Management Based on Data Analysis. Part Two. Artificial Intelligence Data Modeling

Constantin Ilie
Andreea-Daniela Moraru
"Ovidius" University of Constanta, Romania
constantin.ilie@365.univ-ovidius.ro

Abstract

The use of the artificial neural networks in management is common. Employing artificial intelligence allows businesses to gain a commercial advantage. The objective of the present paper is to apply an artificial neural network as a modeling and simulation technique, to determine the importance of influences of four input data on the output data in human resources evaluation. The data are organized in 14,999 datasets: satisfaction level of the employee, the average monthly hours worked, the project number that the employee participated in, and time spent by company for employee care. The used neural network is of feed forward type, with three layers (1 input, 1 hidden, 1 output) and with 9 (13) hidden neurons. Several functions for activation and solver were applied for finding the most accurate one. The simulations concluded that the most important influence is that of the average monthly hours followed closely by the employee level of satisfaction.

Key words: management, artificial neural networks, human resources, evaluation
J.E.L. classification: C45

1. Introduction

The present work follows the first part of the research that applied an open-source software (Python 3) for several types of data visualization, to help management understand how to categorize human resources for obtaining the best evaluation scores (the first part can be found in the same volume of the conference).

Starting initially from 10 data sets, after eliminating two data because of their character type, a correlation function was applied and the data with correlation values greater than 0.1 were preserved. This led to 5 data sets, that were further used: satisfaction level, average monthly hours worked, number of the project, time spent by company, and the last evaluation.

In this second part of the research, the first 4 data sets are considered the input data, and the evaluation is the output data in the training of an artificial neural network (ANN). The ANN's simulation will learn to evaluate the influence of each input data on the output data.

The importance of the paper emerges from the creation of a research base for ANN application on large amounts of data, with the objective to determine the importance of features for management decision making.

2. Theoretical background

AI is fueling product recommendations, targeted advertising, essay grading, employee promotion and retention, risk scoring, image labelling, fraud detection, cybersecurity defenses, and a host of other applications. The broad adoption of algorithmic decision-making has encouraged a great amount of attention and generated a variety of reactions (and a considerable amount of "hype")—especially an excitement about how AI capabilities will augment human decision-making and enhance business performance (Gil, 2019, p. 4)

Artificial Intelligence (AI) is growing in importance in management and suggests the massive changes that the corporate management will encounter in the future (Petrin, 2019, p.965). As Kolbjørnsrud et al. (2016) marks up, managers at all levels are forced to adapt to the world of smart machines. AI can carry out the administrative tasks that consume much of managers' time faster, better, and at a lower cost.

3. Research methodology

The method used is ANN in structure of feed forward. For a better understanding and future work, 4 of activation functions were employed:

1. ReLU - rectified linear unit - function returns 0 if it receives any negative input, but for any positive value of x it returns that value back,
2. tanh - hyperbolic tangent,
3. identity – linear,
4. logistic.

and 2 solver functions:

1. adam - stochastic gradient descent algorithm based on estimation of 1st and 2nd-order moments (Kingma, 2015, p. 1),
2. sgd – simple, efficient approach to fitting linear classifiers and regressors under convex loss functions such as (linear) (https://scikit-learn.org/)

Thus, 8 ANN were compared to determine the best one in terms of accuracy and training loss. Also, Random Forest Regressor was used for comparing the simulated results of each ANN. Random Forest Regression is a supervised learning algorithm that uses ensemble learning method for regression (Bakshi, 2020).

For an additional comparison the all the ANNs will use the 5 data established above, but another ANN will use the initial 8 data that were analyzed in first phase of the research (paper – part one), called from now on "8 data ANN". The results of the first 8 ANNs are compared with the result of the 8 data ANN.

The structure for all 8 ANNs consists of 3 layers: 1 input layer – 4 neurons, 1 hidden layer – 9 neurons, 1 output layer – 1 neuron. The 8 data ANN has 13 neurons in the hidden layer following the calculus of 8 (instead of 4) input data.

