Journal of Pure and Applied Agriculture (2024) 9(1): 48-55
ISSN (Print) 2617-8672, ISSN (Online) 2617-8680
http://jpaa.aiou.edu.pk/

Construction of a modified Cobb-Douglas model using a machine learning approach to optimize wheat productivity

Rabia Siddiqui[1]*, Abdul Qayyum[2], Muhammad Islam[3] and Mirza Waseem Abbas[1]

[1]Deputy Director (Stat), Crop Reporting Service, Agriculture Department, Punjab, Pakistan
[2]Director General, Crop Reporting Service, Agriculture Department, Punjab, Pakistan
[3]Director of Agriculture (Stat), Crop Reporting Service, Agriculture Department, Punjab, Pakistan

*****Corresponding author: Rabia Siddiqui (rabiaislam92@gmail.com)

Abstract

Food availability has become a tangible issue in the face of the prevailing threat of food insecurity. Wheat production is leading significant role to concerns over the country's ability to meet its food needs. In the current study, modified Cobb-Douglas regression model is developed using machine learning model (MLM), which is not developed earlier. MLM is integrate with traditional statistical model (TSM) for multiple linear regression (MLR) with the aims to identify significance of agronomical constrains (factors) for wheat production in Pakistan. The secondary data of crop cut experiments is collected from the Crop Reporting Service (CRS). Python' key library (Scikit Learn) is used to analyze the experiment. The MLM is applied using 80% and 20% randomized partition. The MLM performed better than TSM for MLR. To find the better model, Cobb-Douglas regression is applied using the MLM and TSM for the same dataset. MLM is applied for the train dataset. Highest R^2 and lowest MSE and MAE found for Cobb-Douglas regression using MLM, comparing with TSM. The modified Cobb-Douglas regression using MLM found better fitted model. There is positive but insignificant relation exist for urea, water and adoption of new varieties trends. The harvesting and sowing period shows positive and significant relation, while DAP, other fertilizers, spray pest and soil type shows positive and highly significant relation against wheat productivity. This study can provide deep insights the productivity enhancement practices and can lead to layout the effective strategies to enhance wheat production with the aim to ensure the food availability.

Keywords: Food availability, Machine learning, Modified Cobb-Douglas, Productivity enhancement practices, Significant of agronomical constraints, Statistical model

To cite this article: Siddiqui, R., Qayyum, A., Islam, M., & Abbas, M. W. (2024). Construction of a modified Cobb-Douglas model using a machine learning approach to optimize wheat productivity. *Journal of Pure and Applied Agriculture*, *9*(1), 48-55.

Introduction

The Cobb–Douglas production function (regression) is tested as a statistical model base theory by Charles Cobb and Paul Douglas (Cobb & Douglas, 1928; Biddle, 2012a). The Cobb–Douglas production function is a functional form of factors used to explore the relation between the two or more factors (inputs) and the response (output), and that response is produced by these inputs factors (Cobb & Douglas, 1928). In 1927, Paul Douglas and Charles Cobb presented the papers entitled "A Theory of Production" in which they proposed the Cobb–Douglas production function (regression) as a mathematical relationship for the independent variable (inputs) and output (response) and they estimated its coefficients using the functional form of the least squares regression (Cobb & Douglas, 1928; Berndt & Christensen, 1973; Biddle, 2012b; Tarab, 2014; Laitsou et al., 2017; Khan et al., 2021). Machine learning algorithms is powerful tool of artificial intelligence (AI) applied to extract meaningful information inside from the dataset (Mahesh, 2020; Nazarathy & Klok, 2021; Islam & Shehzad, 2022a). Arthur Samuel, a pioneer in machine learning, defined machine learning in 1959 as: "field of study that gives computers the ability to learn without being explicitly programmed" (Park et al., 2018; Mahesh, 2020; Shehzad et al., 2023). In current study efforts are made to construct the modified Cobb-Douglas Regression model using machine learning algorithms which is not developed earlier.

Why wheat crop is being studied

Food and Agriculture Organization (FAO) reported that there are more than eighty countries around the world are producing wheat crop (Kashish Rastogi, 2022). The International Grains Council reported that in the world, wheat production is reached to 770 million tonnes in 2022-2023, comparing with the 781 million tonnes for the year 2021-2022 with the estimated decrease of 1.40% (Stock, 2023). China is the world's largest wheat producer, accounting 17% in the total production, followed by India 12.5%, Russia 8.4%, USA 8.4%, France 5.4%, Canada 4.0%, Germany 3.5%, Pakistan 3.5%, Australia 3.2% and Ukraine 3.1% (Stock, 2023).

