Brief Analysis of the Evolution of Female Employees in Recent Years. Research Using Mathematical Modelling

Constantin Ilie
Margareta Ilie
"Ovidius" University of Constanta, Romania
ilie.marga@gmail.com

Abstract

The numbers of male against female employment are still an actual and sensitive issue. Thus, the forecast of male and female employment evolution can offer the possibility to make decisions for the minimization of female employment disadvantages. One of the best used methods for model and simulation to forecast is the mathematical applications as artificial neural networks. This kind of method offers the possibility to enhance, understand and forecast the evolutions and influences between different data values. The objective of this paper is to find the proper artificial neural network for modeling and simulating of female employment evolution, under the influence of different indices. The structure considered the best for the simulated values was using hyperbolic tan function and quasi-Newton solver. Also, the ANN structures provide an average feature importance of 51.62% for Total investment, 19.95% for GDP, 15.06% for GERD and 13.37% for Production values.

Key words: female employment, data model, artificial neural network (ANN)
J.E.L. classification: J21

1. Introduction

Artificial neural networks (ANN) have demonstrated their usefulness in forecasting, clustering simulation, etc. ANN is now an important tool that overcome majority of other mathematical models and algorithms in processing speed, level of simulation, accuracy and level of mathematical integration.

Following the analysis of present research, about the employment of Romanian female, the results look to conclude in the same way: the employment is far to equal the men employment and, even more, in the last years the values of Romanian female employment drop further.

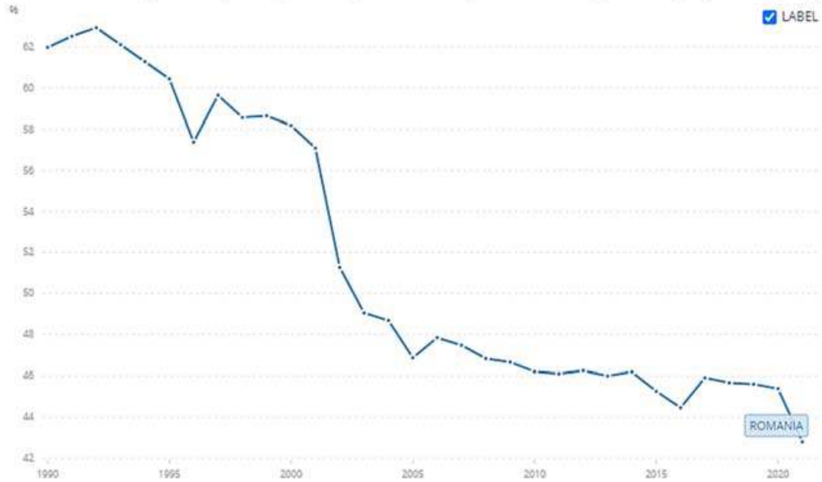
The present data values belong to years 2011-2021 and consider the Romanian employment as total, as males and as females. Also, in order to analyze how these values are influenced, more data were implemented in the model and simulation: GERD by sector of performance (Total intramural expenditure on R&D performed during a specific reference period, broken down by the institutions corresponding to each sector (business enterprise, government, higher education and private non-profit organizations), independent of the source of funds), Gross domestic product at market prices, Total investment in industrial field and Production value.

Thus, the main objective of the present paper is to apply mathematical application in form of artificial neural network. The ANN will be elaborated in several structures to find the one with the best result in training (the smallest training error) and show the feature importance that each ANN give to the input indices have in the training process.

2. Literature review

Females continue to be a disadvantaged social group at regional and national level and their existence in the labor market is insufficiently noticeable (Ştefliac, 2015). According to Word Bank (2022), participation rate of the female to the labor force decreases with 20% from 1990 to 2021, see figure 1.

Figure no. 1. Evolution of labor force participation rate, female (% of female population ages 15+).



Source: World Bank Group, 2022

Also, according to European Institute for Gender Equality (2017), after an analyze on gender equality in Romania between 2005-2015, the equality has seen a slight decline, as gender equality gender gaps in labor market participation and segregation have widened. The employment rate (20-64 years) is 57% for female compared to 74% for male. When the number of hours worked is considered, the full-time equivalent (FTE) employment rate for female is around 41%, compared to 58% for male.

