A Comparative Analysis of Classification Algorithms on Diverse Datasets

Muhammad Alghobiri
Management Information Systems Department
King Khalid University, Abha
Saudi Arabia

Abstract—**Data mining involves the computational process to find patterns from large data sets. Classification, one of the main domains of data mining, involves known structure generalizing to apply to a new dataset and predict its class. There are various classification algorithms being used to classify various data sets. They are based on different methods such as probability, decision tree, neural network, nearest neighbor, boolean and fuzzy logic, kernel-based etc. In this paper, we apply three diverse classification algorithms on ten datasets. The datasets have been selected based on their size and/or number and nature of attributes. Results have been discussed using some performance evaluation measures like precision, accuracy, F-measure, Kappa statistics, mean absolute error, relative absolute error, ROC Area etc. Comparative analysis has been carried out using the performance evaluation measures of accuracy, precision, and F-measure. We specify features and limitations of the classification algorithms for the diverse nature datasets.**

Keywords-data mining; classification algorithms; diverse; dataset

I. INTRODUCTION

Due to the evolving of computer science and the fast development and vast usage of World Wide Web and other electronic data, information extraction is a popular research field. Data mining [1, 2] is a significant method to extract information from data. There are various domains of data mining like classification, clustering, anomaly detection, association rule mining, regression, pattern mining, summarization etc. Data mining has even further involved in many other studies like text mining, social network mining, influence mining, sentiment mining etc. Data mining utilizes the information from existing data to examine the result of a specific problem. It analyzes the data that may have been extracted or gathered from any business. Decision makers opt for data mining to take a decision regarding marketing strategies for their products. Data mining interferes data into real-life investigation and can be applied to enhance sales, new product promotion, or product deletion. Classification [3, 4] is one of the main domains of data mining and has extensively been used for various purposes like decision making, weather forecasting, prediction of customers' attitude, prediction of various social risk analysis as well as official tasks, prediction of influential bloggers [5-10] etc. A classification process can easily be separated into two main steps. In the first step, a part

of data known as training data is used and each row or value of training dataset comprises a set of characteristics. Determination of classes is the main target behind classification. The first phase generates the classification model known as classifiers that depict the relationship between characteristics and classes. In the second phase, the classification accuracy of a classification algorithm that has been generated by the first stage is analyzed. There are various classification algorithms, classified into different categories. The category of classification algorithms includes probability based Bayesian algorithms, decision tree based algorithms, neural network based algorithms and kernel-based algorithms. Most classifiers use probability calculations to make class labels, however, accuracy measure has not been a target. Naive Bayes and the C4.5 learning algorithm are alike in predictive accuracy [11-13].

II. RELATED WORK

Classification is a very vast domain in data mining, having received a great deal of exploring over the last few decades. A comparison of four classification algorithms (logistic regression, Naïve Bayes, C4.5 decision tree and nearest neighbor) has been carried out in standard and boosted forms to predict class members for an online community in [14]. The comparison has been evaluated using two performance measures, one is the area under the curve and another is the accuracy in the standard and boosted forms. Analysis conducted depicts very significant differences among base classification algorithms. Several classification techniques have been empirically compared within the analysis of unbalanced credit scoring data sets. Traditional classification techniques such as neural networks, logistic regression and decision trees have been used in order to find the suitability of support vector machines, gradient boosting and random forests for loan default prediction [15]. The use of data mining is also very common in bio-informatics. Authors in [27] emphasized on the importance of rule-based decision trees as a classification method. There are two types of nondeterministic rules in decision tables known as inhibitory rules and bounded rules. In the former, we have decisions on one right side while in the later we can have fewer decisions on the right side. Two classification algorithms of polynomial time complexity established on deterministic and inhibitory decision rules have been compared. Experiments are executed on five practical

data sets from the predictive toxicology domain and mutagenesis. Later experiments performed on a dataset of artificial datasets constructed to evaluate the performance of the classifiers on the basis of different parameters [16]. In [17], different logical analysis methods were compared for hypothetical target classification. This study reveals how pre-processing can protect result confidence and compares the results between multi-quantization, fuzzy and Boolean techniques. Classification algorithms are used for a variety of purposes including spam filtering [18-20], web page ranking calculation for web spam [21], software defect detection [22], text classification [23], music emotion classification [24], feature-based mining of digital images [25], or annual crop classification [26].

