ISSN:1309 - 0712
İşletme Araştırmaları Dergisi

Journal Of
Business Research
Turk
www.isarder.org

Comparison Of Classification Success Of Human Development Index By Using Ordered Logistic Regression Analysis And Artificial Neural Network Methods

Emre YAKUT
Osmaniye Korkut Ata University, Faculty of Economics and Administrative Sciences, Osmaniye, Turkey
emreyakut@osmaniye.edu.tr

Murat GÜNDÜZ
Uşak University Faculty of Economics and Administrative Sciences, Uşak, Turkey
murat.gunduz@usak.edu.tr

Ayhan DEMİRCİ
Turkish Armed Forces
kho1993@hotmail.com.tr

Extensive Summary

Introduction

Human development is the process of enhancing and improving people's life skills. This process aims to make a positive contribution to human and their living standards by equipping them with skills and capacity (UNDP 1990, p.1). By its *Human Development Index* (HDI) developed in 1990, United Nations Development Program (UNDP) takes a more composite and human oriented perspective by taking education, health and welfare dimensions of development into consideration by widening the perspective which is focused narrowly on economic growth only (Lind, 1992, p.89). Until 2010, GDP calculated per person based on purchasing power parity was taken into consideration for the economic dimension while life expectancy since birth was used for the health dimension and literacy and schooling were used for the education dimension. HDI calculates the arithmetic mean. Both economy and health dimension has one indicator while education dimension has two being literacy (2/3) and schooling (1/3) (Ivanova et al, 1999, pp.159-160). In 2010, index calculation was significantly changed. In this context, index calculation was based on arithmetic average instead of geometric average. With regards to education dimension, literacy among adults was excluded and the average of schooling rate and estimated schooling rate was considered (Morse 2014, p.249). HDI is scored between 0 and 1. 1 shows the highest human development status. The human development report in 2014 stated 4 levels of human development as very high, high, moderate and low. Countries with HDI value lower than 0,550 was classified as low, 0,550–0,699 as moderate, 0,700–0,799 as high and higher than 0,800 as very high (UNDP 2014, p.156). The purpose of this study is to compare the success of multiple classifications and to determine the effective factors by using logistic regression analysis and Elman ANN, multi-layer ANN and LVQ network. This study is comprised of 3 parts. In the first part, logistic regression analysis is introduced while the

second part focuses on Elman ANN and LVQ network. In the third part, application results are compared.

1. Ordered Logistic Regression Model

Logistic regression models are used for modelling the relation between dependent variables measured in different categories and independent variables of categorical or continuous measurement. Ordered logistic regression (OLOGREG) is used when dependent variables consists of at least three categories and measured by ordinal scale (Demirtas v.d., 2009, p.869).

The main features of ordered logistic regression model are as follows (Chen and Hughes, 2004, p.4):

- ✓ Outcome variable of categorical and ordinal measurement is a variable, which can be rearranged multiple times from an unobserved continuous latent variable, however it's not clear whether the space between the categories of this ordinal outcome variable is equal.
- ✓ Ordered logistic regression analysis, uses a correlation function to explain the effects of independent variables on ordered and categorical outcome variable. This model does not require normality and constant variance assumption.
- ✓ Since regression coefficient value is not dependent on the categories of categorical output variable, ordered logistic regression model assumes that the relation between explanatory variables and ordered categorical output variable is independent from categories.

Ordered logistic regression model is actually based on the existence of an continuous and unobserved random Y^* latent variable under a categorical dependent Y variable. The categories of this variable are estimated as sequential intervals on a continuous plane named as cut-off point or threshold value (McCullagh, 1980, p.109).

The most important assumption in ordered logistic regression model is the assumption of parallel curves. According to this assumption, regression parameters obtained in the model is the same in all categories of the dependent variable. In other words, the relation between independent variables and dependent variable does not change according to the categories of dependent variable, and parameter estimations do not change according to different threshold values. Thus, if there's a dependent variable of J category, "β_k " parameter is only one. On the other hand, there is θ_{j-1}cut-off point (threshold value) for J-1 logit comparisons (Akin and Senturk, 2012, p.185).

