***Pak. J. Bot.*, 55(5): DOI: http://dx.doi.org/10.30848/PJB2023-5(13)**

INTEGRATING EXPONENTIAL REGRESSION MODEL OPTIMIZATIONS FOR WHEAT AREA, PRODUCTIVITY AND POPULATION THROUGH STATISTICAL AND MACHINE LEARNING APPROACHES

FARRUKH SHEHZAD[1], MUHAMMAD ISLAM[2], AZEEM ALI[3], ABDUL QAYYUM[4] AND RABIA SIDDIQUI[5]

[1]Assistant Professor, Department of Statistics, The Islamia University of Bahawalpur, Pakistan
[2]Deputy Director (Stat), Crop Reporting Service, Agriculture Department Bahawalpur, Punjab, Pakistan
[3]Assistant Professor, Department of Statistics and Computer Science, University of Veterinary and Animal Sciences Lahore, Pakistan
[4]Director General, Crop Reporting Service, Agriculture Department, Punjab, Pakistan
[5]Assistant Director (Stat), Crop Reporting Service, Agriculture Department Bahawalpur, Punjab, Pakistan
**Corresponding author's email mislam6667@gmail.com*

Abstract

Strategic planning for food security has become the key intention especially for the developing countries like Pakistan. A comparative study is carried out to forecast the wheat area, yield and population eruption in Pakistan using the time series dataset, comprise from 1950-2020.This study layouts the plan to develop the regression model using the compound growth rate, called compound growth exponential regression models (CGREM). CGREM are applied using the machine learning (ML) and statistical approaches to address the food security planning for wheat area, yield and population eruption in Pakistan. Data partition is carried out using 80% and 20% randomized partitions for ML models. The hyper parametric tuning of ML model is further applied for 75%, 25% and 70%, 30% randomized partitions. The Performance of ML models are evaluated based on training and testing datasets. The evaluation metrics (RMSE, R^2) and information criterions (AIC, SIC) are used to measure the performance of models. The decomposition prediction error (P.E) is used to address the variance bias tradeoff and to select the optimum model. The decomposition model is applied to decompose the wheat production into its determinants. CGREM found better fitted model using the machine learning approaches. CGREM predicted, up to 2050, wheat area will rise up to 51.7%, wheat yield will grow up to 109.7%, and population will rise up to 140.6%. It noted that population will likely to upturn 88.9% and 30.9% more from wheat area and yield. Decomposition analysis model depicts that wheat productivity and area sharing 38% and 20% change towards wheat production. This study demonstrated the strong evidences to layout the true policy decisions, which leads to overcome the social dilemma of food security.

Key words: Machine learning, Models optimization, Statistical model, Wheat area, Wheat yield, Population.

Introduction

Food security has become the major concern for the developing countries (Islam & Shehzad, 2022; Nelson *et al.*, 2010; Tilman *et al.*, 2011). With the passage of time, the world's agricultural systems are coming under the sever threat of food security concerns because of upswing trend of population and diminishing productivity of agricultural production. Making true agriculture policies, to feed the nation has become the prime intensions to save the world from expected threat of hunger (Mozumdar, 2012). Machine learning extract suitable facts inside from data using mechanized algorithm and it can trend out, analyzed, identified, clustered and patronized the data (Nelli, 2015). Machine learning (ML) approaches has been categories as an advanced tool applied for the prediction of agriculture production (Alagurajan & Vijayakumaran, 2020; Elavarasan *et al.*, 2018; Mishra *et al.*, 2016; Sanchez *et al.*, 2014; Yadav *et al.*, 2020). Machine learning (ML) algorithms applied in the situation, when the relation between output and input feature are not predictable or difficult to predict in advance (Sanchez *et al.*, 2014). ML algorithms does not make any assumptions about the correct model structure and it is used in complex projection issue such as function form for crop yield prediction (Priya *et al.*, 2018). Arthur Samuel (1901–1990), a pioneer in artificial intelligence was the first who coined the term machine learning in the year 1959. He defined machine learning as "Field of study that gives computers the capability to learn without being explicitly programmed"(Dangeti, 2017; McCarthy & Feigenbaum, 1990). According to Dangeti (2017), ML considered as a branch of study in which a model is learned automatically from the experiences based on data without exclusively being modeled like in statistical models and over a period with more data, the model predictions have become more precise and accurate.