The discussion above reveals that all the existing comparative works on classification algorithms have been done on algorithms of the same category. This study is a novel in the sense that algorithms have been selected with a diverse nature and diverse data sets better than those established on deterministic decision rules.

III. RESEARCH METHODOLOGY

A. Selected Classification Algorithms

There are numerous classification algorithms, but we have focused on algorithms of diverse nature, therefore, three different algorithms have been chosen. C4.5 is the famous algorithm that is based on the decision tree algorithm, whereas the Naïve Bayes is a probabilistic algorithm and the Support Vector Machine algorithm (SVM) is a kernel based algorithm. The diversity of the proposed algorithms can lead decision-making to confusion, therefore, in the following, a brief description will be introduced.

Positioning Figures and Tables

Decision tree is also known as a statistical classifier used in classification. The decision tree is produced by C4.5. After a tree is built, the C4.5 rule induction program is used to produce a set of rules. Trees are shown by C4T and rules by C4R. At each node of the tree, one attribute is efficiently split its example set into subsets. Information gain is used for splitting. Attribute with top value of normalized information gain is selected to make the decision. The C4.5 algorithm is recursive on the smaller sub-lists. Decision tree divides the features of the documents into partitions. Splitting of data reduces the chances of errors at every stage. The root node is used to examine the branches of the tree to predict a label for new data [28]. Data is trained in less time because of the graphical representation. We can examine it quickly from root to child nodes in which root node depicts the input. To calculate decision tree we need to calculate two type of entropies.

- Frequency table of one attribute.

$$E(D) = \sum_{i=1}^{m} -(P_i \log_2 P_i) \qquad (1)$$

- Frequency table of two attributes.

$$E(T, X) = \sum_{e \in X} P(e) E(e) \qquad (2)$$

$$Gain(T, X) = Entropy(T) - Entropy(T, X) \qquad (3)$$

Naïve Bayes (NB)

It is a probabilistic algorithm, supposing that some features are not relevant to other features. Naïve Bayes is used for supervised learning methods, parameter estimation for Naive Bayes models etc. In a practical environment, this classifier performed better than others. Different attempts are taken to improve Naïve Bayes for classification [29]. New text is presented with t* in a document A. it calculates:

$$t *= \text{argmax}_t P(A|As) \qquad (4)$$

The Naïve Bayes (NB) classifier uses Bayes' rule:

$$P(t|A) = \frac{P(t)P(A|t)}{P(A)} \qquad (5)$$

P(A) does not select t*. By supposing conditional independence of features f_i' class d, $P(d|t)$ is being estimated. Training procedure calculates the relative frequency of P(t) and $P(f_i|t)$.

$$P(t|A) = P_{NB}(t)(\prod_{i=1}^{m} P(f_i|t^{n_i^{(d)}})) \qquad (6)$$

Support Vector Machine (SVM)

A learning algorithm which is based on the kernel is used in SVM. It is used to identify pattern regression analysis and classification and it performs well in text classification. It predicts classes by using training data. SVM carries out non-linear classification efficiently [30]. Training points are separated into two categories based on support vectors by using decision surface. SVM optimization is calculated as

$$\vec{\beta} = \arg\min\{-\sum_{i=1}^{n} \beta_i + \sum_{i=1}^{n}\sum_{j=1}^{n} \beta_i y_i (x_i x_j)\beta_j y_j\} \quad (7)$$

$$\sum_{i=1}^{n} \beta_i y_i = 0; 0 \le \beta \le C \qquad (8)$$

B. Selected Datasets

Ten datasets have been considered. Details of each dataset is shown in alphabetical order in Table I.

Contact lenses

The examples in the dataset are complete. The attributes are unable to define all the factors affecting the decision.

CPU

It is used for prediction in numeric form on the basis of instance-based learning with encoding length selection.

Credit

This is a credit related data set which is the largest and consists of 15 attributes and results in whether the credit is positive or negative.

Iris-discretized

A small dataset about iris classification. It is unique in this regard as its values are ranging and given in special characters and unique style.

Labor

This dataset is the most unlike one as the attributes in the dataset are of a unique style. Few are having special characters,

few have enumerated values. There are Boolean attributes as well. The attributes are mainly related to wages, pension, allowances, assistance, plan, duration period etc.

Spambase

This dataset is concerning spam and is a very large set of attributes in real format. Usually based on the frequency of words or characters or case-based sentences in various categories, the attributes can be described as derived values. It is unique as its values are in real format and derived attributes based values.

