International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

ANN and SVM algorithm in Divorce Predictor

Noor Hafidz, Sfenrianto, Yogie Pribadi, Evita Fitri, Ratino

Abstract: ***Classification is a technique used to predict group membership or label for data samples (instances). In order to predict the result, the classification algorithm processes the training set, which contains a set of attributes and corresponding results. One of these classification technique is implemented in order to predict divorce in Turkey. This research is executed by Yöntem, M. K. et al. in 2019. In this research, Yöntem, M. K. concluded that the ANN algorithm combined with correlation-based feature selection has the best performance with an accuracy of 98.82% and Kappa value of 0.9765. Nevertheless, unlike any previous research, ANN is not considered very good in terms of the required training time. In several previous studies, it was also concluded that other classification algorithms, such as SVM, have better prediction accuracy compared to ANN. In this study, prediction accuracy and Kappa value between ANN and SVM algorithms are compared using the same dataset and feature selection as the research done by Yöntem, M. K., to ensure a fair comparison between both of the algorithms. The result obtained from comparing both algorithms is that the SVM algorithm performs better than ANN with an accuracy of 99.8235 and a Kappa value of 0.9964. The training time required by SVM is also better than the ANN training time.***

Keywords: **classification, support vector machine, artificial neural network, divorce prediction.**

I. INTRODUCTION

Nowadays, Data Mining and Machine Learning techniques such as classification, clustering, association rules have played a significant role in extracting undiscovered knowledge from a database. Classification is one of the techniques used to predict group membership or labels for data samples (instances). In order to predict results [1], the classification algorithm processes a training set that contains a set of attributes and corresponding results. One implementation example of this classification technique is the divorce prediction in Turkey conducted by Yöntem, M. K. et al. in 2019. In the research executed by Yöntem, M. K. et al., a comparison between the accuracy of the ANN, RBF Neural Network, and Random Forest algorithms was carried out in order to predict divorce in Turkey. The results of this research, Yöntem, M. K., have confirmed that Divorce Predictors Scale (DPS) developed by Yöntem and İlhan [2] based on Gottman partner therapy [3] [4] can predict divorce. The best model is ANN, which is obtained by using a correlation-based feature selection proposed by Hall in 1999. This has confirmed several previous studies which concluded that the accuracy of ANN classification is better than other classification algorithms such as SVM and RF [5] [6] [7] [8].

However, in several studies, it is also confirmed that ANN requires more significant training time compared to other classification algorithms such as SVM and RF [5] [6] [7] [8] [9] [10]. In several other studies, it is also confirmed that in contrast to research conclusion obtained from Yöntem, MK et al., 2019, SVM has better performance compared to ANN in terms of classification accuracy [10] [11] [12] [13] [14] [15] [16] and its prediction stability [9] [17].

Therefore, this study will compare the accuracy between SVM and ANN algorithms in predicting divorce using the same dataset and feature selection that has been done [19] to ensure a fair comparison between the two algorithms.

Revised Manuscript Received on February 22, 2020.
* Correspondence Author
Noor Hafidz*, Master of Computer Science - Postgraduate Programs STMIK Nusa Mandiri, Indonesia, Email: 14002298@nusamandiri.ac.id
Sfenrianto, Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta 11480 Email : sfenrianto@binus.edu
Yogie Pribadi, Master of Computer Science - Postgraduate Programs STMIK Nusa Mandiri, Indonesia, Email: 14002307@nusamandiri.ac.id
Evita Fitri, Master of Computer Science - Postgraduate Programs STMIK Nusa Mandiri, Indonesia, Email: 14002313@nusamandiri.ac.id
Ratino, Master of Computer Science - Postgraduate Programs STMIK Nusa Mandiri, Indonesia, Email: 14002319@nusamandiri.ac.id

II. LITERATURE REVIEW

2.1 Artificial Neural Network

Artificial Neural Network (ANN) is an information processing system designed by imitating the way human brain work in solving a problem through conducting a learning process via changes in synaptic weights. In theory, ANN has a minimum of 3 (three) processing units [18], including:

1. Input Layer
   This layer states the value of a pattern used for input on the network.
2. Hidden Layer
   This is the connecting layer between the input and output layer, where the resulting output is not directly observed. In some instances, the network may have more than one hidden layer.
3. Output Layer
   This is the last layer in the artificial neural network that works as the output storage. In some applications, the output unit is used to present a pattern.