For evolving effective agricultural strategies, the accurate crop productivity prediction along with population checks has become a crucial task for any economy to attain its food sustainability. There is prerequisite demand of time, is to predict the accurate crop productivity to evoke to handle the uncertainty prevail in food concerns (Jeong *et al.*, 2016). Wheat being a staple food crop of Pakistan, ranks 1st in term of area and production among all others food crops (Farooq *et al.*, 2007). It is alarming that population growth rate in Pakistan is now reaching to 2.40% from 2.05%, while the agriculture growth rate of Pakistan is reached its low level at 0.85% in 2018-19 from 3.9% in 2017-18 (Anonymous. 2020), and it is projected that up to 2050, Pakistan will stands at 4th populous state in the world instead of current status of 6th (Ahmad & Farooq, 2010). Population growth rate in Pakistan is high while the wheat crop productivity is still low and this situation may explode the food conflicts. The accurate wheat crop prediction is significantly required, to check out the future demand of wheat for country with the projection of population explosion to handle the expected threat of food security.

According to Jeong *et al.*, (2016) statistical models gives predictions based on reliable and sufficient dataset for model training within the restrictions of training dataset and these models based on commonly performance measures while the machine learning (ML) used process based algorithms and gives alternatives to traditional statistical model. This study layouts the plan to develop the regression model using the compound growth rate. The prime objective of this study is to integrate the performance of traditional statistical models with machine learning algorithms for the prediction of wheat crop area, productivity along with the projection of population extend in Pakistan. This study will assist the government, to set the goal, planning and strategies to attain food sustainability in term of exploding threat of food security in the region.

Material and Methods

Data collection, measuring scales and analysis tools: The seventy one years historical dataset comprise from 1950 to 2020 are availed from Punjab agriculture marketing information service department (AMIS), Pakistan bureau of statistics (PBS) and various issues of economic surveys of Pakistan. These organizations own by the government of Pakistan and responsible for the valid and sound data collection mechanisms for researchers around the world. The data measuring scale are used as wheat crop area in thousand acres, wheat yield in mds/acre and population in millions. Investigation/experiments/analysis is performed using prominent Python's machine learning (ML) library called Scikit Learn by jupyter notebook https://scikit-learn.org/stable/supervised_learning.html. This library provides beautiful platform to build up useful steps of machine learning model. Data preprocessing (DP) used to improve the quality characteristics of data by cleaning, integration, and transformation technique to increase generalized performance of ML algorithm (Alexandropoulos *et al.*, 2019; Han *et al.*, 2011; Kotsiantis *et al.*, 2006; Rahman, 2019).

Data partition with statistical and machine learning modeling: Datasets partitions are applied by taking the 80% and 20% randomized partitions of train test split. Performance of supervised machine learning model are evaluated based on training and testing/validation datasets and integrated with the performance of traditional statistical models using evaluation metrics and information criterions. The exponential regression model proposed as compound growth exponential regression model (CGREM), and applied to predict the wheat crop area and productivity along with population eruption in Pakistan. The compound growth rate are measured by the following model (Dhakre & Sharma, 2010; Kondal, 2014; Kumar *et al.*, 2017; Qasim *et al.*, 2015).

$$\mathrm{y_t = y_0[1 + r]^{t_i}}$$

where "y_t" are response variables (wheat area/yield /population) at predicted time. "y_0" stands for response value at base year period, "r" stands for compound growth rate. The regression slope is used to investigate the relative change in response for the absolute changes accrue in features covariate time and it predicts the instantaneous growth rate.

$$\mathrm{Ln(y_t) = Ln(y_0) + t\{Ln(1 + r)\}}$$

$$\mathrm{Ln(y_t) = Y,\ Ln(y_0)\ = A\ ,\ Ln(1 + r) = B}$$

$$\mathrm{Y = A + Bt\ + \varepsilon, \quad r = (exp^B - 1) * 100}$$

To predict the parameter the following equation applied as:

$$\mathrm{y_p =\ [y_c\,(1 + B)^n]}$$

where, y_p= value of response variable at projected time, y_c = actual/collected value of response at time "t", B = regression coefficients, n = total no. of year project i.e. t_p - t_c .