Pakistan is ranked 8th in the world in terms of wheat production. Wheat is the 3rd most produced grain crop in the world (Shehzad et al., 2022; Shehzad et al., 2023; Ullah et al., 2023). Pakistan is facing severe food security concerns in spite of that it is an agriculture country. Food availability is one of the essential parts for assuring the food security in Pakistan. Global Hunger Index data 2022, reported that, Pakistan falls under a serious level of hunger and stood at 99th positions out of the 121 countries in world (Dutta et al., 2022). Wheat, maize and rice are the primary food crops of Pakistan (Ali et al., 2017; Rehman et al., 2022). Wheat crop is major staple food crop of Pakistan and it stand 1st among all others food crops (Shah et al., 2020; Tusawar et al., 2021). About 37 % of the total crop area is cultivated under the wheat crops, and about 70% of the food production falls under the wheat crop in Pakistan (Khan et al., 2019; Abro & Awan, 2020). On the average 125 kg, per capita wheat is consumed is Pakistan, and it is providing the 60% of the daily food consumed (Ali et al., 2017; Azam & Shafique, 2017; Mahmood et al., 2019). About 80% of the farmers are sowing the wheat crop in Pakistan (Mann et al., 2008).

Keeping in view the importance of wheat crop being a staple food crop to attain the food sustainability, the government of Pakistan can't allowed its export and import until or unless the accurate wheat production will be determined within the country (Haqqani et al., 2000; Dempewolf et al., 2014; Ali et al., 2017). Figures provided by the economic survey of Pakistan 2021-2022, wheat crop is sown over 22.00 million acres area in the country, and it is accounting 7.8% value added in agriculture and have 1.8% share in GDP of Pakistan. Sustainable wheat production has been remained a one of the important task of every government to avoid any adversity in order to meet food demand of country. Any adversity or shortfall in wheat production may lead to create the inside uncertainty due to increase in the rates of wheat flour. The wheat production was decreased about 2.9%, 3.9% and yield was decreased 1.6%, 1.9% in Pakistan for the years 2018-19 and 2021-22, while our population is increasing rapidly across the years.

There are various factor are affecting the wheat production in Pakistan such as water scarcity, poor irrigation facilities, un-balance used of fertilizers, un-recommended seed rate, adverse climatic conditions, lack of adoption of yielding seed varieties, poor knowledge and soil variation etc. Tirfi (2022) studied the significance of various factors affecting the barley yield in Ethiopia using Cobb-Douglas production regression model through statistical approach. He reported that improved seed and fertilizer had positive and significant relation for the barley yield, while it has moderately responsive for irrigation. The land area had negative and insignificant relation for the barely yield. Zhang et al. (2020) applied Cobb-Douglas production using the statistical approach to identify the significance of the factors affecting the corn yield in China for the factor i.e. chemical fertilizer, planting area, precipitation and pesticide application rates etc. They reported that fertilizer, pesticide application rates, planting area and precipitation had positive affect for corn yield. Chandio et al. (2016) presented the study using Cobb-Douglas production through statistical tools to identify the impact of area, water availability, fertilizer off-take and credit disbursement for the wheat production in Pakistan. They found that area, water availability and credit disbursement had positive and significant affect, while the fertilizer off-take had negative insignificant affect for the wheat production.

Pakistan is blessed by almighty "Allah" with favorable atmospheres and apposite topography conditions with fertile soil, which are suitable for the cultivation of wheat crop. This study is design to integrate, and to compare the magnitude and variation using statistical relationship between factors affecting wheat production and its yield per acre. Sustainable agriculture growth is needed for the progressive development of Pakistan' economy (Anonymous., 2022; Tushaar et al., 2006).

Problem statement and objectives of the study

Pakistan is facing severe food security concerns. The inappropriate and non-recommended use of inputs and their levels are decreasing the crop production and productivity in Pakistan over the years, especially when compared to top-ranking countries and those with similar topography. Wheat crop is important food crop of Pakistan and responsible for the food security and food availability for the country's food demand. Wheat production and productivity is Pakistan has been declining in recent years, leading to concerns over the country's ability to meet its food needs and generate revenue through exports, especially when compared to top-ranking countries and those with similar topography. Wheat is a vital crop in Pakistan, contributing significantly to food security and to the country's economy.