The authors used 14,999 datasets consisting in data about human resource evaluations. The data was downloaded from www.kaggle.com .

4. Findings

Applying 4 activation functions and 2 solvers resulted in 8 ANN variations with different values for training process, accuracy, and Random Forest Regressor. The results are presented in table 1.

Table no. 1 Table title (Times New Roman, 10 pts, italic, left)

		tanh & adam	**tanh & sgd**	**relu & adam**	**relu & sgd**	**logistic & adam**	**logistic & sgd**	**identity & adam**	**identity & sgd**	**relu & adam all**
Training	No. iterations	**150**	**166**	**282**	**198**	**161**	**113**	**141**	**165**	162
After training	Network loss	**0.028**	**0.029**	**0.027**	**0.029**	**0.029**	**0.029**	**0.029**	**0.029**	0.011
	Max. network errors	**0.530**	**0.568**	**0.463**	**0.535**	**0.391**	**0.626**	**0.494**	**0.589**	0.486
	Min. network errors	**-0.667**	**-0.748**	**-0.506**	**-0.569**	**-0.652**	**-0.484**	**-0.468**	**-0.570**	-0.494
Accuracy on testing	Mean absolute error	**0.202**	**0.218**	**0.188**	**0.185**	**0.174**	**0.214**	**0.181**	**0.190**	0.116
	Mean square error	**0.061**	**0.068**	**0.051**	**0.051**	**0.047**	**0.064**	**0.051**	**0.056**	0.021
Random	Model score training	**0.885**	**0.885**	**0.885**	**0.885**	**0.886**	**0.886**	**0.885**	**0.886**	0.896

Forest Regressor	Model score testing	**0.249**	**0.216**	**0.278**	**0.413**	**0.279**	**0.256**	**0.291**	**0.322**	0.294
Legend		min						max		

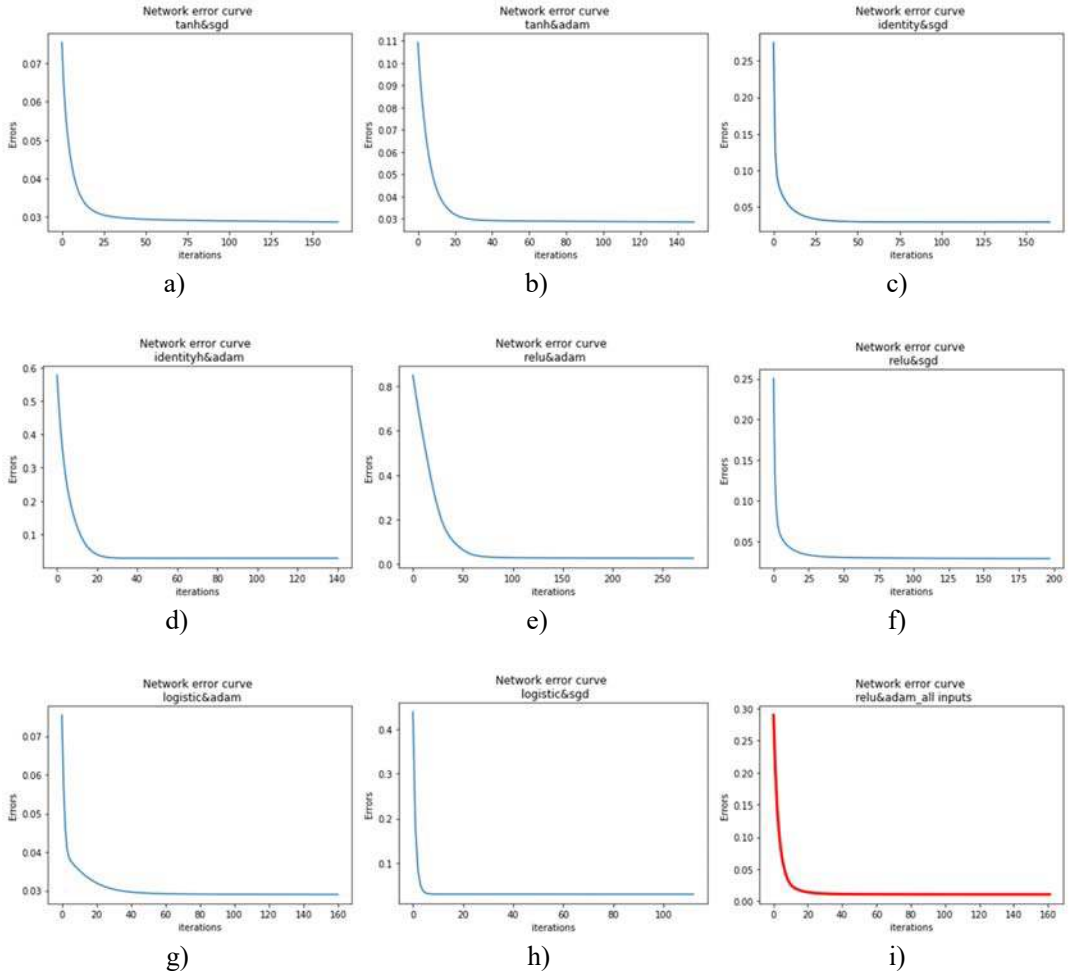
Source: Authors' simulation and representation.