The two types of errors are encountered in machine learning models as the error for training datasets called training error or bias, and the error on testing/unseen datasets called variance (Igual & Segui, 2017). The decomposition of prediction error (P.E) consist on the sum of three components, bias, variance and irreducible error (Dangeti, 2017; Geurts, 2009). The mathematical illustration of variance and bias presented as, the response variables (wheat area, productivity and population) is going to be predict by machine learning model taking the features covariates time by the relation as $y = g(x) + e$, and stipulated as error fallows to normality. The estimated model of $g(x)$ is $\hat{g}(x)$ and the expected squared prediction error at "x" is defined as:

$$\mathrm{P.E\ (x) = E[(y - \hat{g}(x))^2]}$$

Now the prediction error decomposed into bias and variance components as:

$$\mathrm{error\ (x) = Var[\hat{g}(x)] + [Bias\{\hat{g}(x)\}]^2 +\ Var(error)}$$

Prediction error = Variance + Bias^2 + Irreducible error

That term irreducible error known as noise. In machine learning model, the aim is to decrease both the variance and bias terms. However in machine learning model, there exist a tradeoff for minimizing the bias and variance. The optimum model means a models fallows to low prediction error with low variance and low bias and free from over fit and under fit model (Jain, 2016).

Decomposition analysis model: The production is functional form of area and yield and abstracted from the sum of the product of area and yield. The variation in production accrues due to change in area and productivity. The decomposition analysis model is used to measures the relative contribution of area and productivity towards production (Dhakre & Sharma, 2010; Murindahabi *et al.*, 2018; Rehman *et al.*, 2011) and it is measure by the following three effects as area effects, yield effects and their interaction effects.

Change in production = Yield effects + Area effects + Interaction effects

$$\text{Change in production} = \frac{(Y_c - Y_0)\times A_0}{P_c - P_0} \times 100 + \frac{(A_c - A_0)\times Y_0}{P_c - P_0} \times 100 + \frac{(Y_c - Y_0)\times A_c - A_0)}{P_c - P_0} \times 100$$

$$\Delta P = A_0 \Delta Y + Y_0 \Delta A + \Delta Y \Delta A$$

where Δ= Change in area/yield/production over time, P_0 = production at $y_{t=1950}$, P_c = Production at $y_{t=2020}$, Y_0 = yield at $y_{t=1950}$, Y_c = yield at $y_{t=2020}$ A_0 = area at $y_{t=1950}$, A_c = area at $y_{t=2020}$.

Evaluation metrics and model selection criterions: To evaluate the performance of predicted models on the training and testing/validation dataset, the following tools are applied to select the best model.

Lower the root mean square error (RMSE) and high performance score (R^2).

$$\text{RMSE} = \sqrt{\frac{\sum_i^n (y_i - \hat{y}_i)^2}{n}}\ , \quad R^2 = \frac{\sum(\hat{Y} - \bar{Y})^2}{\sum(Y - \bar{Y})^2}$$

The Akaike information criteria, developed by Akaike, 1973 used to select the model which have lower AIC, able to predict the model well and this criterion applied with unified way as log-likelihood functions with simple penalties (Banks & Joyner, 2017; Dziak *et al.*, 2012; Gujarati, 2008).

$$\text{AIC}=n\left\{\text{Ln}\left(\frac{RSS}{n}\right)\right\} + 2k$$

where “k”=no. of regressors including intercept, n = no. observation.

Schwarz information criterion (SIC) applied, and lower the value of SIC lead to select the best model (Gujarati, 2008; Neath & Cavanaugh, 1997).

$$\text{SIC}=n\left\{\text{Ln}\left(\frac{RSS}{n}\right)\right\} + k \times \text{Ln}(n)$$

Results and Discussions

The whole dataset (1950-2020) are used to learn the statistical model, while for the supervised machine learning, the data partition are applied using 80% (1950-2005) dataset as train and 20% (2006-2020) dataset as test datasets. The train data set apply to learn the model, while the test dataset used to test/validate the model. The performance of the traditional statistical and ML algorithms is integrated using model evaluation metrics and information criterions for the prediction of wheat area, productivity and population explosion in Pakistan. The scope of this research is to apply, and to integrate the statistical and machine learning regression tools for the future concerns for the years 2030 and 2050. The compound growth exponential regression model (CGREM) is used for the prediction of the determents.

