Applied Mathematical Sciences, Vol. 16, 2022, no. 1, 27 - 35
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ams.2022.916737

A Comparison between Artificial Neural Networks and Logistic Regression to Determine the General Factors that Effect on Unemployment for Saudi Women

Salwa L. AlKhayyat[1] and Hanan M. Alghamdi[2]

[1] Associate Professor at the Faculty of Science, Department of Statistics, University of Jeddah, Kingdom of Saudi Arabia and Assistant Professor at the Faculty of Commerce, Kafr El-Sheikh University, Department of Statistics, Mathematics and Insurance, Egypt

[2] Department of Statistics, Faculty of Sciences, Jeddah University, Jeddah, Saudi Arabia

Abstract

Objectives: To investigate the factors that effect on unemployment of Saudi women and built a model that predicts the working status for Saudi women who not working.
Methods: This study was conducted on 658 women referred from 13 regions of the kingdom of Saudi Arabia between March 2020 and December 2020. The factors that effect on unemployment of Saudi women who were redirected to different regions, Saudi Arabia. An Artificial Neural Networks (ANNs) method and Logistic Regression (LR) was used to propose a model that predicts working status for Saudi women who suffer from unemployment. Therefore, a random sample of 70.1% of the data was used to build the ANNs model and tested by the remaining 29.9%. LR model were conducted to study relations of multiple independent variables which represent the factors that effect on unemployment of Saudi women and categorical dependent that represent the working status. To determines the performance of a classifier using Area under the curve (AUC).

Results: Out of 658 women, the results indicated that the marital status, sources of income from family, husband and suitability of academic majors with the labor market are the most important of the factors of unemployment in general that affecting on work status to see if they will whether the woman works or not, in both methods. The proposed ANNs and LR models showed a good performance with 86% accuracy using AUC.
Conclusion: Through conducting a comparison between the application of the ANN and the LR models to determine the most important factors that effect of the work status, the study concluded that there is no a difference in the results of applying the AUC on the models. We recommend the use of an ANNs and LR models to predict of work status.

Keywords: Logistic Regression, Unemployment, Artificial Neural Networks, jobs

1 Introduction

According to the Kingdom of Saudi Arabia 2030 vision, which is based on three axes: the resilient society, thriving economy and ambitious country. Through the axe of a thriving economy, one of it is most important goals is to increase employment rates by enabling job creation, Saudi Arabia seeks to grow and diversify the economy through increase the employment and ensure the equal access of job opportunities and the increase of women's participation in the labor market.
The issue of employment of Saudi women is one of the most important issues facing decision makers and planners in the Kingdom due to the large increase in the number of women seeking jobs versus job opportunities that are available in different sectors. Although the unemployment rate in Saudi Arabia has a decline according to the General Authority for Statistics surveys for the second quarter of 2019 (General Authority of Statistics, 2019) [3], women's unemployment remains a challenge in the Kingdom, where the unemployment rate for Saudi women is 31.1%.
In addition, Saudi women are an important and effective element in the development of society and the economy. Therefore, the vision aims to raise Saudi women economic participation rate from 17% to 25%, to increase women's participation in the workforce from 22% to 30%, to lower the rate of unemployment from 11.6% to 7% (Government of Saudi Arabia, 2020) [4].

Based on the Sustainable Development Goals, the most important goals are: decent work, economic growth and gender equality, which aim at creating jobs and eliminate unemployment for both sexes and obstacles to enter the labor market for women (United Nations Development Programme, 1389) [9].

2 Methods

This study was conducted according from a questionnaire. The questionnaire was distributed electronically to the regions of the Kingdom of Saudi Arabia the period from March 2020 to December 2020. It should be noted that the questionnaire was prepared to study the factors of unemployment of Saudi women in the period before the COVID-19. The population of study includes of all Saudi women of working age by administrative area of the Kingdom. The data contained 658 Saudi women with different ages and are referred from different administrative regions of the Kingdom they workless, seriously looked for work and who leaving work. The variables of the study were determined by some previous studies related to the subject of the study and the survey of the General Authority for Statistics for labor market.

Statistical analysis. The Statistical Package for the Social Sciences (SPSS) version 21 (Armonk, NY: IBM Corp.) was used to analyse the data. ANNs and LR ware used to propose a model that can predict the work status and determine the most important factors that effect of the work status. ANNs consists of 3 main layers: input units in the input layer that represent variables measured for each training instance. Those inputs are weighted and fed to a hidden layer at the same time. Output units or neuroses are the units in the hidden layers and output layer. Then, outputs of the hidden layer are fed to different hidden layer. (Johnson et al., 2002; Panchal and Panchal, 2014) [8]. Finally, weighted outputs of the last hidden layer are fed to the output layer, which gives the estimate and prediction for the instances. Each output unit uses a nonlinear (activation function) to the weighted input. (Ciaburro and Venkateswaran, 2017) [2].
The proposed model was built by dividing the data into training and test data. The training set used to build the model represents a random sample of 70.1% of the original data and the test data is based on the remaining 29.9% of the data. The test data is used to measure the performance of the model. Also, the ANNs method identifies the most important variables that have a significant effect on a working status in terms of status (worker, unemployed).

