Data Mining Classification of Cars Based on GPS Shadows in Forward Scatter Radar Systems

Christo Kabakchiev[1], Dorina Kabakchieva[2], Ivan Garvanov[3], Vera Behar[4], K. Kabakchiev[5], H. Rohling[6], K. Kulpa[7], A. Yarovoy[8]

[1]Sofia University "St. Kliment Ohridski", SU, Sofia, Bulgaria
E-mail: ckabakchiev@fmi.uni-sofia.bg
[2]University of National and World Economy, UNWE, Sofia, Bulgaria
E-mail: dkabakchieva@unwe.bg
[3]University of Library Studies and Information Technologies, ULSIT, Sofia, Bulgaria
E-mail: *i.garvanov@unibit.bg*
[4]Institute of Information and Communication Technologies, IICT-BAS, Sofia, Bulgaria
E-mail: behar@bas.bg
[5]University of Birmingham, Birmingham, United Kingdom
E-mail: kalin.kabakchiev@gmail.com
[6]Technical University of Hamburg-Harburg, Hamburg, Germany,
E-mail: rohlingr@tu-harburg.de
[7]Technical University, Warsaw, Poland
E-mail: kulpa@ise.pw.edu.pl
[8]Technical University, Delft, Netherland
E-mail: *a.yarovoy@irctr.tudelft.nl*

Abstract: *The goal of this paper is to introduce a new concept - using Data Mining approach for radar classification of vehicles, based on their GPS shadow signals detected in a GPS L1-based Forward Scatter Radar (FSR) system. Real data is used for the current experiments, recording the GPS radio shadows of several moving vehicles with commercial non-professional GPS equipment. The records are further processed in MATLAB computational environment in order to obtain the estimated parameters of the GPS shadows. Data Mining approach is then implemented for classifying the cars, comparing classifiers generated with different classification algorithms - a decision tree, a neural network, and a Bayesian classifier, for two different variants of the class variable, taking three or two possible values. The best results are achieved for the two values of the class variable and the best performing classifiers are the neural networks.*

1. Introduction

Passive radar systems where GPS satellites are used as transmitters are recently becoming increasingly popular as an alternative to traditional radar systems. The GPS Forward Scatter Radar (GPS FSR) System is a specific case of Forward Scattering Radar (FSR), where GPS satellites are exploited as "transmitters of opportunity".

The goal of this paper is to introduce a new concept - using Data Mining approach for radar classification of vehicles, based on their GPS shadow signals detected in a GPS L1-based Forward Scatter Radar (FSR) system. The presented research results are based on previous research achievements described in [1,2,3,4], devoted to experimental measurements made by using GPS L1-based FSR system and the Software-Defined GPS receiver, developed by the

Aerospace Department in the University of Colorado [5], allowing to observe the geometric shadows (signal blocking) of ground objects of different sizes, mobile and stationary.

The GPS L1-based FSR system, used in the conducted experiments, consists of an USB-based recording system with a small commercial GPS antenna, recording the raw GPS data from GPS L1 satellite and storing it as binary files in a computer, and a Software-Defined GPS receiver processing the recorded data in MATLAB computational environment. The GPS radio shadows created by moving ground targets, vehicles of different sizes and types, are further processed in order to extract their characteristic parameters, which are then used for vehicle classification by applying Data Mining methods and techniques.

The idea of applying Data Mining approach for classification of detected radar targets has been previously tested in [6], for classifying marine targets based on the recorded signals of Bistatic maritime FSR system, developed by the research team of Birmingham University. The real data in that case was collected with an original structure of the Parameter Estimator using a Constant False Alarm Rate (CFAR) processor, for target detection and estimation in the time domains.