The Cobb-Douglas production (regression) function is a widely used to analyze the agronomical constrains (factors) influencing wheat production using traditional statistical approaches. There are variations in the importance of the factors effecting the wheat production. This study is designed to construct the modified Cobb-Douglas regression model using machine learning algorithms, which is not developed earlier with the aims to identify the agronomical constrains (factors), that significantly affect the wheat production. This study can provide deep insights the productivity enhancement practices in the region, and can lead to development the effective policies and strategies to enhance wheat production.

Materials and Methods

Data collection and identification of features

The secondary dataset of crop cut experiments (CCE), is collected from the Crop Reporting Service (CRS) along with different agronomical constrains from year 2019 to 2021. Data pre-processing, a machine learning tool, is applied using the centroid clustering scheme introduced by Islam and Shehzad (2022b) for the 136 tehsils zones of Punjab. The python' key library called Scikit Learn is applied to analyze the experiment. Wheat yield in maunds/acre is used as dependent (labeled) variable and various agronomical features (inputs variables) used as, urea (kg/acre), DAP (kg/acre), other fertilizers (kg/Acre), no. of water, no. of pest spray, soil type (loamy, yes/

no), adoption trend of advanced varieties (yes/no), harvesting wheat crop in April, 1-20 (yes/no) and sowing of wheat crop in November (yes/no). Fig. 1 shows the view of the factors used in the current study.

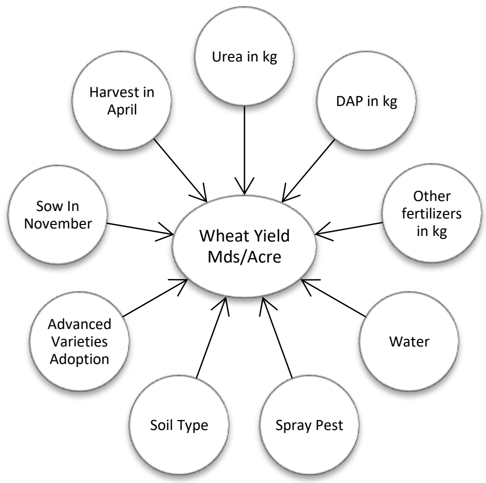

Fig. 1 Identification of features

Regression analysis

In the current study multiple linear regression and Cobb-Douglas regression models are applied to endeavor the relations between the responses (wheat productivity) and agronomical constrains (inputs factors) using the traditional statistical models (TSM) and machine learning models (MLM).

Multiple linear regression (MLR) model

Multiple linear regressions are applied to endeavor the relation between the agronomical constrain and wheat productivity.

$Y_i = \beta_0 + \sum_{i=1}^{n} \beta_i X_i + \varepsilon$ (1)

Where "Y_i" stands the response variable (wheat productivity in Mds/Acre), "β_0" stands the regression intercept, "X_i" stands the agronomical constrains (inputs variables) "β_i" stands the regression coefficient of agronomical constrains and "ε" stands error term.

Cobb–Douglas production (regression) function

The Cobb–Douglas production regression function is tested as a statistical model base theory by Charles Cobb and Paul Douglas (Biddle, 2012a; Cobb & Douglas, 1928). The mathematical transformation of the Cobb–Douglas production (regression) function is as:

$LNY_i = \beta_0 + \sum_{i=1}^{n} \beta_i LNX_i + \sum_{j=1}^{m} \beta_j X_j + \varepsilon$ (2)

Where "Y_i" stands the response variable (wheat productivity in Mds/Acre), "β_0" stands the regression intercept, "X_i" stands the quantitative agronomical constrains (inputs variables), "X_j" stands the binary categorical agronomical constrains (inputs variables) "β_i" stands the regression coefficient of quantitative agronomical constrains, "β_j" stands the regression coefficient of binary categorical agronomical constrains and "ε" stands error term.

