Predicting Fluctuation of Food Price in Bangladesh Using Machine Learning Methods

Md. Sojib Ali[1], MD. Rafiul Hasan[2], Abdullah Al Mamun Tonmoy[3], Nakib Aman Turzo[4]

[1] City University, Dhaka, Bangladesh
[2,3,4]Varendra University, Rajshahi, Bangladesh
sojibali.cse@gmail.com,rafiul1713@gmail.com,
aamtonmoy@outlook.com,nakibaman@gmail.com

Abstract. Food price fluctuation causes adverse effect on economic development in Bangladesh. In contrast with the past, food prices stepping up rigorously. Long term instability in prices initiates immoderate volatility. Several factors are responsible for fetching volatility in the food market. For instance, demand and supply, weather, food production and many more. Learning goal of this paper is to predict food price fluctuation using pre-processed prices of selected foods dataset, machine learning methods, combined with non-linear regressions. Regression models were used as a comparison among MAE, MSE, RMSE, R2, RMSLE, MAPE and time in seconds. Lasso provided least Mean Squared Error compared to other regression model. Contribution of this work goes to various classes i.e., farmers, investors, government etc. By inducing the knowledge of future market in food prices, investors or Farmers can have proper idea of when to invest money on food production.

Keywords: Food, price, prediction, fluctuation, Lasso Regression Model

1 Introduction

Population grows at an alarming rate in a short amount of time and aggregating high demands of food lay out various difficulties when it comes to producing food in a limited cultivable field in Bangladesh. Arrival of disasters back and forth mess up with the fluctuation of food price. Statistically Bangladesh ranks high in terms of malnutrition. More than half (9.5 million) of pre-school age children is underweight [1]. To overcome the rates of malnutrition no alternate choices of foods for Bangladeshi people except consuming mixture of carbo-hydrate and protein kinds but in affordable budget i.e., rice, chicken, egg, milk etc. [2].

As explained by Said Fadlan Asnhari et al [3], abrupt expansion of fluctuation of food price constitute an alarming issue in government stability. Food price fluctuation take places following inclement weather, demand and supply of market and regulation, food production and many more factors. Due to volatility in the market farmers faces difficulties to pay while consumers, conversely must pay high prices to buy food. Stability in life of civilizations among Bangladeshi demolished by the high fluctuation of food price. Motivation of this work is to predicting fluctuation rate, On the contrary, farmers can take suitable production decision and decrease consumers financial reduction [4].

How the prices go up and down can be observed by the chart below. By analyzing dataset, fluctuation rate of chicken's prices found higher compare to other items. Apple prices also fluctuated from August 2009 until June 2020. Rest of the items i.e., rice, egg, milk shows almost equivalent pattern.

Food materials price fluctuation

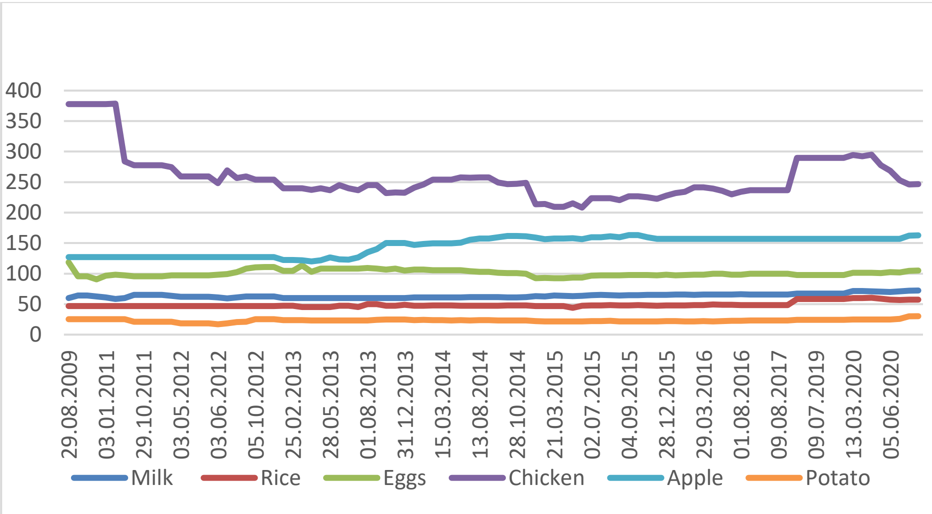

Fig. 1. The price of food items (Milk, Rice, Eggs, Chickens, Apples, Potatoes) in Bangladesh from August 2009 to June 2020

Learning goal of this paper is to gain proper knowledge of food price fluctuation by the utilization of machine learning methods. Here, MAE, RMSE are to be determined in terms of training time depict distinction between the original and predicted values obtained from average over the data set. By predicting fluctuation rate and stability helps farmers to prevent from poverty traps, vendors to understanding market state and government to induce vital strategies to bring stability in food price.