Real data is used for the current experiments, collected by processing the recorded radio shadows of several moving vehicles detected by the above described GPS L1-based FSR system. Target classification is performed by applying data mining techniques on the available target data. Based on the analysis of the target characteristic features, including the target length, the level of energy reflected from the target, etc., classification models are built for a pre-defined number of targets (classes). In the considered case, the targets to be classified based on their GPS shadow signals in Forward Scatter Radar (FSR) system are moving vehicles of different types and sizes. The classification models for predicting the detected target class based on the received and pre-processed target data are achieved by applying different data mining methods. Popular WEKA [7] classifiers (with their default settings unless specified otherwise) are used in the experimental study. The generated classification models are compared and the results are presented in table and graphical formats.

The paper includes the following sections: Introduction, Experiment Description and Target Parameter Estimation, Target Classification with Data Mining Approach, Conclusions, and References.

2. Experiment Description and Target Parameter Estimation

The raw data is collected at the beginning of March 2017, in field experiments outside Sofia city, with the GPS L1-based FSR system. The stationary-based GPS FSR system records the radio shadows formed by cars moving on the road (Fig.1) by using the navigation messages from all visible GPS satellites. Three vehicles of different type and size are used –Kia-Rio (a small-sized target), VW Touran (an average-sized target) and Opel-Vivaro (a large-sized target). The cars are moving into one direction with an average speed of 20 km/h, one after the other at a 3 - 4 meters distance between them. The distance between the GPS receiver and the cars is 1.5 - 4 meters.

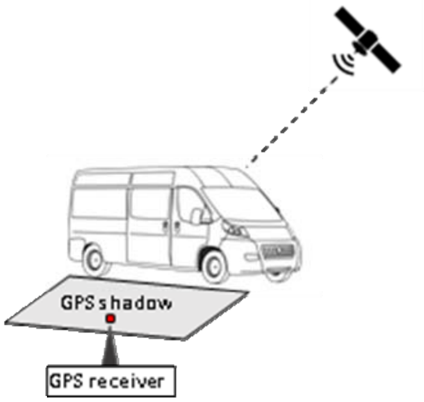

Figure 1. Forward Scatter Radar GPS Shadow

The general block-scheme for target radio shadow processing using a Software-Defined GPS receiver [5] is shown on Fig.2. The Software-Defined GPS receiver contains the Acquisition block to identify satellites, and the Code & Carrier Tracking block to form the navigation message. Next, the obtained navigation message is integrated every hundreds of milliseconds in order to form the radio shadow of the object. The integrated navigation message is then inverted and used for further detection of the object based on its radio shadow. During the experiments, several radio shadows of the same target are received from different GPS transmitters. The operator manually selects the deepest GPS radio shadow of the object, usually resulting from the GPS transmitter with the lowest elevation and whose baseline (the line between a GPS satellite and the GPS receiver) is almost perpendicular to the vehicle trajectory. The pre-processing and detection block includes inversion and integration of the navigation message, and CFAR detection [8]. The CFAR processor performs automatic detection of the target radio shadow, after which the selected parameters of the target radio shadow are manually estimated by an operator. The parameter estimation of the GPS radio shadows of the detected vehicles provides information about the length of the target shadow in the time domain (in seconds and in number of time samples), Signal-to-Noise (SNR) ratio, the average power and the average energy of the signal. These are the parameters used as attributes for the data mining classification of the detected vehicles.

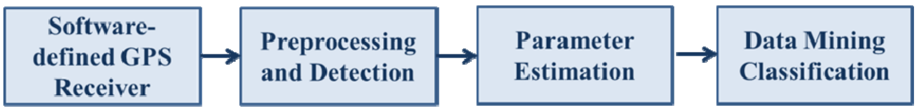

Figure 2. Signal Processing Block-Scheme

The parameter estimation of the received GPS radio shadows of the vehicles are processed in MATLAB computational environment. The processed data is organized in an excel file, containing detailed information about the conducted experiments, including data about the GPS satellites (satellite No, satellite coordinates – time, elevation, azimuth and barrier angle), distance to the detected targets and the estimated parameters of the recorded GPS radio shadows of the vehicles – dT, N, SNR_{peak} in dB, P_{ave}, and E_{ave}. The estimated parameters dT

and N are related to the length of a target shadow recorded in the GPS FSR system, and correspond to the physical size of the object. The length of the target shadow (dT) in seconds is estimated as:

$$dT = T_2 - T_1 \tag{1}$$

where T_1 and T_2 are the initial and final points of the target shadow in the time domain, estimated manually by the operator when processing the experimental records of the target shadow in MATLAB.

