Regression based Price and Yield Prediction of Agricultural Crop

Anjela Diana Corraya
Information and Communication Technology
Mawlana Bhasani Science and Technology University, Bangladesh

Sonia Corraya
School of Engineering and Computer Science
BRAC University, Bangladesh

ABSTRACT

Successfully predicting agricultural economy leads to maximum benefit and best effort utilization for farmers. It is very important and challenging to make correct agricultural decisions for a future stable price and yield balance. For the failure of this prediction, knowledge from historical data could not be used properly for managing agricultural revenue affecting factors. The objective of this research work is to develop an advance and efficient system for predicting future price and yield of agricultural crops considering inflation which is affected by climate conditions and other situational factors. At first, the Auto Regressive Moving Average (ARMA) Model was used for the work but the best estimates would always show a sudden and steep drop in the yield for every crop and would not show an increase until years later. Then the Weighted Linear Regression model was used by giving more weight to recent data which results 78.75% and 83.55% correct prediction for price and yield respectively after comparing the actual and predicted values over previous three years. Initially three types of rice of Bangladesh (Aus, Aman and Boro) are analyzed and there is scope for extensive use of this method.

Keywords

Agricultural Economy, Price, Weighted Linear Regression, Yield, Prediction

1. INTRODUCTION

Economy of South Asian developing countries like Bangladesh is primarily dependent on agriculture[11]. The main reason for this dependency is fertile soil and normally ample water supply [2]. Countries like Bangladesh face challenges in opportunities for agriculture-led growth and food security. There are lots of advancement in recent years to face these challenges. Even after many improvements in cultivation process, irrigation, fertilization, insecticide and hybrid crops production, the profit and longtime benefit from agricultural sector are still considered unpredictable. The main challenge is the intelligent adoption of the unfavorable and favorable factors which are beyond controls like natural disasters, climate change, land fertility change, global warming, political issue, inflation and so on. This unpredictability directly impacts on farmers life standard[9].

One of the main reasons for this is the inability to take right decision in right time considering different factors on agricultural crops and failure to apply extracted knowledge from historical data.

There is no prediction system implemented or applied using the Weighted Linear Regression model in the agricultural sector for facing these challenges.

In this paper the objective is to develop an advance and efficient system for predicting the yield and price of the crops the farmers harvest with an efficient model. Inflation rate is considerd here which is impacted by climate conditions and other situational factors. With this method farmers can know in advance about which crops to plough in the lands in which season so that they could maximize the profit from it and the concerned personnel could successfully forecast the stability of revenue and long term agricultural impacts that a farmer can expect from any crop if they choose to plant it next year. In the past decade, the agriculture sector contributed about three percent per annum to the annual economic growth rate . Proper utilization of the prediction output of our proposed system would multiple this contributions which will increase the annual economic growth rate and decrease the present poverty rate as well.

Rice is the staple food in the everyday diet of most South Asian countries and other countries as well. So the rice was selected to start working with. Initially three types of rice of Bangladesh (Aus, Aman and Boro) is analyzed and there is scope for extensive use of this method for other crops like jute, wheat and tea of different countries. Data set for this research work is collected from Bangladesh Rice Research Institute (BRRI) [4].

At first the Auto Regressive Moving Average (ARMA) [3] Model was used for the work but the best estimates would always show a sudden and steep drop in the yield for every crop and would not show an increase until years later. Then the Weighted Linear Regression model is applied by giving more weight to recent data which results 78.75% and 83.55% correct prediction for price and yield respectively after comparing the actual and predicted values over previous three years.

The remainder of the paper is organized as follows: section 2 summarized the background study for the research work. The research methodology is described in section 3. Description of system structure is discussed in section 4. Section 5 provides the system outcome. Experimental result and discussion are given in section 6 and finally the conclusion is drawn in section 7.

