International Journal of Information, Control and Computer Sciences
ISSN: 2517-9942
Vol:10, No:6, 2016

Case-Based Reasoning: A Hybrid Classification Model Improved with an Expert's Knowledge for High-Dimensional Problems

Bruno Trstenjak, Dzenana Donko

***Abstract*—**Data mining and classification of objects is the process of data analysis, using various machine learning techniques, which is used today in various fields of research. This paper presents a concept of hybrid classification model improved with the expert knowledge. The hybrid model in its algorithm has integrated several machine learning techniques (Information Gain, K-means, and Case-Based Reasoning) and the expert's knowledge into one. The knowledge of experts is used to determine the importance of features. The paper presents the model algorithm and the results of the case study in which the emphasis was put on achieving the maximum classification accuracy without reducing the number of features.

***Keywords*—**Case based reasoning, classification, expert's knowledge, hybrid model.

I. Introduction

TODAY, when the Internet has become inevitable part of daily life, we are faced with the problem of collecting useful information and extracting valuable knowledge from a large amount of data. Data mining and machine learning are research fields that deal with such issues. Major data mining techniques are regression, clustering and classification. In our study, we dealt with the problem of classification and we wanted to create a hybrid model usable for data classification in the various domains of the problem, a model that will enable to achieve high classification accuracy. In this paper, we present a concept of the classifier that can perform excellent classification results from large datasets. The concept of hybrid classification model shows the manner of merging modern machine learning methods and expert's knowledge.

Classification is the process of determining the origin of the object/instances of a class based on its feature values. The task of classification model is to correctly classify new object/instance. Today, for the purpose of data classification, various methods of machine learning are used, such as Neural Networks [1], Support Vector Machines [2] and Naive Bayes [3].

In order to achieve better accuracy of classification, our study focused on the development of the hybrid model concept. The hybrid classification model consists of several methods from the field of machine learning and provides a slightly different classification approach. This hybrid model merges three machine learning techniques: Information Gain (IG), k-means and Case Based Reasoning (CBR).

One of the objectives in developing a new concept of hybrid model was that in the process of classification, the model uses all the features, without any reduction in their number. For this reason, IG method was implemented in the hybrid model. IG is used for ranking the features based on data entropy and certain statistical criteria [4]. IG method calculates the value of the features information. Value is defined as the amount of information, provided by the feature items for the class. With a ranking of the features, IG method determines their importance in the process of classification. The hybrid model uses the obtained rank values for calculating the similarities between instances.

For the purposes of the clustering process, K-means algorithm is used [5]. K-means is one of the simplest unsupervised learning algorithms used for solving clustering problems. Clustering is the process of dividing data into clusters, grouped on the basis of common properties. K-means algorithm is used to optimize the clustering data and preparing for the classification phase.

The third method in the hybrid model is the classification method. For the purposes of classification, CBR method was used. CBR method uses the methodology of solving new problems based on the previous cases. Therefore, past experience is essential for the CBR method. All collected experiences are written in the form of cases. CBR investigates cases from the past, and on the basis of similarity, the method proposes a solution to the current situation [6].

The remainder of the paper is organized as follows; Section II gives an overview of the work related to hybrid models and expert's knowledge. Section III deals with the concept of hybrid model and the implemented algorithm. Section IV presents achieved results of the case study and Section V concludes the paper.

II. Background

The basic idea for the development of a hybrid model was to create a concept of classification based on merging experience and knowledge with the machine learning methods, to develop an algorithm that will preserve all the characteristics of objects in the process of classification and to achieve high classification accuracy without using any

Bruno Trstenjak, is with Department of Computer Engineering, University of Applied Sciences Cakovec, Croatia, B. J. Jelacica 22a, 40000 Cakovec, Croatia (phone: +385 40 396 990; fax: +385 40 396 980; e-mail: btrstenjak@mev.hr).