The rest of this paper is packed in 4 sections. In Section II, the brief explanation about ARIMA, Regression and Fourier model to predict the staple food are given. In Section III, methodology and analysis of results are elaborated. Moreover, the conclusion is drawn in Section IV.

This paper is structured in 5 segments. In segment II, provides brief description of previous work history of linear and non-linear regression models approaches on prediction of food and vegetable prices. Methodology, experimental analysis and result was explained in segment III and IV respectively. In segment V conclusion was added.

2 Literature Review

Various studies have covered up shortcomings regarding to prediction using different machine learning algorithm. As many new Machine learning methods are making their way to be applied, many researchers are also implementing new models to determine forecasting of fluctuation of food price.

Said Fadlan Asnhari, P. H. Gunawan, Yanti Rusmawati [3], used multivariable factors (Linear regression and Fourier regression with ARIMA) to predict staple food materials price. However, in terms of accuracy they obtained Fourier regression with ARIMA performed better compare to multiple linear regression with ARIMA.

Shivani Mahida, Bhargesh Patel [5] reviewed the application of data mining technique for vegetable price prediction. They aim to predict vegetable prices in near future by analyzing time series data on daily arrivals and prices of selected vegetable.

YE Lu et al [4] described vegetable price prediction based on PSO-BP Neural Network. They formulated a comparison with traditional BP method and obtained PSO-BP method could solve several issues and enhance the accuracy of prediction.

Dr Alioune DIENG [6], has explored alternative forecasting techniques for vegetable prices in Senegal. He used two forecasting approaches one with three alternative parametric models and a non-parametric model. His study found naive model, the exponential smoothing and the BOX and Jenkins autoregressive integrated moving average (ARIMA) perform better for forecasting prices.

Gan-qiong Li et al [7], have applied Artificial Neural Network (ANN) to forecasting short-term price for agro-products. Based on comparison with time series model ARIMA, their result showed ANN model outperformed ARIMA in forecasting price with error less than 5.0%

M.Subhasree, Mrs.C.Arun priya [8], worked on their research called vegetable price prediction based on time series analysis. What they attained was genetic based neural network outperformed back propagation neural network.

PURNA CHANDRA PADHAN [9], applied ARIMA model on annual data from 1950 to 2010 to forecast annual productivity of agricultural product. The validation done by minimum of AIC, lowest MAPE values etc. From the selected crops, lowest MAPE and AIC values were given by tea and cardamom respectively.

P.Jasinthan et al [10], conducted vegetables price movement guided with two model. One is gain or loss from the perspective of price movement and the other one is large gain or small gain, large loss or small loss. These model leads to attain transitional probabilities, steady state probabilities and mean recurrence times. By observing their method investor can chose wisely to invest in the vegetable market which can acquire more gain than loss.

Cooray, T.M.J.A [11], has studied on the viability of rule-based forecasting where he explains in which circumstances which method would be preferrable among four extrapolation methods (Decomposition method, Exponential smoothing, Winter's Seasonal smoothing and ARIMA methodology).

K. Assis et al [12], made a comparison among four various kinds of univariate time series methods, namely, the exponential smoothing, autoregressive integrated moving average (ARIMA), generalized autoregressive conditional heteroskedasticity

(GRACH) and the mixed ARIMA/GRACH model. To established the better forecasting model Root mean squared error (RMSE), mean absolute percentage error (MAPE), mean absolute error (MAE) and Theil's inequality coefficient (U-STATISTICS) were used as a basis of election.