Integrating statistical and machine learning modeling: Table 1 shows the results for CGREM using machine learning approach, the performance scores (R^2) found 0.94, 0.94 and 0.99 for the wheat area, yield and population model. The RMSE for wheat area model found as 0.05 and 0.17, for wheat productivity found as 0.103 and 0.17 and for population reported as 0.032 and 0.14, respectively for the train and test models. The AIC and SIC values found as -415.4 and -410.9 for wheat area models, -329.6 and -325.1 for wheat productivity models, -492.7 and -488.2 for population models. The regression coefficient yields significant results and found as 0.014, 0.025 and 0.0297, respectively for wheat area, productivity and population models. The F-statistic yield significant values for CGREM. Table 2 shows the results for the benchmark traditional statistical models. The CGREM model shows the model performance score (R^2) as 0.92, 0.95 and 0.99, RMSE as 0.073, 0.108 and 0.051, AIC as -368.6, -313.8 and -419.6, SIC as -364.4, -308.5 and -415.0, respectively for wheat area, yield and population. The regression coefficients found significant and for wheat area it is reported as 0.012, for wheat yield as 0.023 and for population it is reported as 0.028.The overall F-statistic reveals significance values for both the models. The machine learning models is trained and deployed for the train datasets. All the evaluation metrics depicts lower value of RMSE, AIC and SIC with good performance score for the machine learning models ($RMSE_{ML}$ <$RMSE_{Stat}$, AIC_{ML} <AIC_{Stat}, SIC_{ML}<SIC_{Stat}).

Table 1. Machine learning (CGREM) for wheat area, productivity and population.

Parameters	R^2	RMSE		AIC	SIC	B	F. (sig)	2020 (t_c)	2030 (t_p)	2050 (t_p)
		Training	testing							
Area	0.94	0.05	0.17	-415.4	-410.9	0.014^{**}	925.9	21750.62	24994.8 (15%)	33007.3 (51.7%)
Yield	0.94	0.103	0.17	-329.6	-325.1	0.025^{**}	818.0	29.02	37.15 (28%)	60.87 (109.7)
Population	0.99	0.032	0.14	-492.7	-488.2	0.0297^{**}	12118.8	215.25	288.84 (34.2%)	517.92 (140.6%)

**Denoted for significant results, value in parenthesis shows the predictive relative change across the year

Table 2. Statistical modeling (CGREM) for wheat area, productivity and population.

Parameters	AdjR^2	RMSE	Sig	AIC	SIC	B
Area	0.92	0.073	782.5**	-368.6	-364.4	0.012**
Yield	0.95	0.108	1302.5**	-313.8	-308.5	0.023**
Population	0.99	0.051	8896.5**	-419.6	-415.0	0.028**

**Denoted for significant results

The prevalence of error variation for statistical and machine learning approaches is shown in Fig. 1. For the wheat area, productivity and population models, the RMSE for machine learning model revealed as 0.05, 0.103 and 0.032, while for statistical models these are reported as 0.073, 0.108 and 0.051. The error found lower in machine learning models comparing with statistical models.

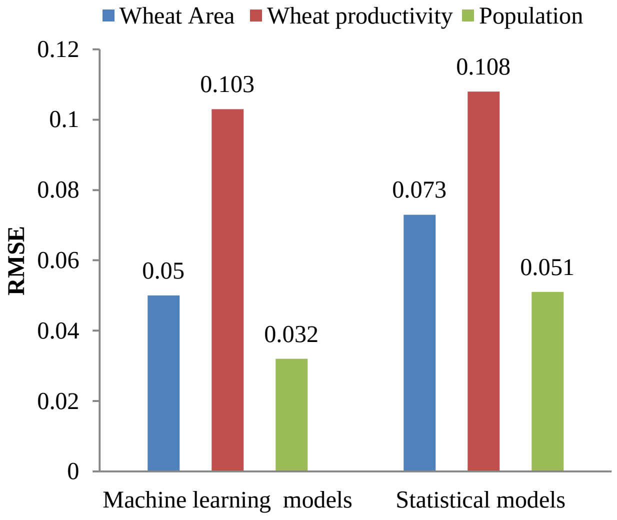

Fig. 1. Integrating the RMSE for machine learning and statistical models.