Dzenana Donko, is with Department of Computer Science, Faculty of Electrical Engineering, Sarajevo, Bosnia and Herzegovina. (e-mail: ddonko@etf.unsa.ba).

techniques of reducing the number of dimensions and the number of object features. The main classification method in the hybrid model is CBR method. The reason for this selection is that CBR uses the previous knowledge collected in the form of cases, a method that enables upgrading a list of cases during solving a new case and conducting a self-learning process.

A. Expert's Knowledge

The specificity of this approach compared to our previous research is in the use of expert opinions on what determines the quality of services [7]. The selection of features can be performed using various algorithms and achieve the different results. Because of this, it is sometimes useful in the process of deciding include the knowledge and experiences of experts from a particular domain of problem. Often, experiences represent corrective measures when the decision is based on the results obtained by a mathematical method. In our study, a relational database was used. The database contains records that represent the opinion of experts about the importance of the features which determine dataset. The value indicates the importance of each prediction feature based on the opinion of experts; strength impact of features on the final result of classification. Various studies in the classification field showed that the features have different influence on prediction results [8]-[10].

The values have a role of a correction factor in the process of determining a feature weight values by the IG method. The values are in the range from 0-1. Value 0 determines a feature which is not important for the process of classification and is not used. That feature is ignored in the classification process. Value 1 determines a feature which has a maximum impact on the result of classification. The feature must come into the process of classification.

B. CBR

The principle of CBR method is based on solving new problems by observing the similarity with the previously solved problems. The CBR method uses a problem-solving approach analogous to the way of problem solving by man when he draws on his experiences [11]. Each CBR system contains an embedded library of past cases that were resolved in the past. This is something like collecting life experiences in the domain of the problem. Each case represents a description of the problem with its associated solution. CBR method with built-in function of similarities tries to find the most similar case from the library. The retrieved cases from the library are used to suggest a solution. If the proposed solution is not satisfactory, the method tries to revise selected cases and find a new solution. The method adds a new revised case to the cases library and thereby expands the knowledge base. The whole execution cycle of algorithm can be divided into four main steps [12]:

1. Retrieve – retrieval is the first step in the cycle. Algorithm tries to find the best matching case(s). The case that will be selected among the retrieved cases depends on the similarity function which is used.
2. Reuse - If the new problem situation is exactly like the previous one, then algorithm reuses the old case. If the retrieved cases do not offer acceptable solution for a new problem, algorithm performs adaptation of the retrieved cases.
3. Revise - This step starts when a solution is proposed to solve a new problem. Revised aim is to assess the acceptability of proposed solutions, a newly formed case.
4. Retain - In the retain step there are new useful cases in the case base for future reuse. In this way the CBR system has learned a new experience and retains the knowledge gained from solving the new problem.

Which cases to be retrieved is decided based on a given similarity threshold. CBR performs measurement of similarity on the local and global level. Local similarity refers to measurement of similarity between pairs of features. Global similarity refers to a comparison of the similarity between all the features that make up the object. Measuring similarity can be shown by (1):

$$Similarity(T,S) = \sum_{i=1}^{n} f\left(T_j, S_i\right) \times w_i \; \}, \tag{1}$$

where T= target case; S= source case; n= number of features in each case; I= individual feature from 1 to n; f= similarity function for features I in cases T and S; w= importance weighting of feature I.

C. Related Work

In the last decade, there has been a great deal of research focused on developing a hybrid classification models. Hybrid models, because of its good features, have found implementation in various fields of research. Reference [13] proposed the hybrid model of SVM and principal component analysis (PCA). The new model uses concept of reducing the data dimensionality using PCA to decrease the complexity of an SVM-based sentiment classification task. The hybrid model has been proved to provide a good classification result in sentiment mining.

