset selection problem by a hybrid metaheuristic. Hybrid Metaheuristics, 59-68.

[26]. García-Torres, M., Gómez-Vela, F., Melián-Batista, B., & Moreno-Vega, J. M. (2016). High-dimensional feature selection via feature grouping: A Variable Neighborhood Search approach. Information Sciences, 326, 102-118.

[27]. Shah, S. C., &Kusiak, A. (2004). Data mining and genetic algorithm based gene/SNP selection. Artificial intelligence in medicine, 31(3), 183-196.

[28]. Long, N., Gianola, D., Rosa, G. J. M., &Weigel, K. A. (2011). Dimension reduction and variable selection for genomic selection: application to predicting milk yield in Holsteins. Journal of Animal Breeding and Genetics, 128(4), 247-257.

[29]. Üstünkar, G., Özöğür-Akyüz, S., Weber, G. W., Friedrich, C. M., & Son, Y. A. (2012). Selection of representative SNP sets for genome-wide association studies: a metaheuristic approach. Optimization Letters, 6(6), 1207-1218.

[30]. Meiri, R., &Zahavi, J. (2006). Using simulated annealing to optimize the feature selection problem in marketing applications. European Journal of Operational Research, 171(3), 842-858.

[31]. Kapetanios, G. (2005). Variable selection using non-standard optimisation of information criteria (No. 533). Working Paper.

[32]. Broadhurst, D., Goodacre, R., Jones, A., Rowland, J. J., &Kell, D. B. (1997). Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. AnalyticaChimicaActa, 348(1-3), 71-86.

[33]. Chuang, L. Y., Yang, C. H., & Yang, C. H. (2009). Tabu search and binary particle swarm optimization for feature selection using microarray data. Journal of computational biology, 16(12), 1689-1703.

[34]. Alba, E., Garcia-Nieto, J., Jourdan, L., &Talbi, E. G. (2007, September). Gene selection in cancer classification using PSO/SVM and GA/SVM hybrid algorithms. In 2007 IEEE Congress on Evolutionary Computation (pp. 284-290). IEEE.

[35]. Duval, B., Hao, J. K., & Hernandez Hernandez, J. C. (2009, July). A memetic algorithm for gene selection and molecular classification of cancer. In Proceedings of the 11th Annual conference on Genetic and evolutionary computation (pp. 201-208). ACM.

[36]. Hans, C., Dobra, A., & West, M. (2007). Shotgun stochastic search for “large p” regression. Journal of the American Statistical Association, 102(478), 507-516.
[37]. Jirapech-Umpai, T., & Aitken, S. (2005). Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes. BMC bioinformatics, 6(1), 148.
[38]. Oh, I. S., Lee, J. S., & Moon, B. R. (2004). Hybrid genetic algorithms for feature selection. IEEE Transactions on pattern analysis and machine intelligence, 26(11), 1424-1437.
[39]. Xuan, P., Guo, M. Z., Wang, J., Wang, C. Y., Liu, X. Y., & Liu, Y. (2011). Genetic algorithm-based efficient feature selection for classification of pre-miRNAs. Genetics and molecular research, 10(2), 588-603.
[40]. Peng, S., Xu, Q., Ling, X. B., Peng, X., Du, W., & Chen, L. (2003). Molecular classification of cancer types from microarray data using the combination of genetic algorithms and support vector machines. FEBS letters, 555(2), 358-362.
[41]. Hernandez, J. C. H., Duval, B., &Hao, J. K. (2007, April). A genetic embedded approach for gene selection and classification of microarray data. In European Conference on Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics (pp. 90-101). Springer, Berlin, Heidelberg.
[42]. Huerta, E. B., Duval, B., &Hao, J. K. (2006, April). A hybrid GA/SVM approach for gene selection and classification of microarray data. In Workshops on Applications of Evolutionary Computation (pp. 34-44). Springer, Berlin, Heidelberg.
[43]. Muni, D. P., Pal, N. R., & Das, J. (2006). Genetic programming for simultaneous feature selection and classifier design.
[44]. Vermeulen-Jourdan, L., Dhaenens, C., &Talbi, E. G. (2005). Linkage disequilibrium study with a parallel adaptive GA. International Journal of Foundations of Computer Science, 16(02), 241-260.
[45]. Zhang, Y., Dong, Z., Phillips, P., Wang, S., Ji, G., Yang, J., & Yuan, T. F. (2015). Detection of subjects and brain regions related to Alzheimer's disease using 3D MRI scans based on eigenbrain and machine learning. Frontiers in computational neuroscience, 9, 66.
[46]. Roffo, G., Melzi, S., &Cristani, M. (2015). Infinite feature selection. In Proceedings of the IEEE International Conference on Computer Vision (pp. 4202-4210).
[47]. Roffo, G., &Melzi, S. (2016). Features selection via eigenvector centrality. Proceedings of new frontiers in mining complex patterns (NFMCP 2016).
[48]. Provost, F., &Kohavi, R. (1998). Guest editors' introduction: On applied research in machine learning. Machine learning, 30(2): 127-132.
[49]. Çelik, Ö.,&Osmanoğlu, U. Ö. (2019). Comparing to Techniques Used in Customer Churn Analysis. Journal of Multidisciplinary Developments, 4(1), 30-38.

Ozer Celik."Examination of Feature Selection Methods and an Application."*American Journal of Engineering Research (AJER),* vol. 9(04), 2020, pp. 33-40.