2.2 Support Vector Machine (SVM)

Support Vector Machine (SVM) is a technique used to compose predictions in either classification or regression case.

SVM is a machine learning system that follows the principle of Structural Risk Minimization (SRM). SRM aims to find the best hyperplane, which separates two classes in the input space. The best hyperplane between two classes can be found by measuring the hyperplane's margin and finding its maximum point. As a supervised classification approach, SVM search for the maximum distance between either class to get better performance of generalization or classification on test data [18]. Margin is the distance between the hyperplane and the closest pattern called a support vector.

2.3 Kappa Statistic

Kappa statistic or Kappa value symbolized by κ [19] is a robust statistic used for either interrater or intrarater reliability testing. It is a metric that compares observations accuracy to expected accuracy. Kappa value is a standard of how close the instances classified by the classification model are to the actual label. Kappa values can not only be used to assess the performance of a classification model but also be used to compare the performance between different classification models for the same classification case.

2.4 Prediction accuracy

A classification model is undoubtedly expected to make an utterly correct prediction, but it is undeniable that the prediction will not reach 100% correct when implemented. This requires performance measurement from the classification model in order to understand various aspects of the test. One of these aspects is the Prediction Accuracy. This aspect is determined by how accurate a model is in predicting its output. Accuracy is one of the most often used measures for the performance of classification, and it is described as a ratio between the correctly classified samples and the total number of samples [20].

The measurement system can be accurate but not exact, or precise but not accurate, or even not both. In this case, a measurement system is said to be valid if it is accurate and precise. Accuracy is a testing method based on the proximity level between predicted and actual value. The amount of correctly classified data can recognize the accuracy of prediction results.

2.5 Correlation-based feature selection

Feature selection is a significant element to optimize classification algorithm performance. Improved accuracy of a classification algorithm can be achieved by implementing the appropriate feature selection algorithm. Elimination of less relevant attributes can also be achieved by using feature selection. Correlation-based feature selection, proposed by Hall in 1999, uses functions and search algorithms to measure the value of information that a group of attributes has on its label [21].

III. METHODOLOGY

The research method used in this study is first calculating the prediction accuracy of the SVM algorithm by tuning its hyperparameter (C and kernel values). Before calculating the prediction accuracy, Correlation-based feature selection, that was first proposed by Hall in 1999, was implemented on the dataset to obtain the most significant attributes [21]. Data analysis was performed using tools widely used in machine learning or data mining research, namely Weka (Waikato Environment for Knowledge Analysis) version 3.8.4, which runs on a Mac with Intel Core i7 1.7 GHz processor specifications, 8 GB memory 1600 MHz DDR3.

The dataset used in this study taken from https://archive.ics.uci.edu/ml/datasets consisted of 170 instances and 54 attributes with two labels or classes, namely divorced and not divorced. The data collection conducted to produce the dataset includes questions about gender, marital status, age, monthly income, family structure, type of marriage, happiness in marriage, and thoughts on divorce. After the Correlation-based feature selection is performed, six most significant attributes were obtained, namely the 2nd, 6th, 11th, 18th, 26th, and 40th attributes [22].

The best prediction accuracy and Kappa statistical values obtained from the SVM hyperparameter tuning process are then compared to accuracy and Kappa values obtained in previous studies [22]. Stratified 10-folds cross validation were conducted through out the experiment as Weka use it by default. The comparison results between prediction accuracy and Kappa values of SVM and ANN are then used as a basis to conclude the study. Figure 1 shows the proposed methodology used in the experiment.