Reference [14] proposed a novel hybrid classification model based on merging genetic algorithms, k-Nearest Neighbor and Backpropagation Neural Network. The main goal of the study was developing the accuracy of the classification model which takes advantage of a synergy that was expected to emerge from the hybridization of the components of the proposed model. The proposed hybrid model was implemented based on combining some methods and algorithms of artificial intelligence in three main stages: features selection, parallel classification with k-NN and the BPNN methods.

Reference [15] presented a hybrid model with a combination of genetic algorithms, Bayesian methods, and k-NN. Their goal was to eliminate the data that are barrier to learning to achieve successful results in classification.

Reference [16] presented a hybrid intelligent system for medical data classification. A hybrid intelligent system consists of the Fuzzy Min–Max neural network, the Classification and Regression Tree, and the Random Forest model. The experimental outcomes positively demonstrate that the hybrid intelligent system is effective in undertaking medical data classification tasks.

Reference [17] proposed hybrid models consist of logistic regression (LR), multivariate adaptive regression splines (MARS), artificial neural network (ANN), and rough set (RS) techniques. The study proposed a hybrid intelligent modeling scheme to obtain different sets of explanatory variables, and the proposed hybrid models effectively classify heart disease.

Reference [18] proposed a hybrid method based on the concepts of ANNs and fuzzy regression models. This model has been proposed for classification purposes, and for achieving higher accuracy and a more generalized application than the traditional ANN models.

Reference [19] proposed a model that can benefit from the merits of the k-nearest neighbor and the extreme learning machine through its novel structure with high robustness particularly for cloud classification. The simulation results demonstrate that the proposed model in this work is practical for cloud classification and outperforms extreme learning machine (ELM) models, ANN, k-nearest neighbor (KNN), hybrid method based on KNN and ANN (KNN–ANN), and support vector machine (SVM).

III. The Concept of CBR Hybrid Model

In this section, we will first explain the work principle of our hybrid model, the structure of model, pseudo code method for building the model and implemented classification algorithm.

A. Building the Classifier

When we analyze the work principle of the CBR hybrid model, it is necessary to emphasize two major situations in model life cycle. The first situation is related on building the model, and the other on its using. Building a model is in line with the characteristics of the dataset is the most important process. Fig. 1 shows the concept of building a CBR hybrid model and Table I shows the pseudo code.

At the beginning, algorithm is determining the weight value for each feature which is found in the dataset. The weight value is determined by aggregation of two values.

IG method based on data entropy determines the rank of each input feature. Measured rank of a feature represents its information importance in the input dataset. After determining the rank of features, algorithm takes the values from the expert's knowledge base. The values indicate the feature influence on the classification process, according to the expert’s opinions. Knowledge of experts was formed in Domain driven model (DDM). In case of irrelevant features, hybrid algorithm excludes them from the further classification process. The algorithm records weight features values in the vector structure. After that, the algorithm begins with clustering of the original instances from data set. To each instance, algorithm assigned label of cluster. The algorithm also records information about all determined cluster centroids.

TABLE I
PSEUDO CODE OF BUILDING ALGORITHM

```
Input data: Dataset data
creating_Hybrid_Model (Dataset data): CBR classifier;
{
 Step 1: Features ranking
   Vector rankVector ← IG(data)
    Determining the rank of features by IG method.
    Determined values write in to vector structure.
Step 2: Determining weight features values
   Vector expertVector ← Creating vector structure from Expert’s
                              database (DDM)
   Vector weightFeatureVector ← Aggregation (rankVector,
                                              expertVector)
    weight value:
          0 ← DDM = 0 The feature is not important, exclude
                        from classification process.
         (rank+DDM) / 2 ← 0 < DDM < 1
         Max ← The feature is very important
Step 3: Clustering dataset instances
   Vector clusterVector ← K-means(data)
     Clustering  the instances of dataset and writing cluster labels
      in to vector structure.
   Vector centroidsVector ← determining the information about
                                clusters centroids
Step 4: Creating a case database
    ForEach( element instance ← data){
     Case caseInstance ← Agg(instance, clasterVector(i))
       Creating a case instance and adjustment to CBR
        classification model. The aggregation of the original data
        and cluster label.
     DB caseDataBase  ← writing new case instance in to
                              database/file
     }
Step 5: Creating a CBR classification model
    File CBR data ← defining a features of future CBR
      classification model base on a dataset characteristics, features
      weight values, number of class labels
    File modelCBR ← creating a classification model in file/object
      structure, suitable for later dynamically load during the
      classification of instance.
  }
```