The length of the target shadow as a number of time samples (N) is estimated as:

$$N = dT/Ts \tag{2}$$

where Ts is the sampling rate of the signal, and coincides with the integration time of the navigation message in the Preprocessing & Detection block.

The peak signal-to-noise ratio (SNR_{peak}) is estimated as the difference between the average noise power in dB and the minimum value of the radio shadow in dB, found in the interval $[T_1, T_2]$:

$$SNR_{peak}[dB] = mean(P_n) - \min(P_s) \tag{3}$$

where P_n is the noise power in dB and P_s is the power of the target shadow in dB.

The mean power of the target shadow (P_{ave}) in dB is estimated as:

$$P_{ave}[dB] = 10\log_{10}(mean(P_{s,i})),\ i = 1 \div N \tag{4}$$

The mean energy of the shadow (E_{ave}) is calculated by multiplying the average power and the length of the shadow in time samples:

$$E_{ave} = P_{ave} N \tag{5}$$

The final dataset used for the implementation of the data mining classification tasks is obtained in excel format. It contains 298 records (excel rows), corresponding to the estimated parameters of the recorded radio GPS shadows of the three vehicles used in the experiments in different experimental conditions (distance between the GPS receiver and the target, satellite positions). The 11 attributes (excel columns) in the dataset include 5 variables describing the experimental conditions, 5 variables corresponding to the estimated parameters of the GPS shadows of targets, and the class variable – the vehicle type. An overview of the dataset is presented in Table 1.

Table 1: Dataset used for the implementation of the radar target classification task

Dataset: 298 instances; 11 attributes			
Variable Name	**VarType**	**Values**	**Missing Values**
Distance to Target	Num	Min=1.5, Max=4, Mean=3.237, StdDev=0.708	4 (1%)
Satellite No	Nom	7 (31), 8 (15), 9 (108), 17 (126), 19 (18)	0 (0%)
Satellite Elevation	Num	Min=2.4, Max=54, Mean=11.056, StdDev=7.478	0 (0%)
Satellite Azimuth	Num	Min=50, Max=161.4, Mean=128.736, StdDev=28.565	0 (0%)
Satellite Baseline – Car Trajectory Angle	Num	Min=0, Max=70, Mean=24.219, StdDev=17.436	0 (0%)
dT	Num	Min=0.244, Max=3.702, Mean=1.44, StdDev=0.387	0 (0%)
N	Num	Min=1, Max=18, Mean=6.738, StdDev=1.979	0 (0%)
SNR Peak [dB]	Num	Min=1.156, Max=22.634, Mean=10.353, StdDev=4.835	0 (0%)

P-Ave	Num	Min=0, Max=10.421, Mean=4.068, StdDev=2.391	0 (0%)
E-Ave	Num	Min=1, Max=88.147, Mean=20.526, StdDev=14.797	0 (0%)
Vehicle Type	Nom	WV Touran (104), Kia-Rio (113), Opel-Vivaro (81)	0 (0%)

The dataset contains 11 attributes - 9 numeric and 2 nominal variables (the class variable and the Satellite No). The “Satellite No” attribute is transformed from numeric to nominal variable since the values corresponding to the numbers of the satellites do not have numeric meaning.