3 Methodology

3.1 Dataset Collection

Fluctuation of selected food price were collected from Numbeo website. Data collection ranges from 2009 to 2020. Data are separated in a month period of time. Since only present-day data available in Numbeo, A website called waybackmachine was used to acquire data of previous years. For this work, according to need of Bangladeshi Citizen, data of milk, rice, chicken, egg were obtained. Data collection procedure can be described as follows:

- Food prices were assembled as text file.
- Data was collected from 29.08.2009-27.07.2020.
- Data was stored in 8 columns 92 rows.
- Wide range of regression algorithms were applied on datasets.
- Missing values were filled by applying mean filtering. Prices of foods were stored in USD, later converted in BDT. Authenticity were checked by contributors.

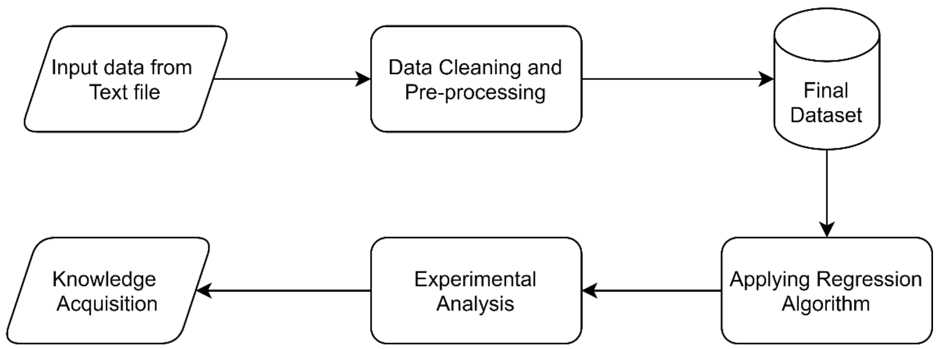

Fig. 2. Virtual Workflow of methodological steps

Table 1. Before preprocessed data

Date	Milk	Rice	Eggs	Chicken	Contributors
29.10.2011	65.219	46.585	95.711	277.816	40
15.11.2011	65.219	46.585	95.711	277.816	40
25.12.2011	65.219	46.585	95.711	277.816	40
25.01.2012	63.525	46.585	97.405	274.428	40
03.05.2012		46.585	97.405		40
05.05.2012	61.831	46.585	97.405	259.182	40
………	………	………	………	………	………

Note: Missing values were found while collecting dataset.

Table 2. After preprocessed data

Date	Milk	Rice	Eggs	Chicken	Contributors
29.10.2011	65.219	46.585	95.711	277.816	40
15.11.2011	65.219	46.585	95.711	277.816	40
25.12.2011	65.219	46.585	95.711	277.816	40
25.01.2012	63.525	46.585	97.405	274.428	40
03.05.2012	61.831	46.585	97.405	259.182	40
05.05.2012	61.831	46.585	97.405	259.182	40
…………	…………	…………	…………	…………	…………

Note: Mean filtering was applied to fill up missing values.

3.2 Dataset details

Table 3. Dataset details

Data source	Category	Amount
Numbeo, wayback machine	Chicken,milk, egg, rice	Total amount

4 Experimental Analysis and Result

MAE estimate average error naturally. For inter comparison similar to food prices, to determine mean absolute error of model performances MAE is preferred. [13]

4.1 Mean Squared Error (MSE)

After looking for regression line substitute a value into linear regression equation to find another value. Determination of subtract will be used for squared the error.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\text{Y}i - \hat{\text{Y}}i)^2$$

4.2 Mean Absolute Error (MAE)

Summation of all the absolute error divided by the number of errors determines the mean absolute error.

$$\text{MAE} = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Yi= prediction
Xi=true value
n=total number of data points

Table 4. Compare Model

	Model	MAE	MSE	TT (Sec)
Chicken	Orthogonal Matching Pursuit	26.9838	36.0852	0.014
	Elastic Net	26.1868	35.0513	0.013
	Light Gradient Boosting Machine	25.9717	35.0013	0.017
	K Neighbors Regressor	27.6461	36.2835	0.061
	Bayesian Ridge	27.2083	36.6901	0.019
	Lasso Least Angle Regression	27.5044	36.9953	0.013
Milk	Lasso Regression	2.7898	3.2825	0.013
	Lasso Least Angle Regression	2.7898	3.2825	0.012
	Light Gradient Boosting Machine	2.7898	3.2825	0.018
	Elastic Net	2.7898	3.2825	0.013
	Bayesian Ridge	2.7902	3.2828	0.014
	AdaBoost Regressor	3.1124	3.6091	0.030
Rice	Lasso Regression	2.7359	3.7728	0.009
	Lasso Least Angle Regression	2.7359	3.7728	0.013
	Light Gradient Boosting Machine	2.7359	3.7728	0.017
	Elastic Net	2.7359	3.7728	0.015
	Bayesian Ridge	2.7361	3.7730	0.014
	Huber Regressor	2.5345	3.8565	0.025

