International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-10 Issue-6, April 2021

Social Mobilization and Migration Predictions by Machine Learning Methods: A study case on Lake Urmia

Fatemeh Dehghan Khangahi, Farzad Kiani

Abstract*: Voluntary or compulsory immigration of people to other regions or countries for different reasons can lead to social, cultural, and economic problems. In recent years, especially the rate of forced migration has increased and it is sometimes chosen as a last resort for social mobilization. Scientists and governments who have gathered data on migration for years have recently realized that Artificial Intelligence (AI) and Machine Learning (ML) methods are important in analyzing this data and developing utility models and systems. It has been gradually understood that these new technologies are very important in recent years, but studies have either only been done in the field of social sciences or only in the field of engineering. In this study, a comprehensive interdisciplinary study covering both dimensions is prepared. In this study, a machine learning-based model is presented by making a multidisciplinary study and exemplifying the Lake Urmia case study. The proposed method can be used in the decision-making process in the migration management. In our study, is proposed a model using three different algorithms (Support Vector Machine (SVM), Random Forest (RF), and K-Nearest Neighbors (KNN)). According to the results, the SVM-based model outperforms others in accuracy and validations. The trained model of SVM has a success rate on mean accuracy as near to 86% with 4,00E-02 standard deviation rate. SVM ranked first and this method was followed by RF and KNN methods, respectively. In this context, this model can make forward-looking predictions and, like an expert system, can guide the relevant researchers and even state or form ideas according to the results obtained from it.

***Keywords*: Classification, Migration, Social Mobilization, Machine Learning, Lake Urmia.**

I. INTRODUCTION

The voluntary or compulsory movement of people across international borders to other countries can have enormous social, cultural, and economic effects, both in the place of host and in the countries of destination. Therefore, this problem remains one of the most important issues in the global policy agenda. In recent years, since these actions have increased due to different reasons, states and scientists have tried to make different solutions and predictions. According to the report of the Global Commission on International Migration (GCIM), there are approximately 200 million migrants in the world. About 60% of this number went to developed countries and 40% to developing countries. This report reveals even more striking results in developed countries. Because approximately one out of every 10 people is a migrant [1]. In general, these migrations have had effects on world balances from various dimensions and continue to do so. For example, economy-wide flows and money transfers have increased [2], and some countries have also benefited positively [3]. Also, from socio-politic perspectives, societies that gather together from different societies and reshaped or try to keep pace by the target country's or region's conditions, are emerging. This creates various advantages and disadvantages. For example, the cheaper and abundant labor force is increasing in target countries. However, it is a more definite determination that the countries/regions of origin are more harmful [4]. However, economic, demographic, and political developments, especially in developed countries, are also heightened by growing concerns about the future of the labor supply [5, 6]. On the other hand, as the countries of origin are increasingly losing their labor resources, their needs for employees are increasing. This problem is much deeper and worrisome than the problem that occurs in the target countries. In addition, the problem of migration and social mobilization should be evaluated not only internationally but also in the national concept. Because people living in different cities and provinces in a country will also get their share from these migrations; especially from the social, cultural and economic aspects. Scientists and governments have been collecting data about migration for years and presenting reports, but it has not become much of a focus to analyze these data easily and effectively and to produce useful models by Artificial Intelligence (AI) and Machine Learning (ML) methods [7-9]. For example, most of her economics-wide studies discuss the determinants and consequences of migration in general. It is very important to analyze the forced migration in national and international dimensions or to make predictions in this field. Unfortunately, there are many examples of this type of migration in our world. One of them is the events and forced migrations in the Lake Urmia basin. This lake is known as the second largest saltwater lake in the Middle East.

Lake Urmia, which hosts many people and creatures, has caused many crises with its inconceivable drying in recent years, one of them is regional migration.

Manuscript received on April 21, 2021.
Revised Manuscript received on April 29, 2021.
Manuscript published on April 30, 2021.

