Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı 44, S. 104-107, Aralık 2022

Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue 44, pp. 104-107, December 2022

Research Article

Violence Detection with Machine Learning: A Sociodemographic Approach

Tolga Ensari[1], Betul Ensari[2], Mustafa Dagtekin[3*]

[1] Arkansas Tech University, College of STEM, Departmant of Engineering and Computing Sciences, Russellville, USA, (ORCID: 0000-0003-0896-3058), tensari@atu.edu

[2] Arkansas Tech University, Graduate College, Information Technology (MSc Program), Russellville, USA (ORCID: 0000-0003-0425-7252), bensari@atu.edu

[3*] Istanbul University-Cerrahpasa, Faculty of Engineering, Departmant of Computer Engineering, Istanbul, Turkey, (ORCID: 0000-0002-0797-9392), dagtekin@iuc.edu.tr

(6th International Symposium on Innovative Approaches in Smart Technologies (ISAS) 2022 – 8-10 December 2022)

(**DOI:** 10.31590/ejosat.1225896)

ATIF/REFERENCE: Ensari T., Ensari B., & Dagtekin M., (2022). Violence Detection with Machine Learning: A Sociodemographic Approach. *European Journal of Science and Technology*, (44), 104-107.

Abstract

This study suggests that by implementing machine learning methods on a sociodemographic data set can be helpful in preventing domestic violence. This approach is important in predicting high-risk factors that an offender may cause and it offers treatment, and financial or mental health aids in order to prevent domestic violence. In this sense, this proposal is critical at a personal and social level in creating a secure and healthy environment as well as empowering an equal society. In our study, we use k-nearest neighbor (k-nn), support vector machine (SVM), decision tree (DT), and Gaussian Naive Bayes (GNB) machine learning algorithms for the prediction analysis. We provide the comparison of the classifiers with precision, recall, F1 score, and accuracy performance measures. According to our analysis, the decision tree (DT) performs the best performance in terms of accuracy.

Keywords: Violence detection, machine learning, sociodemographic, classification, prediction

Makine Öğrenmesi ile Şiddet Tespiti: Sosyodemografik Bir Yaklaşım

Öz

Bu çalışma, sosyodemografik bir veri seti üzerinde makine öğrenimi yöntemlerinin uygulanmasının aile içi şiddeti önlemede yardımcı olabileceğini önermektedir. Bu yaklaşım, failin neden olabileceği yüksek risk faktörlerini öngörmede önemlidir ve aile içi şiddeti önlemek için tedavi, maddi veya manevi sağlık yardımları sunar. Bu anlamda bu öneri, güvenli ve sağlıklı bir çevre yaratmanın yanı sıra eşit bir toplumu güçlendirmede kişisel ve toplumsal düzeyde kritik öneme sahiptir. Çalışmamızda tahmin analizi için k-en yakın komşu (k-EYK), destek vektör makinesi (DVM), karar ağacı (KA) ve Gaussian Naive Bayes (GNB) makine öğrenmesi algoritmalarını kullanılmıştır. Kesinlik, F1 skoru ve doğruluk performans ölçütleri ile sınıflandırıcıların karşılaştırmasını sağlıyoruz. Analizlerimize göre, karar ağacı (KA) doğruluk açısından en iyi performansı sergilemektedir.

Anahtar Kelimeler: Şiddet tespiti, makine öğrenmesi, sosyodemografi, sınıflandırma, tahmin

[*] Corresponding Author: dagtekin@iuc.edu.tr , Dagtekin M., Istanbul University-Cerrahpasa, Computer Engineering, Istanbul, Turkey

1. Introduction

Domestic violence has been one of the most important health issues that raise concerns all over the world. According to data conducted by world health organization (WHO), approximately 1 in 3 women and 1 in 10 men 18 years of age or older experience domestic violence worldwide. In the United States, annually over 1500 deaths happen due to domestic violence [1-3].

There are many studies that aim to decrease violence through maschine learning algorithms. Some of them analyze the messages sent by victims on online platforms such as facebook, tweeter in order to identify victims in critical need by using natural language processing and text mining. [4-6,7] Some studies look into videos to detect physical violence by using classification techniques such as traditional violence detection using machine learning, support vector machine (SVM) and deep Learning. Feature extraction and object detection techniques, k-nearest neighbors (k-nn), random forest classifier [8-14]. Another significant artificial intelligence method in violence detection is using machine learning algorithms in identifying stress, panic, and fear in speech through physiological sensory data, speech, and audio analysis [16].

