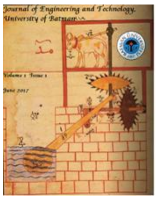


Journal of Engineering and Technology

DergiPark AKADEMİK

Journal of Engineering and Tecnology 1 (2017) 16-22

A Basic and Brief Scheme of an Application of a Machine Learning Process

Ömer Faruk Ertuğrul[1], Mehmet Emin Tağluk[2], Yılmaz Kaya[3]

[1]Batman University, Department of Electrical and Electronics Engineering, 72060, BATMAN/TURKEY

[2]İnönü University, Department of Electrical and Electronics Engineering, 44000, MALATYA/TURKEY

[3]Siirt University, Department of Computer Engineering, 56000, SİİRT/TURKEY

ARTICLE INFO

Article history:
Received 13 December 2017
Received in revised form
27 December 2017
Accepted 30 December 2017
Available online 31 December 2017

Key words:
Machine learning
Feature extracting
Feature selection
Cross-validation

* Corresponding author.
E-mail address:
omerfarukertugrul@gmail.com

ABSTRACT

Machine learning methods are powerful tools in modeling systems or extracting knowledge about a phenomenon from samples. This paper is written in order to make the process of machine learning clearer. To employ a machine learning method, first the features of a sysytem, a phenomenon or a dataset must be exacted. Determining relevant features in a system is still an open issue. Later, relevant features can be determined by a feature selection method. Using a feature selection method may be caused to increase the accuracy of the system. Finally, optimum machine learning method must be determined. This stage is really hard, a machine learning method may be employed uniquely or combined together. Later, the Highleyman dataset was employed in tests in machine learning stages.

1. Introduction

In classical science, a system can be modeled (developing an equation) after many and many experiments. Furthermore, before the experiments, the researcher(s) must have an idea about which physical parameters are relevant to the output of the system. After time consuming and expensive experiments, developed equation can be only defined for some specific conditions.

An alternative solution of these hard process is using a machine learning (ML) method, which is popular research and application concept nowadays [1, 2, 3, 4, 5]. By ML methods, a system or a phenomenon can be modeled by using a few samples. Machine learning is defined as follows.

• Herbert Simon: “Any change in a System that allows it to perform better the second time on repetition of the same task or on another task drawn from the same population.” [6]

• Tom Dietterich: “The goal of machine learning is to build computer systems that can adapt and learn from their experience.”

• Arthur Samuel: “Field of study that gives computers the ability to learn without being explicity programmed”

• Tom Mitchell: “Well-posed Learning Problem: A computer program is said to learn from experience E with respect to some task T, and some performance measure P, if this performance on T as measured by P improves of experience E”

- Computing Dictionary: “The ability of a machine to improve its performance based on previous results”

In ML, a system or a phenomenon is learned. Webster defined “to learn” as follows: “To gain knowledge or understanding of, or skill in by study, instruction or experience". It looks really meaningful and easy to use machines (for example computers). The aim behind writing this paper is to give a brief and frank way to employ a ML method.

2. Materials and Methods

The methodology of using a ML method is simplified and visualized in Figure 1.