* Correspondence Author
Fatemeh Dehghan Khangahi*, Political Science Faculty, International Relation Dept., Istanbul University, Istanbul, Turkey. Email: f.dehghankhangahi@gmail.com
Farzad Kiani, Engineering and Architecture Faculty, Computer Engineering Dept., Istanbul Arel University, Istanbul, Turkey. Email: farzad.kiyani@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

In this paper, this case study is handled and a useful model is created based on the reasons and rates of migration of previous years with various machine learning techniques.

In this way, possible migration rates in the coming years can be presented to the literature as valuable data and results. Briefly, in this study, a machine learning-based model is presented by making a multidisciplinary study and exemplifying the Lake Urmia case study. The proposed method can be used in the decision-making process in the migration management.

The remainder of this paper is organized as follows. Section 2 describes the related studies. Section 3 explains the details of the proposed methods together with the results. Finally, conclusion and future studies are given in the last section.

II. RELATED WORKS

In recent years, artificial intelligence and machine learning concepts have been used in many fields of study. In fact, great investments and incentives are made in many state mechanisms, especially in developed countries. This concept is not only limited to engineering fields; it is also becoming popular in recent ways to conduct various interdisciplinary studies. In this section of the paper, the role of AI and ML in the field of migration and some of the studies that have been carried out are included.

In [10], the authors have tried to use the AI abstract idea to predict Africa's next migration crisis. According to the United Nations Refugee Agency (UNHCR) report, Africa has faced the largest wave of forced migration globally in the past 20 years. In this study, the important positive role of AI in the fields of renewable energies and food production was mentioned. The authors claim that this could lead to economic growth and therefore a decrease in migration. However, they did not offer any data analysis and, moreover, any trained intelligent models.

In [11], the authors have focused on AI role in predicting illegal immigration to the USA. In this study, immigration from Mexico to the USA was discussed as a case study. The authors used a number of pre-immigration variables to estimate the legal status of over 6,281 Mexicans in the United States. For this, they used eight machine learning techniques and, according to their results, presented a model that accurately predicted 80 percent of Mexicans who legally immigrated to the US.

In [12], the authors have focused on the role of AI in Canada's immigration and creating an automated decision-making system. Focusing on this topic, the author has been paid attention to the scale of potential impact. The author claims that an automated decision-making system is of vital importance, according to many data on this, which made an immigration request to Canada. In [13], the author has argued that the UK immigration administration has benefited from an AI-based of a software system, citing the role of AI in the surveillance and control of immigrants. It is possible that such software systems can be used by governments for such needs, and they are generally used vividly. However, the issue that needs to be considered is to deal with the issue from different dimensions and to analyze the parameters that cause migration, and form a reliable model in this direction. Then, thanks to this model, it is used as an expert system. As a matter of fact, our study is based on this subject that many studies in the literature did not focus on.

In [10], a study of international migration management in the age of artificial intelligence is presented. This study has taken into account different dimensions of AI technology in international migration management. Using AI to deepen existing asymmetries between states at the international level; the role of AI in modernizing traditional practices of governments and international organizations; and the importance of AI in migration management and border security. This study, like similar studies in the literature, did not focus on technical and interdisciplinary studies. When looking at the studies in the literature, the studies either only mentioned the importance and role of new technologies such as AI and ML in migration and evaluated from the field of social sciences. On the other hand, discussions, and comments in the social field have not been presented richly in very few studies that approach the subject from a technical and engineering dimension. However, in our paper, by presenting an interdisciplinary study, the subject from both the engineering and social sciences is discussed, analyzes are carried out and comments are made.