2. BACKGROUND

The agriculture sector is the single largest contributor to income and employment generation and a vital element in the countrys challenge to achieve self-sufficiency in food production as well
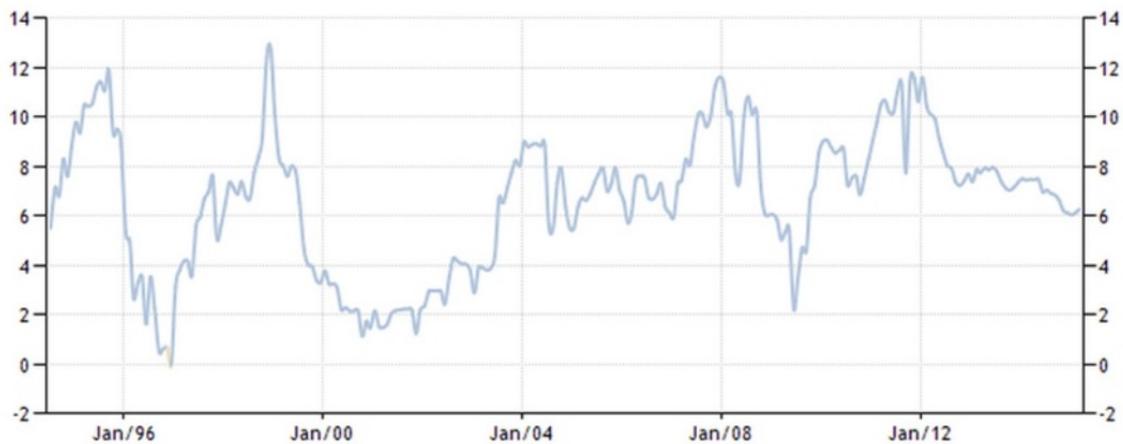
Fig. 1. Bangladesh Inflation Rate

Table 1. Inflation Rates: CPI, GDP, Producer, Export, Import Prices, and Food

Bangladesh Prices	Last	Previous	Highest	Lowest	Unit
Inflaction Rate	6.27	6.14	12.71	-0.02	percent
Consumer Price Index (CPI)	210.74	210.12	201.74	51.99	Index Points
GDP Deflator	174.42	164.26	224.45	126.35	percent
Producer Price	2180	2082	2180	1233	Index Points
Export Prices	223.1	208.5	223.1	78.9	Index Points
Import Prices	318.2	294.6	318.2	89.9	Index Points
Food Inflaction	6.11	6.07	9.09	5.85	percent
Inflation Rate Mom1.13		0.44	2.08	-1.19	percent

as to reduce rural poverty and foster sustainable economic development. The Government of Bangladesh has therefore accorded highest priority to this sector to enable Bangladesh to meet these challenges and to make this sector commercially profitable.

In Bangladesh about 84% of the total population live in rural areas and are directly or indirectly engaged in a wide range of agricultural activities.

The agriculture sector plays a very important role in the economy of the country accounting for 31.6 percent of total GDP in 1997-98 at constant (1984-85) prices[10].

At times due to difference in minds of political leaders, political disturbances knock the economy down. As the economy falls down due to recession, the price inflation occurs in the society. In Bangladesh, the inflation rate measures a broad rise or fall in prices that consumers pay for a standard basket of goods [12].

Focusing more on the food inflation, it can be seen from Table1 that the highest was 9.09% and the lowest was 5.86%. Figure 1 graphically shows the inflation rate of Bangladesh. Figure 2 shows graph (system output) for both historical and predicted yield and price of Rice (Aman, Aus and Boro).

The goal of the research work is to help the farmers the most. The idea was to keep track of all the data in the previous years. The data include price and yield of food. For this research work, initially three types of Rice, which are Aus, Aman and Boro are focused. Figure 2(a) shows the graphical representation of the yield of Rice (Aman) has gradually increased over the years from 1970 to 2014. It could be seen that in the years 1986-1988 and 1992-1998 the production of price was below average. As a result, price went low. Figure 2(b) shows the graphical representation of the price of Rice (Aman). In the graph, it can be seen that there is a drastic falls in

price of Rice (Aman) significantly years, 1970, 1976, 1983, 1984, 1992, 1996, 2008 and 2010. Drastic fall in price of Rice signifies that there was sufficient production of Rice. As a result, the economy was affordable for ordinary people. On the contrary, in the years, 1982, 1987, 1996, 1998, 2004, 2005 the economy was in recession as the prices were too high for ordinary people to afford rice to eat.

