



## Forecasting the Long Term Economics Status of Bangladesh Using Machine Learning Approaches from 2016-2036

Minhaz Uddin <sup>a</sup>, Shraboni Rudra <sup>a</sup>, Mohammed Nazim Uddin <sup>a,\*</sup>

<sup>a</sup> Department of Computer Science, East Delta University, Chittagong, Bangladesh.

\*Corresponding Author  
nazim@eastdelta.edu.bd  
(Mohammed Nazim Uddin)

Received : 22-03-2019  
Accepted : 04-05-2019

**ABSTRACT:** It is a piece of happy news for us that Bangladesh has been now converted to a developing country. The United Nation and World Bank have recognized it. But they have a condition that we need to continue this economic progress till 2024 for getting a permanent recognition. The economic condition depends on many factors like Gross Domestic Product (GDP), Personal Saving, Private Sector Investment, Gross National Income (GNI) per capita, Human Assets Index (HAI) and Economic Vulnerability Index (EVI). This paper portrays the forecast of the long-term economic condition of Bangladesh as an independent variable which is a year and the dependent variables are GDP, private sector investment and personal saving. The living conditions of a country depend on GDP. Personal saving and Private Sector Investment are also important parts of a country's economy. If we will forecast these attributes properly, then we can determine the economic condition of Bangladesh and living status of the people more accurately. Therefore, we can determine that Bangladesh can fulfil the condition of getting permanent recognized as a developing country. For forecasting these attributes, we proposed a model which consists of Karl Pearson's coefficient and modified linear regression techniques. For improving performance, we modify linear regression by gradient boosting. This experiment shows that our model gives us more accurate forecasting about GDP, Private sector investment and Personal Saving.

**Keywords:** Economics, Correlation, Linear Regression, Gradient Boosting, Karl Pearson Coefficient, Statistical Approach.

### 1. Introduction

The economy is a term which is related to production and consumption activities that focus to allocate the nation's resources for the public benefit. The development of a nation mainly depends on its economic status. The main aim of our paper is to focus on forecasting GDP, Personal Saving and Private Sector Investment. GDP is the monetary value of all the finished goods and services produced within a country's borders in a specific time period [1]. GDP is commonly used as an indicator of the economic health of a country [1]. Bangladesh posted a GDP growth rate of 7.9% in FY-18 and is expected to achieve 7.5% in FY-19, according to a report by the Asian Development Bank [2]. Personal saving refers to a person saving money, property, etc for future use or invests in anything in order to get a profit in the future from this investment. People save their money at Banks, Insurance Company or invest in the stock markets, etc. Private sector Investment is also the result of personal saving. Nowadays, Banks, Insurance companies are interested to invest in the private sectors with a low-interest rate. As a result, the unemployed are trying to start

their own business which changes not only their economic condition but also the economic condition of Bangladesh [2].

For forecasting GDP, Personal saving and Private sector Investment, we proposed well known Karl Pearson coefficient and linear regression method [4]. Karl Pearson coefficient is a method which helps us to find a relationship between two or more variables [3]. The relationship is of two types. One is positive and the other one is negative. We found three pairs of a relationship between variables like a year and GDP; a year and Personal Saving; year and private sector investment [1]. We have used this coefficient in the linear regression for forecasting purpose. Linear regression is a method to predict a dependent variable, based on the independent variable. Gradient boosting is an ensemble technique which improves the performance of a weak classifier [5]. After applying gradient boosting, our model forecast variables more accurately [4]. If we forecast these variables, we can say that people saving increased or not, if saving increased than investment increased,