It's challenging to interpret parameters in ordered logistic regression. Methods of calculation of standardizes coefficients, calculation of estimated probabilities, calculation of factor change in estimated probabilities and percentage change in estimated probabilities are used for interpreting parameters. Odds ratio can also be used for interpreting parameters. In the event that all other variables are held constant, $\exp(\beta_k)$ is odd ratio for dummy variable. To standardize odds ratios, s_k: showing standard deviation, $\exp(\beta_k * s_k)$ is calculated provided that all other variables are held constant. For continuous variables; the percentage is found by $[\exp(\beta - 1) * 100]$(Ucdogruk vd., 2001).

2. Artificial Neural Networks

ANNs are cellular systems that can receive, store and use information. ANNs are parallel systems, which are formed by connecting many connecting elements with links of variable weights. Multi-layer artificial neural network is the most popular one among many artificial neural networks (Lippman 1987, p.15). ANN is a system based on simple neural networks, which can receive interconnected information as input, process them and submits to other units, and which can even use the outputs as inputs again (Pissarenko, 2002, p.35). ANN simulates the operation of a simple biological neural system. ANNs provide solutions to problems, which normally requires a person's natural ability to think and observe.Artificial neural networks are computer systems which are developed to perform some characteristics of a human brain automatically without getting any help such as getting new information through learning, creating new information and discovering (Oztemel, 2003, p.29). Artificial neural networks are used to achieve one or more processes including learning through using available data, associating, classification, generalization and optimization (Sen, 2004, p.13).

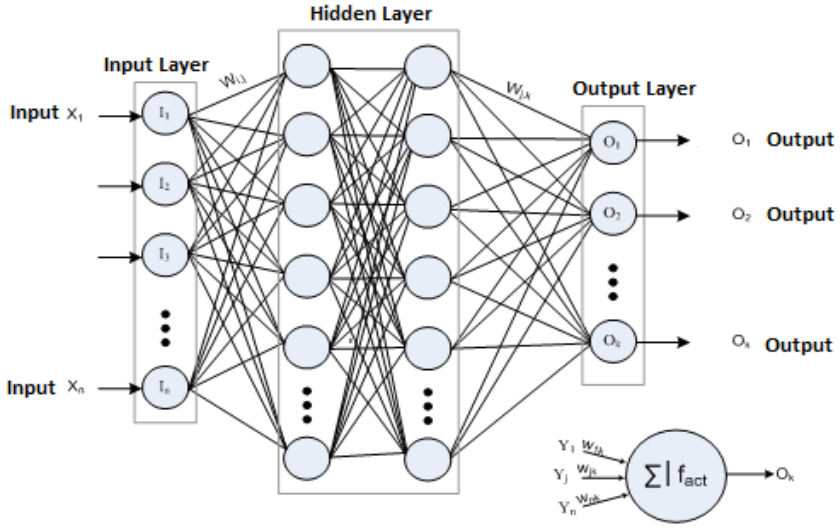


Figure 1. General Structure of an Artificial Neural Network (Bilghan and Turgut, 2010, p.275).

Use of artificial neural networks in such problems whose algorithmic solution could not be found has increased due to the fact that artificial neural networks can find solutions to new occurrences by way of examining former instances and learning the relationship between inputs and outputs of the said occurrence, regardless of whether the relationship is linear or not, from the current instances in hand. The biggest problem in artificial neural networks is that there is a need for such artificial neural networks that contain either very large neurons or multi-layered and a great amount of neurons in order to solve complicated problems (Kohonen,1987, pp. 1-79). An artificial neural network is an intensively parallel-distributed processor which is comprised of simple processing units, has a natural tendency to collecting experiential information, and enabling them to be used (Haykin, 1999, p.2). In a general artificial neural network system, neurons gather on the same direction to form layers (Yildiz, 2001, pp.51-67). There is parallel flow of information from the input layer to the exit in an architectural structure. Such flow is possible with parallel placed cells.

3. Method

Economic development and growth are among the most important objectives for all countries. For this reason, human development index has become a widely preferred and recognized numerical indicator for comparison and classification of countries. The purpose of this research is to compare the classification success of Human Development Index and determine the by using ordered logistic regression as well as Elman ANN, multi layer ANN and LVQ network as artificial neural networks.

In this study, data of highly developed, developed, moderately developed and less developed countries produced by Human Development Report Office of United Nations Development Program (UNDP), which is published annually. This data covers a period of three years and is obtained from 2010-2012 Human Development Index published at United Nations Development web site.Since it was difficult to find data about less developed countries, they are excluded from the analysis. The study classifies 81 countries based on data of three years. With the addition of one more country in 2012, the universe of the study is total of 244 countries.