Analyzing Table 1, the conclusions are:

- The greatest numbers of iterations belong to relu&adam (282), the smallest number is of logistic&sgd (113),
- The smallest (the best) network loss after training pertain to relu&adam,
- The lowest mean absolute and square error belong to logistic&adam,
- The best Random Forest Regressor coefficient fits to logistic&adam.

The results of ANN's error loss and their variations are shown in figure 1.

Figure no. 1. Network error curve for activation&weight function respectively: a) tanh & adam; b) tanh & sgd; c) relu & adam; d) relu & sgd; e) logistic & adam; f) logistic & sgd; g) identity & adam; h) identity & sgd; i) relu & adam all inputs.

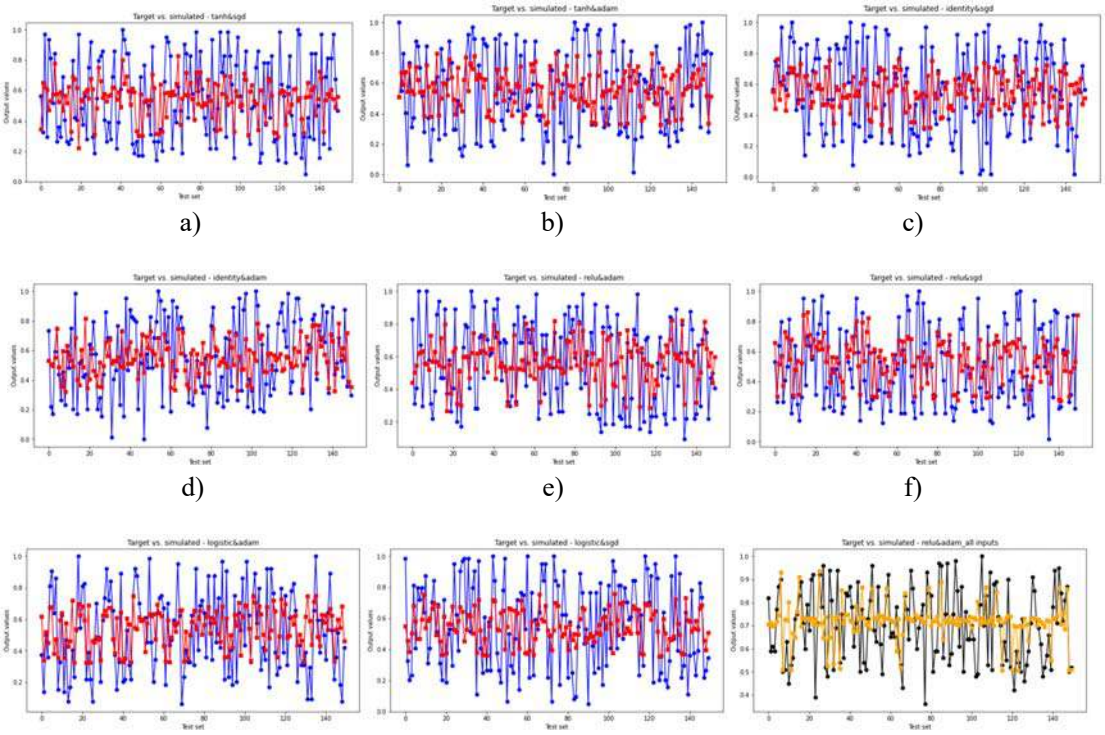


Source: Author's simulation in Python 3.

After evaluating the above values, the best ANN structures are relu&adam and logistic&adam. Considering these values, the authors choose as the best functions to be used in future work: relu and adam. Yet, comparing the ANN with 5 data and the 8 data set ANN, after training and testing with relu and adam functions, the second ANN was better.