Construction of modified Cobb-Douglas regression using machine learning approach

A machine learning (ML) algorithm is powerful tool used to extract meaningful information inside from the data. ML is evolutionary extension of data science that gives the computers, the ability to learn inside the data without being explicitly programmed. In this study efforts are made to construct the modified Cobb-Douglas Regression model using ML algorithms. In machine learning whole data set are split into two randomized partition called train and test datasets. The train data sets used to train the model, while the test dataset used to test the performance of model. Machine learning model deployed for the train models. In the current study 80% of data used as train dataset while the 20% used as test dataset. The train model is applied with 80% randomized partition and test model is applied with 20% randomized partition as:

$LNY_{(i)train} = \beta_0 + \sum_{i=1}^{n} \beta_{(i)train} LNX_{(i)train} + \sum_{j=1}^{m} \beta_{(j)train} X_{(j)train} + \varepsilon$ (3)

$LNY_{(i)test} = \beta_0 + \sum_{i=1}^{n} \beta_{(i)test} LNX_{(i)test} + \sum_{j=1}^{m} \beta_{(j)test} X_{(j)test} + \varepsilon$ (4)

Where "Y_i" stands the response variable (wheat productivity in Mds/Acre) for train and test model, "β_0" stands the regression intercept, "X_i" stands the quantitative agronomical constrains (inputs variables), respectively for train and test model, "X_j" stands the binary categorical agronomical constrains (inputs variables) , respectively for train and test model "β_i" stands the regression coefficient of quantitative agronomical constrains for train model, "β_j" stands the regression coefficient of binary categorical agronomical constrains for train model, and "ε" stands error term.

Testing the significance of regression coefficients

In the current study, an alternative hypothesis (H_i) is built to test the statistical significance of regression coefficient of various agronomical constrains for the response (wheat productivity).

- $H_{1(i)}$: Significant relation exists between agronomical constraints and wheat productivity.
- $H_{0(i)}$: Insignificant relation exists between agronomical constraints and wheat productivity.

The degree of statistical significance may vary by shifting the level of significance. For the current study, hypothesis is tested against the 5% and 1% level of significance. The relation considered highly significance for P-value <0.01, significance for P-values <0.05, and insignificance for P-values >0.05.

Evaluation metrics approach and regression diagnostics

The following evolution metrics are applied to evaluate the performance of models:

- Higher value of performance score (R^2) leads to select the best model.
- Lower the values mean square error (MSE) and mean absolute error (MAE) leads to select the best model.
- Normality of the error term is checked using the graphical presentations.
- The scatter plots of residual and predicted value is used to check the constant variance.
- P-value is used to test the significance of regression coefficients.
- F-statistic is applied to test the overall significant of the model.
- Durban Watson (D.W) test is applied to check the autocorrelation.
- Variance inflation factor (VIF) is used to check the multicollinearity in agronomical constrains.

Results and Discussion

Integrating the performance of MLR using the MLM and TSM

Table 1 show the comparison of MLR for MLM and TSM. The MLR shows the performance score (R^2) as 0.74 for TSM, while for MLM, it is found as 0.80. The MSE found as 10.98 for TSM, while for MLM it is found as 0.0094. For the MLR, highest value of performance score (R^2) and lowest value of MSE found for MLM comparing with TSM. The MLM performed better than TSM for the MLR.

$$R^2{}_{MLM(MLR)} > R^2{}_{TSM(MLR)} \quad (5)$$

$$MSE_{MLM(MLR)} < MSE_{TSM(MLR)} \quad (6)$$

This study offers support for the findings of Islam and Shehzad (2022b) and Islam et al. (2022) regarding the superiority of machine learning over traditional statistical models under similar agronomical constraints. Additionally, this study contributes by evaluating the performance of a modified Cobb-Douglas Regression model.

Integrating the performance of Cobb-Douglas regression model

To find the optimum model, the same dataset is further deployed for the Cobb-Douglas regression model for the MLM and TSM. Table 1 also depicts the comparison of Cobb-Douglas regression for MLM and TSM. For the modified Cobb-Douglas regression using the MLM, the performance score (R^2) found as 0.84 for train data, while for test data, it is found as 0.50. Performance score (R^2) is found as 0.80 for the Cobb-Douglas regression using TSM. MLM is applied for the train datasets. Comparing the performance of models, highest value of performance score (R^2) found for Cobb-Douglas regression for $\mathrm{MLM}_{(\mathrm{train})}$, comparing with the TSM. For the Cobb-Douglas regression using MLM, The MSE and MAE are found as 0.0089 and 0.0753 for the train model, while for the test model, it is reported as 0.0117 and 0.0820. For the TSM the MSE and MAE reported as 0.0094 and 0.760. As the MLM is deployed for the train dataset, the lowest value of MSE and MAE is found for the MLM, comparing with the TSM. The F-statistics revealed the significant value. The value of Durban Watson (D.W) test indicates there are no autocorrelation exist in the dataset.