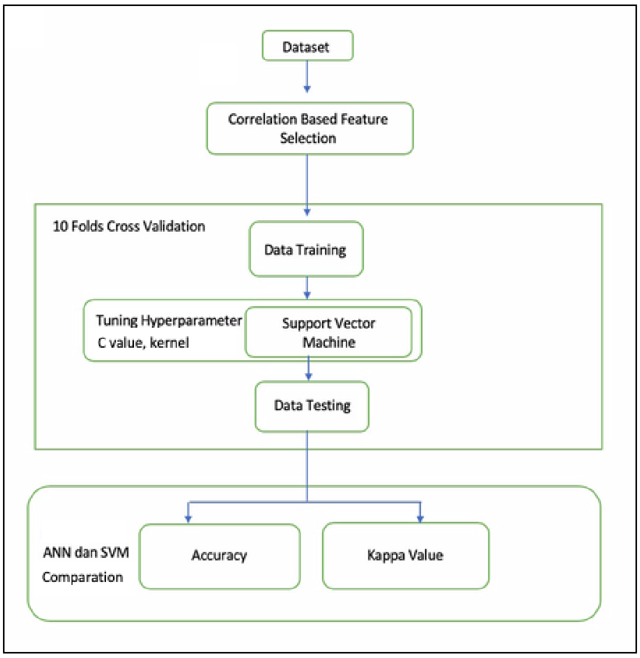

Fig. 1 Proposed methodology in the study

IV. RESULT AND DISCUSSION

The prediction accuracy and Kappa value for SVM using various data pre-procesing techniques and default Weka configuration where C value is 1.0, and polynomial is used as the kernel are given in Table 1. It is seen from Figure 2 and Figure 3 that the best accuracy and Kappa value for SVM with the default configuration in Weka obtained by standardizing the dataset as the data pre-processing technique.

International Journal of Engineering and Advanced Technology (IJEAT)
ISSN: 2249-8958 (Online), Volume-9 Issue-3, February 2020

In Tables 2 and 3, we can see the prediction accuracy and Kappa value resulting from SVM hyperparameter tuning. In the hyperparameter tuning, the C values used are 0.1, 0.3, 0.5, 0.7, 0.9, 1.0, 1.3, 1.5, 1.7, and 2.0. Meanwhile, the kernels used are polynomial and Radial Basis Function (RBF) kernel.

It is seen from Figure 4 and Figure 5 that the best prediction accuracy using SVM was 99.8235, and the highest Kappa value using SVM was 0.9964. Both best prediction classification accuracy and highest Kappa value using SVM was reached when the RBF kernel is used and the value of C is 0.5.

Table I: SVM using Weka default configuration

	Accuracy	Kappa
SVM with normalized data	98,2353	0,9645
SVM with standardized data	98,6471	0,9728
SVM without normalized/ standardized data	98,5294	0,9704