Another possibility to evaluate the training process is to visualize the results of comparing the targeted output (the real data values from database) with the simulated output (data resulted from ANN's simulation, after training process ended). The two data values are presented in figure 2, as the results of testing process, with 150 data sets for all ANNs structures.

Figure no. 2. Target output (in blue, black) against simulated output (in red, orange) for activation&weight function respectively: a) tanh & adam; b) tanh & sgd; c) relu & adam; d) relu & sgd; e) logistic & adam; f) logistic & sgd; g) identity & adam; h) identity & sgd; i) relu & adam all inputs.



a) b) c)

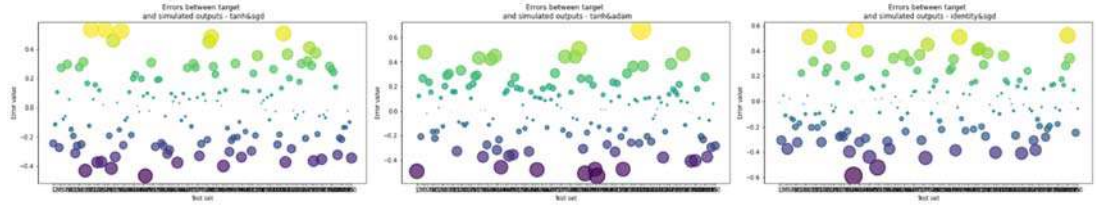
d) e) f)

g) h) i)

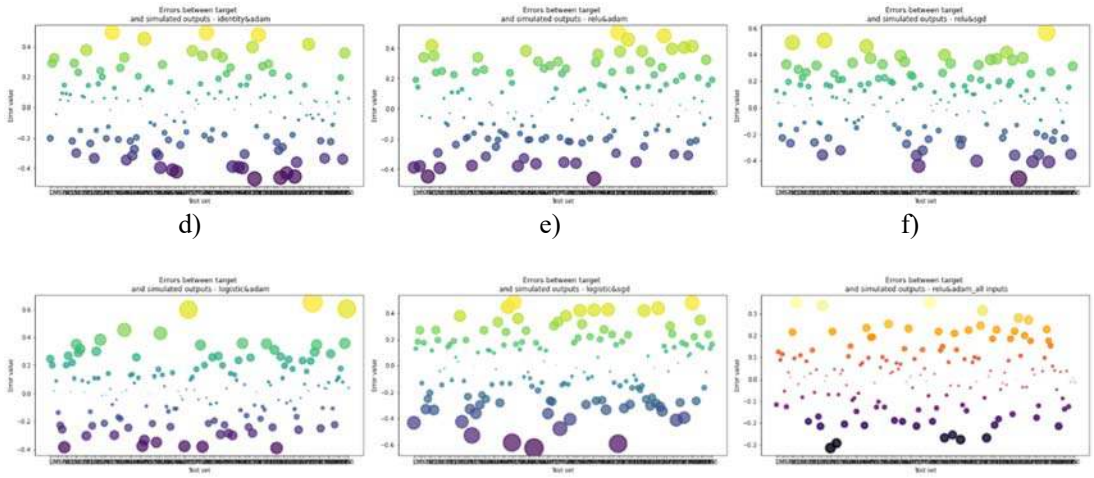
Source: Author's simulation in Python 3.