Titanic

This dataset is related to the famous Titanic sinking event and it predicts a person survival level based on its class, age, and gender.

VO

This dataset is regarding Congress voting. It is very interesting in this regard that it has a number of attributes regarding factors influencing voters to vote either for democrats or republicans.

TABLE I.   DATASETS

	Name	Attributes	Instances	Classes	Remarks
1	Contact lens	4	24	2	Values are in nominal form
2	CPU	8	209	1	Class is in real values
3	Credit	16	490	2	Predicts whether positive or negative
4	Iris-discretized	5	150	3	Attributes in Special characters
5	Labor	17	56	2	Attributed in special format, real as well as anime
6	Spambase	58	4600	2	Predicts whether spam or not
7	Titanic	3	2202	2	Calculates whether individual survived or not
8	VO	17	435	2	Political field data from US elections

C. Software Used

Weka workbench provides the facility of visualizing attributes and algorithms for predictive analysis. It was built in C language, now we found Java-based versions only. It supports many tasks related to data mining like clustering, data

preprocessing, classification, feature selection, visualization and regression. The ARFF file consists of two parts: the header and the data section. As the minimum number of attributes in the datasets is 6, thus, this value has been considered as the number of folds for all the algorithms. The following measures are used to find the results of the given classifiers:

Precision

Precision is denoted as the ratio of retrieved documents that are relevant to the search.

$$\text{Precision} = \frac{TP}{TP+FP} \qquad (9)$$

F-Measure

The f-measure includes both precision and accuracy. It may be considered as the weighted average of both values.

$$F = \frac{2 * \text{precision} * \text{Recall}}{\text{precision} + \text{Recall}} \qquad (10)$$

Recall

Recall is also known as the fraction of relevant instances that have been retrieved over the total amount of retrieved instances.

$$\text{Recall} = \frac{TP}{TP+FN} \qquad (11)$$

MCC

MCC is used for measuring the quality of binary classification. It takes into account TP and FP and is generally known as a balanced measure.

$$|\text{MCC}| = \sqrt{x^2/n} \qquad (12)$$

Kappa Statistic

Kappa is the most robust method of measuring inter-rater-agreement for a qualitative item. K takes into account the possibility of the agreement occurring by chance

$$K = \frac{OA-AG}{1-AG} \qquad K = \frac{OA-AC}{1-AC} \dots (14)$$
$$(13)$$

IV.   RESULTS

All datasets have been taken in ARFF file format, which is used in Weka [31, 32] for data mining. The four main diverse nature algorithms have been compared twice. Exclusively, the datasets have been analyzed extensively as shown in Tables II-VII below.

TABLE II.   COMPARATIVE RESULT ANALYSIS FOR NAÏVE BAYES ALGORITHM

Dataset	Accuracy (%)	Kappa	Mean Absolute Error (%)	Relative Absolute Error (%)	Coverage of Cases (%)	Mean -relevant Region Size (%)
Contact lens	70.83	0.43	0.25	67.0	100.0	84.72
Credit	77.55	0.53	0.23	44.94	87.96	59.49
Iris-discretized	94.0	0.91	0.03	7.15	98.66	35.33
Labor	92.98	0.84	0.12	26.25	98.24	65.78
Spambase	79.28	0.59	0.20	43.45	79.78	50.43
TicDate	84.5	0.16	0.36	144.2	91.41	58.63
Titanic	77.87	0.44	0.32	87.83	100.0	99.9
Voting	90.11	0.79	0.09	20.95	93.1	53.2

TABLE III.   WEIGHTED AVERAGE OF DETAILED ACCURACY RESULTS FOR NAÏVE BAYES ALGORITHM

Dataset	Accuracy (%)	Kappa	Mean Absolute Error (%)	Relative Absolute Error (%)	Coverage of Cases (%)	Mean -relevant Region Size (%)
Contact lens	70.83	0.43	0.25	67.0	100.0	84.72
Credit	77.55	0.53	0.23	44.94	87.96	59.49
Iris-discretized	94.0	0.91	0.03	7.15	98.66	35.33
Labor	92.98	0.84	0.12	26.25	98.24	65.78
Spambase	79.28	0.59	0.20	43.45	79.78	50.43
TicDate	84.5	0.16	0.36	144.2	91.41	58.63
Titanic	77.87	0.44	0.32	87.83	100.0	99.9
Voting	90.11	0.79	0.09	20.95	93.1	53.2