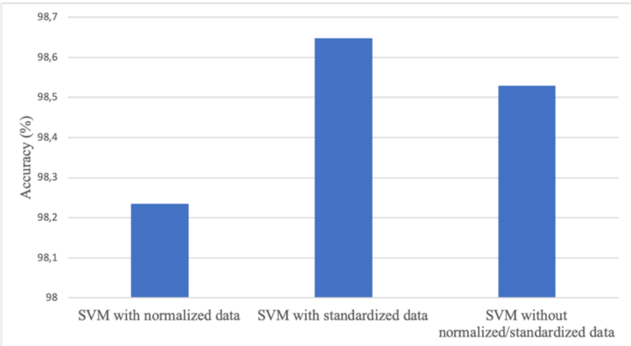

Fig. 2 SVM classification accuracy using Weka default configuration

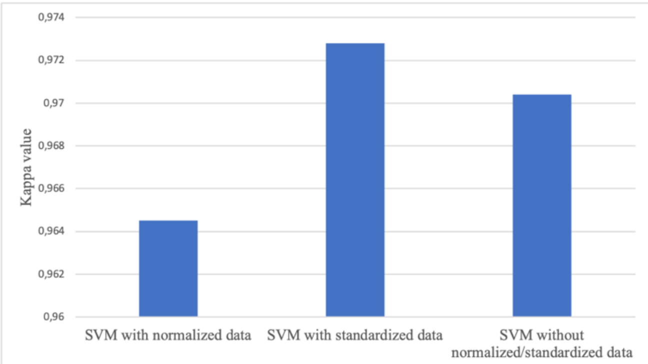

Fig. 3 SVM Kappa value using Weka default configuration

Table II: SVM hyperparameter tuning accuracy

C value	SVM polynomial kernel	SVM RBF kernel
0.1	98,4118	99,1176
0.3	98,8824	99
0.5	98,5294	99,8235
0.7	98,5882	99,2353
0.9	98,5882	99,1176
1.0	98,7647	99,4118
1.3	99,0588	99,0588
1.5	99,2941	98,8824
1.7	99,1765	98,9412
2.0	99,1765	98,9412

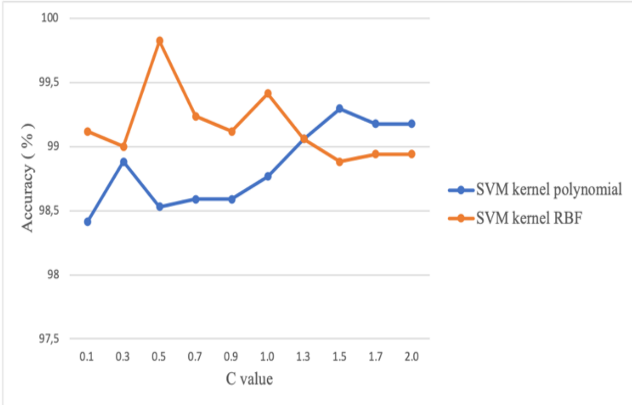

Fig. 4 SVM hypertuning classification accuracy

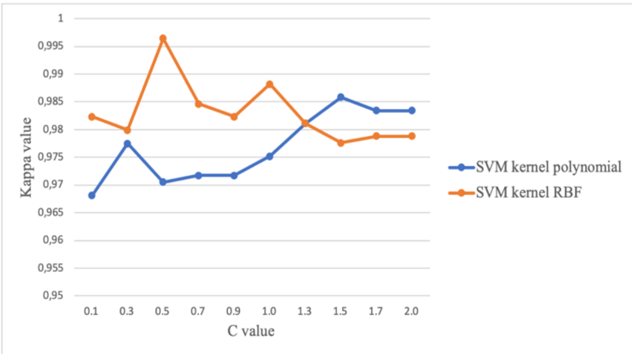

Fig. 5 SVM hypertuning Kappa value

Table III: SVM hyperparameter tuning Kappa value

C value	SVM polynomial kernel	SVM RBF kernel
0.1	0,9681	0,9823
0.3	0,9775	0,9799
0.5	0,9705	0,9964
0.7	0,9717	0,9846
0.9	0,9717	0,9823
1.0	0,9752	0,9882
1.3	0,9811	0,9811

1.5	0,9858	0,9776
1.7	0,9834	0,9788
2.0	0,9834	0,9788

The final comparison of prediction accuracy and Kappa value between ANN and SVM can be seen in table 4. Meanwhile, the comparison of training and testing time between ANN and SVM when both algorithms reach the best accuracy and the highest Kappa value can be seen in table 5. It is seen from the Figure 6 and Figure 7 that the prediction accuracy and Kappa value of the SVM algorithm are better than ANN. From the Figure 8, it is shown that SVM required less training and testing time than ANN.