Income, education and life expectancy are indispensable components in calculation of human development index. Since human development level would be subject to multiple classification process, 11 independent variables were added to these three variables, thus 14 independent variables were studied in order to enable a more detailed examination of methods. The following variables were used in the analysis respectively:

X1: BÖO “infant mortality rate”,
X2: GSMH “gross national product”,
X3: LKO “high school enrolment”,
X4: BY “growth”,
X5: DYY “direct foreign investment”,
X6: ET “energy consumption”,
X7: EÜ “energy production”,
X8: E “inflation”,
X9: IH “export”,
X10: IKS “number of internet users”,
X11:ISZ “unemployment”,
X12: ITH “import”,
X13: MTAS “number of mobile phone subscribers”,
X14: SH “health expenses”

Answer variables for Human Development are coded as follows for Ordered Logistic Regression Analysis:

0: Moderately Developed,
1: Developed,
2: Highly Developed.

3.1. Data Analysis Method

Countries are classified for their Human Development Level by using Stata 11.2 package, Ordered Logistic Regression and Matlab 2012 software with Elman, Multi Layer Neural Networks and LVQ Network methods. In artificial neural network

method, if a country is highly developed in human development, it is valued as 1 while the others' output values are stated as 0.

3.2. Ordered Logistic Regression Analysis

Stata 11.2 statistical analysis program is used to classify Human Development Index of countries.

Results of Ordered Logistic Regression Analysis of Variables Effecting Human Development Level

						Number of obs		244
						LRchi2(12)		**316,95**
						Prob> chi2		**0.0000**
Log likelihood		-110,706				**Pseudo R2**		**0,617**
HDII	**Coef.**	**Std. Err.**	**Wald**	**z**	**P>z**	**Odds Ratio**	**[95% Conf. Interval]**	
BÖO	-0,132	0,035	13,985	-3,740	0,000	0,876	0,817	0,939
GSMH	0,000	0,000	1,840	1,360	0,174	1,000	1,000	1,000
LKO	0,017	0,013	1,749	1,320	0,186	1,017	0,992	1,043
BY	-0,076	0,063	1,482	-1,220	0,223	0,926	0,819	1,048
DYY	0,000	0,000	0,175	-0,420	0,674	1,000	1,000	1,000
ET	0,000	0,000	0,603	-0,780	0,438	1,000	1,000	1,000
EÜ	0,000	0,000	0,011	-0,110	0,916	1,000	1,000	1,000
E	0,051	0,037	1,837	1,360	0,175	1,052	0,978	1,131
IH	0,058	0,023	6,475	2,540	0,011	1,059	1,013	1,108
IKS	0,104	0,020	26,893	5,190	0,000	1,109	1,067	1,154
ISZ	-0,005	0,044	0,012	-0,110	0,914	0,995	0,914	1,084
ITH	-0,062	0,024	6,788	-2,610	0,009	0,940	0,897	0,985
MTAS	-0,006	0,009	0,506	-0,710	0,477	0,994	0,977	1,011
SH	0,237	0,110	4,616	2,150	0,032	1,267	1,021	1,573
/cut1	1,563	1,539					-1,453	4,580
/cut2	6,487	1,614					3,323	9,651

Results of the ordered logistic regression analysis of dependent and independent variables using mlogit command in Stata 11.2 package program are shown in Table 5. As seen in Table 5, the number of observations is 244 in the model and χ^2value is statistically significant ($p < 0.01$). Log likelihood value of the model was found to be -110.71. The first column of Table 5 shows β coefficients of ordered logistic regression analysis. BÖO "infant mortality rate", IKS "number of internet users" and ITH "import" with 0.01 significance level and IH "export" and SH "health expenses" with 0.05 significance level were observed as statistically significant. BÖO and ITH variables above the statistically dependent variable is marked negative while estimated value of İKS, İH and SH variables are marked positive. In addition, marginal effects will be calculated for the change in probabilities of dependent variable pursuant to change in β coefficients.