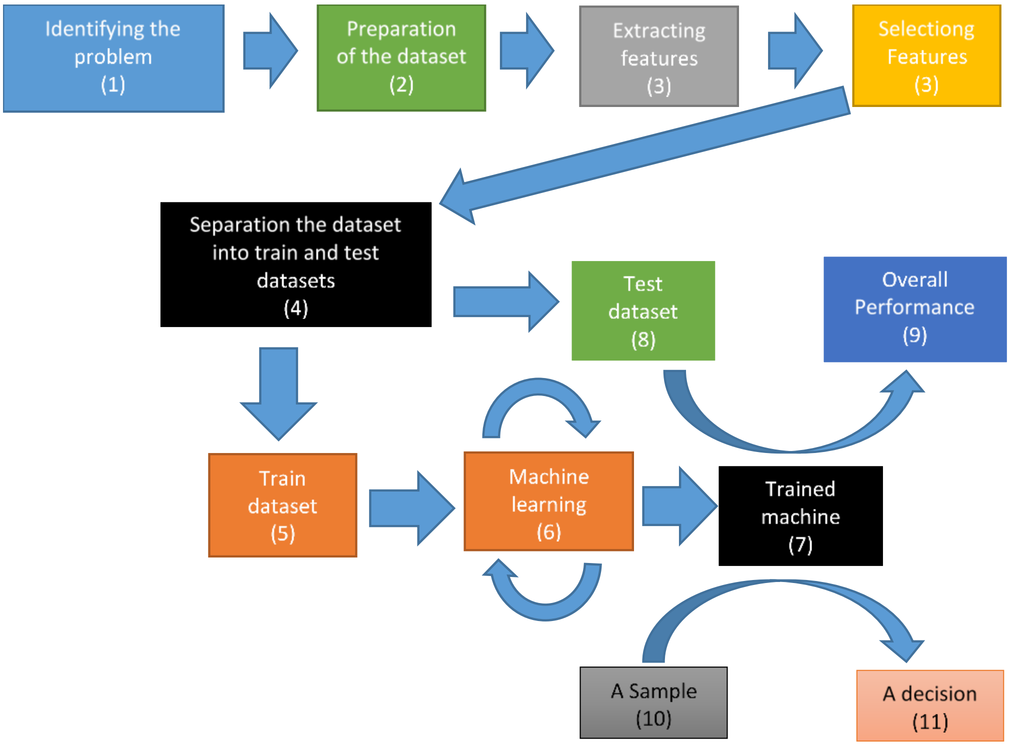


Figure 1. General methodology of employing ML

As seen from Figure 1, in order to achieve successful results, there are some preprocessing stages. The first one is to identify the problem (see Figure 1). It is not true to employ ML in the problems that the relationship between inputs and outputs is clear [7, 8]. Because calculating the outputs via a relationship is easier than employing a ML. After identifying the issue, a dataset must be prepared (see Figure 1). A dataset is a combination of results of experiments. Each phenomenon can be called as a system, which produces some outputs for particular inputs as shown in Figure 2.

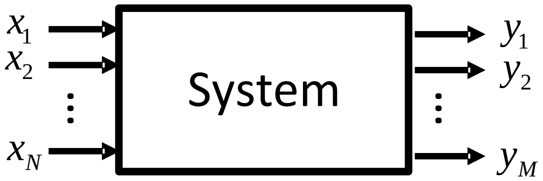


Figure 2. A system

After each experiment, inputs ($x_{1 \dots N}$) and outputs ($y_{1 \dots M}$) must be recorded. Each x_i is a feature of a sample and each y_i is an output. The outputs and inputs must be defined clearly and logically. The inputs may be relevant to the outputs. By ML methods, a relationship between these inputs and the outputs can be established [9].

In preparation of the dataset stage, the samples that recorded in experiments or the datasets that shared on the internet can be used such as UCI [10, 11]. The samples in the datasets must be checked before using them. The samples must be reliable and relevant; furthermore, the samples that are repetitive, noisy, irrelevant or redundant must be discarded. For some cases, the missed samples must be completed via some special methods [12, 13] or outlier samples must be detected [14, 15, 16].

Due to the developed technology, the capacity of the storage, the capability of communication and the dimension and mass of the dataset are all increased. Therefore, instead of using a big dataset, extracting features and using them in ML process causes faster and more stable solutions [17, 18, 19, 20]. There are many feature extraction methods in the literature and it is still an open issue. But the most popular ones are statistical properties for signals in both time and frequency domains and texture detection methods in images.

As it was described before, there are many feature extracting methods and it is hard to know whether the exacted features are relevant or not. Therefore, feature selection methods have been employed to determine relevant ones and discard irrelevant, redundant and noisy features [21, 22, 23]. It was clear from the literature that using only selected features increases obtained success of a ML method [24]. There are basically three types of feature selection algorithms, which are the filter, wrapper and embedded feature selection methods [25, 26].