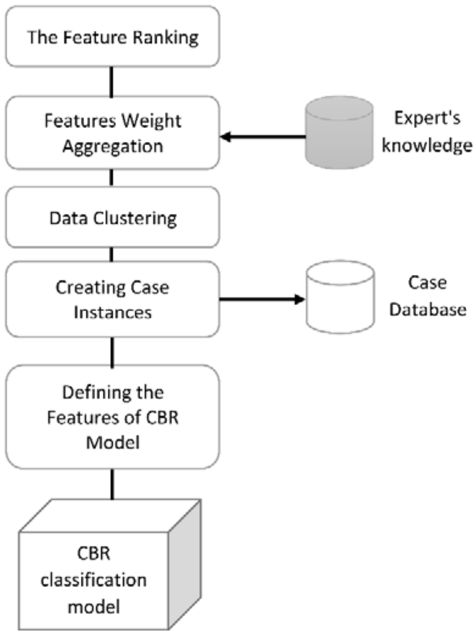

Fig. 1 The concept of building a CBR hybrid model

In step 4, algorithm creates the case instances for CBR case database. Algorithm performs the aggregation, instances values and cluster labels. Aggregation creates new instances, which are recorded in the case database by the algorithm.

After case database was prepared, in the last step algorithm begins with creating a CBR classifier. First, the algorithm creates a file with all information about model classification characteristics. The file contains detail information about how a CBR cycle will be performed, information on an instance data model, a way to perform measurement of local and global similarity, etc. Finally, the algorithm forms CBR classifier.

B. Using the Classifier

The obtained classifier can be used to classify heterogeneous and high dimension data. Fig. 2 shows the way of performing the classification process by a hybrid model. Table II shows the pseudo code of classification algorithm.

TABLE II
PSEUDO CODE OF CLASSIFICATION ALGORITHM

```
Input data: Now instances
classification (instance data): class label [ ];{
 Model initialization
  Step 1: Loading CBR classification model
    Loading a CBR classifier with all established features from
    domain of classification.
  Step 2: Preparing case instance for classification
    Loading a case instances from database in the linear cache
    memory of classifier.
 ForEach( element instance ← data){
  Step 3: Determining cluster label for new instance
    Defining the cluster labels of new instance by using the information
    about the clusters centroids determined during the building hybrid
    model.
  Step 4: Preparation of new case instance
    Instance newInstance ← Instance adaptation to the CBR
    classification model.
  Step 5: Classification
    casesList[ ] ← Retrieve cases
    Distance ← Measuring the similarity.
    if (Distance = 0)
        return class label ← Reuse
    else
     Adapt the retrieved cases to fit the new case.
       newCase ← Revise: Evaluate the solution and revise it.
       cache memory  ← Retain:  newCase in the case memory.
       case database ← write newCase in database
       return class label ← newCase
  }
}
```

Classification process begins with loading the CBR classification model. All information that defines the way of conducting the classification process was loaded with the model. Depending on the domain of the problem, the algorithm loaded case instances from the database in the cache memory of model. The CBR hybrid model uses a cache memory for faster retrieve case instances. Case instances are organized on indexed linear structure. The CBR model prepared in such a way can begin with classification.

In steps 3 and 4, the algorithm determines a cluster label for a new instance. The algorithm defines the cluster labels of new instance by using the information about the clusters centroids determined during the building hybrid model. The algorithm adds this cluster label into data source and forwards it to the CBR cycle.

