Computer Science and Interdisciplinary Research Journal
Volume 1, Issue 1, ISSN: 3033-1218, DOI: 10.70862/CSIR.2024.0101-03, 2024

Analysis of Suicide Risk in European Countries Using Artificial Intelligence Methods

Valentin Georgiev[1], Kostadin Yotov[2]

University of Plovdiv Paisii Hilendarski, 236, Bulgaria Blvd, 4027 Plovdiv, Bulgaria

[1] valentin_georgiev02@abv.bg, [2] kostadin_yotov@uni-plovdiv.bg

Abstract – Suicides are a phenomenon that negatively impacts society. The issues include the loss of human life, negative psychological effects on close ones, increased healthcare costs, lost productivity, and more. Suicides also create social and emotional tensions that can deteriorate the quality of life in affected communities. This is why it is necessary to study and research trends to enable prevention.

In the present study, an analysis of suicide risk in some European countries is conducted using an artificial intelligence system implemented with Neural Networks, Random Forests, Support Vector Machines, and more. The study seeks answers to the following questions: Are suicide rates increasing or decreasing? Are there differences in the number of suicides based on age and gender? Do economic factors affect the number and rate of suicides? Can we predict the number of suicides based on the given data? The machine learning methods used for this purpose are created and modeled in the Orange Data Mining environment, which also offers numerous tools for analyzing and visualizing the obtained results. It is important to note that the study of suicide risk is not equivalent to the diagnosis of suicidal disorders; however, the answers to the questions posed in the study are crucial for building and analyzing the overall picture of suicidal intentions.

Keywords – suicides, suicide risk, Europe, ANN.

DOI: https://doi.org/10.70862/CSIR.2024.0101-03

Corresponding author: Kostadin Yotov,
University of Plovdiv Paisii Hilendarski
Email: kostadin_yotov@uni-plovdiv.bg

Received: 12 July 2024.
Revised: 15 August 2024.
Accepted: 16 August 2024.
Published: 17 August 2024.

© 2024 Valentin Georgiev, Kostadin Yotov. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

The article is published with Open Access at http://csirjournal.com/

1. Significance of issues related to suicidal intentions

Issues related to suicidal intentions are of tremendous significance, as suicide is a serious social and health challenge. It is not only a matter of individual human suffering but also of our societal responsibility towards those battling severe mental disorders often leading to life-threatening actions. Suicidal intentions are frequently the result of prolonged suffering, depression, anxiety, and other mental disorders [1-3]. They have a significant impact not only on the individual but also on their immediate surroundings and the community as a whole [4]. The loss of human life due to suicide represents a grave loss for society from both a human and socio-economic perspective [5]. Therefore, combating suicidal intentions requires a comprehensive approach that includes analysis, prevention, early detection, and effective treatment of mental illnesses.

Preventing suicidal thoughts is essential not only in a medical context but also in educational and societal contexts. Individuals struggling with suicidal intentions often experience intense psychological pain and a sense of hopelessness. Depressive states are among the common symptoms, often accompanied by real physical pain [6, 7]. Hence, support and professional help are crucial for suicide prevention, with the first step being the analysis of these phenomena. Society must engage in various initiatives to raise awareness, reduce the stigma surrounding mental illnesses, and provide support to those in crisis situations. Suicidal intentions should not go unnoticed or ignored, and the issues should be analyzed with understanding. The challenges associated with suicidal intentions require joint efforts from the healthcare system, educational institutions, government bodies, and society as a whole to create better understanding and response to this serious problem.