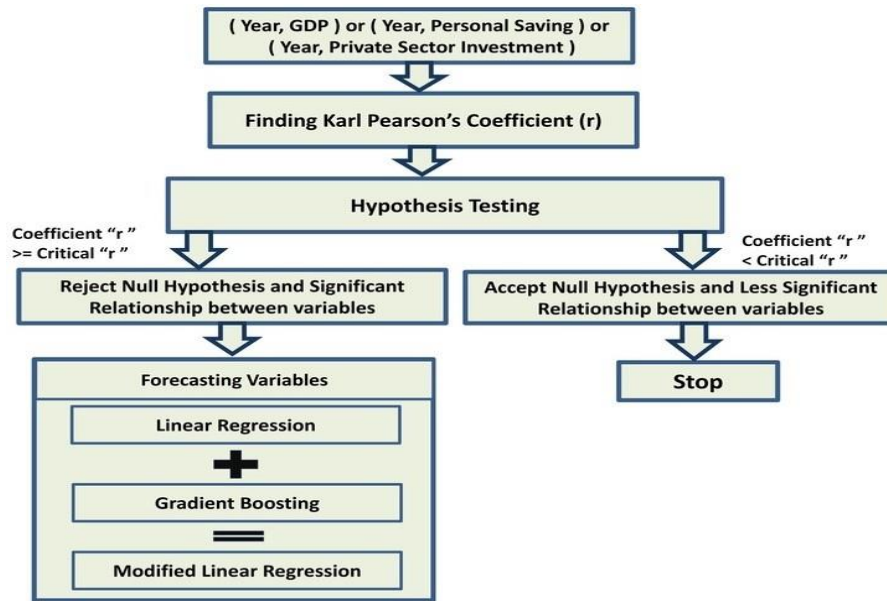


Fig1. The Block Diagram of Proposed Model

if the investment is increased then more jobs field is created which increased the earnings of people. In this situation, the total GDP of a country is also increased [1].

In this paper, Section II describes the Statistical and Machine Learning Algorithms which is used in our model, Section III is a methodology which describes our proposed model, Section IV is an experimental evolution where we describe the description of our data, and Section V is a result analysis step Where, we describe the result of our model, Section VI we concluded our work.

## 2. Statistical And Machine Learning Algorithms

### 2.1 Karl Pearson's Coefficient

It is a widely used mathematical method wherein the numerical expression is used to calculate the degree and direction of the relationship between linear related variables. Pearson's method, popularly known as a Pearsonian Coefficient of Correlation, is the most extensively used quantitative methods in practice [3]. The coefficient of correlation is denoted by "r". The value of the coefficient of correlation (r) always lies between '±1'. Such as, if 'r' is equal to '+1' then perfect positive correlation. If 'r' is equal to '-1' then perfect negative correlation. If 'r' equal to '0', no correlation. The method worked when the relationship between the variables is linear, there are a large number of independent causes that affect the variables under study so as to form a Normal Distribution, The variables are independent of each other [3]. For finding coefficient 'r' we use the following formula (1):

$$r = \frac{\sum(XY)}{\sqrt{\sum(x - \bar{x})^2} \sqrt{\sum(y - \bar{y})^2}} \quad (1)$$

Here, x, y are two continuous variables.

### 2.2. Linear Regression

Linear regression is used for finding a linear relationship between the target and one or more predictors. There are two types of the linear regression- Simple and Multiple. Linear regression is useful for finding a relationship between two continuous variables. One is the predictor or independent variable and another is a response or dependent variable. It looks for a statistical relationship, but a not deterministic Relationship. The relationship between the two variables is said to be deterministic if one variable can be accurately expressed by the other. The core idea is to obtain a line that best fits the data. The best fit line is the one for which total prediction error (all data points) is as small as possible [4]. The error is the distance between the points to the regression line. The general formula of linear regression is (2):

$$y = a + bx \quad (2)$$

Here, y is the dependent variable, x is the independent variable and a, b are coefficient.