TABLE IV.   COMPARATIVE RESULT ANALYSIS FOR C4.5 ALGORITHM

Dataset	Accuracy (%)	Kappa	Mean Absolute Error (%)	Relative Absolute Error (%)	Coverage of Cases (%)	Mean relevant Region Size (%)
Contact lens	83.3	0.71	0.15	0.32	91.0	45.0
Credit	85.3	0.70	0.18	37.8	96.0	86.0
Iris-discretized	94.0	0.91	0.05	13.4	98.0	44.2
Labor	57.9	-0.04	0.46	100.0	92.0	93.0
Spambase	92.9	0.85	0.08	18.6	94.9	55.9
TicDate	93	-0.01	0.11	100.0	99.9	99.9
Titanic	78.9	0.42	0.31	71.3	99.7	96.3
Voting	96	0.92	0.06	12.0	98.6	54.8

TABLE V.   WEIGHTED AVERAGE OF DETAILED ACCURACY RESULTS FOR C4.5 ALGORITHM

Dataset	True Positive Rate	False Positive Rate	Precision	Recall	F-Measure	MCC	ROC Area	RPC Area
Contact lens	0.83	0.09	0.85	0.83	0.83	0.70	0.84	0.81
Credit	0.85	0.15	0.85	0.85	0.85	0.70	0.87	0.83
Iris-discretize	0.94	0.03	0.94	0.94	0.94	0.91	0.96	0.90
Labor	0.57	0.62	0.53	0.57	0.54	-0.05	0.49	0.55
Spambase	0.93	0.078	0.93	0.93	0.93	0.85	0.93	0.91
TicData	0.94	0.94	0.88	0.94	0.91	-0.06	0.49	0.88
Titanic	0.78	0.42	0.82	0.78	0.75	0.50	0.74	0.77
Voting	0.96	0.04	0.96	0.96	0.96	0.92	0.97	0.95

TABLE VI.   COMPARATIVE RESULT ANALYSIS FOR SVM ALGORITHM

Dataset	Accuracy (%)	Kappa	Mean Absolute Error (%)	Relative Absolute Error (%)	Coverage of Cases (%)	Mean relevant Region Size (%)
Contact lens	70.8	0.43	0.31	88.32	87.5	66.6
Credit	86.3	0.72	0.13	27.6	86.3	50.0
Iris-discretized	94	0.91	0.23	53	100	66.6
Labor	85.9	0.68	0.14	30.6	85.9	50.0
Spambase	90.6	0.79	0.09	20.2	90.4	50.0
TicDate	98.9	0.10	0.98	0.98	98	50.0
Titanic	77	0.43	0.22	51	77.6	50.0
Voting	96.0	0.91	0.039	8.2	96	50.0

TABLE VII.   WEIGHTED AVERAGE OF DETAILED ACCURACY RESULTS FOR SVM ALGORITHM

Dataset	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	RPC Area
Contact lens	0.80	0.05	0.80	0.80	0.80	0.74	0.92	0.71
Credit	0.86	0.12	0.87	0.86	0.86	0.73	0.86	0.82
Iris-discretized	0.94	0.03	0.94	0.94	0.94	0.91	0.96	0.90
Labor	0.86	0.19	0.85	0.86	0.85	0.68	0.83	0.80
Spambase	0.90	0.12	0.90	0.90	0.90	0.79	0.89	0.86
TicData	0.98	0.11	0.98	0.98	0.98	0.92	0.89	0.92
Titanic	0.77	0.37	0.77	0.77	0.76	0.45	0.70	0.69
Voting	0.96	0.04	0.96	0.96	0.96	0.91	0.96	0.94