III. METHOD AND MATERIAL

A. Lake Urmia and Datasets

Lake Urmia, which is known as a Biosphere Reserve, is the largest lake in Iran in terms of area. In 1976, Rahim Hoveyda [14] said in his work on Chichast Geography or Rezaiyye Lake, "In the past, the basin and heights around Lake Urmia had been covered with forest, the width of the lake was twice its current state". When this situation is compared with the last 15 years, the striking situation of the lake and the struggle for survival under adverse conditions become more visible. Looking at the existing ruins on the slopes and mountainous areas, the border and width has advanced from one side to the cities of Dilman (today's Salmas City Region), Urmia, Maragheh, and Goshachay, and on the one hand, Tabriz and Julfa. Later, as some rivers dried out as a result of the destruction of forests and river waters decreased, eventually the width of the lake was reduced. In general, this basin is involved three provinces (East Azerbaijan, West Azerbaijan and Kurdistan) in Iran. Although Lake Urmia has a long history, it points to a very old and important location in terms of water quality, medical features, historical, geographical, strategic-military features. With the drying of the lake, the issue of human migration has become a very important problem in addition to the health issue within the scope of social consequences. Besides this situation it creates various economic problems.

The data of this study consists of a dataset containing various parameters. These data have been obtained from different sources. The population data of Urmia and the other five cities mentioned in the last 40 years is one of the parameters used. Other sample parameters are the precipitation rate (seasonal) in the lake basin, the salt rate, and education level, migration rates. These parameters are used as dataset features. We used 25 features in overall in train and validation phases. This dataset contains more than 10 thousand sample data.

© Copyright: All rights reserved.

B. Proposed Method

In this study, Support Vector Machine (SVM) [15], Random Forest (RF) [16] and K-Nearest Neighbors (KNN) [17] methods are used. Also, Grid-search method is used to optimize metrics for RF and improves the performance of RF classifier.

The RF method fits surprisingly well into the test dataset, especially with nonlinear data. In other words, this method has been the most successful general purpose prediction algorithm. In tree-based methods, the forecast area is divided into several regions or sections. To make predictions in a test observation, we use the average or mode of training observations in the region to which the respective test observation belongs. The major disadvantages of decision trees are their high variance. The revised RF method may be a suitable method for this. In the RF method, a boot sample is selected from the training observations and the corresponding tree is enlarged. In the second phase, a random sample of the predictors is chosen at each node of the tree to be able to make the next decision. These two phases are repeated over and over again until a RF is formed, and at the end, the majority of votes are used for final prediction. The RF and GridSearch-based RF methods are represented in Fig. 1., left and right, respectively.

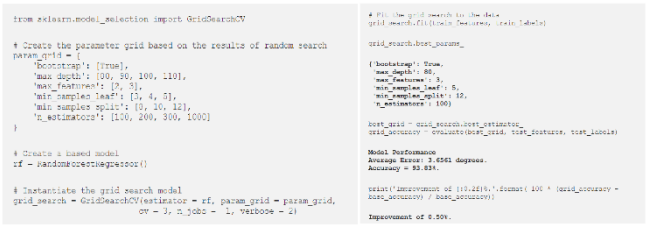

Fig 1. A sample of RF and GridSearch-based RF method [16]

The other model used in this study to data analysis is the Support Vector Classifier (SVC). This method classifies a test observation based on which side of the hyperplane it belongs to. However, SVC works better when there is a linear boundary between classes. SVM is an extension of SVC that includes a nonlinear boundary between classes. The SVM method is based on statistical learning theory [18] and basically classifies the model, class and tags in this method into two different groups and its purpose is to create optimal hyper planes and increase the flexibility of the relevant model [19]. SVM is used to solve classification problems, find the sub-plane with the largest separation space (delimited by margins) among the categories of the selected variable. Fig. 2 shows the general representation of an SVM. The lower plane is the equation that separates the two classes, while the support vectors are the data points closest to the margin boundaries.

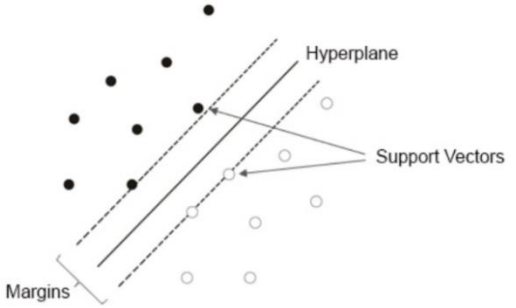

Fig. 2. Represented SVM by showing the support vectors, hyperplane and margins [9, 17].