### 2.3. Gradient Boosting

It is introduced by Friedman in 2001. It is also known as MART (Multiple Additive Regression Trees) and GBRT (Gradient Boosted Regression Trees). Gradient boosting is widely used in industry and has won many Kaggle competitions. GBM constructs a forward stage-wise additive model by implementing gradient descent in function space [5]. The first step of the GBM is calculated Error. For this we use (3):

$$\text{Error} = \text{Actual Value} - \text{Predicted value} \quad (3)$$

Then we use (4) for getting the new predicted value.

$$\text{New Predicted Value} = \text{Predicted Value} + \text{Error} \quad (4)$$

### 3. Methodology

For the purpose of forecasting, a forecasting model was defined. The working principle of the proposed model has been shown in Figure 1. The basic modules of our model are described below:

#### 3.1. Finding Karl Pearson's Coefficient

According to Fig.1, the first step of our model is to find the coefficient (r). For this purpose, we use (1). Before using (1), we need to find the value of the equipment of the equation [3]. So, we are finding the mean of variables using (5) and (6) which is denoted by ' $\bar{x}$ ' and ' $\bar{y}$ '. Where, 'x' and 'y' are the values and 'N' is the total number of individuals. Then we have found the value of "XY" using (7). At last, we put all the values in (1) and find the value of "r".

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} \quad (5)$$

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N} \quad (6)$$

$$XY = (x - \bar{x}) * (y - \bar{y}) \quad (7)$$

**Table 1.** Level Of Significance For A Two-Tailed Test

df	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.02$
1	$\pm 0.988$	$\pm 0.997$	$\pm 1.000$
2	$\pm 0.900$	$\pm 0.950$	$\pm 0.980$
3	$\pm 0.805$	$\pm 0.878$	$\pm 0.934$
4	$\pm 0.729$	$\pm 0.811$	$\pm 0.882$
5	$\pm 0.669$	$\pm 0.754$	$\pm 0.833$
6	$\pm 0.622$	$\pm 0.707$	$\pm 0.789$
7	$\pm 0.582$	$\pm 0.666$	$\pm 0.750$
8	$\pm 0.540$	$\pm 0.632$	$\pm 0.716$
9	$\pm 0.521$	$\pm 0.602$	$\pm 0.655$
10	$\pm 0.497$	$\pm 0.576$	$\pm 0.658$

#### 3.2. Hypothesis Testing

The second step of our model is hypothesis testing. We use hypothesis testing for finding that the relationship between variables is significant or not. For this purpose, we need to choose an appropriate significant level ' $\alpha$ '. The most common use value of the " $\alpha$ " is 0.05. Then, we have found the degree of freedom using (8). Then we compare the value of coefficient "r" with a critical value of "r" using Table 1. If coefficient "r" is greater than or equal critical "r", then reject the null hypothesis and the relationship between variables are significant. If coefficient "r" is less than critical "r", then accept the null hypothesis and the relationship

between variables are less significant. If the relationship between Variables is significant then we go to the forecasting step. If the relationship between variables is less significant, then we stop our model.

$$df = N - 2 \quad (8)$$

#### 3.3. Forecasting Variables

##### 1. Linear Regression:

At forecasting step, firstly we are using linear regression for forecast variables. For this purpose, we use (2). For finding the value of 'a' and 'b' we basically use (9) and (10). After finding the values of 'a' and 'b', we formulate an equation like (2). But this equation can't give us a more accurate result, so we need to modify it.

$$b = r \frac{\sqrt{\frac{\sum (y - \bar{y})^2}{N-1}}}{\sqrt{\frac{\sum (x - \bar{x})^2}{N-1}}} \quad (9)$$

$$a = \bar{y} - b\bar{x} \quad (10)$$

##### 2. Gradient Boosting:

By using this method, we can improve the performance of linear regression. For this purpose, first, we are using (4) to find the error. Then this value is added with a predicted value according (5). This process gives 100% accurate result. But we can only apply this method when the actual value is known. In a single sentence, we can say that it is only applicable to training data. We try to modify this method to apply to test data [4][5].

##### 3. Modified Linear Regression:

Using (2), we can predict the value of a dependent variable based on an independent variable. Using this method first we try to predict the value of training data. From this value, we choose one value as an ideal value and calculated the error for this value using (3). If we predict GDP from 2016-2036 then we chose 2016 as an ideal value because we know its actual value. So we calculate an error for this value. Then we can add to the predicted value. But, we can't use the method for 2018-2036 because the actual value of this year is unknown. So in this case, we proposed (11).

$$y = a + bx + nError \quad (11)$$

Here, 'Error' is found in an ideal case using (3), 'n' is a variable which value is 1, 2, 3... a number of the test case. It is increased with 'x'. If starting from 2016 then 'n' is one for 2016, 'n' is two for 2017, which is increased with 'x'. We are always taking the first test case as an ideal case. We can choose a first test case in a way that its actual value is known to us. This approach gives a better result than linear regression [4].