Figure 2(c) shows the graphical representation of the yield of Rice (Aus) has gradually increased over the years from 1970 to 1996. Then there was a better production from 1998 until 2014 and still on going. As a result, price went low. Figure 2(d) shows the graphical representation of the price of Rice (Aus). This graph shows that there are numerous falls in the price of Rice (Aus) prominently in years 1973, 1987, 1992, 1996, 2011 and many more. Drastic fall in price of Rice signifies that there was sufficient production of Rice. On the contrary, in the years, 1978, 1984, 1998, 2004 and 2008 the economy was in recession as the prices were too high for ordinary people to afford rice to survive. Figure 2(e) shows the graphical representation of the yield of Rice (Boro) has gradually increased over the years from 1977 to 2014. There was a small fall in the year 1976 and 1993. As a result, price remained less most of the times. Exceptional cases are the years when there is a fall. Figure 2(f) shows the graphical representation of the price of Rice (Boro). There is a few fall in the price of Rice in years 1983, 1984 and 2010 in this graph. On the contrary, in the years, 1977, 1980, 1987, 1991, 1994, 1995, 1997, 1998, 2004, 2005 the economy was in recession as the prices were apparently too high for ordinary people to afford rice to survive.

The circled orange line in the graph is the prediction of the future price and yield based on previous years data analysis. Red line rep-
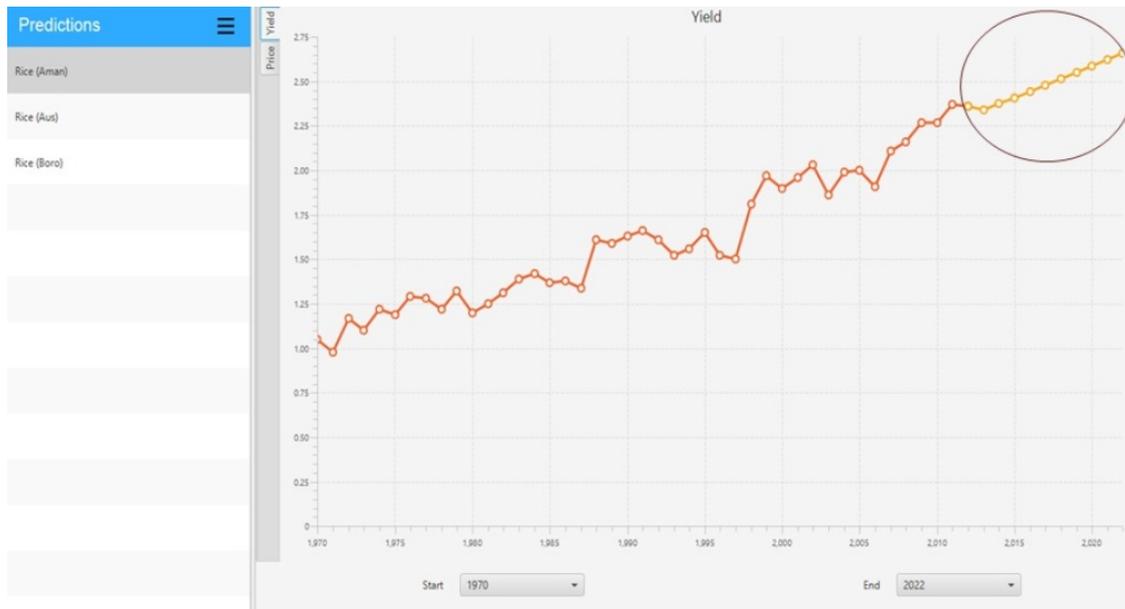
Fig. 2(a)
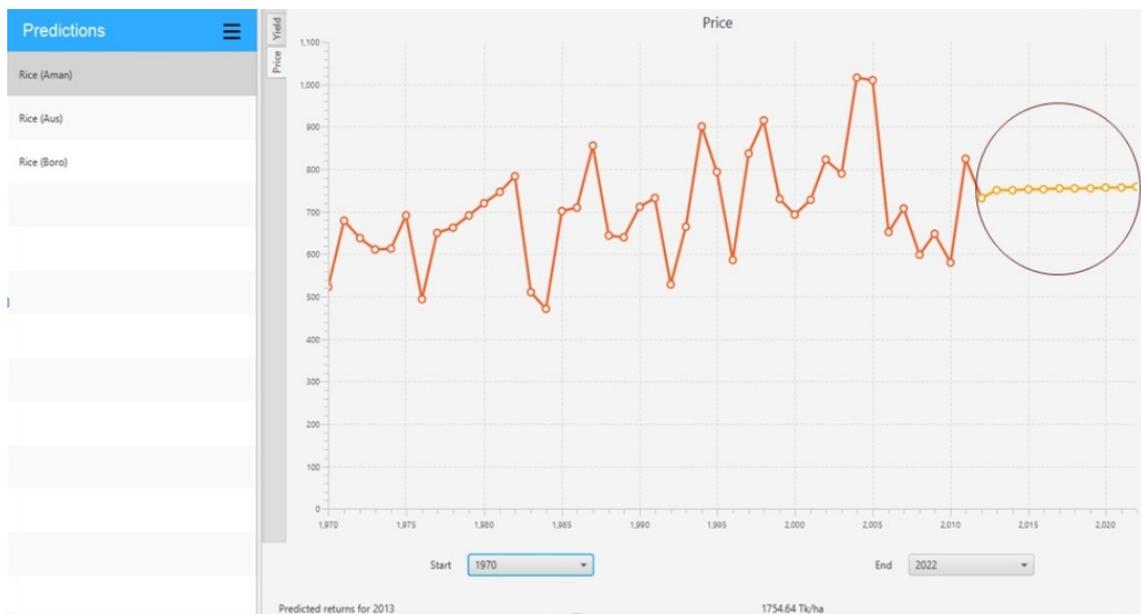
Fig. 2(b)

