Prediction of Gross Domestic Product for South Asian Countries

Dr. Kavita Pabreja
Maharaja Surajmal Institute (affiliated to Guru Gobind Singh Indraprastha University),
New Delhi, India
kavita_pabreja@rediffmail.com

***Abstract*— Gross Domestic Product (GDP) is one of the most significant indicators to understand the economical status of an economy. Economic analysis provides an insight into the essentials of an economy. It is a systematic process for determining the optimum use of scarce resources and selecting the best alternative to achieve the economic goal. The economic health of a country depends upon many factors *viz.* human resources, natural resources, capita formation, technological development and social and political factors. The objective of this study is to find out correlation among health, climate, and education related indicators of developing South-Asian countries, select the reduced subset of indicators and then forecast GDP. The datasets have been collected from officially documented international sources.**

***Keywords*— *Gross Domestic Product, Health, Education, Development, Correlation, Artificial Neural Network*.**

I. INTRODUCTION

The economic development of every nation is closely related to the increase in the utilization and burning of fossil fuels, coal, oil, and natural gas by factories and electric power plants, motor vehicles, and family units. The consequential carbon dioxide (CO_2) emissions have turned into the largest source of greenhouse gases that do not allow the infrared radiation from the earth to leave the atmosphere and build the risk of global warming [1]. It has been recommended that every country must put in efforts to reduce these harmful CO_2 emissions for the sake of its citizens.

These emissions affect the agriculture directly that causes adverse effects on the economy of a country. As suggested by Smith [2], a little enhancement in average temperature worldwide (approximate change of 2°C, calculated with reference to year 1990 temperature readings) would cause in net negative impacts on GDP of various developing countries and net positive market impacts on GDP of several developed economies. As a result, this would rise gaps in the income inequality among countries across the globe.

Health of people of a country plays a vital role in its growth and development. The importance of Adult Survival Rate on growth rates for poor countries has been discussed by [3]. The authors discussed that other parameters of health such as disease occurrence rates and cognitive functioning are imperative for sustaining a balanced provision of trained manpower which is a significant component for growth of economy. In another study, the authors have examined the role of health to economic growth [4] and concluded that good health affects the economic growth of a nation in a constructive and numerically significant way.

Education is one of the foremost mechanisms for decreasing poverty and inequality and lays a basis for continual economic growth. The effect of education on GDP has been studied by [5]. The authors argued that the skills available in the labour force and the value of those skills are important determinants of economic performance of any country. There is requirement of workers with higher levels of education in order to handle complex services and production systems [6]. The data of over six five year periods for 65 countries has been utilized by [7] and applying Bayesian models, the authors recommended that with the improvement in education, the economic growth of a country improves.

In the past, various Soft Computing techniques have been utilized by many researchers for forecasting GDP of a country. Neural networks have been applied for forecasting of macro-economical variables and an evaluation of different linear and non-linear models has been done [8]. The authors have found that multivariate linear models are better. Different variants of neural networks have been used for anticipating the future of Egypt's cereal and it was concluded that ANN performed better than ARIMA [9]. GDP of Britain has been forecasted by authors using ANN and they have compared two different training algorithms [10]. The forecasting ability of ANN to anticipate financial output increase based on financial variables has been found to be better than linear models, as discussed in [11]. The GDP of Malaysia has been forecasted based on various economical indicators [12]. The authors have also compared econometric approaches with ANN and have demonstrated that ANN has better performance. A combination of ANN and ARIMA has been also been experimented and in comparison with ANN or ARIMA, this hybrid model has produced more convincing results [13].