As figure 2 shows, the differences between the target output and the simulated output are significant enough. These differences are supported by the high values in absolute mean and square error and the small values in Random Forest Regressor coefficient. The graphic representation can be successfully used for estimating the ANN training and testing error (Ilie et al., 2017, p. 401) For a better visualization of these differences, in figure 3, there are presented the values.

Figure no. 3. Differences between target output and simulated output for activation&weight function respectively: a) tanh & adam; b) tanh & sgd; c) relu & adam; d) relu & sgd; e) logistic & adam; f) logistic & sgd; g) identity & adam; h) identity & sgd; i) relu & adam all inputs.



a) b) c)



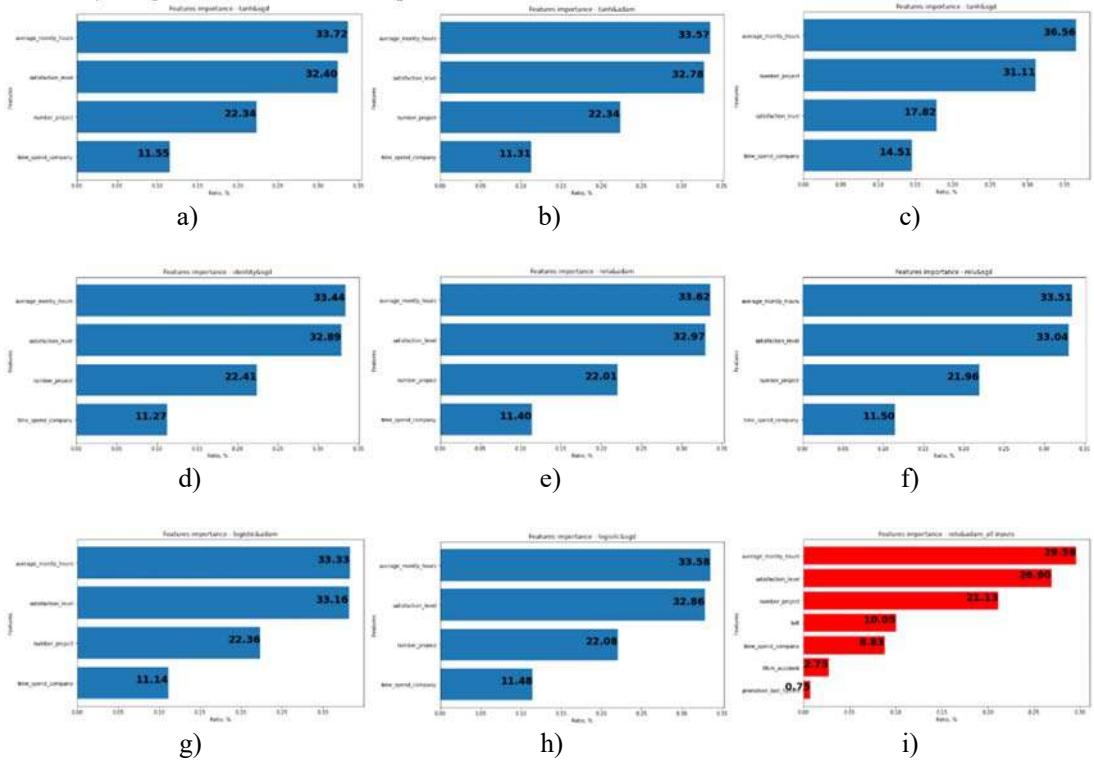
d) e) f)

g) h) i)

Source: Author's simulation in Python 3.