It is predicted from models that the wheat area will rise up to 15.0% and 51.7%, wheat yield will grow up to 28.0% and 109.7%, population will rise up to 34.2% and 140.6%, up to 2030 and 2050 (Fig. 2). Wheat area will touch to 24994.8 (“000” acre) in 2030 and 33007.3 (“000”acre) in 2050 from 21750.62 (“000” acre) in 2020, wheat productivity predicted as 37.15 (mds/acre) in 2030 and 60.87 (mds/acre) in 2050 from 29.02 (mds/acre) in 2020 and population predicted as 288.84 (millions) in 2030 and 517.92 (millions) in 2050 comparing with 215.25 (millions) in 2020. It is relatively estimated that up to 2030 and 2050, the population will increase about 19.2% and 6.2%, 88.9.% and 30.9% more than that from wheat crop area and yield and this situation might depress the Pakistan due to shortage of food. To overcome the food demand and to attain the sustainable agriculture, it is needed to increase the wheat yield to major extent through intensive agriculture farming and from getting precision agriculture. From ML model, highest growth rate reported from population about 3.014% while for area and yield it is reported 1.409% and 2.53%.

Hyper parametric tuning of machine learning models: Table 3 shows the hyper parametric tuning of machine learning CGREM. The models are further deployed for various sub-fold of train test split as 75%, 25% and 70%, 30% randomized data partitions. For the wheat area at 80%, 75% and 70% train phase, the RMSE found as 0.05, 0.049 and 0.041 and for test phase, it is reported as 0.17, 0.198 and 0.22. For wheat yield, the RMSE reported as 0.103 and 0.17 for 80% and 20% train test split, found 0.105 and 0.168 for 75% and 25% partitions and found as 0.108 and 0.163 for 70% and 30% partitions, respectively for train and test/validation subsets. The CGREM model for population at various sub folds reveals RMSE as 0.0320 and 0.141 for partitions 80% and 20%, 0.032 and 0.149 for partitions 75% and 25% and for 70%, 30% train test split, it is reported as 0.0243 and 0.159, respectively for train and test datasets. It is reported that as slight changes observed in train and test phase for both the determinants of prediction. The learning curves for the prediction error (Fig. 3), depicts that data partitions with 80% train and 20% test subsets fallows lowest prediction error, which elaborates at 80% and 20% train test split model’s performance is better comparing with validation matrices to avoid from over and under fit models.