resents data of the previous years. Apparently, the orange line is a continuation of the red line, which describes all the graphs in this section.

3. METHODOLOGY

3.1 The ARMA model

The Auto Regressive Moving Average (ARMA) Model is usually used in signal processing to interpolate, or to fill in, dropped data. However, it can also be used to model and forecast stochastic pro-cesses such as stock market fluctuations and seismic waves [8]. This made the ARMA Model seem like the perfect model to do the price and yield forecasting in the application. However, with this method the best estimates would always show a sudden and steep drop in the yield for every product and would not show an increase until years later. Rather than accept these, it was decided to stop using the ARMA model and switch to the much simpler Weighted Linear Regression model[7].
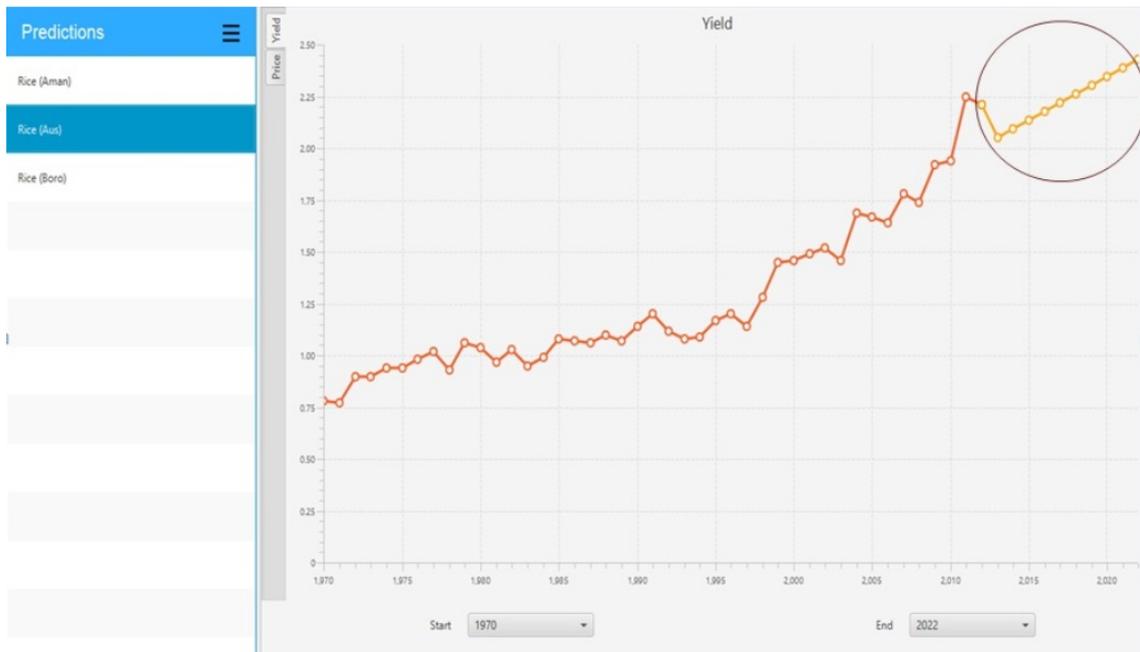
Fig. 2(c)
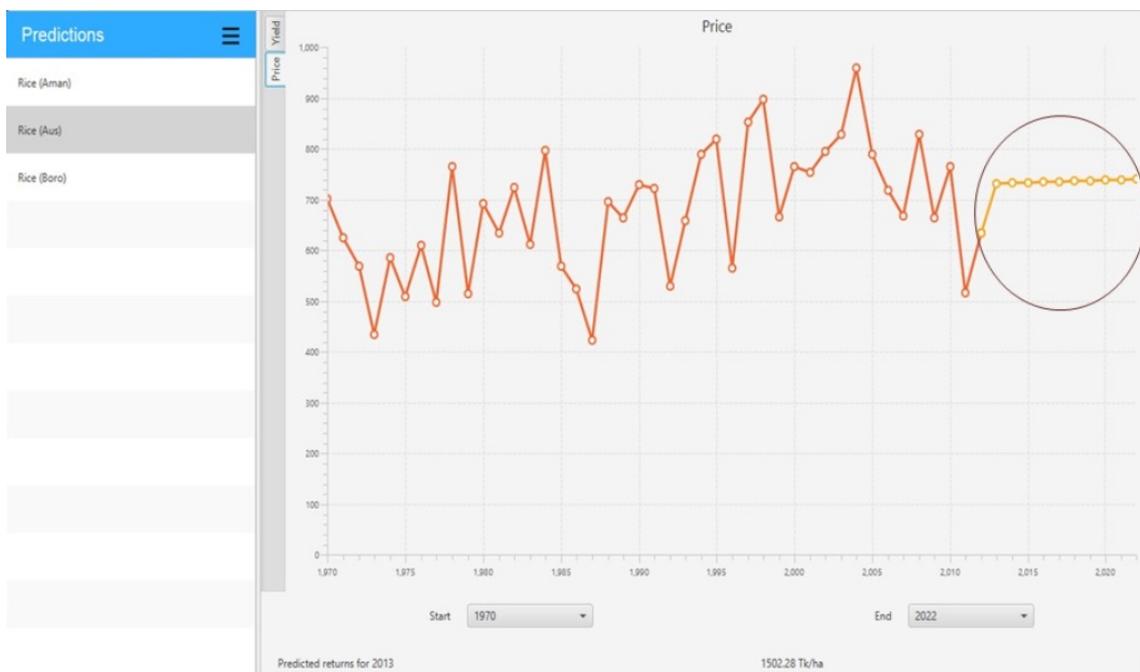
Fig. 2(d)