The main objective of the research can be best determined in figure 4, which contains the feature importance for each of the data, as the ANN considers the importance in accordance with how it learns the connections and influences during the training process.

Figure no. 4. Simulated feature importance for activation&weight function respectively: a) tanh & adam; b) tanh & sgd; c) relu & adam; d) relu & sgd; e) logistic & adam; f) logistic & sgd; g) identity & adam; h) identity & sgd; i) relu & adam all inputs.



a) b) c)

d) e) f)

g) h) i)

Source: Author's simulation in Python 3.

Figure 4 shows that all the ANNs determined the same hierarchy, with the data limits, as follows:

1. average monthly hours [33.33; 33.72]%, and 29.58 in case of 8 data ANN,
2. satisfaction level [31.11; 33.16] %, and 26.90 in case of 8 data ANN,
3. project number [17.82; 22.41] %, and 21.13,

4. time spent by the company [11.14; 14.51] %, and 8.83,

only for 8 data ANN other 4 data are represented:

5. employees that left - 10.05 %,
6. employees implicated in work incidents – 2.75 %,
7. employees that were promoted in the last 5 years – 0.75 %.

There is one major difference only encountered on the 8 data ANN, which places data "employees that left" before "time spent by the company", as the other ANNs' structures do not contain this data set.

5. Conclusions

Uses of the artificial neural network is demonstrated by many applications. In the present research the ANN was used to determine, through AI simulation, which of the four data sets (average monthly hours spent by the employee, employee's satisfaction level, the project number that the employee participated in, time spent by the company for employee) is more important for employee evaluation. The result points to the number of hours worked by the employee followed shortly by employee's satisfaction level.

Also, the research shows the possibility to see how the ANN runs with high amounts of data (14,988 datasets, 8 data each). The results also encourage using all the data available, even though it does not appear to be any direct connections between input and output data, as the ANN will mathematically find these connections.

6. References

- Bakshi, C., 2020, *Random Forest Regression*, Available at: https://levelup.gitconnected.com/random-forest-regression-209c0f354c84 [Accessed 08 December 2021]
- Gil, D., et al., 2019, *AI for Management: An Overview*, The Future of Management in an AI World, ISBN: 978-3-030-20680-2, pp. 3-19
- Kingma, D., Lei Ba, J, 2015, A*DAM: A Method for Stochastic Optimization*, UC Berkeley Computer Vision Review Letters (CVRL), Available at: https://theberkeleyview.wordpress.com/2015/11/19/berkeleyview-for-adam-a-method-for-stochastic-optimization/ [Accessed 08 December 2021]
- KAGGLE: https://www.kaggle.com/giripujar/hr-analytics
- Kolbjørnsrud, V., Amico, R., Thomas, R.J., 2016, *How Artificial Intelligence Will Redefine Management*, Harvard Business Review, Leadership Development, Available at: https://hbr.org/2016/11/how-artificial-intelligence-will-redefine-management [Accessed 08 December 2021]
- Ilie, M., et al., 2017, *The Hierarchical Determination of Customer Satisfaction With Banking Services Using an Artificial Neural Network*, TRANSFORMATIONS IN BUSINESS & ECONOMICS, Vol. 16, No 2A (41A), pp. 401-420
- Petrin, M., 2019, *Corporate Management in the Age of AI*, Colum. Bus., p.965
- https://scikit-learn.org/stable/modules/sgd.html