Most of these studies are based on either statistical relationship between different indicators or application of soft computing techniques using economical indicators as input variables. This paper differs from all of these studies as it is based not only on economical indicators but also on Education statistics and Health. Also this piece of research is for five south-asian countries *viz.* India, Bhutan, Bangladesh, Sri Lanka and Pakistan. In this paper, we have analysed the

most current and precise global development datasets accessible at The World Bank's official website and using strong correlations as the basis of selection of a small subset of features, we have forecasted the GDP of south-asian countries. The paper is arranged as follows: - section 2 discusses Data and Methodology including data collection, countries under study, indicators used for the analysis and feature reduction. Section 3 discusses the concept of Artificial Neural Network used for forecasting GDP, Results and discussions are explained in section 4 followed by conclusions as section 5.

II. DATA AND METHODOLOGY

Datasets employed for the study have been downloaded from The World Bank website. Various steps followed *viz.* selection of countries, different indicators responsible for economic growth and development, pre-processing of datasets and finally reduction of datasets, have been explained in following sub-sections.

A. Data Collection

There is huge data available for download for 249 countries, for 1343 attributes for 55years for the time period 1960 till date [6]. The countries as per their development status have been categorized by The World Bank in four groups: High Income group, Upper Middle Income group, Lower Middle Income group and Low Income group.

B. Countries under study

This study is based on the developing countries of South Asian region. The countries under study are India, Bhutan, Bangladesh, Pakistan and Sri Lanka.

*C. **Indicators under study***

In this paper, data related to three important determinants of GDP have been selected. The records corresponding to attributes describing the World Development Indicators viz. Education statistics, Health, and environmental statistics have been collected for the time period 1998-2013 [6]. The list of attributes under consideration, under each of these two groups has been explained below.

World Development Indicators

World Development Indicators (WDI) is the key World Bank repository of development data, assembled from officially distinguished global resources. It includes the latest and precise global development data, at national, regional and global level. The indicators considered under the study are as follows: Agriculture, value added (% of GDP); CO_2 emissions (metric tons per capita); Domestic credit provided by financial sector (% of GDP); Electric power expenditure (kWh per capita); Energy utilization (kg of oil equivalent per capita); Exports of goods and services (% of GDP); External liability stocks, total (DOD, current US$); Foreign direct investment, net inflows (BoP, current US$); GDP (current US$); GDP growth (annual %); GNI per capita, Atlas method (current US$); GNI, Atlas method (current US$); Gross capital formation (% of GDP); Imports of goods and services (% of GDP); Inflation, GDP deflator (annual %); Internet users (per 100 people); Life expectancy at birth, total (years); Merchandise trade (% of GDP); Mobile cellular subscriptions (per 100 people); Mortality rate, under-5 (per 1,000 live births); Population density (people per sq. km of land area); Population growth (annual %); Population, total; Surface area (sq. km); Total debt service (% of exports of goods, services and primary income); Urban population growth (annual %).

Education statistics

The World Bank EdStats Query holds approximately 2,500 internationally comparable education indicators for contact purpose, evolution, conclusion, literacy, educators, population, and money spent by countries. The indicators elaborate the education sequence from pre-primary to tertiary education. The inquiry placed on huge repository of datasets provides data from international learning assessments, equity data from domestic analysis, and protuberance data till year 2050. The attributes selected for this study are as follows: Government expenditure on teaching and learning as % of GDP (%); Gross registration ratio, primary, female (%); Gross registration ratio, primary, gender equality index; Net registration rate, pre-primary, female (%); Net registration rate, primary, female (%); Out-of-school children of primary school age, both sexes (number); Percentage of all students in tertiary education registered in ISCED 6 and 7, both sexes (%); Primary achievement rate, both sexes (%); Ratio of girls to boys in primary and secondary education (%); School registration, primary (% gross); School registration, secondary (% gross).