#### 4. Experimental Evolution

**Table 2.** Gdp, Personal Saving And Private Sector Investment With Respect To Year From 2008-2017 [6], [7], [8].

Year	GDP (USD Billion)	Personal Saving (BDT Billion)	Private Sector Investment (Percentage of GDP)
2008	91.631	1210.4	33.82
2009	102.478	1439	35.99
2010	115.279	165.1	40.77
2011	128.638	1897.6	42.26
2012	133.356	2239.5	42.78
2013	149.99	2642	41.58
2014	172.885	2968.2	43.51
2015	195.079	3358.8	44.2
2016	221.415	4328.3	45.09
2017	249.72	5004.6	47.41

For forecasting purposes, we choose four attributes where the year is an independent variable and the dependent variables are GDP, personal saving and private sector investment. The value of year is a positive integer number. The value of other variables is a positive floating point number. We collect all the value of a variable against with year. The relationship between the year and other variables is linear. That is indicated in almost case the value of the variables are linearly increased which are linearly increased with year. This type of data is also known as time series data. The time series model is basically an econometric model. It helps us to forecast data based on previous data. The value of data is changing with time. In Table 2, We see the GDP, personal saving and private sector investment data from '2008' to '2017'. We use this data because it is available for non-commercial research purpose. The value of GDP is USD billion, personal saving is BDT billion and private sector investment is the percentage of GDP published [7]. GDP indicates the total amount of Bangladesh GDP in the year.

Personal saving indicates the total money saving by total Bangladeshi people from their income in a year. Private sector investment indicates how many percentages of total GDP invest in the private sector.

#### 5. Result Analysis

According to Fig.1, First, we are finding the coefficient between variables. The Karl Pearson's Coefficient between year and GDP is '0.97', between a year and Personal Saving is '0.971', between a year and Private Sector Investment is '0.92'. All the Values pass the Hypothesis testing. In the hypothesis testing, the value of the critical 'r' is '0.632' which is less than Coefficient. In the forecasting step, first, we find the linear regression between variables. We found three equations (12), (13) and (14).

$$\text{GDP} = 16.98 \text{ Year} - 34016.2 \quad (12)$$

$$\text{Personal\_Saving} = 406.57 \text{ Year} - 815546.7 \quad (13)$$

$$\text{Private\_Sector\_Investment} = 1.25 \text{ Year} - 2476.3 \quad (14)$$

Here, we put the value of year is equal to '2016' and '2017' for finding values and compare results. We can see in Table 3 that the error rate of this case is very high. It is also increased with a year[7]. In this case, we need to modify linear regression like (11). We take the year '2016' as an ideal case. Finding the error for GDP, personal saving and private sector investment by taking a year is equal to '2016'[4]. The error for GDP is '5.93', Personal saving is '229.88' and Private Sector Investment is '1.39'. Using this error, we formulate (15), (16) and (17).

$$\text{GDP} = 16.98 \text{ Year} - 34016.2 + 5.93n \quad (15)$$

$$\text{Personal\_Saving} = 406.57 \text{ Year} - 815546.7 + 229.88n \quad (16)$$

$$\text{Private\_Sector\_Investment} = 1.25 \text{ Year} - 2476.3 + 1.39n \quad (17)$$

Again, we put the value of year is equal to 2016 and 2017 for finding values and compare results.

We can see in Table 4 that the error rate, in this case, is very lower than the previous case.