The use of artificial neural networks for the mathematical model of employment evolution overcomes the dependency on index choice in the model process of the traditional econometric model, but also overcoming the flaw that the linear econometric model cannot incorporate the nonlinear interaction of variables. The interest for the use of artificial neural network in employment forecast has several objectives. Some of them are directed towards forecast, like computing short-term forecasts of regional employment patterns (Gopalakrishnan et Konstantina, 2013) or regional employment Blien el al., 2006) in Germany or forecast of the employment situation of college graduates (Xing Li and Yang, 2021) or prediction of employment index for college students (Wu, 2022). Other research directions cover subjects like what factors affecting employment of women (Karimi et al., 2013) or employee turnover (Randall et al., 2005).

3. Research methodology

The applied method is construction of code modules in Python, using Colab Notebooks from Google Drive, for the visualization of data representation and emphasis of evolutions. The python modules used were:

1. sklearn.model_selection with train_test_split
2. numpy as np
3. MLPRegressor
4. matplotlib.pyplot as plt
5. torch.nn as nn

The used data are the following indices:

- GERD by sector of performance (Total intramural expenditure on R&D performed during a specific reference period, broken down by the institutions corresponding to each sector (business enterprise, government, higher education and private non-profit organizations), independent of the source of funds);
- Gross domestic product at market prices;
- Total investment in industrial field;
- Production value;

- Female employment.

Example of the applied Python code is:

Figure no. 2. Code example for Training process (than_adam).

```
Network = MLPRegressor(max_iter=5000,
                        verbose=True,
                        tol=0.00001,
                        activation = 'tanh',
                        solver = 'adam',
                        learning_rate = 'constant',
                        learning_rate_init = 0.001,
                        batch_size = 2,
                        hidden_layer_sizes = (9),
                        #early_stopping = True,
                        n_iter_no_change = 200)
network.fit(X_train, y_train)
```

Source: Author's model in Python.

Data values belong to years 2011-2021 and consider the Romanian indices values. In order to determine a forecast through mathematical model of artificial neural networks, the values data for GERD, GDP, Total investments and Production were used as one year behind the values for female employment (e.g. GERD, GDP, Total investments and Production values from 2011 were used for the values of female employment from 2012). This link between annual values allows to forecast the female employment values for the next year, using values from the present year for all the other indices.

To determine the best structure for the ANN several activation functions and solvers for weight optimization were applied. Activation function for the hidden layer.

- 'identity', no-op activation, for linear bottleneck implementation, $f(x) = x$.
- 'logistic', logistic sigmoid function, $f(x) = \frac{1}{1+\exp(-x)}$.
- 'tanh', hyperbolic tan function, $f(x) = tahn(x)$.
- 'relu', rectified linear unit function, $f(x) = \max(0, x)$.

The solver for weight optimization.

- 'lbfgs' is a quasi-Newton methods optimizer.
- 'sgd' is the stochastic gradient descent.
- 'adam' is a stochastic gradient-based optimizer proposed by Kingma, Diederik, and Jimmy Ba