The application of new technologies and methods such as artificial intelligence can play a crucial role in

the prevention and management of suicidal intentions through better understanding and risk analysis [8-10]. Building effective systems for analyzing suicidal intentions can provide us with a real picture of the situation and help in making the right decisions to address these problems. *The aim of the presented research is to create machine learning models to analyze data on suicide rates in European countries accumulated during the period 1992-2022.* Throughout the research, various approaches were used, including Support Vector Machines (SVM), Linear Regression (LR), Decision Trees (DT), k-Nearest Neighbors (kNN), Random Forest (RF), and Artificial Neural Networks (ANNs). These were implemented using data analysis software with extensive capabilities for graphical and tabular visualization of the analysis – Orange Data Mining. This article presents the results of two methods that interpolate the data with the least error and the highest reliability in the analysis. Predictions of suicidal intentions in the studied regions were made using RF and ANNs.

2. Machine learning methods used in the study

Several different machine learning methods were used for the purposes of the study to assess suicide risk in European countries. All the models used were created in the Orange Data Mining application, which offers excellent capabilities for analyzing and visualizing the results.

2.1. Support Vector Machine

SVM is a machine learning algorithm used for both classification and regression tasks. The main idea of SVM is to find a hyperplane in a multidimensional space that separates different classes of data. However, in many cases, the classes are not linearly separable, making it difficult to construct and use such a hyperplane initially.

In SVM, so-called "induced feature spaces" are used through of kernels, which transform the data into higher-dimensional feature spaces where they can become more easily separable. The selection of kernels in Orange Data Mining can be easily done through settings window (fig. 1). The kernel can be:

- *Linear:* Used for linearly separable data.
- *Polynomial:* Suitable for more complex, nonlinear separations.
- *Radial Basis Function (RBF)*: Preferred when the data are not linearly separable.
- *Sigmoid:* Used less frequently but is an option for certain specific tasks.

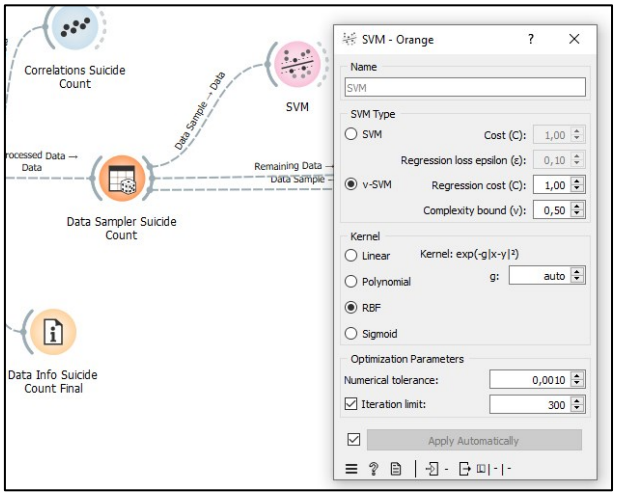


Figure 1. Setting the parameters of SVM in Orange

2.2. Random Forest

RF is a machine learning method that can be used for various tasks related to data analysis, predictions, classification, regression, and more. By combining decisions from multiple trees, RF provide high accuracy and robustness while reducing the risk of overfitting.

Each tree is trained on a different subset of data, which helps increase diversity and reduce correlation between individual trees. In the Orange, various parameters such as the number of trees, the number of attributes in each partition, the maximum depth of individual trees, etc. can be adjusted to optimize RF performance (fig. 2)

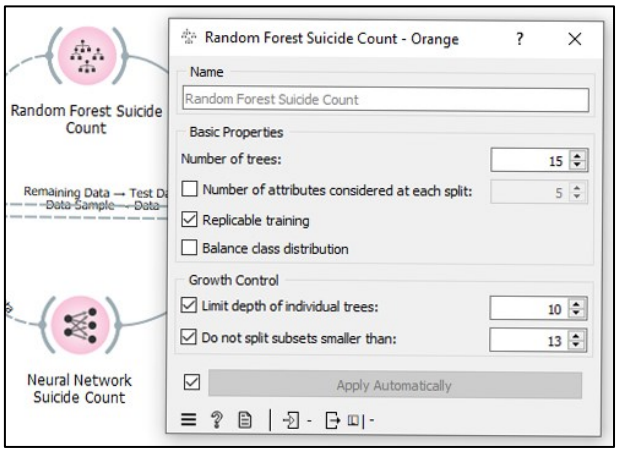


Figure 2. Setting the parameters of RF

2.3. Artificial Neural Networks

ANNs find applications in many areas such as computer vision, speech recognition, natural language processing, and more.