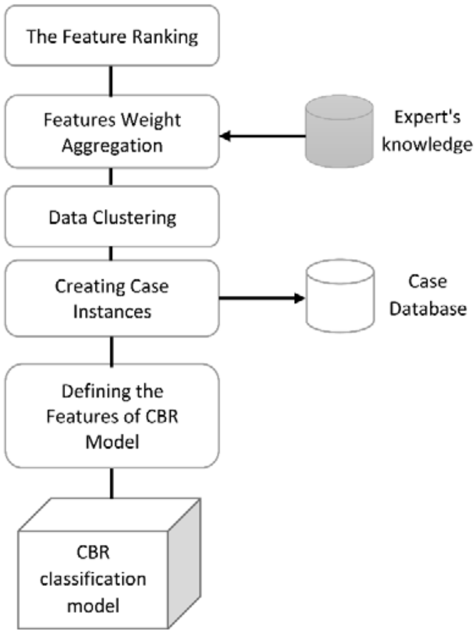

Fig. 2 The concept of classification process

The classification is carried out by executing a standard CBR cycle. If the CBR model, using the similarity function, retrieves an identical case from a database, algorithm returns a class label of retrieved instance. If the compared instances are identical, the calculated value of distance is 0. Otherwise, model starts with the instance adjustment and updating case database. If multiple instances are sent into the model for classification, the initialization of the model is performed only once.

C. Case Representation

Case database is the base element of the CBR hybrid model. The cases represent a collection of data used in the process of machine learning. All data from the dataset are transformed into a structure that represents the knowledge base. Each formed case consists of input data and a result. The result is represented by a class label.

CBR models commonly use three basic ways to organize cases: flat, structured and unstructured [12]. Our CBR hybrid model uses a flat organization of cases. The flat organization is the simplest to design and implement, and most suitable for a small number of cases. In such an organization there is no relation between the cases, a case is described with feature-value pairs. Each feature value pair represents a part of an instance in a case database. The case structure is recorded in the cache memory and forms a linear structure. Therefore, the time of classification is proportional to the total number of instances.

IV. Case Study

A. Data Sets

In the development of the hybrid model and evaluation of the prediction quality, a large number of data sets was used. The objective of evaluation was to determine the quality of the classification and the sensitivity of the model based on the datasets character. For the purpose of evaluation, 20 data sets from different problem domains were used. All used datasets have evaluated data, made publicly available on UC Irvine Machine Learning Repository via Internet. Table III shows the list of the used data sets and their basic properties.

TABLE III
Overview of Data Sets and Their Properties

No	Dataset	Feature No.	Size	Class No.
1	autos.arff	26	205	7
2	abalone.arff	9	4177	29
3	bank_marketing.arff	17	4521	2
4	breast-cancer.arff	10	286	2
5	breast-w.arff	10	699	2
6	car.arff	7	1728	4
7	colic.arff	23	368	2
8	contact-lenses.arff	5	24	3
9	credit-g.arff	21	1000	2
10	dermatology.arff	35	366	6
11	diabetes.arff	9	768	2
12	glass.arff	10	214	7
13	haberman.arff	4	306	2
14	hepatitis.arff	20	155	2
15	sonar.arff	61	208	2
16	tae.arff	6	151	3
17	vehicle.arff	19	846	4
18	vowel.arff	14	990	10
19	winequality.arff	12	1469	11
20	zoo.arff	18	101	7

B. Evaluation of CBR Hybrid Model

The first part of the case study focused on measuring the classification accuracy of the CBR hybrid model. In order to achieve the high quality testing of the hybrid model, 20 public datasets were used in the case study.

Before testing the properties of CBR hybrid model, the evaluation of the classification accuracy with a Support Vector Machine method (SVM), Naive Bayes (NB) and Decision Tree (DT-J48) was performed using WEKA[1] tools. After that, under the same conditions was performed classification using CBR hybrid model. Table IV shows the shorter list of used datasets and the results achieved by classification.