Note: When it comes to find less MAE, RMSE in terms of Training Time (Sec) Lasso Regression performs well compare to rest of the model in the following table except chicken and egg prices. Every model was tested in the same way.

Table 5. Tuned Model

		MAE	MSE	RMSE	R2	RMSLE	MAPE
Chicken	0	31.7106	1697.8195	41.2046	0.3887	0.1384	0.1110
	1	15.9211	396.1053	19.9024	-0.2593	0.0756	0.0597
	2	26.3930	1305.5164	36.1319	-1.6222	0.1302	0.1033
	3	12.0733	312.4182	17.6754	-0.0035	0.0672	0.0463
	4	24.2539	661.9170	25.7277	-0.0274	0.1020	0.0983
	5	29.1468	2082.3389	45.6327	-6.9113	0.1637	0.1220
	6	34.9577	2068.8587	45.4847	-2.5528	0.1641	0.1425
	7	34.1556	1878.3209	43.3396	0.4023	0.1475	0.1228
	8	33.5213	2167.9984	46.5618	0.1770	0.1515	0.1165
	9	27.7045	1535.9709	39.1915	0.2266	0.1248	0.0872
	Mean	26.9838	1410.7264	36.0852	-1.0182	0.1265	0.1010
	SD	7.3335	676.9001	10.4203	2.1682	0.0328	0.0281
Milk	0	3.2316	17.4374	4.1758	-0.3143	0.0613	0.0473
	1	4.0545	22.1528	4.7067	-0.0462	0.0699	0.0610
	2	2.8542	10.3093	3.2108	-3.7662	0.0505	0.0468
	3	3.9276	19.7594	4.4452	-0.1023	0.0686	0.0627
	4	3.3136	15.7726	3.9715	-0.3154	0.0592	0.0495
	5	3.0031	11.3836	3.3740	-1.1394	0.0532	0.0493
	6	1.4255	2.7559	1.6601	-1.1485	0.0250	0.0216
	7	1.6657	4.2243	2.0553	-0.0024	0.0316	0.0260
	8	1.6245	3.6649	1.9144	-0.0169	0.0297	0.0257
	9	2.7978	10.9636	3.3111	-0.1006	0.0500	0.0426
	Mean	2.7898	11.8424	3.2825	-0.6952	0.0499	0.0433
	SD	0.8886	6.5310	1.0333	1.1041	0.0153	0.0137
Rice	0	2.0850	5.6923	2.3858	-3.2320	0.0492	0.0450
	1	2.4486	11.7954	3.4344	-0.0010	0.0650	0.0476
	2	3.0140	16.9842	4.1212	-0.0056	0.0781	0.0590
	3	2.6173	12.4845	3.5333	-0.0120	0.0681	0.0521
	4	2.3214	16.1467	4.0183	-0.0497	0.0743	0.0426
	5	1.5725	3.1310	1.7695	-0.9170	0.0362	0.0333
	6	3.7851	35.3083	5.9421	-0.3258	0.1091	0.0662
	7	3.0331	23.6831	4.8665	-0.0255	0.0891	0.0557
	8	1.4929	2.5561	1.5988	-6.8084	0.0326	0.0316
	9	4.9895	36.7004	6.0581	-0.0596	0.1139	0.0942
	Mean	2.7359	16.4482	3.7728	-1.1437	0.0716	0.0527
	SD	0.9968	11.5843	1.4880	2.1136	0.0262	0.0172

Note: To get most accurate outcomes and valuable perception of dataset, tuning the model is necessary. After tuning the model Lasso Regression shows least Mean Absolute Error.