$$R^2{}_{\text{Cobb-Douglas(Train)}_{(MLM)}} > R^2{}_{\text{Cobb-Douglas}_{(TSM)}} \quad (7)$$

$$MSE_{\text{Cobb-Douglas(Train)}_{(MLM)}} < MSE_{\text{Cobb-Douglas}_{(TSM)}} \quad (8)$$

$$MAE_{\text{Cobb-Douglas(Train)}_{(MLM)}} < MAE_{\text{Cobb-Douglas}_{(TSM)}} \quad (9)$$

Table 1 Comparative analysis of Cobb-Douglas Regression model

Evaluation metrics	Cobb-Douglas regression using MLM		Cobb-Douglas regression using TSM	MLR using TSM
	Train Model	Test Model	--	--
R^2	0.84	0.50	0.80	0.74
MSE	0.0089	0.0117	0.0094	10.98
MAE	0.0753	0.0820	0.0760	--
D.W	--	--	1.22	--
F- Statistic	--	--	11.63**	--

** Depicts highly significant values

Fig. 2 shows the learning bars for the R^2 and Fig. 3 shows the learning bars for MSE and MAE. It revealed that modified Cobb-Douglas regression using the MLM found optimized model as it fallows lowest values of MSE and MAE with highest value of coefficient of determination.

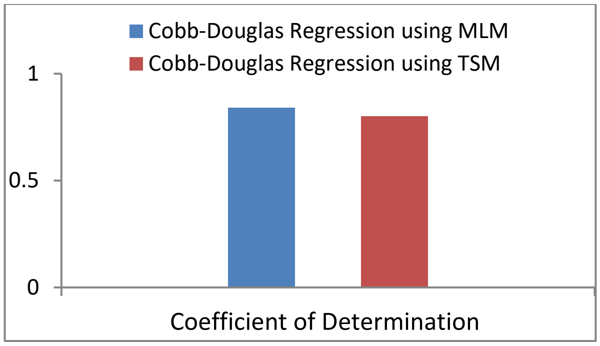

Fig. 2 Coefficient of determination for MLM and TSM

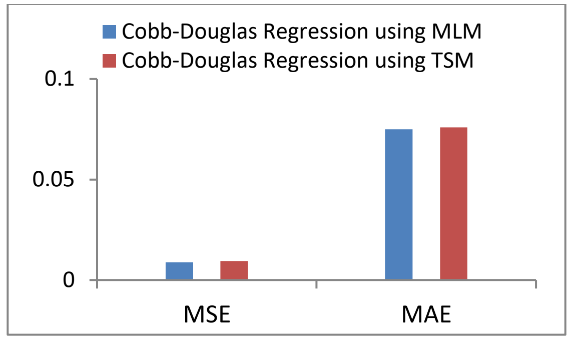

Fig. 3 MSE and MAE for MLM and TSM

Regression diagnostics are checked for the residual analysis in Fig. 4-6. Fig. 4 show the histogram with normal curve and Fig. 5 shows the P-P plot for the residual, and indicates that it fallows the normality. Fig. 6 depicts the scatter plot for residual and predicted values and it depicts that error term fulfill the assumption of constant variance. All the values of variance inflation factor (VIF) found less than 10 and predicted that there is no multicollinearity exists between the agronomical constrain (Table 2). Alamri and Mark (2018), Croppenstedt (2005), and Wang et al. (2019) conducted studies similar to this one, employing statistical models. However, this study innovatively introduced the modification of Cobb-Douglas using machine learning, providing evidence that machine learning outperformed, comparing with traditional statistical models.