The experimental conditions are described by the attributes Distance to Target, Satellite No, and the three attributes defining the satellite position, including Satellite Elevation, Satellite Azimuth, and Satellite Baseline – Car Trajectory Angle (the angle between the Satellite baseline, i.e. the line between the GPS satellite and the GPS receiver, and the car trajectory). The estimated parameters of the GPS shadows of the targets include dT, N, SNR Peak [dB], P-Ave and E-Ave, explained above. The class variable is Vehicle Type taking three possible values corresponding to ground targets of different type and size – Kia-Rio (small-sized vehicle), WV Touran (average-sized vehicle) and Opel-Vivaro (large-sized vehicle).

3. Target Classification with Data Mining Approach

The target classification task is implemented following the CRISP-DM (Cross-Industry Standard Process for Data Mining) model [9]. The CRISP-DM is chosen as a research approach because it is non-propriety, freely available, and application-neutral standard for data mining projects, and it is widely used by researchers and practitioners. It is a cyclic approach, including six main phases – Business understanding, Data understanding, Data preparation, Modelling, Evaluation and Deployment. There are a number of internal feedback loops between the phases, resulting from the very complex non-linear nature of the data mining process and ensuring the achievement of consistent and reliable results.

The software tool that is used for the task implementation is the open source software WEKA, offering a wide range of classification methods for data mining [7].

During the Modelling Phase, the methods for building a model that would classify the GPS shadows of the targets into the defined three classes (the three types of vehicles – small, average and large), are considered and selected. Several different classification algorithms are applied during the performed research work, selected because they have potential to yield good results. Popular WEKA classifiers (with their default settings unless specified otherwise) are used in the experimental study, including a decision tree algorithm (WEKA J48 based on the C4.5 algorithm), a Neural Network (Multilayer Perceptron), and a Bayesian classifier (NaiveBayes).

The data mining classification task is performed as shown in the diagram on Fig.3, for two testing options – 10-fold cross validation and percentage split. The 10-fold cross validation test option is usually more effective when the available data is very limited. Every time an algorithm is run, the available data is distributed in two data sets – training data containing 9/10 of the whole data, and test data including the other 1/10 of the data. Each algorithm is run ten times and that is how the final results for the selected classification algorithms are achieved. For the percentage split test option, the dataset is divided into two parts – 2/3 is used as training data and 1/3 as test data.

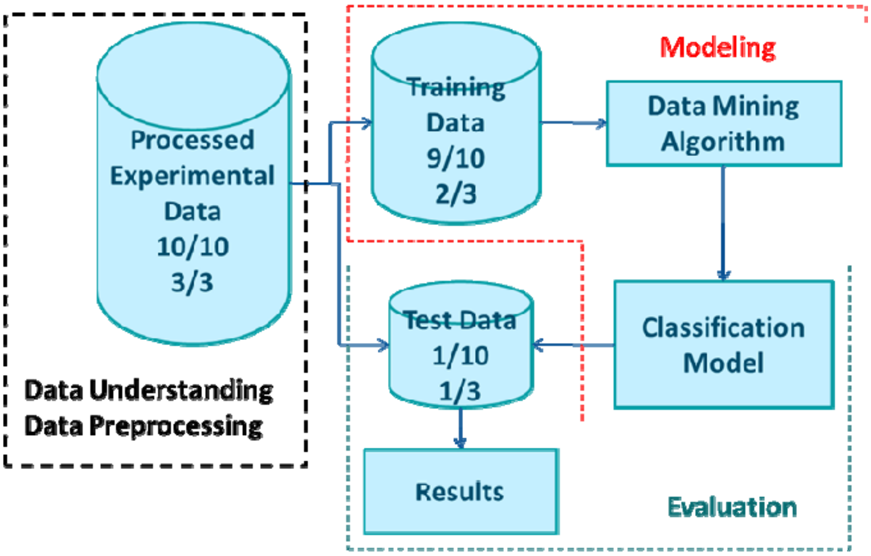

Figure 3. Data Mining Classification Task Implementation

The achieved results from the evaluation of the three classifiers, generated by applying three different methods for classification – decision tree (WEKA J48, for M – min number of records in a leaf), NaiveBayes algorithm and neural network (NN), on the dataset described in Table 1, for two testing options – 10-fold cross validation and % split, are presented on Fig.4 and Fig.5.