The last ML method used in this study is the KNN algorithm. This algorithm is a non-parametric classification method. The output depends on whether k-NN is used for classification or regression. In both cases, the input consists of the k closest training examples in dataset. It can be a useful technique to weigh on the contributions of neighbors so that closer neighbors contribute more than those that are further away on average. A simple sample of working mechanism of KNN is shown in Fig. 3.

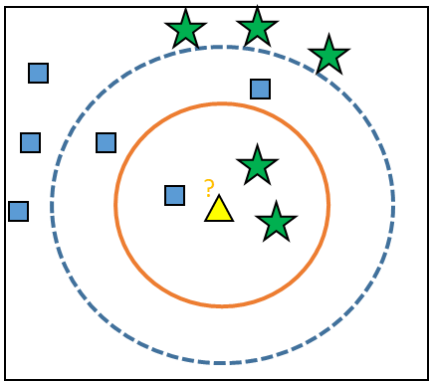

Fig 3. The KNN classification example

Each fitness function trains a model and evaluates its performance. Accuracy is one of the most important metrics in classification and it is calculated using Equation 1 for each class.

$$accuracy_i = \frac{TP_i+TN_i}{TP_i + FP_i + TN_i + FN_i} \quad (1)$$

Where i is the index of the class (label), TP_i is the count of correct predictions that actually belong to class i, and FP_i is the count of wrong predictions that actually belong to classes else than class i. In addition, TN_i is the count of correct predictions that actually belongs to classes else than class i, and FN_i is the count of wrong predictions that actually belong to class i.

Another measurement parameter of our study is the misclassification error, which is obtained as *'1-accuracy'*. A Receiver Operating Characteristic (ROC) curve is a graphical plot showing the ability of a binary classifier as the discrimination threshold changes. With this we can reach to Area Under the ROC Curve (AUC). The Other measurement parameters are also maximum accuracy, mean accuracy, standard deviation, and Average Precision (AP).

In the working mechanism of our study, first of all, there is the data cleaning and normalization process. Then there is the feature extraction and data preprocessing phase. Finally, three different classifier algorithms will be used, which are the most important and fundamental phase in our method. The schematic working mechanism of the proposed method is presented in Fig. 4.

© Copyright: All rights reserved.

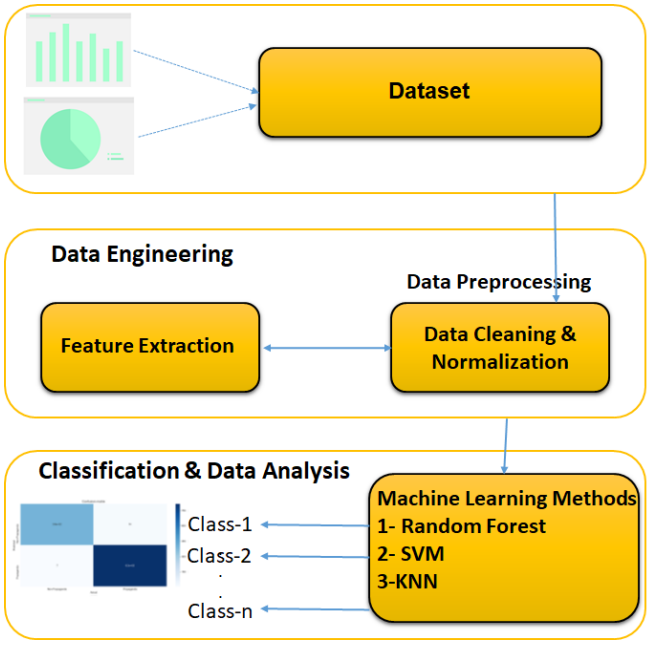

Fig 4. Proposed system mechanism

The results of the study are shown in the confusion matrix as Fig. 5, 6, and 7. According to the results, it is determined that the SVM method gives better results than the other two algorithms. During the training and testing of the proposed model, it was observed that there was neither insufficient fit nor overfitting. Percentage results with other output parameters are also presented in detail in Table 1.