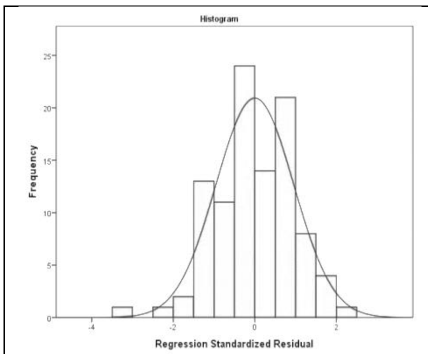

Fig. 4 Histogram with normal curve for the residual

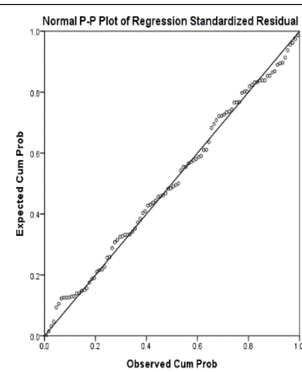

Fig. 5 P-P plot for the residual

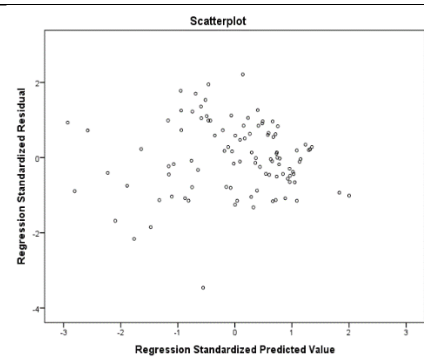

Fig. 6 Scatter plot for residual and predicted values for constant variance analysis

Testing the significance of regression coefficients

An alternative hypothesis (H_i) is testing for the statistically significance relationship for the use of agronomical constrains and wheat productivity. There is positive but insignificant ($P > 0.05$) relation exists between the urea and wheat productivity. Its coefficient says that wheat productivity increased about 0.117 Mds/Acre by a unit increase in the application of fertilizer urea. For the DAP, there is positive and highly statistically significant ($P < 0.01$) relation exists for the DAP and wheat productivity. Its coefficient found as 0.365, which indicates that wheat productivity increased about 0.365 Mds/Acre by a unit increase in the application of fertilizer DAP. The coefficient of other fertilizers found positive (0.025) and highly significant ($P < 0.01$), while coefficient of water found positive (0.026) and insignificant ($P > 0.05$). This study reinforces the findings made by Khan et al. (2021) highlighting the beneficial impact of fertilizer on wheat productivity, as indicated by coefficients of 0.1646. However, while DAP fertilizer demonstrated a significant positive effect, the analysis suggests that urea and other fertilizers may have a positive impact but lack statistical significance. The same conclusions drawn by Islam et al. (2021) that coefficient of spray pest and soil found positive and highly significant ($P < 0.01$). For the adoption of new varieties trends, regression coefficient reported as 0.119 and it is found insignificant ($P > 0.05$). The coefficient of harvesting and sowing period found positive as 0.116 and 0.29 and both found statistically significant ($P < 0.05$).

Table 2 The regression coefficient agronomical constrains and it significance

	B	t-statistic	P-Value	VIF
Constant	1.064	1.986	--	--
Urea	0.117	1.694	0.094	2.43
DAP	0.365	2.767	0.007**	1.545
Other fertilizers	0.025	2.689	0.009**	1.18
Water	0.026	0.668	0.506	2.124
Spray pest	0.027	2.964	0.004**	1.128
Soil type	0.256	5.772	0.000**	1.12
Adoption of advanced varieties	0.119	1.486	0.141	1.36
Harvest in April	0.116	2.466	0.016*	1.974
Sow in November	0.29	2.455	0.016*	1.205

*depicts significant values; **depicts highly significant values

Conclusion

Food availability has become a tangible issue in the face of the prevailing threat of food insecurity. Wheat crop is significantly contributing to food availability concerns in Pakistan. The non-recommended levels of inputs are reducing crop production and productivity in Pakistan over the years, particularly when compared to leading countries and those with similar topography. This study is designed to construct the modified Cobb-Douglas regression using machine learning model (MLM), which is not developed earlier. MLM is integrated with traditional statistical model (TSM) for multiple linear regression (MLR), to identify the various agronomical constrains (factors), which significantly affected the wheat production in Pakistan. The secondary data of crop cut experiments is collected from the Crop Reporting Service for various agronomical features (inputs variables) and the python' key library (Scikit Learn) is applied to analyze the experiment. The MLM train model is applied using 80% randomized partition and MLM test model is applied using 20% randomized partition of dataset. The MLM outperformed the TSM in the MLR, as it achieved the highest R^2 value and the lowest error. To search the optimum model, the same dataset is deployed for the Cobb-Douglas (C-D) regression using MLM and TSM. MLM is applied for train model. When comparing the models' performance, the Cobb-Douglas regression for $\mathrm{MLM}_{(\mathrm{train})}$ showed the highest R^2 values and the lowest MSE, outperforming the TSM. The modified Cobb-Douglas regression using the MLM found better fitted model. The statistically significance is tested using alternative hypothesis for wheat productivity against various agronomical constrains. There is positive but insignificant relation exist for wheat productivity against urea, water and adoption of new varieties trends. The harvesting and sowing period shows positive and significant relation for wheat productivity. The positive and highly significant relation exists for wheat productivity against DAP, other fertilizers, spray pest and soil type. This study can provide deep insights the productivity enhancement practices and can lead to layout the effective strategies to enhance wheat production with the aim to ensure the food availability.