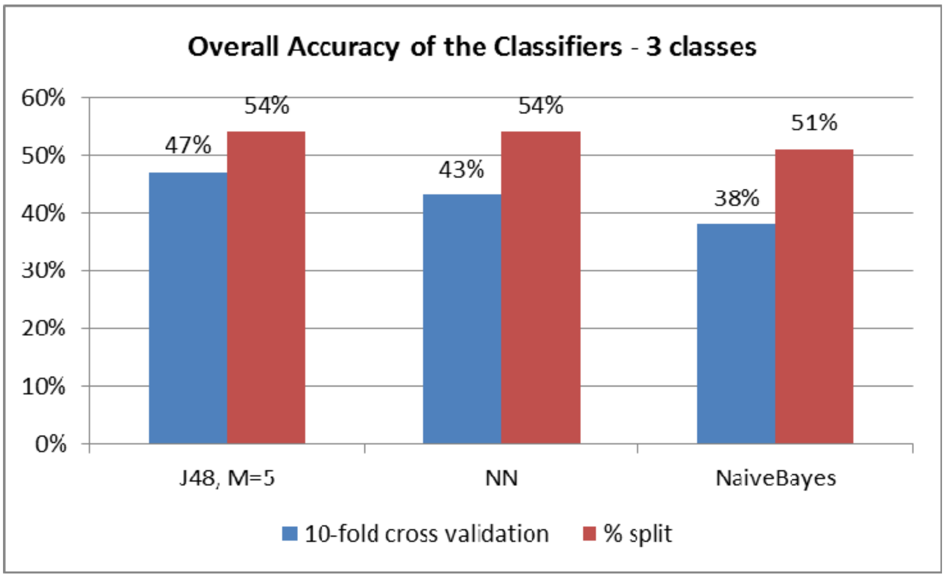

Figure 4. Overall Accuracy of Classifiers (for 3 classes)

The overall accuracy (Fig.4) is higher for the classifiers generated in the % split test option and is 54% for the decision tree and the neural network classifiers. The value 54% is not high, but nevertheless it is much higher that the naive classification (probability for random guessing of the class, without using a classifier for prediction, which in the case of three class values is about 33%).

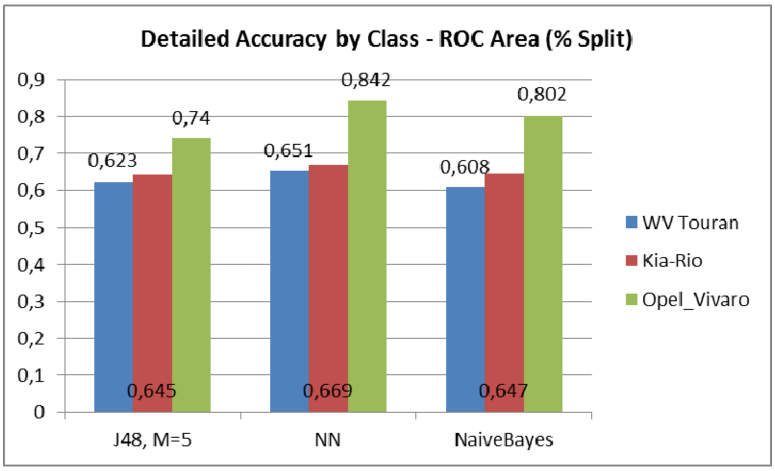

Figure 5. Detailed Accuracy by Class – ROC Area (% Split)

The results for the detailed accuracy by class for the achieved classifiers, in the % split test option (the better classifiers compared to those achieved in the 10-fold cross validation test option), are presented on Fig.5. The detailed accuracy by class is measured by the area under the ROC curve – an "optimal" classifier will have ROC area values approaching 1, with 0.5 being comparable to "random guessing". The ROC value for the Opel-Vivaro class is 0.84 for the neural network and 0.802 for the NaiveBayes which means that this class is very accurate predicted. The other two classes are less accurately predicted but still better than random guessing.