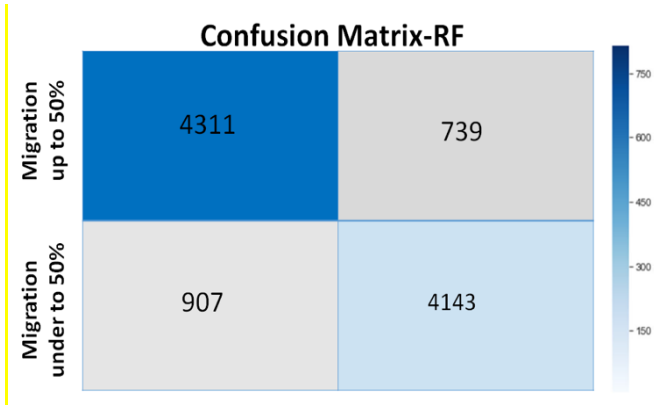

Fig. 5. Confusion matrix of RF

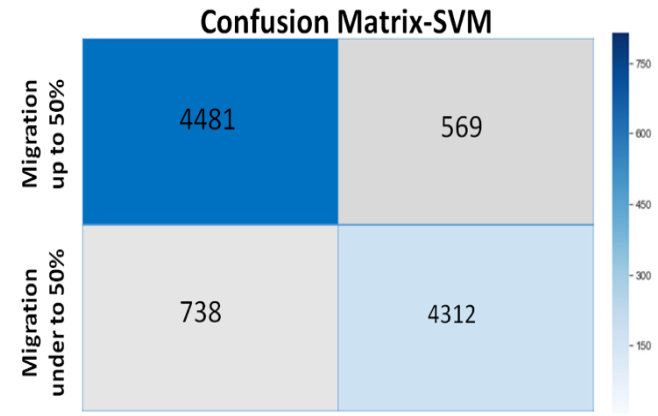

Fig. 6. Confusion matrix of SVM

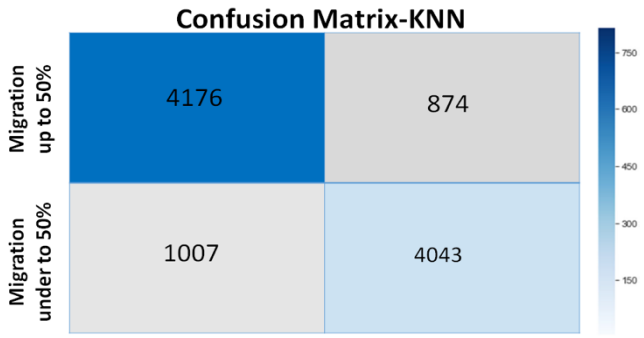

Fig. 7**. Confusion matrix of KNN**

The results of our method are presented in Table IV. As a comparison with other two machine learning methods (decision tree and naïve Bayes). In this comparison for discussing has been used six parameters. They are maximum accuracy, mean of accuracy, standard deviation, area under the curve rate, average precision rate and F1 score. The comparison of three methods are also presented in Fig. 6. The result of ROC curve is also shown in Fig. 8.