Table 6. Parameter

	Parameters	Values
Chicken	Fit_intercept	True
	n_nonzero_coefs	None
	Normalize	True
	Precompute	Auto
	Tol	None
MIlk	alpha	1.0
	copy_X	True
	fit_intercept	True
	max_iter	1000
	normalize	False
	positive	False
	precompute	False
	random_state	6724
	selection	cyclic
	tol	0.0001
	warm_start	False
Rice	alpha	1.0
	copy_X	True
	fit_intercept	True
	max_iter	1000
	normalize	False
	positive	False
	precompute	False
	random_state	6724
	selection	cyclic
	tol	0.0001
	warm_start	False

Note: Condense of data represent parameters. The above-mentioned parameters have been found after tuning the model.

4.3 Training Time

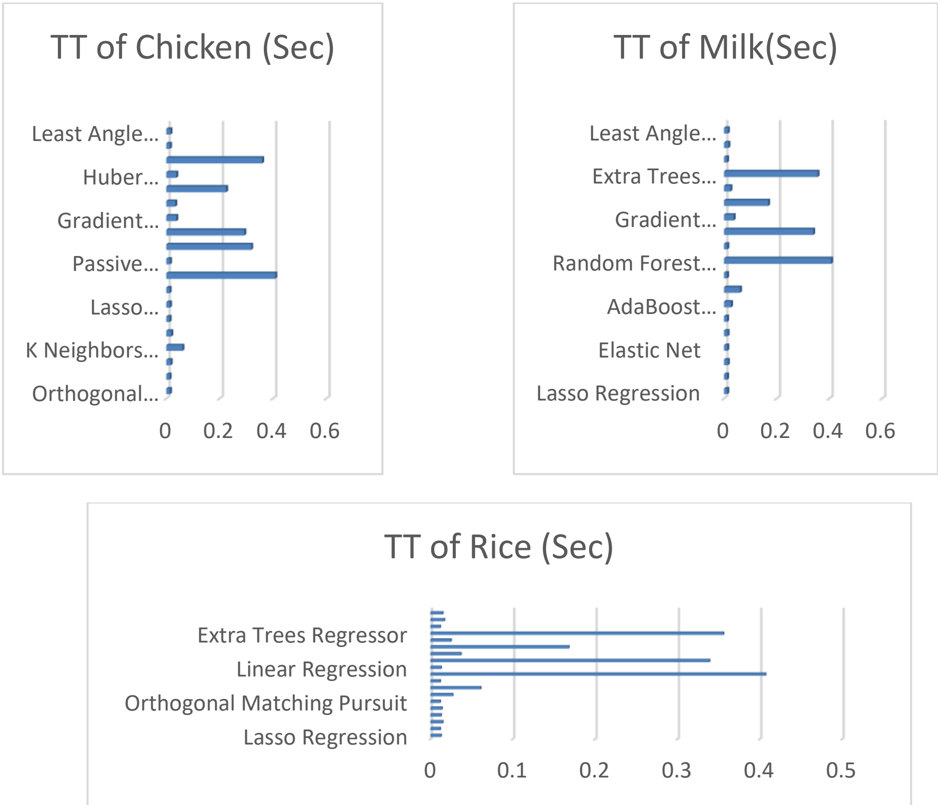

Fig. 3. Comparison between regression models of chicken, milk and rice price's in terms of training time

4.4 Learning and Validation Graph

Learning curve, a model which measure performance over experience of time. Learning curve for Lasso Regression has been shown below. In the following, validation graph is represented. It works on parameters of dataset and shows changes in some parameters of the model. Here, cross validation score rises above with changes in parameters alpha.