In order to obtain a fair success rate, the samples that used in training must not be used in tests [27]. Additionally, each sample must be used in test dataset in different folds. Three different basic cross-validation scheme have been proposed in the literature, which are leave-one-out [28, 29, 30, 31], n-folds cross-validation [30, 31, 32] and Monte Carlo cross-validation scheme [33].

After the dataset is divided into two partitions: training and test datasets, a ML method must be employed. The ML methods can be employed in classification, regression, clustering and reinforcement problems [34, 35, 36, 37, 38]. There are many and many numbers of ML methods in the literature [7, 8, 39]. It is really hard to determine the optimal ML method for each of the datasets. Therefore, the optimal ML method must be tested for each dataset. Finally, there are some tools that can be used in ML process such as WEKA, Dlib-ml and machine learning toolkit on the internet [8, 40, 41, 42].

3. Results and Discussion

In order to apply a ML method, Highleyman dataset, which was generated by Prtools, was employed in the evaluation and validation of the process. Each test was employed based on 5-folds cross-validation scheme. The Highleyman dataset was illustrated in Figure 3.

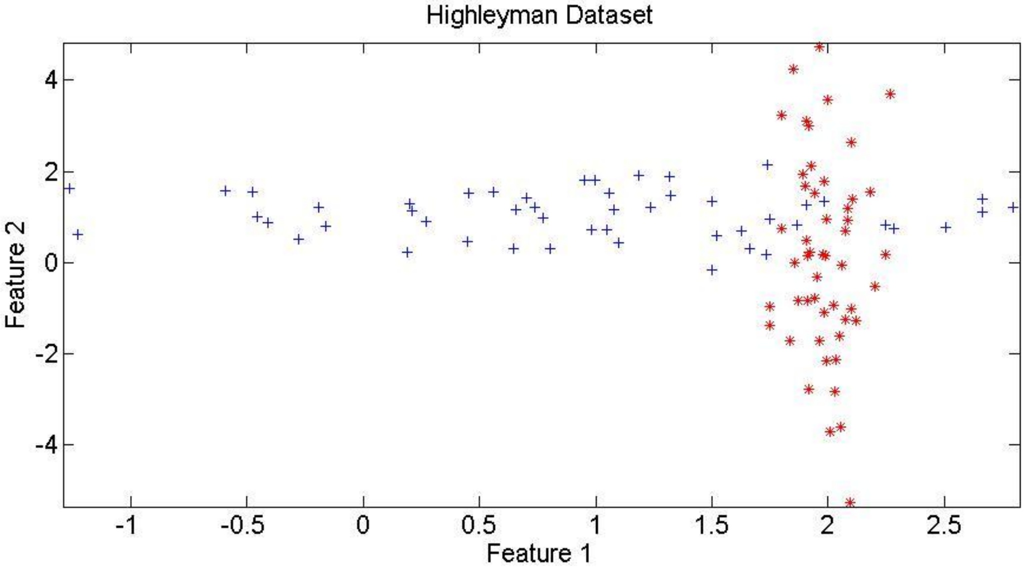


Figure 3. Highleyman Dataset

As seen in Figure 3, in the Highleyman Dataset, some samples can easily be distinguished on the other hand classifying a part of this dataset, which is in the intersection of two classes, is really hard. Obtained success rates by k nearest neighbors (kNN), Naïve Bayes (NB) and artificial neural network (ANN) in this dataset are given in Table 1.

Table 1. Obtained success rates in Highleyman Dataset

kNN	**# of Nearest Neighbors**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
	Accuracy (%)	91.6	86.7	90.0	83.3	90.0	85.0	83.3	83.3	83.3	75.0
NB	**Bin**	2	4	6	8	10	12	14	16	18	20
	Accuracy (%)	73.3	81.7	93.3	88.3	91.7	90.0	90.0	90.0	90.0	88.3
ANN	**Hidden Layer**	[1]	[2]	[5]	[10]	[20]	[5 5]	[5 10 5]	[5 10 10 5]	[5 10 20 10 5]	[5 10 20 20 10 5]
	Accuracy (%)	88.3	73.3	93.3	96.7	95.0	93.3	90.0	91.7	93.3	86.7

As seen from the results given in Table 1, the parameters of the employed methods must be determined. These parameters are generally determined by trials. Obtained regions for each class by the employed kNN, NB, and ANN are given in Figure 4.