Additional research is performed after some changes in the dataset – the class variable is transformed and takes only two values – Small and Large (the values Kia-Rio and WV Touran are combined in one value – Small, and the Opel-Vivaro value is replaced with the value Large). The results from the evaluation of the achieved classifiers, with the same classification methods and test options, are presented on Fig.6 and Fig.7.

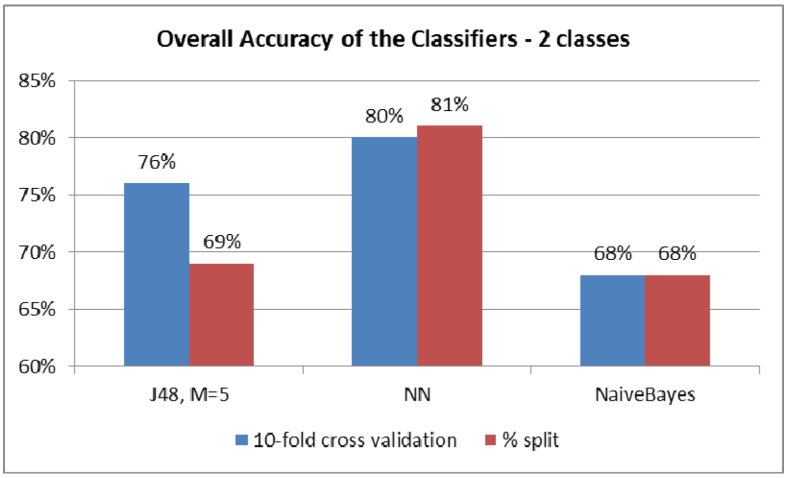

Figure 6. Overall Accuracy of Classifiers (for 2 classes)

The overall accuracy in this case (Fig.6) is higher for the classifiers generated in the 10-fold cross validation test option, and is much higher if compared to the accuracy of classification into three classes - 80% for the neural network classifier and 76% for the decision tree classifier.

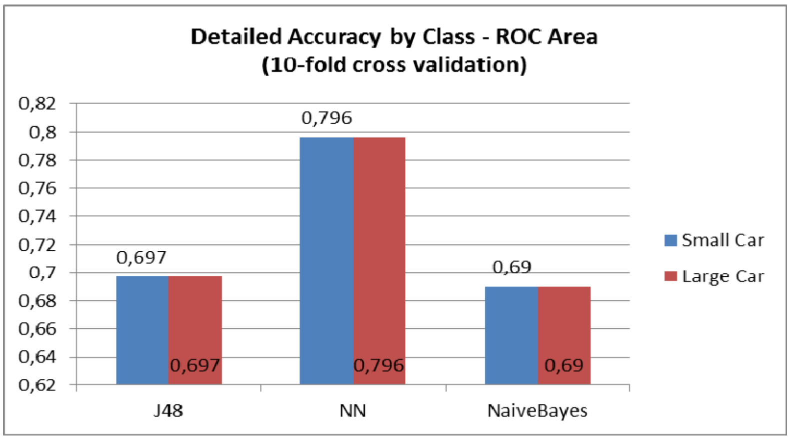

Figure 7. Detailed Accuracy by Class – ROC Area (10-fold cross validation)

The results for the detailed accuracy by class for the classifiers, in the 10-fold cross validation test option (the better classifiers compared to those achieved in the % split test option), is presented on Fig.7. The ROC values for both classes are equal and are highest for the neural network – 0.796, which means that this classifier accurately predicts Small and Large vehicles. The other two classifiers are also performing much better than random guessing – the ROC area values of 0.69 are higher than 0.5 – the "random guessing".