Table-I: Classification results of attributes obtained-Comparison

Classifier	Max Acc	Mean Acc	Std	AUC	AP
RF	0.85	0.82	0.08	0.81	0.82
SVM	0.89	0.86	0.04	0.85	0.88
KNN	0.83	0.81	0.04	0.80	0.82

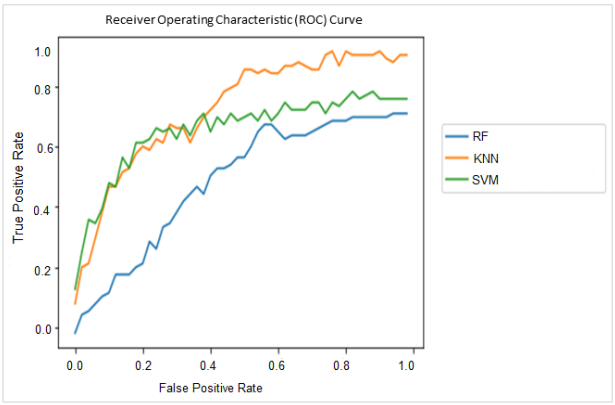

Fig. 8. ROC curve- RF, SVM, and KNN

IV. RESULT AND DISCUSSION

According to the results, humans in the relevant region have been exposed to migration for various reasons. These types of migration were evaluated at domestic and international levels. Problems such as unemployment and economic difficulties, the emergence of various diseases, the decrease in drinking water, and air pollution have caused these migrations. In our proposed method, we tried to create the best model using three different algorithms. According to the results, the SVM-based model learned better and proved its reliability in the validations too. In this direction, this model can make forward-looking predictions and, like an expert system, can guide the relevant researchers and even state or form ideas according to the results obtained from it.

© Copyright: All rights reserved.

The causes and processes that lead to migration can gradually upset the ecological and indigenous balance and lead to dangerous consequences. Accordingly, sustainable development, management of basic resources, protection of resources and realization of technical improvements in this area, introduction and implementation of appropriate organizational structures are very important. Revising sustainable agriculture and environmental policies is also of great importance in overcoming these problems. A population of approximately 5.700.000 lives in the lake basin, most of which are engaged in agriculture and some of them livestock, which is closely related to farming. However, it is possible to talk about an agricultural sector that almost consumes 94% of the basin water. When the agricultural life of the people of the region, who have no other source of income and whose economic activities are largely dependent on water, face a problem such as water scarcity, they will continue to live in the region to the extent withstand at best.

V. CONCLUSION

In this study, the problem of migration has been discussed in general terms. For this, Lake Urmia was considered as a case study. After the relevant data (10.100) was cleaned and normalized, three different algorithms were used in our proposed method at the last stage. According to the results, our SVM based model showed the best performance with 88% acc. Accordingly, this model can also be used in forward-looking predictions and can be useful as an expert system that can guide researchers (especially social scientists) interested in these areas. It can also declare or form opinions to those in government mechanisms. In future studies, a suitable application will be developed by using different machine learning methods (e.g. reinforcement learning-based methods [20]).