The obtained results show a high degree of accuracy achieved by a CBR hybrid model during classification using different datasets. The model showed stability in the process of classification, regardless of the size of various datasets and different numbers of features and classes. The CBR hybrid model has achieved over 92% average classification accuracy, which is exceptionally good result compared to the results achieved with other classification methods used.

[1] WEKA- Machine learning tool developed at the University of Waikato, New Zealand.

In order to obtain complete details about the characteristics of the CBR hybrid model, additional analysis of model performances using the ROC curve was performed in the study. ROC (Receiver Operating Characteristic) analysis is usually associated with the classifier and used as an alternative metric to compare the performance of the classifier.

The ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. The ROC curve shows the balance, trade-off between the sensitivity and specificity shown by a classifier. The ROC curve plots the false positive rate (100-Specificity) on the x-axis and the true positive rate (Sensitivity) on the y-axis [20]. The classifier whose ROC curve is closer to the upper left corner of the chart, shows better classification accuracy. Fig. 3 shows the ROC curves, measurements of classification accuracy over a few datasets noted in Table IV. The shapes of the ROC curves indicate a very good performance of CBR hybrid model regardless of the number of features, the size of the dataset and the number of classes.

TABLE IV
Overview of Achieved Classification Accuracy Results

Dataset	Accuracy %			
	SVM	NB	DT	CBR
autos.arff	70	56	71	**72**
abalone.arff	25	23	21	**87**
bank_marketing.arff	88	88	89	**95**
breast-cancer.arff	72	71	75	**88**
breast-w.arff	97	95	94	**100**
car.arff	85	85	92	**94**
colic.arff	81	78	85	**89**
contact-lenses.arff	70	70	83	**85**
credit-g.arff	75	75	70	**81**
dermatology.arff	97	97	93	**97**
diabetes.arff	74	76	73	**100**
glass.arff	70	48	66	**92**
haberman.arff	72	76	72	**95**
hepatitis.arff	83	85	83	**91**
sonar.arff	76	67	71	**95**
tae.arff	47	54	59	**98**
vehicle.arff	60	45	72	**99**
vowel.arff	61	63	81	**96**
winequality.arff	50	45	48	**96**
zoo.arff	96	95	92	**97**

Except the measurement of specificity and sensitivity, the evaluation of the classification accuracy depending on the size of the data set was performed. A method of random selection of instances from the original data set was used in the test. During the testing, on each step the original data set was reduced by 20%. The results of the classification accuracy were compared with the initial value which was measured for the original size. Four data sets with different properties and different sizes were used in the evaluation. Table V shows the achieved results, which confirm that the quality classification of the CBR hybrid model does not depend on the size of the data set. There have been only a few degree accuracy deviations.

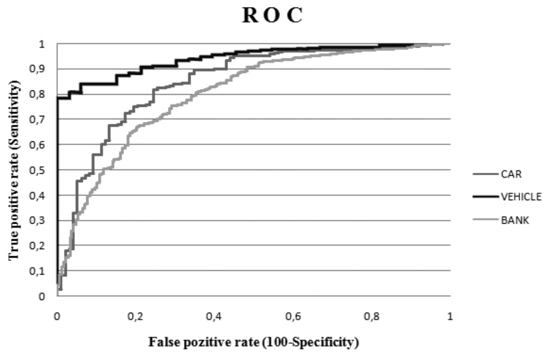

Fig. 3 ROC curves

TABLE V
ACHIEVED CLASSIFICATIONS ACCURACY DEPENDING ON THE SIZE OF THE DATASET

Dataset	Accuracy %					
	% data set size					
	Size	**20**	**40**	**60**	**80**	**100**
car.arff	1728	91	92	92	93	94
haberman.arff	306	96	96	95	96	95
winequality.arff	1469	96	97	97	97	96
hepatitis.arff	155	90	91	92	90	90

In the evaluation, we tried to determine the impact of expert's knowledge on the classification accuracy. Evaluation has offered an answer to the question how much the hybrid model is sensitive to incorrect definition of weight values of the features. In order to solve a potential problem, the CBR hybrid model is merged with expert's knowledge (DDM) and IG method.