In odds ratio, one unit increase in BÖO variable decreases odds of high level human development rate by 12.4% while one unit increase in ITH variable decreases it by 6% provided that all other independent variables are held constant against moderate and low level of human development rate. One unit increase in IH variable increases odds of high level human development rate by 5.9% while one unit increase in IKS variable increases it by 10.9% and one unit increase in SH variable increases it by 26.7% against moderate and low level of human development rate. Thus, the most important variable that has a positive effect on human development level is SH "health expenses" variable, the second one is IKS "number of internet users" variable and the

third one is IH "export" variable while the most important variable with negative effect is BÖO "infant mortality rate". Table 7 shows the results of 244 countries' human development level classification results by using ordered logistic regression analysis.

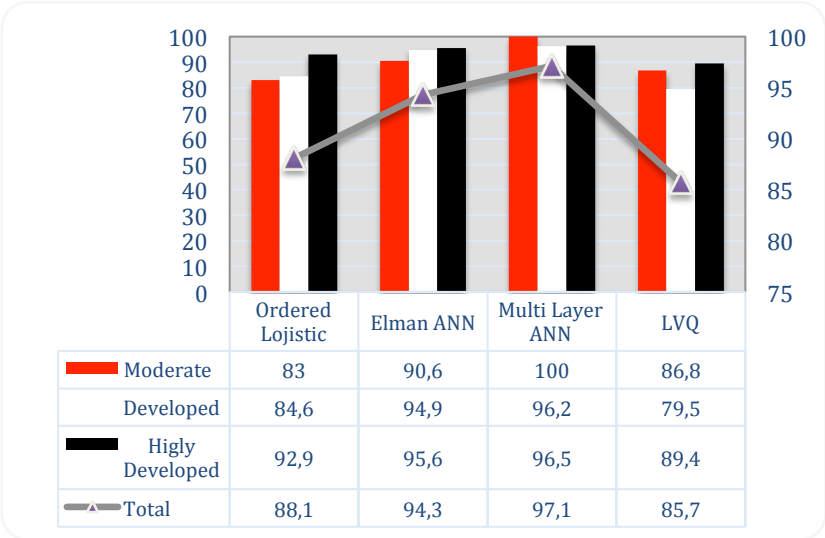
3.3. Artificial Neural Networks Analysis

Matlab R2012a computer software was used to establish artificial neural network models. In order to determine the appropriate artificial neural network method, trial-and-error method is commonly used and many tests are performed. In this context, different combinations of parameters including number of hidden layers, number of nodes in hidden layers, momentum term, activation function, number of cycles were tested on both the training set and the test set to find the best performing network. Elman artificial neural network, multi layer artificial neural network and LVQ network were used as artificial neural networks in this study.

3.4. Findings and Conclusion

In this study, human development level of countries were classified by using ordered logistic regression, Elman ANN, multi layer ANN and LVQ network. Multiple classification methods were used to determine the relation between moderately developed, developed and highly developed countries as the dependent variable with ordered variable of 3 categories and 14 independent variables. The data of 81 countries, which has United Nations Development Program's Human Development Index, between the years of 2010-2012 were used in this study.

In ordered logistic regression analysis among 14 independent variables, it was observed that BÖO "infant mortality rate" and ITH "import" had a statistically significant negative effect while IKS "number of internet users", IH "export" and SH "health expenses" a statistically significant positive effect. Thus, the most important variable that has a positive effect on human development level was SH "health expenses" variable and the second one was IKS "number of internet users" varible while the most important variable with negative effect was BÖO "infant mortality rate". The classification of Turkey, which was classified as a developed country with regards to human development in 2010, 2011 and 2012, was successfully estimated by using ordered logistic regression analysis. Chart 1 shows percentage values of all four methods of analysis for moderate, development and high development levels and the overall classification accuracy.



	Ordered Lojistic	Elman ANN	Multi Layer ANN	LVQ
Moderate	83	90,6	100	86,8
Developed	84,6	94,9	96,2	79,5
Higly Developed	92,9	95,6	96,5	89,4
Total	88,1	94,3	97,1	85,7

Chart 1. Performance Criterion for All Four Methods of Analysis

Multi Layer ANN analysis had the best performance among all. Its classification success was 100% for moderately developed countries, 96.2% for developed countries and 96.5% for highly developed countries. As a result of comparison of analyses, it's seen that Muli Layer ANN provides results with a higher accuracy percentage compared to Elman ANN while ordered logistic regression analysis provides results with a higher accuracy compared to LVQ network. In all four methods of analysis, it was proved that Multi Layer ANN had a better performance compared to the other three methods with regards to total classification results of moderately, developed and highly developed countries.