Health, Nutrition and population statistics

Data about Key health, nutrition and population statistics gathered from a variety of international sources has been utilized for the study. The indicators used are as follows: Adolescent fertility rate (births per 1,000 women ages 15-19); Health expenditure per capita (current US$); Health expenditure, total (% of GDP); Health expenditure, total (current US$); Immunization, DPT, measles, polio; Improved cleanliness facilities (% of population with access); Improved cleanliness facilities, urban (% of urban population with access); Improved water resource (% of population with access); Improved water resource, urban (% of urban population with access); occurrence of tuberculosis (per 100,000 people); Out-of-pocket health spending (% of private expenditure on health); Out-of-pocket health spending (% of total spending on health); occurrence of anemia amid children (% of children under 5); occurrence of anemia amid non-pregnant women (% of women ages 15-49); occurrence of anemia amid pregnant women (%); occurrence of tuberculosis (per 100,000 population); Tuberculosis case recognition rate (all forms); Tuberculosis death rate (per 100,000 people).

D. Data Pre-processing and feature reduction

Data for the analysis is approximately 90% complete. The missing values have been filled up with estimated values using linear regression model.

Pearson's correlation coefficient has been calculated to find out the statistical relationships between two or more variables of health, climate change, education and other development related indicators so as to perform feature reduction. In positively correlated variables, increase or decrease in the value of one variable, causes the value of other variable also to increase or decrease accordingly. In negatively correlated variables, the value of one variable decreases as the value of the other increases and vice-versa. Out of the 57 attributes, there are many that are strongly correlated and there are certain findings that are quite unexpected and are discussed in Results and Discussions section. Further only following eight attributes as given in Table I, have been selected based on correlation coefficient values, for anticipation of GDP corresponding to South Asian countries, using Artificial Neural Networks.

TABLE I LIST OF 8 SELECTED INDICATORS FOR FORECASTING GDP (CURRENT US$)

Time	Pupil-teacher ratio in primary education (headcount basis)
CO2 emissions (metric tons per capita)	Pupil-teacher ratio in secondary education (headcount basis)
Government expenditure on education as % of GDP (%)	Health expenditure, total (% of GDP)
Percentage of students in secondary vocational education who are female (%)	GDP growth (annual %)

III. ABOUT ARTIFICIAL NEURAL NETWORK

An Artificial Neural Network is a numerical representation designed with motivation from the composition of biological neural networks. A neural network comprises of an interrelated collection of artificial neurons, and the neurons process information using a connectionist method to calculation [14,15].

In this paper, ANN has been implemented using Matlab 7.0 software (matrix laboratory). A two layer MLP Back Propagation network with default settings has been applied for the training and testing of artificial neural network. In hidden layer of the back propagation network, Tangent-sigmoid transfer function is employed. A pure linear transfer function is utilized in the output layer. The Back propagation learning algorithms *viz.* trainlm has been used for prediction of percentage of GDP. The input dataset comprises of attributes selected based on correlation coefficient, mentioned in Table I. The output data corresponds to GDP of the country under consideration. A snapshot of dataset is illustrated in table II. Columns 1 to 8 of this table are used as input and column 9 corresponds to target. The range of inputs and outputs have been adjusted and hence lies in the range [-1,1]. A pre-defined function premnmx() provided by Matlab has been used for the purpose.