The Orange Data Mining software offers capabilities for constructing feedforward ANNs, with options for defining multiple hidden layers, selecting the number of neurons in each layer, choosing the

training method, the number of epochs, and more (fig. 3). The following transfer functions are available:

- Linear: $f(x) = x$;
- Logistic: $f(x) = \frac{1}{1+e^{-x}}$;
- Hyperbolic tangent: $f(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}}$, and
- Rectified Linear Unit (RELU): $\mathrm{f(x)} = \max(0, \mathrm{x})$.

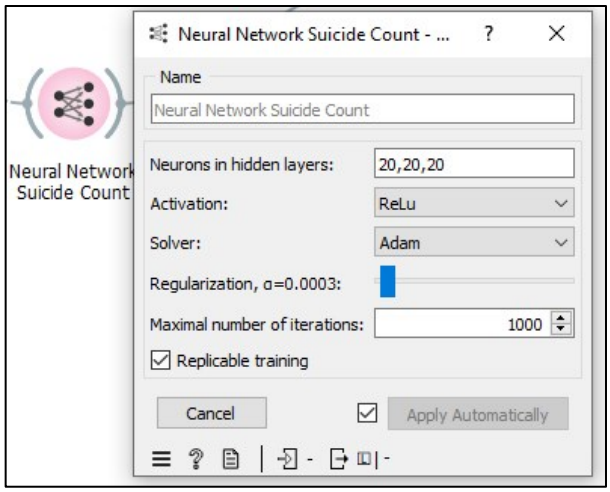


Figure 3. Setting the parameters of an ANN in Orange

The supported training methods for ANNs are:

- Limited-memory Broyden–Fletcher–Goldfarb–Shanno (L-BFGS-B) algorithm;
- Stochastic Gradient Descent (SGD);
- Adam optimization algorithm, an extension of Stochastic Gradient Descent [11].

3. Criteria for the effectiveness of the methods used

To determine the effectiveness of the methods used for analyzing and interpolating the levels of suicide risk, the following errors were used:

- Mean Squared Error

$$(1) \quad MSE = \frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2$$

- Root Mean Squared Error ($RMSE$),

$$(2) \quad RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}$$

- Mean Absolute Error (MAE)

$$(3) \quad MAE = \frac{\sum_{t=1}^{n}|y_t - \hat{y}_t|}{n}$$

Additionally, we use the coefficient of determination (R^2) related to the correlation between the model responses and the actual data.

$$(4) \quad R^2 = 1 - \frac{\sum_{i=1}^{n}(y_t - \hat{y}_t)^2}{\sum_{i=1}^{n}(y_t - \bar{y}_t)^2}$$

In the given formulas, *n* is the sample size available (the number of input-output training samples), $(y_t - \hat{y}_t)^2$ are the squares of the differences during the interpolation of the respective nodes, and $(y_t - \bar{y}_t)^2$ are the squares of the residuals in the calculation of the R^2 coefficient.

4. Analysis and assessment of suicide intent levels in European countries

For the purposes of the study on suicide risk analysis, several different methods were used: LR, SVM, kNN, DT, RF, and ANN. The approaches with the best performance in data interpolation and result analysis were found to be RF and ANN.

The dataset used [12] contains nearly 120,000 records collected from the World Health Organization and the World Bank, featuring various social and economic characteristics. It includes the following columns: *RegionCode, RegionName, CountryCode, CountryName, Year, Sex, AgeGroup, Generation, SuicideCount, CauseSpecificDeathRate, DeathRatePer100K, Population, GDP, GDPPerCapita, GrossNationalIncome, GNIPerCapita, InflationRate, EmploymentPopulationRatio*. The columns *RegionCode* and *CountryCode* were removed at the beginning of the study due to duplication with other characteristics. The columns *GrossNationalIncome* and *GNIPerCapita* were removed later on as they had high correlation with GDP and GDPPerCapita.