Differently from the above mentioned methods, this study proposes to use social-demographic data set to pre-detect high risk factor in domestic violence. according to some research, domestic violence is triggered by some social and psychological reasons such as poverty, lack of education, cultural norms, post-traumatic stress disorder (PTSD), etc. [17]. The research signifies that by healing and treating offenders, domestic violence can be abolished. Considering the approach underlined by these researches, in this study, it is suggested that using a socio-demographic data set can be also helpful in preventing domestic violence [18-22]. Based on the data set, it can be predicted how high the rate at an intimate partner may cause a violent action. To demonstrate how our proposed method/system makes early predictions, a synthetic data set has been created. By identifying the rate of violence tendency in an intimate partner, this method/system aims to give early feedback to the decision-makers in social workers, law enforcement, and clinics where these individuals can be provided with education, financial aid, and medical care.

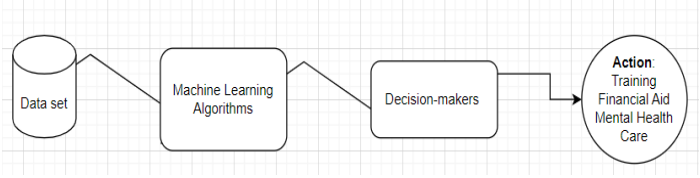

Fig. 1. The proposed method/system for domestic violence detection

We also illustrate the proposed system in Fig. 1. This platform can work with this architecture or can be improved adding new stakeholders into the design. Machine learning algorithms work as an automatic detector as an artificial intelligence engine.

2. Material and Method

2.1. Data Set

There is no data set in the literature for this kind detection system. Therefore, we created a new synthetic data set to use in the prediction and early detection analysis. This data set contains a sociodemographic information of the individuals. We use mental health, income, drug use, and education level in the data set. We rate mental health from 0 to 5, income from 0 to 5, drug use from 0 to 2, and education level from 1 to 5. Highest values show the best conditions for the individuals.

mental_health_0_5	income_0_5	drug_0_2	Educ_1_5	class
2	3	1	2	High-risk
4	0	2	3	High-risk
0	3	1	4	Low-risk
1	5	1	2	Low-risk
5	4	0	5	No-risk
5	4	0	5	No-risk

Fig. 2. The sample tuples (lines) from the data set

Sample tuples from the data set can be seen in Fig. 2. We cathegorize it with high risk, low risk, and no risk labels. There are 240 individuals in this synthetic data set. In real world, governmental facilities, health clinics, law enforcement units, counselors, family centers, educational institutes, and other official centres can collect and process this new data set in their systems. After collection of the real world data sets, they can be proceed with machine learning approaches to have their early warning/detection platform.

2.2. Machine Learning Methods

We use k-nearest neighbours (k-nn), support vector machines (SVM), decision tree (DT), and Gaussian Naive Bayes (GNB) algorithms as a classifier to predict the violence with our data set. We use Python programming language for the experimental analysis.

The k-nearest neighbour (k-nn) method is a non-parametric classification algorithm that groups data with neighbour distance measures. An unknown data point is assigned the mojority of the class labels. The k value shows the number of the neoighbour for that spesific unknown data point. The support vector machine method is an supervised algorithm that uses borders (hyper planes) between the classes. The margins betwwen hyperlanes are optimized to have a better classification result. Also, a kernel function can also be used for the SVM method. The decision tree algorithm is also another non-parametric method that has a tree structure with nodes and leaves. It is one of the supervised learning algorithms. The Gaussian Naive Bayes method is another supervised learning algorithm and it uses a Gaussian distribution for the classification and prediction tasks.