TABLE II SAMPLE OF DATASET FOR ANTICIPATION OF GDP

country	Time	CO2 emissions (metric tons per capita)	Government expenditure on education as % of GDP (%)	Percentage of students in secondary vocational education who	Pupil-teacher ratio in primary education (head count basis)	Pupil-teacher ratio in secondary education (head count basis)	Health expenditure, total (% of GDP)	GDP growth (annual %)
Bangladesh	1998	0.19	2.10	23.42	43.50	36.31	2.25	5.18
Bangladesh	1999	0.20	2.13	23.83	44.00	37.41	2.29	4.67
Bangladesh	2002	0.25	2.02	25.59	45.50	34.37	2.59	3.83
Bhutan	2005	0.61	7.08	33.65	31.05	28.11	5.28	7.12
Bhutan	2006	0.59	6.00	32.96	29.18	21.88	5.27	6.85
Bhutan	2007	0.58	5.00	33.96	29.50	22.70	5.88	17.93
India	2005	1.23	3.13	20.08	39.30	30.90	4.28	9.28
India	2006	1.29	3.09	22.54	38.60	29.30	4.25	9.26
India	2007	1.37	3.12	25.00	37.90	27.90	4.23	8.61
Pakistan	2011	0.94	2.22	41.64	39.83	21.30	3.01	2.75
Pakistan	2012	0.93	2.14	42.76	41.35	21.04	2.76	3.51
Pakistan	2013	0.92	2.49	42.99	42.55	20.17	2.70	4.37
Sri Lanka	1998	0.42	3.05	35.00	27.10	16.86	3.71	4.70
Sri Lanka	1999	0.47	2.90	35.80	26.80	16.89	3.65	4.30
Sri Lanka	2000	0.55	2.81	36.30	26.50	16.92	3.77	6.00

TABLE III DETAILS OF VALUES OF R CORRESPONDING TO DIFFERENT DATASETS UNDER CONSIDERATION

S.No.	Datasets	Correlation Coefficient R
1	Training	0.99986
2	Validation	0.99196
3	Testing	0.99402
4	All	0.99694

IV. RESULTS AND DISCUSSIONS

There are 57 variables under various categories and we have calculated correlation coefficient amongst them. The variables having positive correlation of more than 0.9 have been selected. Since a modification in the value of one variable will be able to forecast the change in the similar way in the second variable, we have selected a subset of variables with 8 features to be used for forecasting GDP in case of South-Asian countries. There are certain correlations that are implied but there are certain findings which are quite innovative, infrequent and interesting.

It has been found that there exists very strong correlation coefficient of more than 0.9 between External debt stocks, Health expenditure and Foreign direct investment, net inflows. The Health Expenditure is strongly correlated with CO_2 emissions, Electric Power consumption and Energy use.

Hence after selecting the important features based on correlation coefficient, we have trained the ANN with the mentioned datasets, so as to forecast GDP. The performance measurement in terms of Mean Square Error between actual and forecasted GDP stands at 0.0377 at 14 epochs, which is quite convincing, shown in Fig. 1.

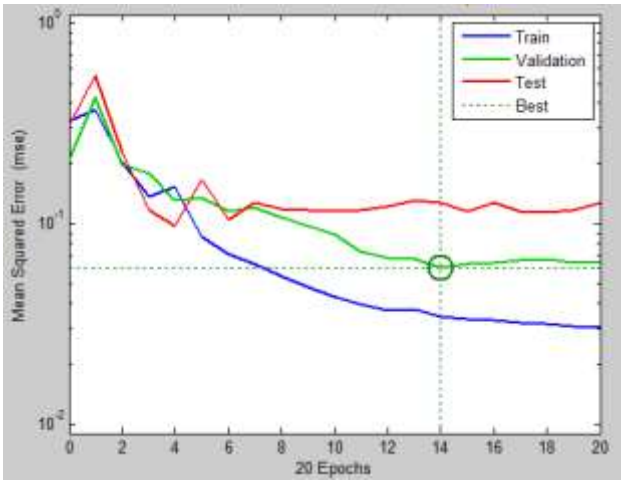

Fig. 1. Result of training, testing and Validating ANN for South-Asian countries data using learning function trainlm.

The correlation coefficient (R) representing the measure of linear relationship between actual GDP and the predicted GDP has been found as mentioned in table III. A very strong correlation coefficient has been reported that advocates of the fact that the ANN with trainlm back propagation learning algorithm, transfer function namely tangent-sigmoid applied in hidden layer of the back propagation network and pure linear transfer function utilized in the output layer is able to predict the GDP correctly.