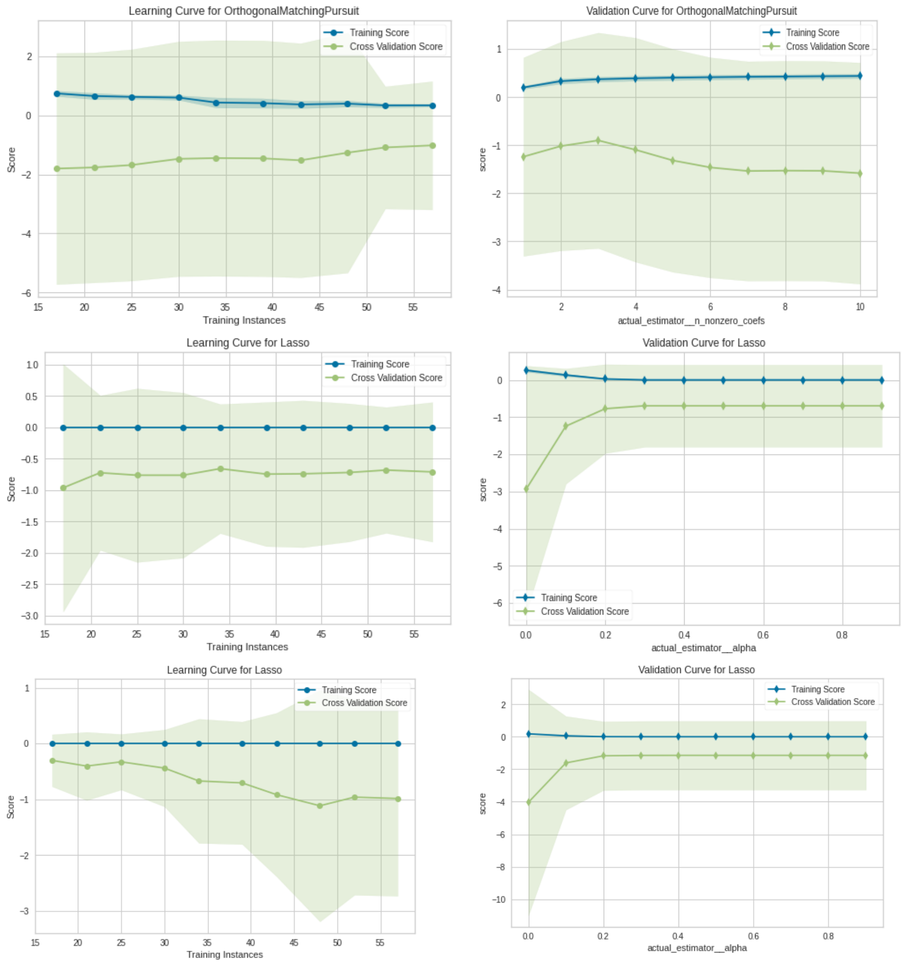

Fig. 4. Learning and validation curve of chicken, milk and rice price.

4.5 Prediction Error plot

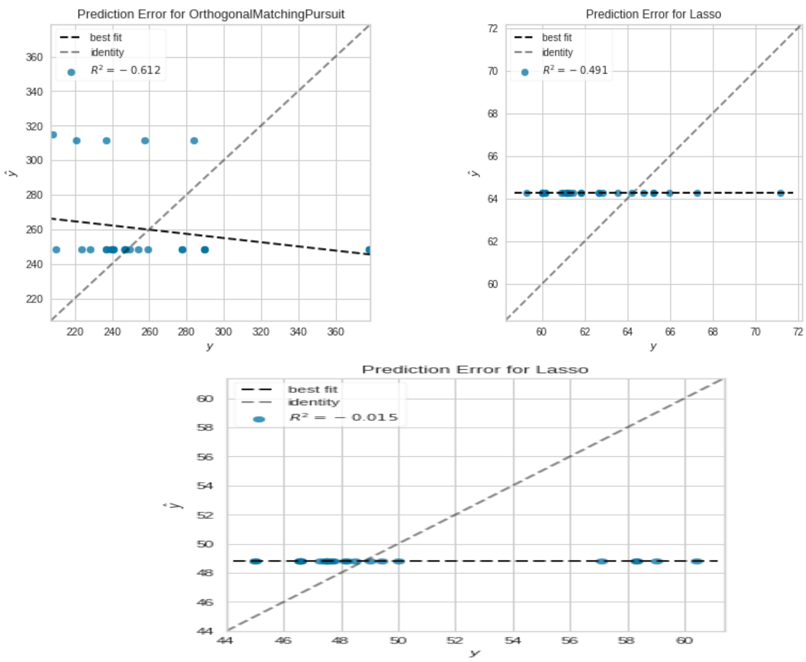

Fig. 5. Predictive model failure could take place at any moment. Prediction error plot for Lasso Regression model.