Table IV: Comparison of prediction accuracy and Kappa value between ANN and SVM

	Accuracy	Kappa value
ANN	98,82	0,9765
SVM	99,8235	0,9964

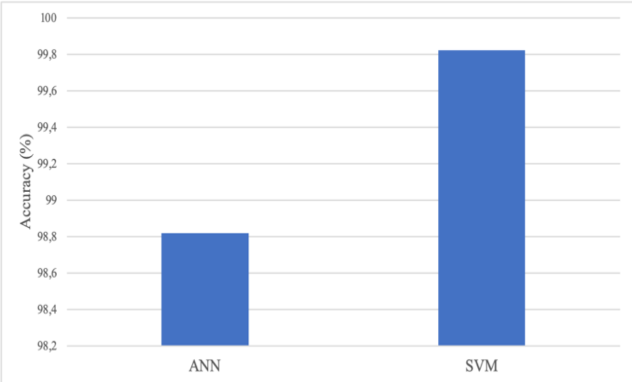

Fig. 6 ANN and SVM prediction accuracy

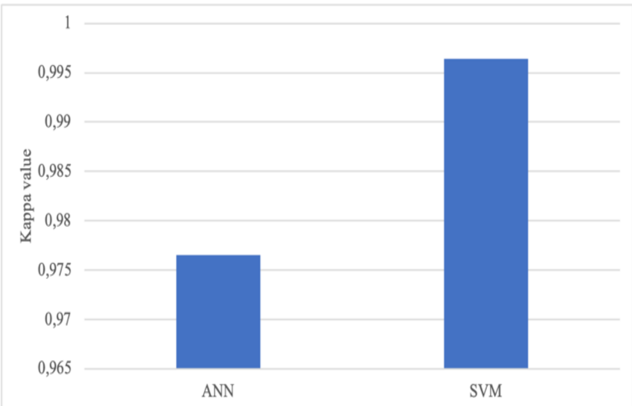

Fig. 7 ANN and SVM Kappa value

Table V: Training and testing time between ANN and SVM

	Training time	Testing time
ANN	0,1039	0,0069
SVM	0,0001	0,0001

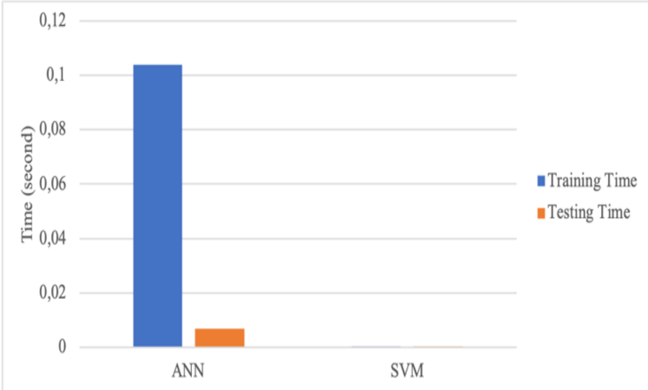

Fig. 8 ANN and SVM training and testing time

Based on the comparison result between ANN and SVM algorithms above, it can be concluded that SVM algorithm outperformed ANN in the prediction accuracy, Kappa value, and training and testing time required in terms of divorce prediction using the Divorce Predictors dataset taken from https://archive.ics.uci.edu/ml/datasets.

V. CONCLUSION

The comparison result between ANN and SVM classification algorithm in terms of divorce prediction cases using Divorce Predictors dataset taken from https://archive.ics.uci.edu/ml/datasets is that the prediction accuracy and Kappa value obtained from SVM are better than ANN. This confirms the previous research which states that SVM has a better performance compared to ANN in terms of classification accuracy [10] [11] [12] [13] [14] [15] [16].

Also, the training time required by SVM is better than ANN. This confirms previous research which states that ANN requires more significant training time compared to other classification algorithms such as SVM and RF [5] [6] [7] [8] [9] [10]. From the experiment it can be concluded that SVM outperform ANN in terms of divorce prediction using the Divorce Predictors dataset taken from https://archive.ics.uci.edu/ml/datasets.