Conflict of Interest: All authors declared no conflict of interest.

Data Availability: The datasets used to support the findings of this study are available from the corresponding author upon request.

Acknowledgement: All authors acknowledge the strong data collection mechanism of Crop Reporting Service, Punjab being a sole source of agricultural crop statistics in Punjab, Pakistan.

References

Abro, A. A., & Awan, N. W. (2020). Comparative analysis of profitability of major and minor crops in Pakistan. *Journal of the Saudi Society of Agricultural Sciences, 19*(7), 476-481.

Alamri, Y., & Mark, T. (2018). Functions of wheat supply and demand in saudi arabia. *Alamri Yosef, Mark T (2018). Functions of Wheat Supply and Demand in Saudi Arabia. Journal of Agricultural Economics and Rural Development, 4*(2), 461-468.

Ali, S., Liu, Y., Ishaq, M., Shah, T., Ilyas, A., & Din, I. U. (2017). Climate change and its impact on the yield of major food crops: Evidence from Pakistan. *Foods, 6*(6), 39.

Anonymous. (2022). Economic survey of Pakistan. *Ministry of Finance, Govt. of Pakistan, Islamabad.*

Azam, A., & Shafique, M. (2017). Agriculture in Pakistan and its Impact on Economy. A Review. *International Journal of Advanced Science and Technology, 103*, 47-60.

Berndt, E. R., & Christensen, L. R. (1973). The translog function and the substitution of equipment, structures, and labor in US manufacturing 1929-68. *Journal of Econometrics, 1*(1), 81-113.

Biddle, J. (2012a). The Cobb–Douglas Regression and the Measurement of Economic Growth and its Causes: Working paper.

Biddle, J. (2012b). Retrospectives: The introduction of the cobb–Douglas regression. *Journal of Economic Perspectives, 26*(2), 223-236.

Chandio, A. A., Jiang, Y., Joyo, M. A., & Rehman, A. (2016). Impact of area under cultivation, water availability, credit disbursement, and fertilizer off-take on wheat production in Pakistan. *Journal of Applied Environmental and*

Biological Sciences, *6*(10), 10-18.

Cobb, C. W., & Douglas, P. H. (1928). A theory of production. *The American Economic Review, 18*(1),139-165.

Croppenstedt, A. (2005). Measuring technical efficiency of wheat farmers in Egypt. *The ESA Working Papers are produced by the Agricultural and Development Economics Division (ESA) of the Economic and Social Department of the United Nations Food and Agriculture Organization (FAO)*.

Dempewolf, J., Adusei, B., Becker-Reshef, I., Hansen, M., Potapov, P., Khan, A., & Barker, B. (2014). Wheat yield forecasting for Punjab Province from vegetation index time series and historic crop statistics. *Remote Sensing*, *6*(10), 9653-9675.

Dutta, S., Lanvin, B., Wunsch-Vincent, S., & León, L. R. (2022). *Global Innovation Index 2022: What is the Future of Innovation-driven Growth?* (Vol. 2000): WIPO.

Haqqani, A., Zahid, M., & Malik, M. (2000). Legumes in Pakistan. *Legumes in rice cropping system of the Indo-Genetic Planes-constraints and opportunities ICRISAT. India, 230,* 98-128.

Islam, M., & Shehzad, F. (2022a). A Prediction Model Optimization Critiques through Centroid Clustering by Reducing the Sample Size, Integrating Statistical and Machine Learning Techniques for Wheat Productivity. *Scientifica*, *2022*, 1-11.

Islam, M., & Shehzad, F. (2022b). 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. doi: 10.1155/2022/7271293

Islam, M., Shehzad, F., & Omar, M. (2021). Modeling Wheat Productivity using Hierarchical Regression: A way to Address Food Security Concerns. *Elementary Education Online, 20*(2), 1184-1195.

Islam, M., Shehzad, F., Omar, M., Qayyum, A., & Siddiqui, R. (2022). Integrating machine learning models for linear and exponential regression to predict wheat area, productivity and population. *Sarhad Journal of Agriculture, 38*(3), 894-901.