2.3. Results

We use k-NN, SVM, DT, and GNB supervised learning algorithms in our experiments. Our system finds the prediciton results of the individuals (candiates) as the high risk, low risk, and no risk groups (labels). After the training and testing process, these algorithms have results for the proposed scenario. We use precision, recall, F1 score and accuracy metrics to measure performance of the classification methods. According to

the experiments, decision tree classifier performs the best result in terms of the accuracy. Also other performance metric values and accuracy values can be seen from Table 1 below. SVM gives the second best, GNB gives the third best, and k-nn gives the worst accuracy values in our analysis. The algorithms/methods can also be classified in terms of the other metrics in Table 1.

Table 1. Prediction results of machine learning algorithms (classifiers) for the violence detection

Machine learning methods (algorithms)		*Precision*	*Recall*	*F1 score*	*Accuracy (%)*
k-nearest neighbours (k-nn)	High-risk	0.57	0.67	0.62	75.0
	Low-risk	0.75	0.75	0.75	
	No-risk	0.89	0.80	0.84	
Support vector machines (SVM)	High-risk	0.62	0.83	0.71	79.16
	Low-risk	0.75	0.75	0.75	
	No-risk	1.00	0.80	0.89	
Decision trees (DT)	High-risk	1.00	1.00	1.00	**91.66**
	Low-risk	0.80	1.00	0.89	
	No-risk	1.00	0.80	0.89	
Gaussian Naive Bayes (GNB)	High-risk	0.71	0.83	0.77	79.10
	Low-risk	0.71	0.62	0.67	
	No-risk	0.90	0.90	0.90	

We also illustrate the accuracy prediction performance in Fig. 2. The result values can be seen from this figure.

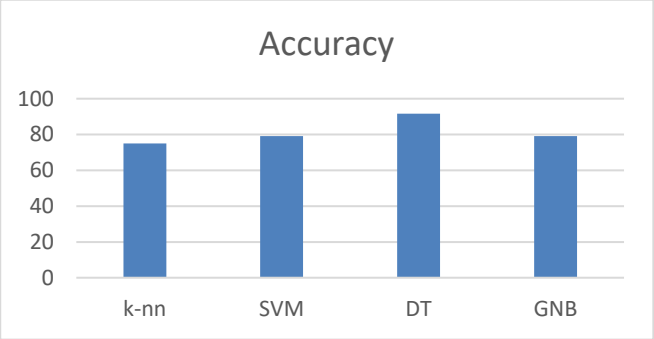

Fig. 2. Accuracy performance comparison of the machine learning algorithms

4. Conclusions and Recommendations

This study presented an approach to using machine learning methods on a sociodemographic data set to predict how high a risk factor an offender may pose to their family members or intimate partners. Our goal was to provide some help and healing to the offender in order to prevent domestic violence by collecting a socio demografic data to understand the sociological and psychological reasons that trigger violent actions. Based on the prediction value, the machine learning algorithms give feedbacks to social workers, law enforcement, and clinics. According to the feedback drawn from the data set, social workers, law enforcement, and clinics can provide the offender some help such as training, and financial or mental health aids. In this sense, this proposal is critical at personal and social level in creating a secure and healthy environment as well as empowering an equal society. In our study, we use k-nearest neighbour (k-nn), support vector machine (SVM), decision tree (DT), and Gaussian Naive Bayes machine learning algorithms for the prediction analysis. The performance comparison of the classifiers has been given with precision, recall, F1 score, and accuracy performance measures. Decision tree (DT) performs the best performance in terms of accuracy.

References

[1] Boserup, B., McKenney, M., & Elkbuli, A. (2020). Alarming trends in US domestic violence during the COVID-19 pandemic. *The American Journal of Emergency Medicine*, *38*(12), 2753-2755.

[2] Wahi A, Zaleski KL, Lampe J, Bevan P, Koski A. The Lived Experience of Child Marriage in the United States. Soc Work Public Health. 2019; 34(3):201-213.

[3] Harland KK, Peek-Asa C, Saftlas AF. Intimate Partner Violence and Controlling Behaviors Experienced by Emergency Department Patients: Differences by Sexual Orientation and Gender Identification. J Interpers Violence. 2021 Jun; 36(11-12): NP6125-NP6143.

[4] Subramani, S., Wang, H., Vu, H. Q., & Li, G. (2018). Domestic violence crisis identification from facebook posts based on deep learning. *IEEE Access*, *6*, 54075-54085.