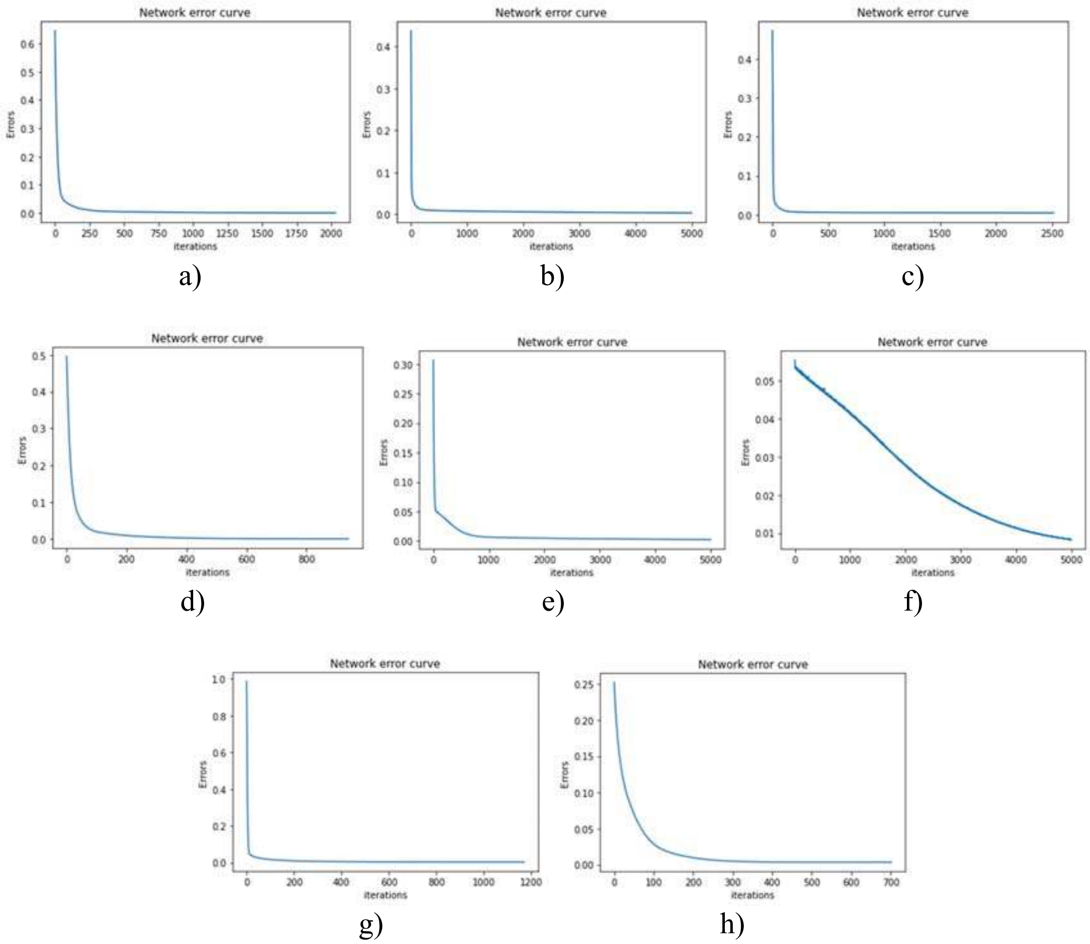
Also, some values for training process were preserved constant for all the different functions and solvers that were used: the learning rate was 0.001; the number of hidden layers was 1, with 9 hidden neurons; batch size was 2; numbers of maxim iterations were 5000, with maxim number of iterations with no change in network error equal with 200. These values can be seen in figure 2.

The application of different functions and solvers for weight optimization in several figures, that show the errors of training results as error curve (figure 2) and the feature importance of the indices in the training process as ratio influence (in figure 3). Also, in table 1, the results of different activation functions and solvers in training process can be traced, in conditional color table and in table 2 the feature importance values for each ANN structure.

4. Findings

After the application of Python code to the data values the following results were obtained. Figure 2 shows the evolution of network error curve on the training process. Analyzing the graphics, we saw that the smallest numbers of iterations belong to the identity_adam structure, as the highest number goes to tanh_sgd and logistic_adam structures. Also, we must emphasize that structures with lbfgs solver does not have graphics representations of network error curve, as the structure does not calculate evolution of this index.

Figure no. 2. Network error curve under the use of different ANN structures: a) tanh_adam; b) tanh_sgd; c) relu_sgd; d) relu_adam; e) logistic_adam; f) logistic_sgd; g) identity_sgd; h) identity_adam.



Source: Author's model in Python.

Table no. 1 Results of different activation functions and solvers in training process.

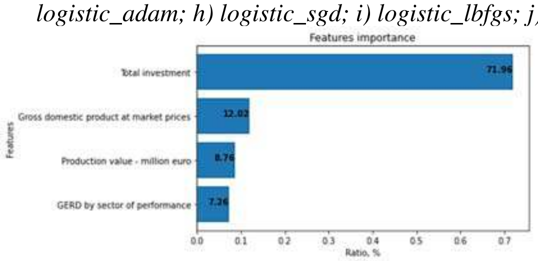
Activation & solver functions	Network loss	Mean absolute error testing	Mean squared error testing	Mean absolute error training	Mean squared error training	Model score on training data
tanh_adam	0.00037	0.30209	0.11534	0.00559	7.87339	0.8248
tanh_sgd	0.00315	0.08632	0.00773	0.05732	0.00573	0.7803
tanh_lbfgs	0.00035	0.30621	0.12354	0.00086	0.000002	0.85941
relu_lbfgs	0.00081	0.28207	0.14073	0.0023	0.000007	0.81071
relu_sgd	0.00419	0.2942	0.08658	0.05345	0.00792	0.81882
relu_adam	0.00038	0.23652	0.06204	0.01466	0.00031	0.81752
logistic_adam	0.12603	0.0163	0.03458	0.00293	0.12603	0.79638
logistic_sgd	0.00828	0.03381	0.00128	0.11247	0.01578	0.80327
logistic_lbfgs	0.00024	0.40975	0.20596	0.00218	0.000007	0.8067
identity_lbfgs	0.00647	0.09413	0.01441	0.10548	0.01285	0.75288
identity_sgd	0.0033	0.2549	0.07271	0.06796	0.00609	0.90154
identity_adam	0.00329	0.24971	0.07128	0.06603	0.00608	0.88468
Minimum	0.00024	0.0163	0.00128	0.00086	0.000002	0.75288