Evaluation of hybrid model was performed in three steps. In the first step of evaluation, equal importance was defined for all features. In the next step, model performs correction of the importance based on the IG method and DDM. Table VI shows the achieved results. The achieved results confirm the assumption that weight values have a significant impact on the accuracy classification. The best result was achieved when hybrid model used DDM as a correction factor in defining the final weight value of features. Using DDM, the hybrid model has achieved significantly better classification result.

TABLE VI
ACCURACY OF PREDICTIONS DEPENDS ON THE IMPORTANCE OF FEATURES

Dataset	Accuracy %					
	% data set size					
	Size	**20**	**40**	**60**	**80**	**100**
car.arff	1728	91	92	92	93	94
haberman.arff	306	96	96	95	96	95
winequality.arff	1469	96	97	97	97	96
hepatitis.arff	155	90	91	92	90	90

V. CONCLUSION AND FUTURE RESEARCH

In this paper is presented a concept of hybrid model for classification. The aim of the research was the development of a hybrid and DDM based on the CBR method. According to the results obtained with the hybrid model, it can be concluded that the concept of a hybrid algorithm based on CBR method has been successfully.

A new hybrid algorithm that merges several machine learning methods has been implemented in the CBR hybrid model. The pseudo code of the hybrid algorithm describes the way of simultaneous use of a classic learning methods and the knowledge base of experts. The obtained results show the significant dominance of the hybrid algorithm in comparison to classic machine learning methods that were used during the evaluation. The achieved classification results emphasize the stability of CBR hybrid model. During testing, the hybrid model showed some sensitivity in classification accuracy based on the defined weight values. For this reason, it is necessary to implement further and deeper testing. One of the possibilities is to include another method for ranking or selection features, which will be shown by future measurements.

Our concept is certainly a good starting point for further development with various types of classification. Future testing of different types of datasets will indicate the elements that will need to be modified, with the main purpose to achieve better results.

REFERENCES

[1] P Ravisankar, V Ravi, and I Bose, "Failure prediction of dotcom companies using neural network–genetic programming hybrids," Information Sciences, vol. 180, no. 8, pp. 1257–1267, 2010.
[2] Rudrajeet Pal, Karel Kupkab, Arun, P. Anejac, and Jiri Militkyd, "Business health characterization: A hybrid regression and support vector machine analysis," Expert Systems with Applications, vol. 49, pp. 48-59, 2016.
[3] Guozhong Fenga, Jianhua Guo, Bing-Yi Jingc, and Tieli Sun, "Feature subset selection using naive Bayes for text classification," Pattern Recognition Letters, vol. 65, pp. 109–115, 2015.
[4] Sunita Beniwal and Jitender Arora, "Classification and feature selection techniques in data mining," International Journal of Engineering Research & Technology, vol. 1, pp. 1-6, 2012.
[5] Y. An, J. Baek, S. Shin, M. Chang, and J. Park, "Classification of feature set using K-means clustering from Histogram Refinement method," in Networked Computing and Advanced Information Management, Gyeongju, 2008, pp. 320 - 324.
[6] E. Hullermeier and W. Cheng, "Preference-based CBR: General ideas and basic principles," in Proceedings of the Twenty-Third international joint conference on Artificial Intelligence, 2013, pp. 3012-3016.
[7] B. Trstenjak and D. Donko, "Predicting quality of web service using IKS hybrid model," in Telecommunications (BIHTEL), 2014 X International Symposium, Sarajevo, BiH, 2014, pp. 1 - 6.
[8] Andreas Janecek, Wilfried Gansterer, Michael Demel, and Gerhard Ecker, "On the relationship between feature selection and classification accuracy," in Journal of Machine Learning Research: Workshop and Conference Proceedings 4, 2008, pp. 90-105.
[9] B. Trstenjak and D. Donko, "Determining the impact of demographic features in predicting student success in Croatia," in Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 2014.
[10] M. Sahu, S. Sharma, V. Raj, N. Nagwani and S. Verma, "Impact of Ranked Ordered Feature List (ROFL) on classification with visual data mining techniques," in International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT), Chennai, Tamilnadu, India, 2016.
[11] G., A. Klein, How people make decisions. London, Cambridge: 1st MIT Press pbk, 1999.
[12] M., M. Richter and R. Weber, "Case-Based Reasoning: A Textbook," in Basic CBR Elements: Springer Science & Business Media, 2013, pp. 17-34.