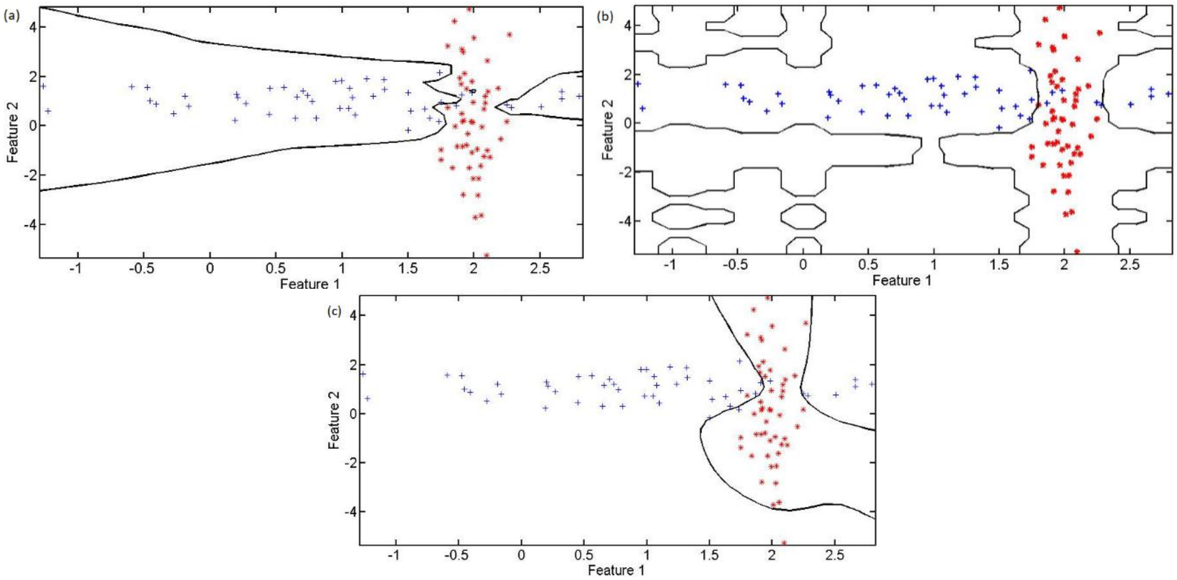
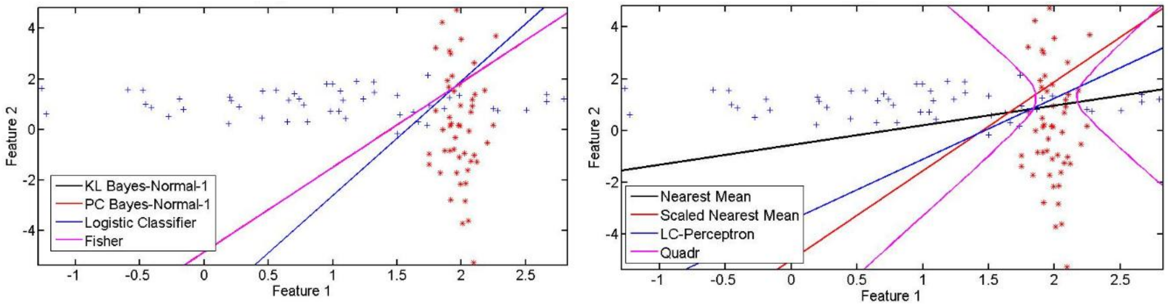


Figure 4. Obtained classification regions by kNN, (b) Naïve Bayes, and (c) ANN

As seen from Figure 4, the regions of the classes were changed according to employed method. Better classifying regions yields higher accuracies. In order to show the regions some other ML methods were employed and obtained accuracies by these methods are given in Table 3 and the regions are in Figure 5.