REFERENCES

1. Ramya. S, Senthil Kumar. R, (2014) "Survey of Accident Severity Estimation Using Data Mining Techniques," International Journal of Computer Science & Engineering Technology, Vol. 5 No. 10 Oct page no 1041- 1044
2. Yöntem, M.K., and İlhan, T. (2018). Boşanma Göstergeleri Ölçeğinin Geliştirilmesi. [Development of the Divorce Predictors Scale]. Sosyal Polika Çalışmaları Dergisi. 41, 339-358
3. Gottman, J. M. (2014). What Predicts Divorce? The Relationship Between Marital Processes and Marital Outcomes. New York: Psychology Press
4. Gottman, J. M., and Gottman, J.S. (2012). Çiftler Arasında Köprüyü İnşa Etmek: Gottman Çift Terapisi Eğitimi 1. Düzey Kitabı, [Level 1 Clinical Training. Gottman Method Couples Therapy. Bringing to Couple Chasm.] İstanbul: Psikoloji İstanbul

5. Edwin Raczko & Bogdan Zagajewski (2017) Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images, European Journal of Remote Sensing, 50:1, 144-154
6. Vijayarani & Dayanand. (2015). Kidney Disease Prediction Using SVM And ANN Algorithms. *International Journal of Computing and Business Research (IJCBR)*
7. Rodrigo Moraes, João Francisco Valiati ⇑, Wilson P. Gavião Neto, Document-level sentiment classification: An empirical comparison between SVM and ANN, Programa Interdisciplinar de Pós-Graduação em Computação Aplicada – PIPCA, Universidade do Vale do Rio dos Sinos – UNISINOS, Av. Unisinos, 950 São Leopoldo, RS, Brazil, Expert Systems with Applications 40 (2013) 621–633
8. Majid Moavenian, Hamid Khorrami *, A qualitative comparison of Artificial Neural Networks and Support Vector Machines in ECG arrhythmias classification, Department of Mechanical Engineering, Ferdowsi University of Mashhad, Iran, Expert Systems with Applications 37 (2010) 3088–3093
9. Kabiru O. Akande[1], Taoreed O. Owolabi[2], Ssennoga Twaha[1] Sunday O. Olatunji[3] [1], Performance Comparison of SVM and ANN in Predicting Compressive Strength of Concrete, (Electrical Engineering, King Fahd University of Petroleum and Minerals, Saudi Arabia) [2] (Physics, King Fahd University of Petroleum and Minerals, Saudi Arabia) [3] (Computer Science, University of Dammam, Saudi Arabia), IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 5, Ver. I (Sep – Oct. 2014), PP 88-94 www.iosrjournals.org
10. Scott Doniger,[1] Thomas Hofmann,[2] And Joanne Yeh[1;3], Predicting CNS Permeability of Drug Molecules: Comparison of Neural Network and Support Vector Machine Algorithms, Journal of Computational Biology Volume 9, Number 6, 2002© Mary Ann Liebert, Inc.Pp. 849–864
11. Bahareh Kalantar, Biswajeet Pradhan, Seyed Amir Naghibi, Alireza Motevalli & Shattri Mansor (2017): Assessment of the effects of training data selection on the landslide susceptibility mapping: a comparison between support vector machine (SVM), logistic regression (LR) and artificial neural networks (ANN), Geomatics, Natural Hazards and Risk, DOI: 10.1080/19475705.2017.1407368
12. Pradeep Kumar, Dileep Kumar Gupta, Varun Narayan Mishra & Rajendra Prasad (2015) Comparison of support vector machine, artificial neural network, and spectral angle mapper algorithms for crop classification using LISS IV data, International Journal of Remote Sensing, 36:6, 1604-1617, DOI: 10.1080/2150704X.2015.1019015
13. Vinita Chandani, Romi Satria Wahono, Purwanto, Komparasi Algoritma Klasifikasi Machine Learning Dan Feature Selection pada Analisis Sentimen Review Film, Fakultas Ilmu Komputer, Universitas Dian Nuswantoro, Journal of Intelligent Systems, Vol. 1, No. 1, February 2015