4. Conclusions

The presented research focused on the application of Data Mining approach for classifying cars based on their GPS shadows recorded in a FSR system with commercial non-professional GPS equipment. Several different classification algorithms are applied for building a model that would classify the GPS shadows of the targets into the defined classes, including a decision tree, a Neural Network, and a Bayesian classifier. Experiments are performed for two variants of the class variable Vehicle Type. In the first case, the Vehicle Type variable accepted three possible values, corresponding to three types of cars with different sizes (small, average, and large). In the second case, the Vehicle Type variable accepted two possible values (small, large).

The generated classifiers are evaluated and compared. In the case of the class variable with three values, the achieved overall accuracy is higher for the decision tree and neural network classifiers. The results for the detailed accuracy by class reveal that the neural network and the NaiveBayes classifier accurately predict the large-sized car class (Opel-Vivaro), but could not well differentiate the other two classes. Since the achieved accuracy of prediction is not high, experiments are also conducted with the class variables with two values – the small and medium-sized cars are combined in one class (small) and the second class is the large-sized car. In that case the achieved overall accuracy of prediction is much higher if compared to the accuracy of classification into three classes. High overall accuracy is achieved for the neural network and decision tree classifiers. The neural network is the classifier which accurately predicts both classes.

The proposed approach and results can be used in various systems for security and surveillance facilities and traffic control.

Acknowledgments

The conducted research is supported by the Bulgarian Science Fund, project No. DFN-IT 02/3/2014.

References

[1] Garvanov I., C. Kabakchiev, V. Behar, P. Daskalov: "Air target detection with a GPS forward-scattering radar", 19th International Symposium on Electrical Apparatus and Technologies (SIELA), Burgas, Bulgaria, 2016.

[2] Behar, V., Chr. Kabakchiev: "Detectability of Air Target Detection using Bistatic Radar Based on GPS L5 Signals", 12th International Radar Symposium, Leipzig, Germany, 212-217, 2011.

[3] Behar, V., Chr. Kabakchiev, H. Rohling: "Air Target Detection Using Navigation Receivers Based on GPS L5 Signals", ION GNSS, Portland OR, 333-337, 2011.

[4] Liu Changjiang, Hu Cheng, Zeng Tao, Wang Li, Long Teng: "Signal modeling and experimental verification in GNSS forward scatter radar", 17th International Radar Symposium, Krakow, Poland, 2016.

[5] Borre, K., D. Akos, N. Bertelsen, P. Rinder, S. Jensen: "A Software-Defined GPS and Galileo Receiver: Single-Frequency Approach", Birkhäuser, Boston, MA, 2006.

[6] Kabakchiev, H., Kabakchieva, D., Cherniakov, M., Gashinova, M., Behar, V., Garvanov, I. (2011). Maritime Target Detection, Estimation and Classification in Bistatic Ultra Wideband Forward Scattering Radar. Conference Proceedings of the International Radar Symposium (IRS 2011), 7-9 September 2011, Leipzig, Germany, pp.79-84.

[7] Witten, I., Frank, E. (2005). Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann Publishers, Elsevier Inc. 2005.

[8] Kabakchiev, C., I. Garvanov, V. Behar, M. Cherniakov, M. Gashinova, A. Kabakchiev, "CFAR Detection and Parameter Estimation of Moving Marine Targets using Forward Scattering Radar ", Proc. of the International Radar Symposium – IRS'11, Leipzig, Germany, pp. 85-90, 2011.

[9] Chapman, P., et al. (2000). CRISP-DM 1.0: Step-by-step data mining guide. 2000 SPSS Inc. CRISPWP-0800. Available at: http://www.crisp-dm.org/CRISPWP-0800.pdf.