Figure 5. Obtained classification regions by some other popular methods



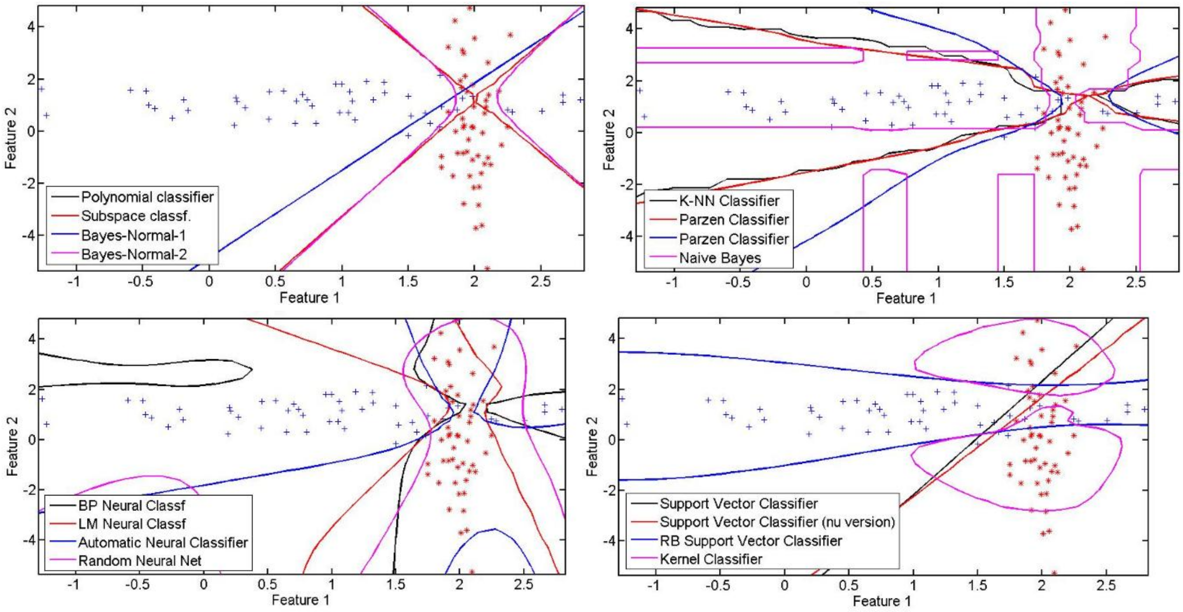


Table 3. Obtained Accuracies (%) in Highleyman Dataset

ML Method	Accuracy	ML Method	Accuracy	ML Method	Accuracy
KL Bayes-Normal-1	76.7	Polynomial	76.7	BP Neural	91.7
PC Bayes-Normal-1	76.7	Subspace .	91.7	LM Neural	91.7
Logistic Regression	80.0	Bayes-Normal-1	76.7	Automatic Neural	93.3
Fisher	76.7	Bayes-Normal-2	93.3	Random Neural Net	95.0
Nearest Mean	73.3	kNN	90.0	Support Vector	98.0
Scaled Nearest Mean	78.3	Parzen-1	90.0	nu-Support Vector	76.7
LC-Perceptron	76.7	Parzen-2	91.7	RB Support Vector	81.7
Quadr	93.3	Naive Bayes	91.7	Kernel	78.3

As seen from Table 3, the success of a method does not depend on the complexity of the method, it is highly related with the structure or geometric complexity of the dataset. Furthermore, these ML methods can be combined in order to achieve higher accuracies. Obtained classification regions by combined ML methods are given in Figure 6 and achieved accuracies are given in Table 4.