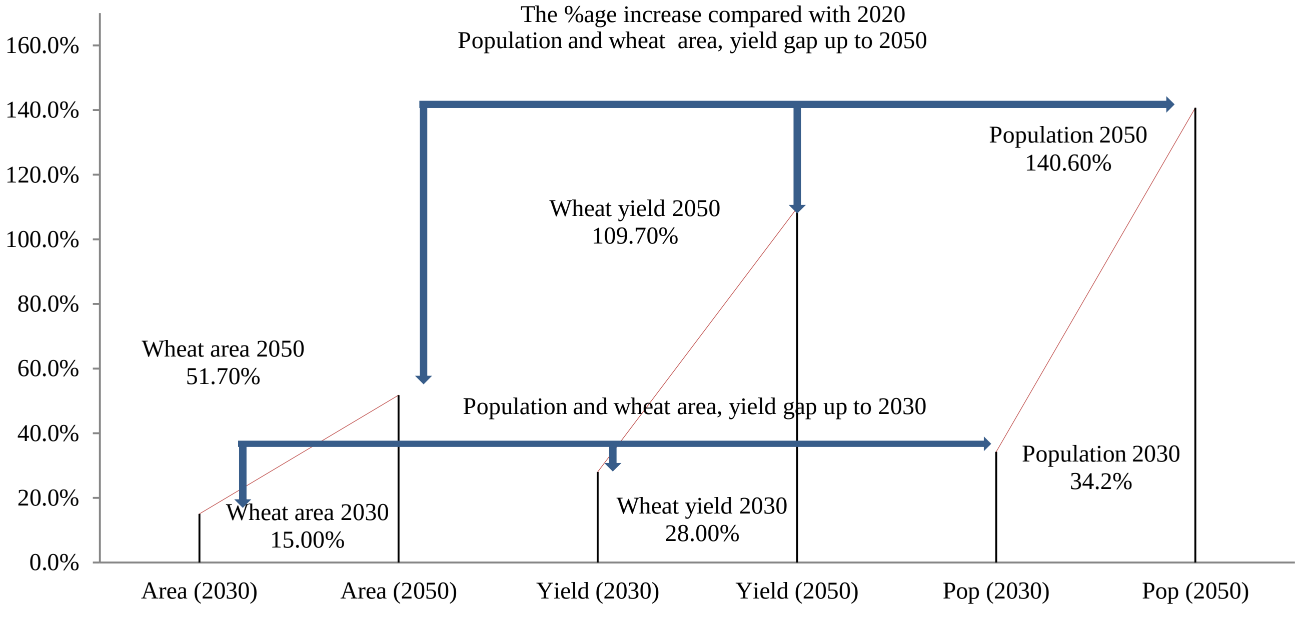

Fig. 2. Gap analysis for wheat area, yield with the population using ML (CGREM).

Table 3. Hyper- parametric tuning for error of various sub-fold of ML CGREM models.

CGREM	Train (80%) 1950-2005	Test (20%) (2006-2020)	Train (75%) 1950-2005	Test (25%) (2006-2020)	Train (70%) 1950-2005	Test (30%) (2006-2020)
Wheat area	0.05	0.17	0.049	0.198	0.041	0.22
Wheat productivity	0.103	0.17	0.105	0.168	0.108	0.163
Population	0.0320	0.141	0.032	0.149	0.0243	0.159

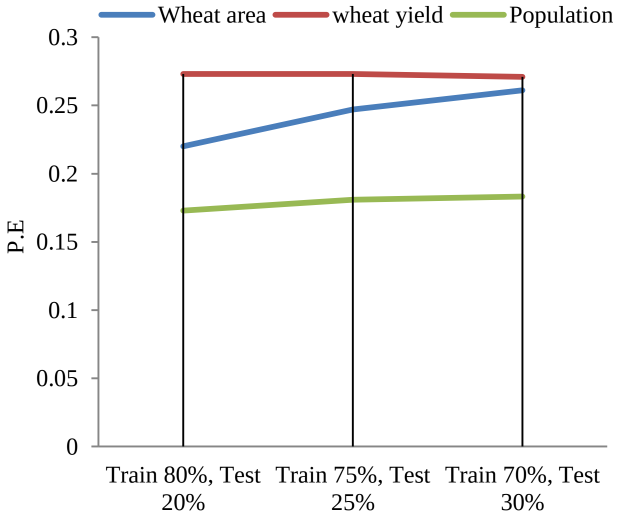

Fig. 3. Prediction error learning curves for machine learning CGREM models.

Table 4. Decomposition analysis model for wheat production.

	Area effects	Productivity effects	Interaction effects
Wheat	20%	38%	42%

Decomposition analysis model: The decomposition analysis model is applied to determine the contribution of area and productivity towards production. The productivity is the main contributor for change of wheat crops production about 38% as compared to area about 20% (Table 4), which shows productivity effect is larger than area, to change the production (productivity effects> area effects). The interaction effects for production found as 42%. The low effect of area and high productivity effect indicates yield is contributing majorly towards production and it is also elaborated by growth rate (2.53%) as compared to area (1.409%) (CGR_{area}<CGR_{yield}).