3.2 Weighted Linear Regression

The Weighted Linear Regression model finds a linear relation between the dependent and independent variables, but rather than treating all data points equally, it gives more weight to some than others [1]. Here, for prediction more weight is given to recent data.

Going through the years, the weights of the data points increase exponentially.
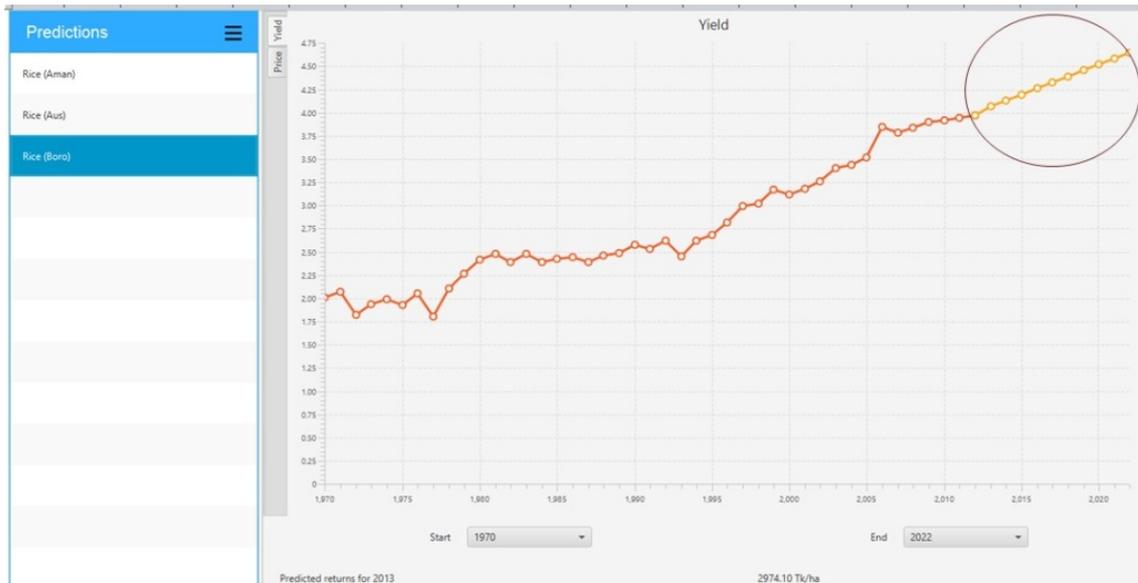
Fig. 2(e)
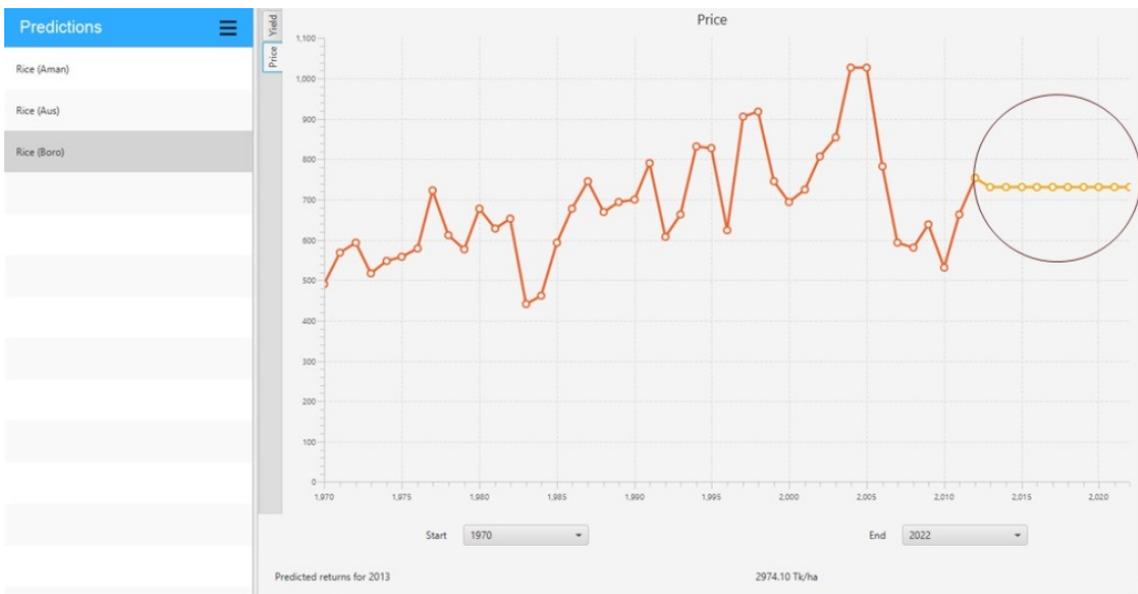
Fig. 2(f)
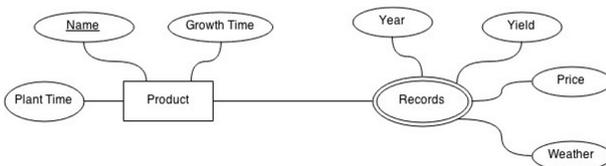
Fig. 3. Database overview of the proposed system