LR is one of the most important analytic tools in the social and natural sciences. LR is the baseline supervised machine learning algorithm for classification, and also has a very close relationship with ANNs. It studies the association between a categorical dependent variable and a set of independent (explanatory) variables. The model form for predicted probabilities is expressed as a natural logarithm (ln) of the odds ratio:

$$P(Y) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\cdots+\beta_k X_k}}$$

where, $\ln\left[\frac{P(Y)}{1-P(Y)}\right]$ is the log (odds) of the outcomes, Y is the dichotomous outcome; $X_1.X_2.\ldots.X_k$ are the predictor variables, $\beta_0.\beta_1.\beta_2.\ldots.\beta_k$ are the regression (model) coefficients and β_0 is the intercept. In Equation, the LR model directly relates the probability of Y to the predictor variables. The goal of LR is to estimate the k + 1 unknown parameters β in Equation.

The regression coefficients indicate the degree of association between each independent variable and the outcome. Each coefficient represents the amount of change we would expect in the response variable if there was a one unit change in the predictor variable. The objective of LR is to correctly predict the category of outcome for individual cases using the best model. To accomplish this goal a model is created that include all predictor variables that are useful in predicting the response variable. LR calculates the probability of success over probability of failure. The results of the analysis are in the form of an odds ratio. (Pample, 2000) [7]

Finally, an AUC is calculated to estimate the accuracy and performance of the predictive model. (Bewick et al., 2004) [1].

3 Results

The classification (**Table 1**) shows the practical results of using the ANNs. Of the cases used to create the model, 290 of the 313 women that unemployed is classified correctly. 84 of the 124 women that worker is classified correctly. Overall, 84.8% of the training cases are classified correctly, corresponding to the 15.2%. A better model should correctly identify a higher percentage of the cases. The testing sample helps to validate the model; here 80.2% of these cases were correctly classified by the model.

Table 1 Classification Table Using ANNs

Sample	**Observed**	**Predicted**		
		Worker	**Unemployed**	**Percent Correct**
Training	**Worker**	84	44	65.6%
	Unemployed	23	290	92.7%
	Overall Percent	24.3%	75.7%	84.8%
Testing	**Worker**	34	31	52.3%
	Unemployed	12	140	92.1%
	Overall Percent	21.2%	78.8%	80.2%

The classification in (**Table 2**) using LR shows: The sensitivity of the model, which is the rate of the correct predictions in the cases of women that is classified unemployed, is 89.5%. The specificity of the model, which is the correct predictions in the cases of women that is classified worker is 62.2%. In general, the accuracy of the model is the percentage of the correct total classification, which is the number of correct predictions on the total number of respondents of the study sample, is 81.5%.

Table 2 Classification Table Using LR

Observed		**Predicted**		
		Work status		**Percentage Correct**
		Worker	**Unemployed**	
Work status	**Worker**	120	73	62.2
	Unemployed	49	416	89.5
Overall Percentage				81.5

The results in (**Table 3**) indicate that (Sources of Income from family husband) has the most significant effect on work status followed by (Suitability of academic majors with the labor market) and the (Marital status) with a Normalized importance of 100 %, 90.3%and 42.3% using ANNs.

Table 3 Independent Variables Importance Using ANNs

Variables	**Importance**	**Normalized Importance**
Age	.059	29.6%
Marital status	.085	42.3%
Educational level	.038	19.1%
Academic Degree	.034	16.9%
Education Field	.047	23.4%
Administrative Area	.079	39.3%
Suitability of academic majors with the labor market	.181	90.3%
Requirement of years of experience	.059	29.3%
English language requirement	.066	32.9%

Table 3 (Continued): Independent Variables Importance Using ANNs

Rejection low salary	.061	30.3%
Sources of Income from family husband	.201	100.0%
Sources of Income from Assets, Projects	.025	12.6%
Sources of Income from citizen account	.005	2.4%
Sources of Income from (Hafez account)	.014	7.1%
Waiting for jobs in the public sector	.044	21.9%

From (**Table 4**) can be seen the independent variables of the determinants of unemployment in general that have an impact on work status (Sig < 0.05, confidence level 95%).
They: Age, Marital status, Suitability of academic majors with the labor market, requirement of years of experience, rejection low salary, Sources of income from family: husband, citizen account, Hafez account.