Conclusion and Policy recommendations

Food security has been evolved as the foremost global threat. Strategic planning for food security has become the key intention especially for the developing countries. A comparative study is carried out, to forecast the wheat area, yield and population eruption in Pakistan using the time series dataset, comprise from 1950-2020.This study layouts the plan to develop the regression model using the compound growth rate, called compound growth exponential regression models (CGREM). The current study integrated the efficacies of machine learning and statistical models using CGREM based on 80% and 20% train test split. It is demonstrated here that machine learning CGREM delivers better prediction capability for wheat area, yield and population with performance score as 0.94, 0.94 and 0.99, comparing with benchmark statistical models as 0.92, 0.95 and 0.99. Lower error reported for machine learning models as 0.05, 0.103 and 0.032, comparing with benchmark statistical model as 0.073, 0.108 and 0.051, respectively for wheat area, productivity and population in Pakistan. The information criterion i.e. AIC and SIC found lower in machine learning approaches. The growth rate reported high as 3.014% for population, comparing with wheat area as 1.409% and productivity as 2.53%. The CGREM predicted for 2050, the wheat area will rise up to 51.7%, wheat yield will grow up to 109.7% and population will rise up to 140.6%. It noted that population will upturn 88.9% and 30.9% more from wheat area and yield and it might critiques the food threat. The 80% and 20% train test split found superior then other models as lowest prediction error reported for the 80% and 20% train test split, comparing with others randomized partitions. Decomposition analysis model depicts that wheat productivity and area sharing 38% and 20% change towards production. It is foremost need of time to upstairs the wheat productivity through precision agriculture and intensive agricultural program to ensure food sustainability. This study demonstrated the strong evidences to layout the true policy decisions, which leads to overcome the social dilemma of food security in the region.

Ethics Statement

This research is completely determined by the consent and confirmations from all the respondent authors and they are fully aware and agreed with the policies of journal.

Acknowledgments

The authors would like to thanks all the participants who were willing to give their good response for this research.

References

Ahmad, M. and U. Farooq. 2010. The state of food security in Pakistan: Future challenges and coping strategies. *Pak. Dev. Rev.*, 49(4): 903-923.

Alagurajan, M. and C. Vijayakumaran. 2020. ML methods for crop yield prediction and estimation: An exploration. *IJEAT*, 9(3): 3506-3507.

Alexandropoulos, S.A.N., S.B. Kotsiantis and M.N. Vrahatis. 2019. Data preprocessing in predictive data mining. *Knowl. Eng. Rev.*, 34(1): 1-33.

Anonymous. 2020. Economic survey of Pakistan. Ministry of Finance, Govt. of Pakistan, Islamabad.

Banks, H.T. and M.L. Joyner. 2017. AIC under the framework of least squares estimation. *Appl. Math. Lett.*, 74: 33-45.

Dangeti, P. 2017. *Statistics for machine learning*. Packt Publishing Ltd.

Dhakre, D. and A. Sharma. 2010. Growth analysis of area, production and productivity of maize in Nagaland. *Agric. Sci. Digest.,* 30(2): 142-144.

Dziak, J.J., D.L. Coffman, S.T. Lanza and R. Li. 2012. Sensitivity and specificity of information criteria: *The Methodology Center, Penn State University, Tech. Rep. Ser.* 12-119: 1-30.

Elavarasan, D., D.R. Vincent, V. Sharma, A.Y. Zomaya and K. Srinivasan. 2018. Forecasting yield by integrating agrarian factors and machine learning models: A survey. *Comput. Electron. Agric.,* 155: 257-282.

Farooq, A., M. Ishaq, S. Yaqoob and K.N. Sadozai. 2007. Varietal adoption effect on wheat crop production in irrigated areas of NWFP. *Sarhad J. Agric.,* 23(3): 807-814.

Geurts, P. 2009. *Bias vs variance decomposition for regression and classification*. In: (Eds.): Maimon, O. & L. Rokach. Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA., pp. 733-746

Gujarati, D. 2008. *Basic Econometrics.* New York, MeGraw-hill.

Han, J., M. Kamber and J. Pei. 2011. *Data mining concepts and techniques third edition.* The morgan kaufmann series in data management systems.

Igual, L. and S. Segui. 2017. *Introduction to data science: A python approach to concepts,techniques and applications.* Springer international publishing Switzerland, pp. 1-4.