4. SYSTEM STRUCTURE

The system takes data from past years as inputs and uses Weighted Linear Regression Model to produce two graphs as shown in Figure 2. One graph is for the crop yielding and the other is for the price of crops. For the first graph with years as x-axis and the amount of crop yielding as y-axis, the system generates future crop yielding prediction by applying Weighted Linear Regression Model on the data of past years, which are collected from a reliable source (Bangladesh Rice Research Institute). The same function applies with the second graph. The years are considered as the x-axis and the price as y-axis. Input is the past years price value
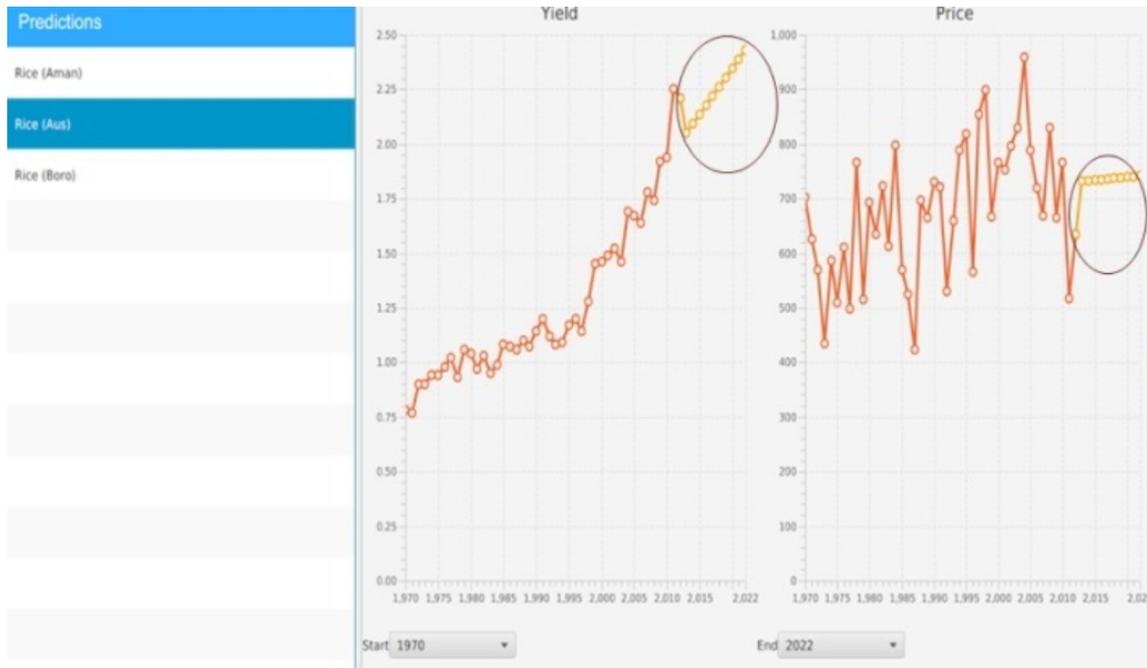
Fig. 4. Yield and Price of Rice (Aus)

Table 2. Actual and predicted yield and price values for three rice types in previous three years.

Rice	Year	Actual Yield (approx.) (Tons per Hactare)	Predicted Yield (approx.) (Tons per Hactare)	Actual Price (approx.) (per quintal)	Predicted Price (approx.) (per quintal)	Difference in yield	Difference in yield %	Difference in price	Difference in price %
Anus	2010	1.93	1.3	760	650	0.63	32.67	110	14.47
	2007	1.67	1.9	660	800	0.23	13.77	140	21.21
	2004	1.77	1.95	785	650	0.18	10.17	135	17.2
Aman	2010	2.28	2.5	580	400	0.22	9.65	180	31.03
	2007	1.9	2.2	710	850	0.3	15.79	140	19.72
	2004	2	1.7	1020	1180	0.3	15	160	15.69
Boro	2010	3.95	4.4	540	430	0.45	11.39	110	20.37
	2007	3.8	4.7	580	790	0.9	23.68	210	36.21
	2004	3.45	2.9	1040	880	0.55	15.94	160	15.38
Avg difference							16.45		21.25
Avg similarity							83.55		78.75

and the output is graph, which shows the price variation throughout the years and tries to predict which way the price tends to go in future. The effect of change in a variable on crop in successive periods would not be an abrupt or erratic change but an orderly one that follows some mathematical law. These effects are composed of the terms of a polynomial function of time. Further, the value of the variable in with period, X_w was also expressed in terms of orthogonal functions of time. Substituting these in usual regression equation and utilizing the properties of orthogonal and normalized functions, the following can be obtained

$$Y = A_0 + \alpha_0\rho_0 + \alpha_1\rho_1 + \ldots\ldots\ldots + \alpha_k\rho_k \quad (1)$$

where Y denoted yield, A_0, α_0, α_1, α_2,.... α_k are constants to be determined and ρ_i (i=1,k) are distribution constants of Xw. For this experiment, 5 is used as K value for most of the practical situations. In fitting this equation for k = 5, the number of constants to be evaluated is 7; e is the residual.

Database overview of the proposed system is given in Figure 3. This model is used for studying the influence of inflation factor on the yield and price of rice. The calculation process considers the facts like natural calamities, economy which affects the price inflation and the overall situation specially the political situation

that has affected or can affect out crop yielding and price. GIT[5] is used as the source control software and the Gradle build tool [6] for the research work.

5. SYSTEM OUTCOME

Proposed system is for building a better future of agricultural industry with cultivation process by predicting the yield and price of crops. An example of how the system works is shown in Figure 2; when data is inputed in the system, two graphs are generated as per the given data. The circled orange line in the graph is the prediction of the future price and yield based on previous years data analysis. Red line represents data of the previous years. Apparently, the orange line is a continuation of the red line, which describes all the graphs presented in Figure 2 and Figure 4.