The dataset contains three features related to the number of suicides: *SuicideCount* (the number of suicides in one year), *DeathRatePer100K* (the number of suicides per 100,000 people in one year), and *CauseSpecificDeathPercentage* (the percentage of deaths in one year due to suicide). All three features have different values depending on gender, age, country, and other factors for a given year.

Although the dataset contains data for the entire world, we have conducted this study focusing on Europe, which reduces the data volume to approximately 45,000 records. The main questions we sought to answer are:

1. *What is the trend for suicides - are they increasing or decreasing?* Figures 4 and 5 respectively show the number and rate of suicides. In these figures, the size of the symbols corresponds to the magnitude of the characteristic being examined, the color represents gender, and the type of symbol represents the age group. We can see that both the number of suicides and the suicide rate have been decreasing over the years.

2. *Is there a correlation between age and gender on one hand, and the number and rate of suicides on the other?* Tables 1 and 2 provide descriptive statistics for the data from 2015 to 2022. Both the figures and Tables 1 and 2 show clear answers to these questions. Men tend to commit suicide more frequently, both in terms of quantity and rate. Regarding age, the highest number of suicides is observed among people aged 25 to 54, but this age group has the lowest rate. For the population over 55 years old, we observe a lower number of suicides but the highest rate. The youngest age group has the lowest number of suicides but ranks second in terms of the suicide rate.

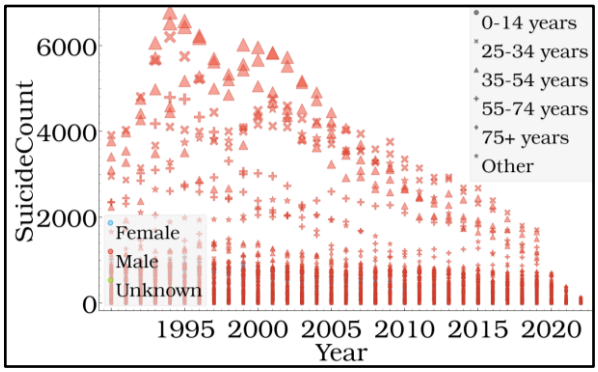


Figure 4. Number of suicides in Europe

Table 1. Statistics on the number of suicides in Europe

Number of suicides for:	Mean	Median
Men	84.34	25
Women	23.38	7
Ages 15 to 24	36.26	11
Ages 25 to 54	85.88	26
Ages over 55	69.28	26

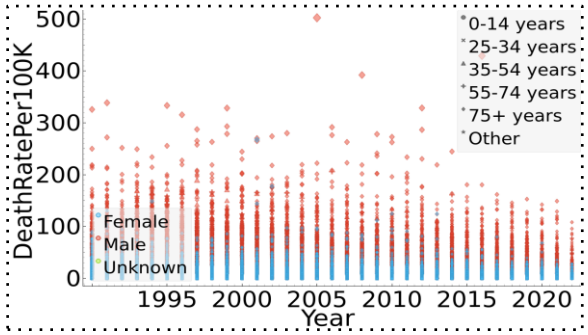


Figure 5. Suicide rate in Europe per 100,000 People

Table 2. Statistics on the suicide rate in Europe per 100,000 People

Number of suicides for:	Mean	Median
Men	20.58	16.95
Women	5.17	4.10
Ages 15 to 24	12.82	9.17
Ages 25 to 54	7.33	5.19
Ages over 55	21.74	14.00

3. *Do economic factors affect the number and rate of suicides?* Table 3 presents correlation coefficients and the ranking of characteristics (with highly correlated characteristics removed) according to RReliefF. From the data, we can conclude that there is no linear or monotonic relationship between economic factors and suicide metrics. According to their ranks, we see that the inflation rate does not have a significant impact, GDP and GDP per capita have an influence, although not the most significant, while the employment rate has a noticeable significance regarding all three suicide metrics. It should be noted that due to the low correlation values, this significance is nonlinear.