5 Conclusion

Different factors, seasons hit market any moment makes food prices fluctuation high above. It almost becomes difficult for consumers let alone investors, farmers to plan either buy products or sell. Collection of datasets from Numbeo with Wayback Machine website of six food items which are mandatory to increase nutrition for Bangladeshi people was a challenging task as data cleaning process required. Lasso Regression model shows best result when it comes to apple, milk, potato and rice prices. But for chicken price light gradient boosting machine and for egg price orthogonal machine pursuit performs well for Mean Absolute Error. For better experimental result, time training plot, learning and validation curve, prediction error, tuned model and parameters were analyzed.

This work fulfils a portion of knowledge in the field of machine learning. As broad as machine learning this work can be expressed in many internal areas of data science. For further work, neural network can be applied on the increased dataset provided by many trustworthy contributors. Discovers of any new factors can affect huge on food price fluctuation.

6 References

[1] "Nutrition country profiles: Bangladesh summary." http://www.fao.org/ag/agn/nutrition/bgd_en.stm (accessed Dec. 07, 2020).

[2] "Food-based nutrition strategies in Bangladesh." http://www.fao.org/3/ag126e/AG126E05.htm (accessed Dec. 07, 2020).

[3] S. F. Asnhari, P. H. Gunawan, and Y. Rusmawati, "Predicting staple food materials price using multivariables factors (regression and fourier models with ARIMA)," *2019 7th Int. Conf. Inf. Commun. Technol. ICoICT 2019*, pp. 1–5, 2019, doi: 10.1109/ICoICT.2019.8835193.

[4] L. Ye, Y. Li, W. Liang, Q. Song, Y. Liu, and X. Qin, "Vegetable Price Prediction Based on PSO-BP Neural Network," *Proc. - 8th Int. Conf. Intell. Comput. Technol. Autom. ICICTA 2015*, pp. 1093–1096, 2016, doi: 10.1109/ICICTA.2015.274.

[5] S. Mahida and B. Patel, "A Review of the Application of Data Mining Techniques for Vegetable Price Prediction," *Int. J. Sci. Res. Comput. Sci. Eng. Inf. Technol. © 2018 IJSRCSEIT*, vol. 3, no. 3, pp. 35–39, 2018, [Online]. Available: http://ijsrcseit.com/CSEIT1831408.

[6] D. Alioune, "Alternative Forecasting Techniques for Vegetable Prices in Senegal," *Bur. d'analyses macro-économiques*, vol. 1, no. 3, pp. 5–10, 2008, [Online]. Available: http://www.bameinfopol.info/IMG/pdf/Dieng_MP_1_.pdf.

[7] G. Q. Li, S. W. Xu, and Z. M. Li, "Short-term price forecasting for agro-products using artificial neural networks," *Agric. Agric. Sci. Procedia*, vol. 1, pp. 278–287, 2010, doi: 10.1016/j.aaspro.2010.09.035.

[8] M. Subhasree and C. Arun Priya, "Vegetable Price Prediction Based on Time Series Analysis," *Int. J. Comput. Sci. Inf. Technol. Secur.*, vol. 6, no. December, pp. 399–405, 2015.

[9] F. L. Article, "Application of ARIMA Model for Forecasting Agricultural Productivity in India," pp. 50–56, 2012.

[10] P. Jasinthan, A. Laheetharan, and N. Satkunanathan, "A Markov Chain Model for Vegetable Price Movement in Jaffna," *Sri Lankan J. Appl. Stat.*, vol. 16, no. 2, p. 93, 2015, doi: 10.4038/sljastats.v16i2.7825.

[11] T. M. J. A. Cooray, "Statistical Anlysis and Forecasting of Main Agriculture Output of Sri Lanka : Rule-Based Approach Statistical Anlysis and Forecasting of Main Agriculture Output of Sri Lanka : Rule-Based Approach," pp. 1–9, 2006.

[12] K. Assis, A. Amran, Y. Remali, H. A.-T. in Agricultural, and undefined 2010, "A comparison of univariate time series methods for forecasting cocoa bean prices," *agris.fao.org*, Accessed: Dec. 07, 2020. [Online]. Available: http://agris.fao.org/agris-search/search.do?recordID=DJ2012055924.

[13] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," Clim. *Res.*, vol. 30, no. 1, pp. 79–82, 2005, doi: 10.3354/cr030079.