Source: Author's model in Python.

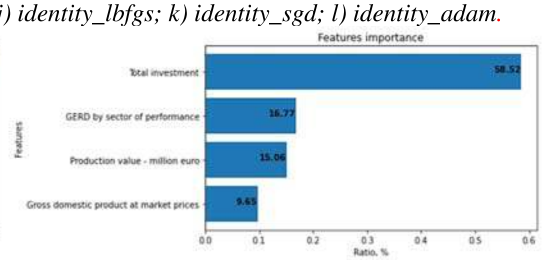
The results of training are presented in table no. 1. To simplify the evaluation of ANN’s results and reading easier the errors, the higher values are in red and smallest ones in green. All other values and colors show the values between min and max.

Analyzing table no. 1 one can see that the smallest network loss belongs to logistic_lbfgs structure. Also, the mean absolute error and mean squared error on testing data set belong to logistic_adam and logistic_sgd, respectively. The smallest errors for mean absolute error and mean squared error for training data set can be found in tanh_lbfgs structure, and the best model score on training data is on identity_sgd (the closer to 1.00, the better the model score is). This makes it difficult to choose the best structure. To help the selecting process we narrow down the error that we should consider by concentrating the attention on error on training data set. This choice was motivated by the fact that the test set was too small to consider (only 2 values). Thus, the best structure was tanh_lbfgs with the best mean absolute error and mean squared error and the 3rd model score.

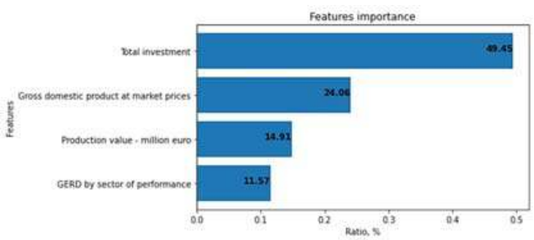
Figure no. 3. Feature importance of indices, as influence over the training process for different structures: a) tanh_adam; b) tanh_sgd; c) tanh_lbfgs; d) relu_lbfgs; e) relu_sgd; f) relu_adam; g) logistic_adam; h) logistic_sgd; i) logistic_lbfgs; j) identity_lbfgs; k) identity_sgd; l) identity_adam.



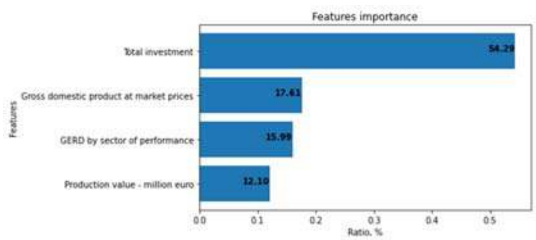
a)



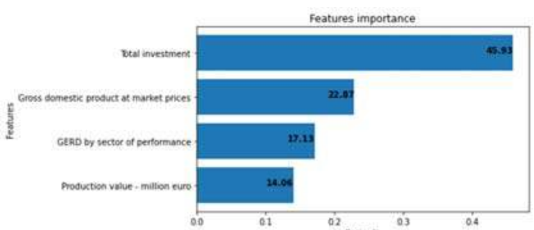
b)