Figure 6. Obtained classification regions by combining some other popular methods

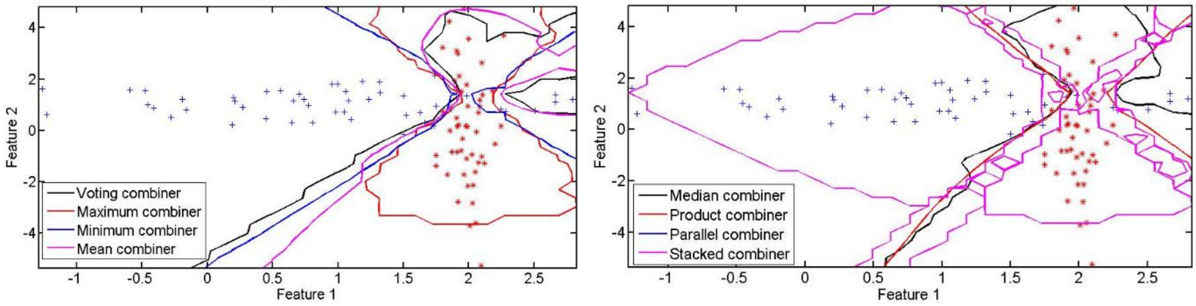


Table 4. Obtained Accuracies (%) by combined ML methods in Highleyman Dataset

Combined ML Method	Accuracy (%)	Combined ML Method	Accuracy (%)
Voting combiner	86.7	Median combiner	88.3
Maximum combiner	90.0	Product combiner	91.7
Minimum combiner	91.7	Parallel combiner	90.0
Mean combiner	91.7	Stacked combiner	90.0

It can be seen from Table 3 and 4 that combining methods do not guarantiee the success of ML methods. Furthermore, it must be noted that determining the optimal ML method and its optimal parameters are really hard stage and there is not still a simple way to solve this complexity and determine them simply.

4. Conclusions

This paper was written in order to give an example of a ML process. Each stage in ML process was described briefly and there are still some open issues in each of these stages. Later, Highleyman dataset was employed in order to show that it is really hard to know or predict which ML method is optimal for a dataset. But better ML methods can be predicted based on the characteristics of the dataset.

References

[1] Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. AI magazine, 17(3), 37.

[2] Yang, Q., & Wu, X. (2006). 10 challenging problems in data mining research. International Journal of Information Technology & Decision Making, 5(04), 597-604.

[3] Rasmussen, C. E., & Williams, C. K. (2006). Gaussian processes for machine learning (Vol. 1). Cambridge: MIT press.

[4] Gorunescu, F. (2011). Data Mining: Concepts, models and techniques (Vol. 12). Springer Science & Business Media.

[5] Liao, S. H., Chu, P. H., & Hsiao, P. Y. (2012). Data mining techniques and applications–A decade review from 2000 to 2011. Expert systems with applications, 39(12), 11303-11311.

[6] Luger, G. F. & Stubblefield, W. A. (1989). Artificial Intelligence: Structures and Strategies for Complex Problem Solving, The Benjamin/Cummings Publishing Company, Inc.

[7] Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2007). Supervised machine learning: A review of classification techniques.

[8] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). Data Mining: Practical machine learning tools and techniques. Morgan Kaufmann.

[9] Larose, D. T. (2014). Discovering knowledge in data: an introduction to data mining. John Wiley & Sons.

[10] Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. Applied Artificial Intelligence, 17(5-6), 375-381.

[11] Refaat, M. (2010). Data preparation for data mining using SAS. Morgan Kaufmann.

[12] Lakshminarayan, K., Harp, S. A., Goldman, R. P., & Samad, T. (1996). Imputation of Missing Data Using Machine Learning Techniques. In KDD (pp. 140-145).

[13] Jerez, J. M., Molina, I., García-Laencina, P. J., Alba, E., Ribelles, N., Martín, M., & Franco, L. (2010). Missing data imputation using statistical and machine learning methods in a real breast cancer problem. Artificial intelligence in medicine, 50(2), 105-115.