6. EXPERIMENTAL RESULT AND DISCUSSIONS

Experimental result for three types of rice (Aus, Amon and Boro) is compared in Table 2 for analyzing the proposed systems performance. From three previous years' (2010, 2007 and 2004) information and system output, the average difference between the actual and predicted values are 16.45% and 21.25% for yield and price respectively. This means the performance accuracy is 78.75% and 83.55% for predicting price and yield respectively. The proposed method can be applied not only for rice, but also for other types of crop for predicting their yield and price.

As a human work this system has some limitations. There is always a chance that the prediction data can go wrong. There are lots of factors like natural disaster that cannot be controlled or foresee the damage it can cost. Future plan of this work is to come up with an improved algorithm that can perform the future prediction considering more facts so that the prediction can be more accurate. This system can be launched into an android app version so that people who are associated with agricultural activities can access the system from their mobile phones. This would increase the use of the system and help farmers understand the importance of using technology in every aspects of life in this 21st century. In the improved version, other features will be included in the system so that a managing role could also contribute to the information system. They can correct any data that is wrong in system and can notify concerned system [13]. Industry level implementation of the proposed system also includes establishing local stores or offices with trained people who can help the farmers or villagers to understand the system and its advantages.

7. CONCLUSION

In this century, developing countries like Bangladesh strive to predict a stable price and yield balance in agricultural economy. In this paper an effective Weighted Liner Regression Based method is proposed for price and yield prediction of agricultural crops. The system takes data from past years as inputs and uses Weighted Linear Regression model to produce two graphs representing both old and predicted price and yield for any particular crop. Experimental results show that the method is quite successful in predicting price and yield considering inflation factor.
This technique will help, to grow agricultural economy, to create more workplace and to improve the living standard for farmers and other people who are associated with any kind of agricultural work. This system would help farmer know about the future prediction beforehand so they can prepare for any kind of unwanted disaster. In this way, they can work more efficiently and produce more crops.

The proposed method can be applied not only for rice, but also for certain types to crops in predicting their yield and price.

8. REFERENCES

[1] *Anonymous. ARIMA model for time series forecasting,*, 2011(accessed April-2016). `http://people.duke.edu/~rnau/411arim.htm`.

[2] *Anonymous. Agriculture Retrieved,*, (accessed April-2015). `http://www.sdnbd.org/sdi/issues/agriculture/`.

[3] *Autoregressive Moving Average Model,*, (accessed April-2016). `https://www.mathworks.com/help/econ/arma-model.html`.

[4] *Bangladesh Rice Research Institute,*, (accessed April-2016). `http://www.brri.gov.bd/index.php?lang=en`.

[5] *Git,*, (accessed April-2016). `https://git-scm.com/`.

[6] *Gradle Build Tool I Modern Open Source Build Automation,*, (accessed April-2016). `https://gradle.org/`.

[7] *Weighted Linear Regression,*, (accessed April-2016). `http://www.real-statistics.com/multiple-regression/weighted-linear-regression/`.

[8] *Auto Regression Analysis(AR),*, Novembor 1998(accessed April-2016). `http://paulbourke.net/miscellaneous/ar/`.

[9] Mansur Ahmed and Barry Goodwin. Agricultural mechanization and non-farm labor supply of farm households: Evidence from bangladesh. Technical report, Agricultural and Applied Economics Association, 2016.

[10] Mohammad Alauddin. Recent developments in the bangladesh economy. In *Economic Growth, Economic Performance and Welfare in South Asia*, pages 11–27. Springer, 2005.

[11] Pramod K Joshi, Ashok Gulati, Pratap S Birthal, and Laxmi Tewari. Agriculture diversification in south asia: patterns, determinants and policy implications. *Economic and Political Weekly*, pages 2457–2467, 2004.

[12] AKM Golam Rabbani Mondal, Md Abdul Khalek, Md Mostafizur Rahman, and Dedar Mahmud. Price escalation of rice in bangladesh: A time series approach.

[13] BM Whelan, AB McBratney, and B Minasny. Vesper 1.5–spatial prediction software for precision agriculture. In *Precision Agriculture, Proc. 6th Int. Conf. on Precision Agriculture, ASA/CSSA/SSSA, Madison, WI, USA*, volume 179, 2002.