4. *Can we predict the three suicide-related characteristics using machine learning methods?*

To answer the last question, we selected the three characteristics *SuicideCount, DeathRatePer100K,* and *CauseSpecificDeathPercentage,* and built the models three times, each time choosing one of the characteristics as the target and excluding the others.

In preparing the data, we first excluded all characteristics that duplicated information in any way or had a high correlation with other characteristics, as previously mentioned. We removed all records with missing values due to the large volume of data and normalized the predictor data to the range [-1,1]. Finally, we split the remaining data into training and testing sets in a 70/30 ratio.

For the analysis of suicide risk, we used several different methods: LR, SVM, RF, Gradient Boosting, and ANN. For training and parameter tuning, we used cross-validation, splitting the test data into 5 parts.

The approaches with the best performance in data interpolation and result analysis were found to be RF and ANNs. The parameters of individual machine learning models are determined using the features provided by the software, which are specific to each widget. In order to achieve the necessary accuracy when using neural networks, experiments were conducted by changing and adjusting the following parameters offered by the Orange Data Mining software:

- Neurons per hidden layer**:** defined as the i-th element represents the number of neurons in the i-th hidden layer;
- Activation function for the hidden layer;
- Solver for weight optimization;
- L2 penalty (regularization term) parameter;
- Maximum number of iterations;

The main parameters of the RF used for model tuning are as follows:

- Number of trees;
- Number of trees considered at each split;
- Replicable training;

- Balance class distribution;
- Growth control: Limit depth of individual trees;

In our case, the most suitable model turned out to be an artificial neural network with three hidden layers, each containing 20 neurons. All neurons in the hidden layers are equipped with the ReLU transfer function, and the network is trained using the Adam method. The optimal random forest contains 15 trees, and the limit depth of individual trees is 10.

The results of these models are shown in Table 3. As we can see, the results are satisfactory for the models predicting the number and rate of suicides. The models for the third characteristic, *CauseSpecificDeathPercentage*, perform noticeably worse. This is understandable, as this metric is significantly more complex in nature than the others. It turns out that with the information available to us, we cannot predict what percentage of deaths in a given year are due to suicide when limited to a specific country, age group, and gender.

Table 3. Correlation and rank according to RReliefF. The first value is Pearson correlation, the second is Spearman correlation, and the third is the rank according to RReliefF

	SuicideCount			**DeathRatePer100K**			**CauseSpecific DeathPercentage**		
GDP	+0.189	+0.367	5	-0.079	+0.008	6	+0.006	+0.106	5
GDPPerCapita	-0.096	-0.043	4	-0.124	-0.065	5	+0.164	+0.094	3
InflationRate	+0.049	+0.059	8	+0.063	+0.094	8	-0.022	-0.040	7
Employment PopulationRatio	+0.017	-0.011	3	-0.015	+0.028	4	+0.124	+0.069	2

Table 4. Results of the best models on the test set
Abbreviations used: SC – the target characteristic is SuicideCount; DR – the target characteristic is DeathRatePer100K; CSDP – the target characteristic is CauseSpecificDeathPercentage

№	**Model**	**RMSE**	**MAE**	R^2
1	SC ANN	79.579	20.737	0.931
2	SC RF	86.422	17.842	0.919
3	DR ANN	9.191	4.450	0.855
4	DR RF	9.554	4.594	0.843
5	CSDP ANN	4.585	2.213	0.686
6	CSDP RF	4.742	2.297	0.664

5. Conclusion

In this study, we examined the socially significant issue of suicides. We used social and economic data for Europe, covering the period from 1992 to 2022. We found that the number and rate of suicides are decreasing. Men are more prone to suicide compared to women. The highest number of suicides is observed in the middle-aged population, but the rate is the lowest. The highest rate is observed in the elderly population. For all three suicide metrics, there is no linear or monotonic relationship. The most significant characteristic is the employment rate, followed by GDP and GDP per capita. The inflation rate is relatively insignificant.