Islam, M. and F. Shehzad. 2022. A prediction model optimization critiques through centroid clustering by reducing the sample size: Integrating statistical and machine learning techniques for wheat productivity. *Scientifica*, 2022 (7271293): 1-11.

Jain, A. 2016. *Complete guide to parameter tuning in XGBoost (with codes in Python).* analyticsvidhya.

Jeong, J.H., J.P. Resop, N.D. Mueller, D.H. Fleisher, K.Yun, E.E. Butler, D.J. Timlin, K.M. Shim, J.S. Gerber, V.R. Reddy and S.H. Kim. 2016. Random forests for global and regional crop yield predictions. *PLoS One,* 11(6): 1-15.

Kondal, K. 2014. Growth rate of area, production and productivity of onion crop in Andhra Pradesh. *Indian J. Appl. Res.,* 4(3): 4-6.

Kotsiantis, S.B., D. Kanellopoulos and P.E. Pintelas. 2006. Data preprocessing for supervised leaning. *IJCS,* 1(2): 111-117.

Kumar, N.S., B. Joseph and M. Jaslam. 2017. Growth and instability in area, production, and productivity of cassava (Manihot Esculenta) in Kerala. *IJARIIT,* 4(1): 446-448.

McCarthy, J. and E.A. Feigenbaum. 1990. In memoriam, Arthur samuel, pioneer in machine learning. *AI Magazine,* 11(3): 10-10.

Mishra, S., D. Mishra and G.H. Santra. 2016. Applications of machine learning techniques in agricultural crop production: A review paper. *Indian J. Sci. Technol.,* 9(38): 1-14.

Mozumdar, L. 2012. Agricultural productivity and food security in the developing world. *Bangladesh J. Agric. Econs,* 35(1-2): 53-69.

Murindahabi. T, Q. Li and E.M.B.P. Ekanayake. 2018. Economic analysis of growth performance of various grains crops during agricultural reform in Rwanda. *IISTE,* 9(2): 32-42.

Neath, A.A. and J.E. Cavanaugh. 1997. Regression and time series model selection using variants of the schwarz information criterion. *Commun. Statist. Theory Meth.,* 26(3): 559-580.

Nelli, F. 2015. *Python data analytics, data analysis and science using PANDAs, Matplotlib and the Python programming language*. Apress.

Nelson, G.C., M.W. Rosegrant, A. Palazzo, I. Gray, C. Ingersoll, R. Robertson, S. Tokgoz, T.Zhu, T.B. Sulser, C. Ringler, S. Msangi and L. You. 2010. *Food security, farming, and climate change to 2050: Scenarios, results, policy options.* Intl Food Policy Res Inst.

Priya, P., U. Muthaiah and M. Balamurugan. 2018. Predicting yield of the crop using machine learning algorithm. *IJESRT*. 7(1): 1-7.

Qasim, M., S. Hassan, A. Bashir, H.Z. Mahmood and I. Mehmood. 2015. Analyzing production potential of selected food and legume crops for food security in Punjab, *Pakistan. J. Agric. Res.,* 28(3): 255-262.

Rahman, A. 2019. Statistics based data preprocessing methods and machine learning algorithms for big data analysis. *Int. J. Artif. Intell.,* 17(2): 44-65.

Rehman, F. U., I. Saeed and A. Salam. 2011. Estimating growth rates and decomposition analysis of agriculture production in Pakistan: Pre and post sap analysis. *Sarhad J. Agric.,* 27(1): 125-131.

Sanchez, A.G., J.F. Solis and O.W. Bustamante. 2014. Predictive ability of machine learning methods for massive crop yield prediction. *Span. J. Agric. Res.,* 12(02): 313-328.

Tilman, D., C. Balzer, J. Hill and B.L. Befort. 2011. Global food demand and the sustainable intensification of agriculture. *PNAS,* 108(50): 20260-20264.

Yadav, N., S.M. Alfayeed and A. Wadhawan. 2020. Machine learning in agriculture: Techniques and applications. *IJEAST,* 5(7): 118-122.

(Received for publication 25 November 2021)