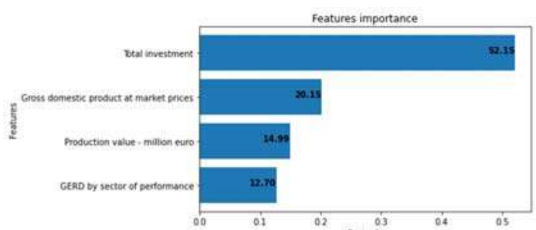
c)



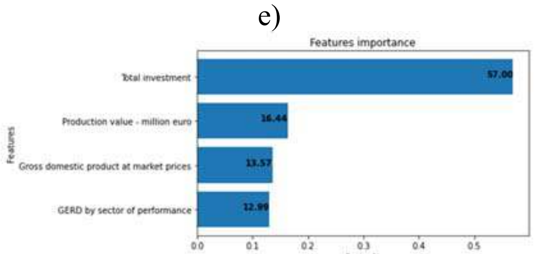
d)



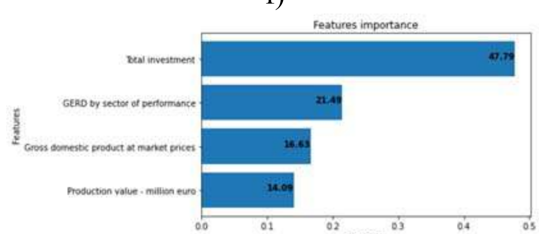
e)



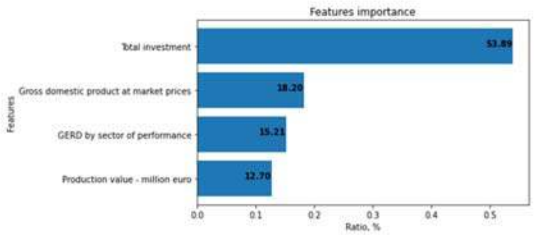
f)



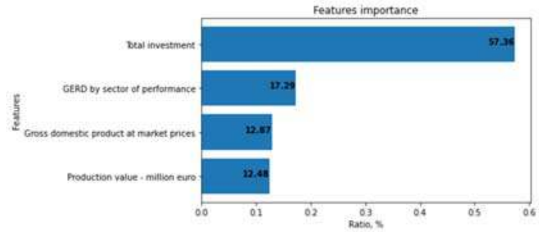
g)



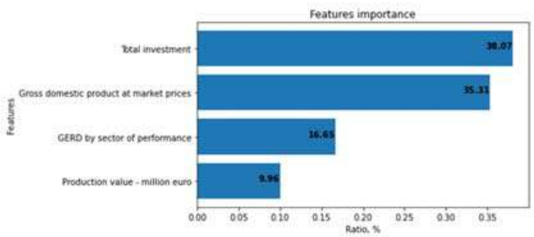
h)



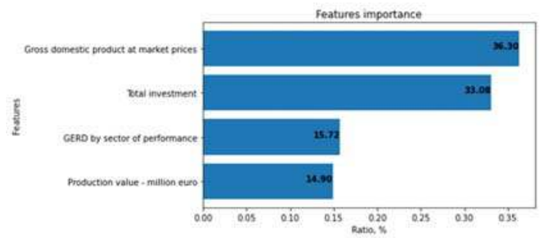
i)



j)



k)



l)

Source: Author's model in Python.

Analyzing figure 3 we can conclude that the values of feature importance is very different form one structure to another. As the hyperbolic tan function give the more importance to the Total investments (between 49.45 and 71.96%) and the less to the GERD (between 7.26 and 16.77%) or Production values (between 8.76 and 15.06%). The identity function put on the smallest values (compared with the other functions) to the Total investments and increases the importance for the GDP. The chosen structure tanh_lbfgs had the following feature importance values: Total investments - 49.45%, GDP - 24.06%, GERD - 11.57%, Production values - 14.91%. This puts the Total investment to all most 50% influence over the training process results, followed by the GDP with almost a quarter influence. The last are Production values and GERD as the last.

For an easier reding and analysis of figure 3, the values were translated in table no. 2.

Table no. 2 Feature importance for different activation functions and solvers in training process (%).