Using machine learning methods, we built various models. The most successful ones, Artificial neural networks and Random Forests, performed satisfactorily in predicting the number and rate of suicides. Predicting the percentage of deaths due to suicide proved to be too complex to forecast with the available information.

Acknowledgements

This work was partly funded by the MUPD23-FMI-021 project of the Research Fund of the University of Plovdiv "Paisii Hilendarski".

Data Source

The original data used in the study are openly available at the websites of the World Health Organization and the World Bank, as follows:
https://www.who.int/health-topics/suicide
https://data.worldbank.org

References

[1] P. Dome, Z. Rihmer and X. Gonda, "Suicide Risk in Bipolar Disorder: A Brief Review", Medicina, vol. 55(8), 403, 2019. https://doi.org/10.3390/medicina55080403

[2] K. Wetherall, K. Robb and R. O'Connor, "Social rank theory of depression: A systematic review of self-perceptions of social rank and their relationship with depressive symptoms and suicide risk", Journal of Affective Disorders, vol. 246, pp. 300-319, 2019. https://doi.org/10.1016/j.jad.2018.12.045

[3] V. Fox, Ch. Dalman, H. Dal, A.-C. Hollander, J. Kirkbride and A. Pitman, "Suicide risk in people with post-traumatic stress disorder: A cohort study of 3.1 million people in Sweden", Journal of Affective Disorders, vol. 279, pp. 609-616, 2021. https://doi.org/10.1016/j.jad.2020.10.009

[4] M. Sajan, K. Kakar and U. Majid, "Social Interactions While Grieving Suicide Loss: A Qualitative Review of 58 Studies", The Family Journal, vol. 30 (3), 401-410, 2021. https://doi.org/10.1177/10664807211052492

[5] T. Bak, K. Kardis and D. Trong, "Suicide and Society: The Sociological Approach", Clinical Social Work and Health Intervention, vol. 10 (3), pp. 63-69, 2019. https://doi.org/10.22359/cswhi_10_3_02

[6] M. Trivedi, "The Link Between Depression and Physical Symptoms", The Primary Care Companion to The Journal of Clinical Psychiatry, vol. 6, pp. 12-16, 2004. PMID: 16001092.

[7] C. Arango-Dávilaa, H. Rincón-Hoyos, "Depressive disorder, anxiety disorder and chronic pain: Multiple manifestations of a common clinical and pathophysiological core", Revista Colombiana de Psiquiatría, vol. 47(1), pp. 46-55, 2018. https://doi.org/10.1016/j.rcpeng.2017.12.003

[8] V. Kumar, K. Sznajder and S. Kumara, "Machine learning based suicide prediction and development of suicide vulnerability index for US counties", Npj Mental Health Research, vol. 1, 2022. https://doi.org/10.1038/s44184-022-00002-x

[9] M. Ansari, S. Tiwari, S. Gupta, H. Vardhan, V. Shrivastava, "Suicide Rate Prediction Using Machine Learning", International Journal of Advances in Engineering and Management, vol. 4(5), pp. 2731-2737, 2022.

[10] N. Ismail, F. Razak and S. Masrom, "Suicide Prediction Using Machine Learning Techniques and Interactive Visualization of Suicide Information", International Journal of Academic Research in Business and Social Sciences, vol. 13(5), pp. 1905 - 1917, 2023.

[11] D. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization", The 3rd International Conference for Learning Representations, San Diego, May 7-9, 2015.

[12] R. Onyango, Suicide Rates & Socioeconomic Factors (1990 - 22) [Data set]. Kaggle, 2024. https://doi.org/10.34740/KAGGLE/DS/4597596