Table 4 Independent Variables Using LR

Variables	B	Sig.
Age	.384	.000
Marital status	.651	.001
Educational level	.277	.142
Academic Degree	-.006-	.957
Education Field	-.018-	.688
Administrative Area	.015	.679
Suitability of academic majors with the labor market	.439	.000
Requirement of years of experience	.604	.011
English language requirement	.315	.175
Rejection low salary	-.604-	.010
Sources of Income from family husband	2.459	.000
Sources of Income from Assets, Projects	.417	.455

Table 4 (Continued): Independent Variables Using LR

Sources of Income from citizen account	1.254	.012
Sources of Income from (Hafez account)	1.687	.012
Waiting for jobs in the public sector	.040	.653
Constant	-4.445-	.000

According to the previous results, the LR model can be formulated as follows:

$$\text{Log Odd} = \text{Log}\left[\frac{p}{1-p}\right]$$

$$= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_5 + \beta_6 x_6 + \beta_7 x_7 + \beta_8 x_8$$

$$= -4.445 + 0.384x_1 + 0.651x_2 + 0.439x_3 + 0.604x_4 - 0.604x_5 + 2.459x_6 + 1.254x_7 + 1.687x_8$$

Whereas:
odd: weighting coefficient is the probability of an event occurs divided by the probability of non-occurrence.

x_1: Age

x_2: Marital status

x_3: Suitability of academic majors with the labor market

x_4: Requirement of years of experience

x_5: Rejection low salary

x_6: Sources of income from family or husband

x_7: Sources of income from citizen account

x_8: Sources of income from Hafez account.

From (**Table 5**) The AUC is a numerical summary of the Receiver Operating Characteristic (ROC) curve, and the values in the table represent, for each category. There is a 0.864 probability that the model predicted when using ANNs and 0.857 probability that the model predicted when using LR.

Table 5 AUC Using ANNs and LR

Quality of the Model	**AUC**
ANNs	0.864
LR	0.857

4 Conclusions and Recommendation

This study was performed to measure the most important factors that effect of the work status using the ANNs and the LR, the study concluded that there is no a difference in the results of applying the AUC on the models.
Also, the results indicated that the marital status, sources of income from family, husband and suitability of academic majors with the labor market are the most important of the determinants of unemployment in general that affecting on work status to see if they will whether the woman works or not, in the both methods.
After getting these results, we recommend ANNs and LR provides an alternative good methodology for predicting and classify the work status for Saudi women and proved to be the best method for classification and predicting. There is still a high need to conduct more research and comparisons the LR and ANNs models in studies to analyze the determinants of Saudi women's unemployment. Raising awareness of the importance of financial independence for women to achieve the aspirations of their goals at the personal and family levels.

References

[1] V. Bewick, L. Cheek and J. Ball, (2004) Statistics Review 13: Receiver Operating Characteristic Curves, *Critical Care*, **8** 508-512. https://doi.org/10.1186/cc3000

[2] G. Ciaburro and B. Venkateswaran, Neural Networks with R. Packt Publishing Ltd., (2017).

[3] General Authority of Statistics. (2019). Labour Market Survey. Saudi Arabia. Retrieved from https://www.stats.gov.sa/sites/default/files/labour_market_second_quarter_2019ar.pdf

[4] Government of Saudi Arabia. (2020). Vision 2030 Kingdom of Saudi Arabia. Vision 2030 Kingdom of Saudi Arabia, 1–85. https://vision2030.gov.sa/download/file/fid/417

[5] T. Honkela, W. Odzis Aw Duch, & M. Girolami, (Eds.), *Artificial Neural Networks and Machine Learning: ICANN 2011*, part 1-2. Springer, 2011.

[6] R. A. Johnson, D. W. Wichern, et al. *Applied Multivariate Statistical Analysis*, Volume 5, Prentice hall Upper Saddle River, NJ, 2002.

[7] C. Pample, *Logistic Regression: A Primer* (Quantitative Applications in the Social Sciences), Beverly halls, CA: SAGE Publications, Inc., 2000.

[8] F. S. Panchal and M. Panchal, Review on methods of selecting number of hidden nodes in artificial neural network, *International Journal of Computer Science and Mobile Computing*, **3** (2014), no. 11, 455–464.

[9] United Nations Development Programme. (1389). Sustainable Development Goals. Retrieved From https://www.arabstates.undp.org/content/rbas/en/home/sustainable-development-goals.html

Received: January 15, 2022; Published: January 28, 2022