V. COMPARATIVE ANALYSIS

Thorough analysis of the above Tables reveals clear evidence of the three algorithms' effectiveness to different data sets. For comparative explicit evaluation, we have compared the algorithms only with respect to the three most commonly used classification evaluation measures of accuracy, precision, and F-measure. The comparative analysis concludes that SVM outperforms the other classification algorithms. Comparisons on accuracy and F-measure are given in Figures 1 and 2 respectively.
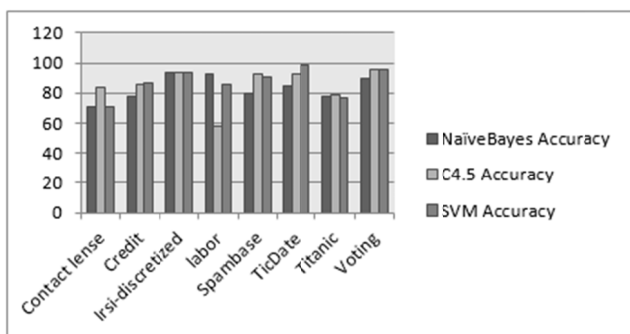
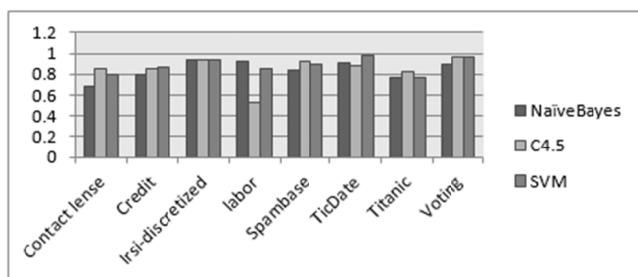Fig. 1.   Comparative analysis based on accuracy

Fig. 2.   Comparative analysis based on F-measure

VI. CONCLUSION

Classification is a main domain of data mining that deals with the prediction of classes based on existing data. There are many classification algorithms, each established on different logic and methods. We have gathered different data sets and have applied three widely used classification algorithms. We have worked extensively and found out various performance measures like accuracy, kappa statistic, mean statistic, precision, recall, F-measure, ROC area, etc. In addition, we carried out the comparative analysis using three state of the art and most widely used performance evaluation measures.

REFERENCES

[1]   N. M. Ramos, J. M. Delgado, R. M. Almeida, M. L. Simoes, S. Manuel, Appliation of Data Mining Techniques in the Analysis of Indoor Hygrothermal Conditions, Springer, 2015

[2]   B. Bakhshinategh, O. R. Zaiane, S. ElAtia, D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years", Education and Information Technologies, Vol. 23, No. 1, pp. 537-553, 2018

[3]   F. Ahmed, M. Samorani, C. Bellinger, O. R. Zaiane, "Advantage of integration in big data: Feature generation in multi-relational databases for imbalanced learning", IEEE International Conference on Big Data, Washington, DC, USA, pp. 532-539, December 5-8, 2016

[4]   P. G. Clark, C. Gao, J. W. Grzymala-Busse, "MLEM2 Rule Induction Algorithm with Multiple Scanning Discretization", Smart Innovation, Systems and Technologies, Vol. 72, pp. 218-227, Springer, 2017

[5]   H. U. Khan, A. Daud, U. Ishfaq, T. Amjad, N. Aljohani, R. A. Abbasi, J. S. Alowibdi, "Modelling to identify influential bloggers in the blogosphere: a survey", Computers in Human Behavior, Vol. 68, pp. 64-82, 2017

[6]   H. U. Khan, A. Daud, T. A. Malik, "MIIB: A Metric to identify top influential bloggers in a community", PloS One, Vol. 10, p. e0138359, 2015

[7]   U. Ishfaq, H. U. Khan, K. Iqbal, "Modeling to find the top bloggers using sentiment features", International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, pp. 227-233, April 11-12, 2016

[8]   U. Ishfaq, H. U. Khan, K. Iqbal, "Identifying the influential bloggers: a modular approach based on sentiment analysis", Journal of Web Engineering, Vol. 16, pp. 505-523, 2017

[9]   H. U. Khan, "Mixed-sentiment classification of web forum posts using lexical and non-lexical features", Journal of Web Engineering, Vol. 16, pp. 161-176, 2017

[10] H. U. Khan, A. Daud, "Using machine learning techniques for subjectivity analysis based on lexical and non-lexical features", International Arab Journal of Information Technology, Vol. 14, No. 4, 2017

[11] A. Patel, S. Gandhi, S. Shetty, B. Tekwani, "Heart Disease Prediction Using Data Mining", International Research Journal of Engineering and Technology, Vol. 4, No. 1, pp. 1705-1707, 2017

[12] T. Pranckevicius, V. Marcinkevicius, "Comparison of Naïve Bayes, Random Forest, Decision Tree, Support Vector Machines, and Logistic Regression Classifiers for Text Reviews Classification", Baltic Journal of Modern Computing, Vol. 5, No. 2, pp. 221-232, 2017

[13] P. V. Ngoc, C. V. T. Ngoc, T. V. T. Ngoc, D. N. Duy, "A C4. 5 algorithm for english emotional classification", in: Evolving Systems, pp. 1-27, Springer Berlin Heidelberg, 2017