[14] Aggarwal, C. C., & Yu, P. S. (2001). Outlier detection for high dimensional data. In ACM Sigmod Record (Vol. 30, No. 2, pp. 37-46). ACM.

[15] Laurikkala, J., Juhola, M., Kentala, E., Lavrac, N., Miksch, S., & Kavsek, B. (2000). Informal identification of outliers in medical data. In Fifth International Workshop on Intelligent Data Analysis in Medicine and Pharmacology (Vol. 1, pp. 20-24).

[16] Hodge, V., & Austin, J. (2004). A survey of outlier detection methodologies. Artificial intelligence review, 22(2), 85-126.

[17] Liu, H., & Motoda, H. (Eds.). (1998). Feature extraction, construction and selection: A data mining perspective (Vol. 453). Springer Science & Business Media.

[18] Mörchen, F. (2003). Time series feature extraction for data mining using DWT and DFT.

[19] Sophian, A., Tian, G. Y., Taylor, D., & Rudlin, J. (2003). A feature extraction technique based on principal component analysis for pulsed eddy current NDT. NDT & e International, 36(1), 37-41.

[20] Guyon, I., & Elisseeff, A. (2006). An introduction to feature extraction. Feature extraction, 1-25.

[21] Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. Artificial intelligence, 97(1), 245-271.

[22] Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent data analysis, 1(1-4), 131-156.

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

[24] Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. bioinformatics, 23(19), 2507-2517.

[25] Das, S. (2001). Filters, wrappers and a boosting-based hybrid for feature selection. In ICML 1, 74-81.

[26] Swiniarski, R. W., & Skowron, A. (2003). Rough set methods in feature selection and recognition. Pattern recognition letters, 24(6), 833-849.

[27] Van der Aalst, W. M. (2011). Data Mining. In Process Mining (pp. 59-91). Springer, Berlin, Heidelberg.

[28] Cawley, G. C., & Talbot, N. L. (2003). Efficient leave-one-out cross-validation of kernel fisher discriminant classifiers. Pattern Recognition, 36(11), 2585-2592.

[29] Cawley, G. C., & Talbot, N. L. (2004). Fast exact leave-one-out cross-validation of sparse least-squares support vector machines. Neural networks, 17(10), 1467-1475.

[30] Meijer, R. J., & Goeman, J. J. (2013). Efficient approximate k- fold and leave- one- out cross-validation for ridge regression. Biometrical Journal, 55(2), 141-155.

[31] Wong, T. T. (2015). Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. Pattern Recognition, 48(9), 2839-2846.

[32] Fushiki, T. (2011). Estimation of prediction error by using K-fold cross-validation. Statistics and Computing, 21(2), 137-146.

[33] Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In Encyclopedia of database systems (pp. 532-538). Springer US.

[34] Dietterich, T. G. (2000). Ensemble methods in machine learning. Multiple classifier systems, 1857, 1-15.

[35] Ikonomakis, M., Kotsiantis, S., & Tampakas, V. (2005). Text classification using machine learning techniques. WSEAS transactions on computers, 4(8), 966-974.

[36] Hyvärinen, A., Gutmann, M., & Entner, D. (2010). Unsupervised Machine Learning.

[37] Han, J., Pei, J., & Kamber, M. (2011). Data mining: concepts and techniques. Elsevier.

[38] Hodeghatta, U. R., & Nayak, U. (2017). Unsupervised Machine Learning. In Business Analytics Using R-A Practical Approach(pp. 161-186). Apress.

[39] Hastie, T., Tibshirani, R., & Friedman, J. (2009). Overview of supervised learning. In The elements of statistical learning (pp. 9-41). Springer New York.

[40] Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G., & Cunningham, S. J. (1999). Weka: Practical machine learning tools and techniques with Java implementations.

[41] King, D. E. (2009). Dlib-ml: A machine learning toolkit. Journal of Machine Learning Research, 10(Jul), 1755-1758.

[42] Quinlan, J. R. (2014). C4. 5: programs for machine learning. Elsevier.