[5] Subramani, S., Michalska, S., Wang, H., Du, J., Zhang, Y., & Shakeel, H. (2019). Deep learning for multi-class identification from domestic violence online posts. *IEEE Access*, *7*, 46210-46224.

[6] Le Glaz, A., Haralambous, Y., Kim-Dufor, D. H., Lenca, P., Billot, R., Ryan, T. C., and Lemey, C. (2021). Machine learning and natural language processing in mental health: systematic review. *Journal of Medical Internet Research*, *23*(5).

[7] Xue, J., Chen, J., and Gelles, R. (2017). Using Data Mining Techniques to Examine Domestic Violence Topics on Twitter. Violence and Gender, 6 (2), 105–114.

[8] Siddique, M., Islam, M., Sinthy, R., Mohima, K., Kabir, M., Jibon, A. H., and Biswas, M. (2022). State-of-the-Art Violence Detection Techniques: A review. *Asian Journal of Research in Computer Science*, 29-42.

[9] Ye, L., Wang, L., Wang, P., Ferdinando, H., Seppänen, T., Alasaarela, E. (2018). Physical Violence Detection with

Movement Sensors. Machine Learning and Intelligent Communications. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 251. Springer, Cham.
[10] Vashistha, P., Singh, J.P., Khan, M.A. (2020). A Comparative Analysis of Different Violence Detection Algorithms from Videos. In: Kolhe, M., Tiwari, S., Trivedi, M., Mishra, K. (eds) Advances in Data and Information Sciences. Lecture Notes in Networks and Systems, vol 94.
[11] Deepak, K., Vignesh, L. K. P., and Chandrakala, S. J. I. E. (2020). Autocorrelation of gradients based violence detection in surveillance videos. *ICT Express*, *6*(3), pp 155-159.
[12] S. Das, A. Sarker and T. Mahmud, "Violence Detection from Videos using HOG Features," *2019 4th International Conference on Electrical Information and Communication Technology*, 2019, pp. 1-5.
[13] A. Jain and D. K. Vishwakarma, "State-of-the-arts Violence Detection using ConvNets," *2020 International Conference on Communication and Signal Processing*, 2020, pp. 0813-0817.
[14] Chaudhary, D., Kumar, S., and Dhaka, V. S. (2022). Video based human crowd analysis using machine learning: a survey. *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, *10*(2), 113-131.
[15] Jiang Y, DeBare D, Colomer I, Wesley J, Seaberry J, Viner-Brown S. Characteristics of Victims and Suspects in Domestic Violence-Related Homicide- Rhode Island Violent Death Reporting System, 2004-2015. R I Med J (2013). 2018 Dec 03; 101(10):58-61.
[16] Rituerto-González, E., Mínguez-Sánchez, A., Gallardo-Antolín, A., & Peláez-Moreno, C. (2019). Data augmentation for speaker identification under stress conditions to combat gender-based violence. *Applied Sciences*, *9*(11), 2298.
[17] Evans, S. (2005). Beyond gender: Class, poverty and domestic violence. *Australian Social Work*, *58*(1), 36-43.
[18] Wallach, H. S., and Sela, T. (2008). The importance of male batters' attributions in understanding and preventing domestic violence. *Journal of Family Violence*, *23*(7), 655-660.
[19] Guerrero, A., Cárdenas, J. G., Romero, V., & Ayma, V. H. (2020). Comparison of Classifiers Models for Prediction of Intimate Partner Violence. In Proceedings of the Future Technologies Conference, pp. 469-488, Springer, Cham.
[20] Abramsky, T., Watts, C. H., Garcia-Moreno, C., Devries, K., Kiss, L., Ellsberg, M., and Heise, L. (2011). What factors are associated with recent intimate partner violence, Findings from the WHO multi-country study on women's health and domestic violence. BMC public health, 11(1), 1-17.
[21] Vyas, S., and Watts, C. (2009). How does economic empowerment affect women's risk of intimate partner violence in low and middle income countries. A systematic review of published evidence. Journal of International Development: The Journal of the Development Studies Association, 21(5), 577-602.
[22] Leonard, K. E. (2005). Alcohol and intimate partner violence: when can we say that heavy drinking is a contributing cause of violence, Addiction, vol 100. no. 4, pp 422-425.