Activation & solver functions	**Total investments**	**GDP**	**GERD**	**Production values**
tanh_adam	71.96	12.02	7.26	8.76
tanh_sgd	58.52	9.65	16.77	15.06
tanh_lbfgs	49.45	24.06	11.57	14.91
relu_lbfgs	54.29	17.61	15.99	12.1
relu_sgd	45.93	22.87	17.13	14.06
relu_adam	52.15	20.15	12.7	14.99
logistic_adam	57	13.67	12.99	16.44
logistic_sgd	47.79	16.63	21.49	14.09
logistic_lbfgs	53.89	18.2	15.21	12.7
identity_lbfgs	57.36	12.87	17.29	12.48
identity_sgd	38.07	35.31	16.65	9.96
identity_adam	33.08	36.3	15.72	14.9
Average value	51.62	19.95	15.06	13.37

Source: Author's model in Python.

Table no. 2 also, shows the big difference on how the different structure evaluate the feature importance. This big gaps between values are due to the small amount of data and the decision of combining the different years for inputs and output.

5. Conclusions

Even that the research is based on a small number of data sets and the data have a small heterogeneity level, the results are satisfying, considering the very small values of errors that were calculated. From the representation of the network error curve, we can show that the smallest numbers of iterations belong to the identity_adam structure, as the highest number goes to tanh_sgd and logistic_adam structures.

Comparing different structures in terms of activation functions and solvers for weights, we concluded that the best structure was tanh_lbfgs with the best mean absolute error and mean squared error and the 3rd model score. Also, this decision was taken in view of a difficult choice of the best structure, as the different type of errors were split between different structures.

For the feature importance, the chosen structure tanh_lbfgs had the following values: Total investments - 49.45%, GDP - 24.06%, GERD - 11.57%, Production values - 14.91%. Also, hyperbolic tan function gives the more importance to the Total investments (between 49.45 and 71.96%) and the identity function put on the smallest values (compared with the other functions) to the Total investments and increases the importance for the GDP. So, the feature importance depends on the ANN's structure.

For future research more data sets are needed, and more indices should be introduced. This will make the training more accurate but will make the feature importance more difficult to determine.

6. References

- Ştelіac, N., 2015. *What Are the Female Employment and Unemployment Patterns on The Romanian Labour Market?,* [online] Available at: https://www.researchgate.net/publication/280306700_WHAT_ARE_THE_FEMALE_EMPLOYMENT_AND_UNEMPLOYMENT_PATTERNS_ON_THE_ROMANIAN_LABOUR_MARKET , [Accessed June 2022]
- Gopalakrishnan, K., Konstantina Gkritza, "Nadia" K., 2013. Forecasting transportation infrastructure impacts of renewable energy industry using neural networks. *Technological and Economic Development of Economy* 19:sup1, pages S157-S175.
- Blien, U., Peter, N., Patuelli, R., & Reggiani, A., 2006. New Neural Network Methods for Forecasting Regional Employment: an Analysis of German Labour Markets, *Spatial Economic Analysis*, Vol 1, 7-30, 10.1080/17421770600661568.
- Xing Li, X., Yang, T., 2021. Forecast of the Employment Situation of College Graduates Based on the LSTM Neural Network, *Neural Network-Based Machine Learning in Data Mining for Big Data Systems*, Volume 2021, Article ID 5787355,| https://doi.org/10.1155/2021/5787355.
- Wu, D., 2022. Prediction of Employment Index for College Students by Deep Neural Network, *Mathematical Problems in Engineering,* vol. 2022, Article ID 3170454, 8 pages, 2022. https://doi.org/10.1155/2022/3170454
- Karimi Moughari, Z., Nazifi Nainie, M., Abbaspour, S., 2013. 'Evaluating the Economic factors affecting employment of women in Iran Using artificial neural network approach', *Women's Studies Sociological and Psychological*, 11(3), pp. 53-80. doi: 10.22051/jwsps.2014.1446.
- Randall, S., McMurtrey, S., Michalopoulos, J., Smith, A., 2005. Employee turnover: A neural network solution, *Computers & Operations Research*, 32. 2635-2651. 10.1016/j.cor.2004.06.022.
- World Bank Group, 2022. Labor force participation rate, female (% of female population ages 15+) (modeled ILO estimate) – Romania, *International Labour Organization, ILOSTAT database*. [online] Available at: https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?end=2021&locations=RO&start=1990&view=chart [Accessed June 2022]
- European Institute for Gender Equality, 2017. Gender Equality Index 2017: Romania, *Gender Statistics Database*, [online] Available at: https://eige.europa.eu/gender-statistics/dgs, [Accessed June 2022]