[13] G. Vinodhini and R., M. Chandrasekaran, "Sentiment mining using SVM based hybrid classification model," Cyber Security and Computational Models, vol. 246, pp. 155-162, 2013.
[14] Nader Salari, Shamarina Shohaimi, Farid Najafi, Meenakshii Nallappan, and Isthrinayagy Karishnarajah, "A novel hybrid classification model of Genetic algorithms, modified k-Nearest Neighbor and developed Backpropagation Neural Network," PLOS ONE, vol. 9, no. 11, pp. 1-50, 2014.
[15] M Aci, C Inan, and M Avci, "A hybrid classification method of K nearest neighbor, bayesian methods and genetic algorithm," Expert Systems with Applications, vol. 37, pp. 5061–5067, 2010.
[16] M Seera and CP Lim, "A hybrid intelligent system for medical data classification," Expert Systems with Applications, vol. 41, pp. 2239–2249, 2014.
[17] Ye Shao, C,D Hou, and C,C Chiu, "Hybrid intelligent modeling schemes for heart disease classification," Applied Soft Computing, vol. 14, pp. 47–52, 2014.
[18] M Khashei, Hejazi, S Reza, and M Bijari, "A new hybrid artificial neural networks and fuzzy regression model for time series forecasting," Fuzzy Sets and Systems, vol. 159, pp. 769–786, 2008.
[19] Min Xiaa et al., "A hybrid method based on extreme learning machine and k-nearest neighbor for cloud classification of ground-based visible cloud image," Neurocomputing, vol. 160, pp. 238–249, 2015.
[20] Simon Bernard, Clément Chatelain, Sébastien Adama, and Robert Sabourin, "The multiclass ROC front method for cost-sensitive classification," Pattern Recognition, vol. 52, pp. 46-60, 2015.

Bruno Trstenjak is a lecturer at the University of Applied Sciences in Cakovec, Croatia. In 2012, he completed MSc degree in the field of Computer sciences at the Faculty of Electrical Engineering in Sarajevo, BiH. Since 2008, he has worked as an academic member at the Medjimurje University of Applied Sciences, where he has been involved in lecturing several courses in fields of Computer Sciences. His area of interest covers: programming and application development, field of artificial intelligence, machine learning and data mining. He has been involved in the several EU projects in the field of education, as a project manager or a member. Recently, he is head of department at professional study program - Computer Engineering.

Dzenana Donko is an associated professor at the Faculty of Electrical Engineering University of Sarajevo, BiH. She received her Ph.D. from the Faculty of Electrical Engineering, University of Sarajevo in 2004. Ms. Donko participated in many project and studies, some of them related to the UNDP development projects in BiH. Her current research interest's areas of object oriented analysis and data mining. She is author of many papers published at different journals and conferences. In several occasions she participated as a plenary speaker at the conferences presenting subject on data mining. She is also a member of many scientific and program committees on the different international conferences.