V. CONCLUSION

Important correlations between various World Development Indicators have been found and discussed. It is concluded that ANN has capability to provide conclusive results and is very appropriate for anticipating Gross Domestic Product. Using the input parameters describing Education, Health, Climate change indicators, the ANN has been trained to forecast the GDP for five countries of South Asia. This research has noticeably demonstrated that application of Soft Computing techniques can facilitate in providing advanced details for forecast of GDP.

This paper differs from all of the earlier studies explained in Introduction section, as it is based not only on economical indicators but also on Education statistics and Health, also the anticipation of GDP is not done just for a single country but for five countries of South Asian segment of economies. In this paper, selection of a small subset of features has been done on the basis of strong correlations between more than 50 indicators. Also, a few unusual correlations have been observed that have already been discussed in Results section.

The study is further to be extended for finding correlations between different development indicators for remaining categories of economies.

Acknowledgment

The author would like to express deepest sense of gratitude to Prof.(Dr.) R. K. Datta, a renowned personality in the field of Information Sciences and Meteorology, currently Director, Mohyal Educational Research Institute of Technology, for his encouragement, guidance and mentoring. Without his support, it would not have been possible to take up research in this challenging field.

References

[1]. Economic Development and the Risk of Global Climate Change, Available from:- www.worldbank.org/depweb/beyond/beyondbw/begbw_14.pd

[2]. J.B.Smith,, Vulnerability to Climate Change and Reasons for Concern: A Synthesis. In: Climate Change 2001: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change [J.J. McCarthy *et al.* Eds.]. Cambridge University Press, Cambridge, UK, and New York, N.Y., U.S.A., (2001).

[3].A. Bhargava,, D.T.Jamison, L. Lau, C.J.L.Murray, Modeling The Effects Of Health On Economic Growth, GPE Discussion Paper Series: No. 33 Evidence and Information for Policy World Health Organization.

[4]. E.David, D.E.Bloom, D. Canning, J.Sevilla,, The Effect of Health on Economic Growth: Theory and Evidence, NBER Working Paper No. 8587, Issued in November 2001

[5]. From: Education at a Glance 2012 Highlights Access the complete publication at: http://dx.doi.org/10.1787/eag_highlights-2012-en

[6]. E.A.Hanushek,, D.T.Jamison, E. Jamison, L. Woessmann, Education and Economic Growth, Education and Economic growth, Education Next, spring 2008, Vol.8, No.2.

[7]. J.C. Cuaresma,, T. Mishra, The role of age-structured education data for economic growth forecasts, Available from :- www.isid.ac.in/~pu/conference/dec_09_conf/Papers/TapasMishra.pdf

[8]. N.R. Swanson, H. White, A model selection approach to real time macroeconomic forecasting using linear models and artificial neural networks, Review of Economics and Statistics, Vol.79, pp. 540-550, (1997).

[9]. N. Kohzadi, S.B. Milton, I. Kaastra, B.S.Kermanshahi, D. Scuse, Neural Networks for forecasting : An introduction, Canadian journal of Agricultural Economics, Vol. 43, pp.463-474, (1995).

[10].Y. Li, Macroeconomics modeling on UK GDP growth by neural computing, technical report, CSC-95009, 1995

[11].G. Tkacz, S. Hu, Forecasting GDP growth using artificial neural networks, Working paper 1999-3/ Bank of Canada, pp.1-24, (1999).

[12].M.Z.H.M. Junoh, Predicting GDP growth in Malaysia using knowledge based economy indicators: a comparison between neural networks and econometric approach, Sunway college journal, vol. 1, pp. 39-50, (2004).

[13].G. P. Zhang, Time series forecasting using a hybrid ARIMA and neural network model, neurocomputing, Vol. 50, pp. 159-175, (2003).

[14].S.N. Sivanandam S. Sumathi S.N. Deepa, Introduction to Neural Networks using Matlab, Tata McGraw Hill Education Private Ltd., 2009.

[15].B. Kosko, Neural Networks and Fuzzy Systems, Prentice Hall of India Ltd., 2005.