14. Yang Shao [a,*], Ross S. Lunetta [b], Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points, [a] US Environmental Protection Agency, National Research Council, National Exposure Research Laboratory, 109 T.W. Alexander Drive, Research Triangle Park, NC 27711, USA, [b] US Environmental Protection Agency, National Exposure Research Laboratory, 109, T.W. Alexander Drive, Research Triangle Park, NC 27711, USA, ISPRS Journal of Photogrammetry and Remote Sensing 70 (2012) 78–87
15. Muhammad Hussain1, Summrina Kanwal Wajid2, Ali Elzaart1, Mohammed Berbar1 Department of Computer Science, King Saud University, Riyadh, KSA1 Alyamamah University, Riyadh, KSA2, A Comparison of SVM Kernel Functions for Breast Cancer Detection, 2011 Eighth International Conference Computer Graphics, Imaging, and Visualization
16. Evgeny Byvatov,[†] Uli Fechner,[†] Jens Sadowski,[‡] and Gisbert Schneider[*,†], Comparison of Support Vector Machine and Artificial Neural Network Systems for Drug/Nondrug Classification, Institut fu ̈r Organische Chemie und Chemische Biologie, Johann Wolfgang Goethe-Universita ̈t, Marie-Curie-Strasse 11, D-60439 Frankfurt, Germany, and AstraZeneca R&D Mo ̈lndal, SC 264, S-431 83 Mo ̈lndal, Sweden, J. Chem. Inf. Comput. Sci. 2003, 43, 1882-1889
17. Xianfeng Song, Zheng Duan & Xiaoguang Jiang (2012): Comparison of artificial neural networks and support vector machine classifiers for land cover classification in Northern China using a SPOT-5 HRG image, International Journal of Remote Sensing, 33:10, 3301-3320
18. Hastie, T., Tibshirani, R., & Friedman, J. (2001). The Elements of Statistical Learning. New York, NY, USA: Springer New York Inc
19. Marston L. Introductory Statistics for Health and Nursing Using SPSS. Sage Publications, Ltd.; 2010.
20. M. Sokolova, N. Japkowicz, S.Szpakowicz. Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. Australasian Joint Conference on Artificial Intelligence, Springer (2006), pp. 1015-1021
21. Hall, M. (1999). Correlation-Based Feature Selection for Machine Learning. PhD Thesis, Department Of Computer Science, Waikato University, New Zealand, 26-28.
22. Yöntem, M. K., Adem, K., İlhan, T. ve Kılıçarslan, S. (2019). Divorce Prediction Using Correlation Based Feature Selection and Artificial Neural Networks. Nevşehir Hacı Bektaş Veli Üniversitesi SBE Dergisi, 9(1), 259-273

AUTHORS PROFILE

Noor Hafidz, is a student of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri, Indonesia. (e-mail: 14002298@nusamandiri.ac.id). Research interests: network security, business intelligence, and Computer Science.

Dr. Sfenrianto, S, Kom, M.Kom, is a Faculty Member of the Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta 11480, Indonesia. (e-mail: sfenrianto@binus.edu). With lecturing subject: Digital Business and E-Commerce Management. Research interest in Digital Business, e-Commerce, business intelligence, E-Learning and Information System

Yogie Pribadi is a student of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri, Indonesia. (e-mail: 14002307@nusamandiri.ac.id). Research interests: business intelligence, and Computer Science.

Evita Fitri is a student of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri, Indonesia. (e-mail: 14002307@nusamandiri.ac.id). Research interests: business intelligence, and Computer Science.

Ratino, is a student of Computer Science - Postgraduate Programs, STMIK Nusa Mandiri, Indonesia. (e-mail: 14002307@nusamandiri.ac.id). Research interests: business intelligence, and Computer Science.