**Table 3.** Comparison Between Actual Data and Forecasted Data Using Linear Regression

Year	GDP (USD Billion)			Personal Saving (BDT Billion)			Private Sector Investment (Percentage of GDP)		
	Actual Value	Forecasted Value	Error	Actual Value	Forecasted Value	Error	Actual Value	Forecasted Value	Error
2016	221.415	215.48	2.68%	4328.3	4098.42	5.31 %	45.09	43.7	3.08 %
2017	249.72	232.46	6.91%	5004.6	4504.99	9.98%	47.41	44.95	5.18 %

**Table 4.** Comparison Between Actual Data And Forecasted Data Using Modified Linear Regression

Year	GDP (USD Billion)			Personal Saving (BDT Billion)			Private Sector Investment (Percentage of GDP)		
	Actual Value	Forecasted Value	Error	Actual Value	Forecasted Value	Error	Actual Value	Forecasted Value	Error
2016	221.415	221.415	0%	4328.3	4328.3	0 %	45.09	45.09	0%
2017	249.72	244.32	2.16%	5004.6	4964.75	0.79%	47.41	47.73	0.67 %

**Table 5.** The Forecasted Data Of Gdp, Personal Saving And Privatesector Investment With Respect To Year From 2016-2036.

Year	GDP (USD Billion)	Personal Saving (BDT Billion)	Private Sector Investment (Percentage of GDP)
2016	221.41	4328.3	45.09
2017	244.32	4964.75	47.73
2018	267.23	5601.2	50.37
2019	290.14	6237.65	53.01
2020	313.05	6874.1	55.65
2021	335.96	7510.55	58.29
2022	358.87	8147	60.93
2023	381.78	8783.45	63.57
2024	404.69	9419.9	66.21
2025	427.6	10056.35	68.85
2026	450.51	10692.8	71.49
2027	473.42	11329.25	74.13
2028	496.33	11965.7	76.77
2029	519.24	12602.15	79.41
2030	542.16	13238.6	82.05
2031	565.06	13875.05	84.69
2032	587.97	14511.5	87.33
2033	610.88	15147.95	89.97
2034	633.79	15784.4	92.61
2035	656.7	16420.85	95.25
2036	679.61	17057.3	97.89

So, we use (15), (16) and (17) for long-term forecasting purpose. Using this equation we find the value for '2016-2036' which we can see in Table 5.

In Fig.2 we can see that GDP is increased with a year. If GDP is increased then personal saving and private sector investment should be increased with a year. In Fig.3 and Fig.4 we can see that personal saving and private sector investment are increasing with a year. We see that all these



three attributes are increasing with a year; as a result, the economic condition of Bangladesh is progressing [8]. If the economic condition is improved then the living status of peoples are improved. By using this forecast, we find the value of GDP, personal saving and private sector investment from 2016-2036 which help us to take any decision to improve these values.

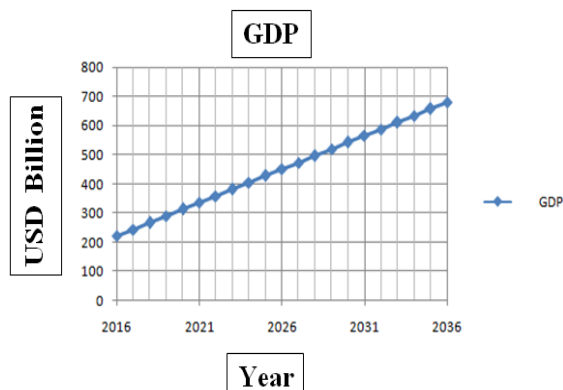


Fig2. Forecast GDP from 2016-2036.

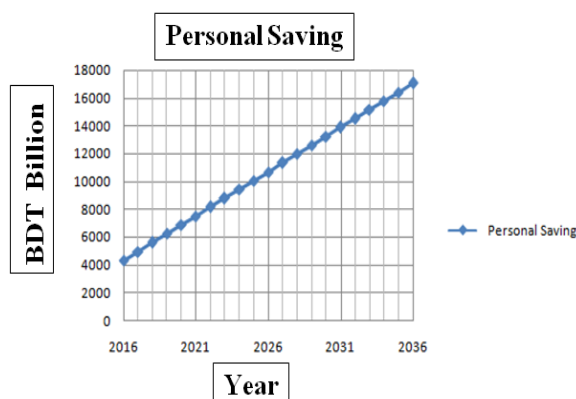


Fig3. Forecast Personal Saving from 2016-2036.