REFERENCES

1. S. SRO-Kundig, “Global commission on international migration, ‘migration in an interconnected world: new directions for action’,”. Geneva: Switzerland, 2005, pp. 2-19.
2. P. Fajnzylber and J. Humberto López, “Global economic prospects 2006: international remittances and migration”. The International Bank for Reconstruction and Development, The World Bank, Washington, DC: United States, 2008, pp. 21-51.
3. R., Dilip, “Workers’ Remittances: An important and stable source of external development finance”. Global Development Finance 2003, Washington, DC: United States, 2005, pp. 157-175.
4. Krause, K. and T. Liebig, “The labor market integration of immigrants in OECD countries and temporary migration,” OECD Working Party on Migration (2005), Paris: France, Vol. 12, 2011, pp.18-32.
5. D. Ushakov and T. A. Auliandri, “international labor migration in South Asia: current situation and the problems of efficient national regulation”, IOP Conf. Ser.: Materials Science and Engineering, Vol. 753, 2020, pp. 1-11.
6. D König, “Labor shortages and the need for immigrants: a review of recent studies,” OECD, Paris: France, 2006, pp. 2-17.
7. A. S. Miranda, “The shape of segregation: The role of urban form in immigrant assimilation”, Cities, Elsevier, Vol. 106, 2020, pp.1-18.
8. O. Kucuk and F. Kiani, “Smart advisor: An intelligent inventory prediction based on regression model”, International Journal of Machine Learning and Networked Collaborative Engineering, Vol. 2, No.3, pp. 86-94.
9. F. D. Khangahi and F. Kiani, “The role of social networks in the formation of social lifestyle changes caused by the Covid-19”, International Journal of Recent Technology and Engineering (IJRTE), Vol. 9, No. 5, 2021, pp. 263-267.
10. A. Beduschi, “International migration management in the age of artificial intelligence”, Migration Studies, 2020, pp. 1-21.
11. S. Azizi and K. Yektansani, “artificial intelligence and predicting illegal immigration to the USA”, International Migration, IOM, Wiley, 2020, pp. 1-12.
12. P. Molnar and L. Gill, “A human rights analysis of automated decision-making in Canada’s immigration and refugee system”. Bots at the Gate, 2018, pp. 1-14.
13. N. B. Lindström, D. Koutsikouri, J. Stier and M. Arvidsson, “Migrant employment integration and Artificial Intelligence (AI)”. The 32nd annual workshop of the Swedish Artificial Intelligence Society (SAIS) will be held as an online conference on June 16-17, 2020, pp. 1-5.
14. J.de Morgan, “Heyat-ı elmiye Fransa dar Iran, motaleate jografiyayi {Fransa’nın Iranda’ki akademik heyeti, cografya araştırmaları}”. Cev. Kazem Vadiei, Tebriz Çehr yayınları, Vol. 1, 1959, pp.355-356.
15. C. Cortes, and V. Vapnik. “Support vector networks”. Mach. Learn., vol. 20, 1995, pp.273-297.
16. J. Wu, X. Chen, H. Zhang, L. Xiong, H. Lei and S. Deng, “Hyperparameter optimization for machine learning models based on bayesian optimization”. Journal of Electronic Science and Technology, Vol. 17, No. 1, 2019, pp. 26-40.
17. H.D. Zhang, Z.H. Xing and L. Chen, “Efficient metric all-k-nearest-neighbor search on datasets without any index”. Journal of Computing Science Technology, Vol. 31, 2016, pp. 1194-1211.
18. V.N. Vapnik. “An overview of statistical learning theory”. IEEE Trans. Neural Network, vol. 10, 1999, pp.988-999.
19. W. Wang, Z. Xu, W. Lu, and X. Zhang. “Determination of the spread parameter in the Gaussian kernel for classification and regression”. Neurocomputing, vol. 55, 2003, pp.643-663.
20. F. Kiani, “Reinforcement learning based routing protocol for wireless body sensor networks”. IEEE 7th International Symposium on Cloud and Service Computing (SC2), 2017, pp. 71-78

AUTHORS PROFILE

Fatemeh Dehghan Khangahi, received bachelor, master and PhD at Azad university of Zanjan and Istanbul University, in 2006, 2009, and 2020. respectively. Her interest area is the influence of social media on social movements and human rights. In addition, migration, social mobilization, and ecosystem issues are included in her researches. She is a part time lecturer doctor in political science and government management department of Istinye University. Also, she is advisor in various research centers and human rights organizations and foundations. Besides these, other areas of Dr. Fatemeh's work is environment/water problems, Middle east and Africa studies, comparative political ideologies, race and racism, gender studies and feminism.

Farzad Kiani , received bachelor, master, and Ph.D. degrees in computer engineering department. He finished on 2006, 2009 and 2014 years, respectively. He is an Associate Prof. in computer engineering dept. at Istanbul Arel University. Dr. Farzad, who has many publications works on machine learning and trend technologies. His interest area is machine learning, wireless sensor networks, Internet of Things, Industrial 4.0 and IoT, Internet of Food and agriculture, vehicle ad-hoc networks, autonomous vehicles, reinforcement learning, optimization algorithms, metaheuristic, robot path planning, and data analysis. He is an active member in various indexed famous journals in IEEE, Elsevier and Springer. Also, he has some published scientific books.

© Copyright: All rights reserved.