[14] C. Sibona, J. Brickey, "A Statistical Comparison of Classification Algorithms on a Single Data Set", AMCIS 2012 Proceedings, pp. 1-13, AIS Electronic Library, 2012

[15] A. Beque, K. Coussement, R. Gayler, S. Lessmann, "Approaches for credit scorecard calibration: An empirical analysis", Knowledge-Based Systems, Vol. 134, pp. 213-227, 2017

[16] N. S. Ketkar, L. B. Holder, D. J. Cook, "Empirical comparison of graph classification algorithms", IEEE Symposium on Computational Intelligence and Data Mining, Nashville, USA, pp. 259-266, March 30-April 2, 2009

[17] R. Dixit, H. Singh, "Comparison of detection and classification algorithms using boolean and fuzzy techniques", Advances in Fuzzy Systems, Vol. 2012, Article No. 406204, 2012

[18] T. R. Patil, V. Thakare, S. Sherekar, "A Combined Naïve Bayes and URL Analysis Based Adaptive Technique for Email Classification", International Journal of Electronics, Communication and Soft Computing Science & Engineering, Special Issue: International Conference on "Advances In Computing, Communication and Intelligence", pp. 88-90, 2014

[19] M. Esmaeili, A. Arjomandzadeh, R. Shams, M. Zahedi, "An Anti-Spam System using Naive Bayes Method and Feature Selection Methods", International Journal of Computer Applications, Vol. 165, No. 4, pp. 1-5, 2017

[20] D. D. Arifin, M. A. Bijaksana, "Enhancing spam detection on mobile phone Short Message Service (SMS) performance using FP-growth and Naive Bayes Classifier", IEEE Asia Pacific Conference on Wireless and Mobile, Bandung, Indonesia, pp. 80-84, September 13-15, 2016

[21] X. Zhuang, Y. Zhu, C.-C. Chang, Q. Peng, F. Khurshid, "A unified score propagation model for web spam demotion algorithm", Information Retrieval Journal, Vol. 20, No. 6, pp. 547-574, 2017

[22] O. F. Arar, K. Ayan, "A Feature Dependent Naive Bayes Approach and Its Application to the Software Defect Prediction Problem", Applied Soft Computing, Vol. 59, pp. 197-209, 2017

[23] L. Jiang, C. Li, S. Wang, L. Zhang, "Deep feature weighting for naive Bayes and its application to text classification", Engineering Applications of Artificial Intelligence, Vol. 52, pp. 26-39, 2016

[24] Y. An, S. Sun, S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics", IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS), Wuhan, China, pp. 635-638, May 24-26, 2017

[25] H. Lad, M. A. Mehta, "Feature Based Object Mining and Tagging Algorithm for Digital Images", in: Proceedings of International Conference on Communication and Networks, Singapore, Advances in Intelligent Systems and Computing, Vol. 508, pp. 345-352, 2017

[26] H. Zhang, Q. Li, J. Liu, J. Shang, X. Du, L. Zhao, N. Wang, T. Dong, "Crop classification and acreage estimation in North Korea using phenology features", GIScience & Remote Sensing, Vol. 54, No. 3, pp. 381-406, 2017

[27] P. Delimata, B. Marszał-Paszek, M. Moshkov, P. Paszek, A. Skowron, Z. Suraj, "Comparison of some classification algorithms based on deterministic and nondeterministic decision rules", in: Transactions on Rough Sets XII, Springer, pp. 90-105, 2010

[28] D. Oreski, S. Oreski, B. Klicek, "Effects of dataset characteristics on the performance of feature selection techniques", Applied Soft Computing, Vol. 52, pp. 109-119, 2017

[29] L. Jiang, D. Wang, Z. Cai, X. Yan, "Survey of improving naive bayes for classification", Lecture Notes in Computer Science, Vol. 4632, Springer, Berlin, Heidelberg, pp. 134-145, 2007

[30] C. Cortes, V. Vapnik, "Support-vector networks", Machine learning, Vol. 20, pp. 273-297, 1995

[31] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "The WEKA data mining software: an update", ACM SIGKDD Explorations, Vol. 11, No. 1, pp. 10-18, 2009

[32] R. R. Bouckaert, E. Frank, M. A. Hall, G. Holmes, B. Pfahringer, P. Reutemann, I. H. Witten, "WEKA-Experiences with a Java Open-Source Project", Journal of Machine Learning Research, Vol. 11, pp. 2533-2541, 2010