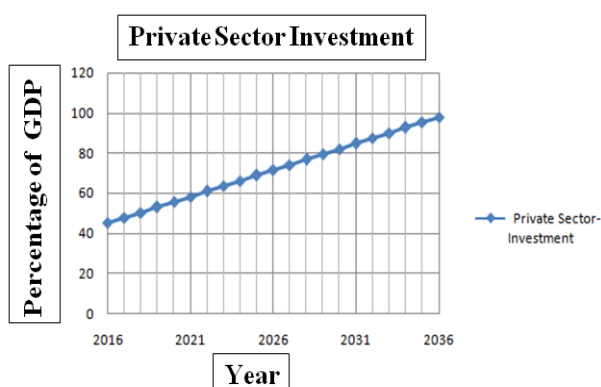


Fig4. Forecast Private Sector Investment from 2016-2036.

In Fig.2 we can see that GDP is increased with a year. If GDP is increased then personal saving and private sector investment should be increased with a year. In Fig.3 and Fig.4 we can see that personal saving and private sector investment are increasing with a year. We see that all these three attributes are increasing with a year; as a result, the economic condition of Bangladesh is progressing. If the

economic condition is improved then the living status of peoples are improved. By using this forecast, we find the value of GDP, personal saving and private sector investment from 2016-2036 which help us to take any decision to improve these values.

## 6. Conclusion

Long-term forecasting about anything is a very difficult task, especially in the economy. Because the economic condition of a country depend on many factors. That factor again influenced by its surrounding condition. We forecast on GDP, private sector investment and personal saving of their values and assuming it is linearly increased. But these values are influenced by the political condition of the country, the environmental condition of a country, etc. But using our model we try to provide a more accurate forecast about this variable. From this forecasting, we can see that the value of these variables is increased by the year. It is a positive sign for our country. If GDP of our country is increased in a way that we can say that the future living conditions of the peoples will be improved. Personal saving is increased so that they will invest more money. As a result, private sector investment is also increased. For these, the unemployment rate will be decreased. It is a good sign for our, the value of GDP, personal saving and private sector investment are increasing but we need to also ensure that we use it for properly for developing in the country. We try to use this model to forecast other variables which influence the economic condition of our nation in the future.

## 7. References

- [1] Martin Feldstein, Underestimating the Real Growth of GDP, Personal Income, and Productivity, J. Econ. Perspe. Spring, (2017) 145–164
- [2] Nicole Maestas, Kathleen J. Mullen, David Powell, The Effect of Population Aging on Economic Growth, the Labor Force and Productivity, NBER (2016)
- [3] Bernard J. Norton, Karl Pearson and Statistics: The Social Origins of Scientific Innovation, Soc. Stud. Sci., (1978) 3-34.
- [4] MAI Navid, NH Niloy, Multiple Linear Regressions for Predicting Rainfall for Bangladesh, science publishing group, (2018) 1-4
- [5] T. Chen, H. Li, Q. Yang, and Y. Yu. General functional matrix factorization using gradient boosting. In Proc of 30<sup>th</sup> Inter. Conf. Mach. Lear. (ICML'13), 1 (2013) 436–444.
- [6] Parves sultan, Trade, Industry and economic Growth in Bangladesh, J. Econ. Coop. 2008, 71-92.

- [7] S Edwards, Trade Orientation, Distortions and Growth in Developing Countries, J. Devel. Econ. (2009) 31-57
- [8] Md Niaz Murshed Chowdhury, nation in the future, Bangladesh, J. world Econ. Res. (2018,) 52-63.

#### About The License

© 2019 The Authors. This work is licensed under a Creative Commons Attribution 4.0 International License which permits unrestricted use, provided the original author and source are credited.