Kashish Rastogi, C. A. (2022). Visualizing Global Wheat Production by Country (2000-2020). Retrieved from *https://www.weforum.org/agenda/authors/kashish-rastogi, FAO*.

Khan, N., Bano, A., & Babar, M. A. (2019). The stimulatory effects of plant growth promoting rhizobacteria and plant growth regulators on wheat physiology grown in sandy soil. *Archives of microbiology*, *201*(6), 769-785.

Khan, S., Khan, M., Alam, A., Shah, I., Khan, M., & Khan, F. M. (2021). Factors affecting wheat productivity of small farm households in the rural district Charsadda. *Sarhad Journal of Agriculture, 37*, 1089-1097.

Khan, Z., Jan, K., & Zaib, L. (2021). Impact of human capital on economic growth-a growth accounting equation for Pakistan. *Journal of Educational Management and Social Sciences*, *1*(1), 14-30.

Laitsou, E., Kiriakidis, M., Kargas, A., & Varoutas, D. (2017). Economies of scale, cost minimization and productivity in telecom markets under economic crisis: evidence from Greece. *NETNOMICS: Economic Research and Electronic Networking, 18*, 169-182.

Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research, 9*, 381-386.

Mahmood, N., Arshad, M., Kächele, H., Ma, H., Ullah, A., & Müller, K. (2019). Wheat yield response to input and socioeconomic factors under changing climate: Evidence from rainfed environments of Pakistan. *Science of the Total Environment, 688*, 1275-1285.

Mann, R. A., Ramzan, M., & Munir, A. (2008). Improving the sustainability of wheat production in irrigated areas of Punjab, Pakistan through conservation tillage technology. *International Journal Of Agriculture & Biology*, *10*(3), 249-254.

Nazarathy, Y., & Klok, H. (2021). *Statistics with Julia: Fundamentals for Data Science, Machine Learning and Artificial Intelligence*: Springer Nature.

Park, C., Took, C. C., & Seong, J.-K. (2018). Machine learning in biomedical engineering. *Biomedical Engineering Letters, 8*, 1-3.

Rehman, A., Ma, H., Ozturk, I., & Ahmad, M. I. (2022). Examining the carbon emissions and climate impacts on main agricultural crops production and land use: updated evidence from Pakistan. *Environmental Science and Pollution Research, 29*(1), 868-882.

Shah, M. A. A., Ozel, G., Chesneau, C., Mohsin, M., Jamal, F., & Bhatti, M. F. (2020). A statistical study of the determinants of rice crop production in Pakistan. *Pakistan Journal of Agricultural Research, 33*(1), 97-105.

Shehzad, R. A., Sarwar, G., Shah, S. H., Tahir, M. A., Sabah, N. U., Muhammad, S., Aftab, M., Manzoor, M. Z., & Shehzad, I. (2022). Efficacy of P enriched organic manures to improve soil health and nutrient acquisition of wheat. *Pakistan Journal of Agricultural Research, 35*(2), 266-273.

Shehzad, R. A., Sarwar, G., Shah, S. H., Tahir, M. A., Sabah, N.-u.-S., Muhammad, S., Aftab, M., Manzoor, M. Z., Shehzad, I., & Saleem, U. (2023). Growth and yield response of wheat to organic manures (farm yard manure, phospho-compost (PROM), and press mud) alone and in combination with mineral fertilizer. *Pakistan Journal of Agricultural Research, 36*(1), 1-8.

Shehzad, F., Islam, M., Ali, A., Qayyum, A., & Siddiqui, R. (2023). Integrating exponential regression model optimizations for wheat area, productivity and population through statistical and machine learning approaches. *Pakistan Journal of Botany, 55*(5), 1-6.

Stock, A. (2023). Top 10 Wheat Producing Countries *International Grains Council*.

Tarab, G. A. B. E. (2014). *Economic Analysis of Sugar Cane Crop Production in Elguneid Scheme, Sudan (1991-*

2011). University of Gezira.

Tirfi, A. G. (2022). Modeling factors influencing barley yield in ethiopia: Augmented cobb-douglas production function approach. *Journal of Agricultural, Food and Environmental Sciences, JAFES, 76*(1), 48-57.

Tusawar, Ahmad., Rana E. A. Khan, Mariam A. Soharwardi, Muhammad N. Shafiq, & Gillani, S. (2021). Socioeconomics and agronomy of wheat yield in cotton-wheat cropping system in Punjab, Pakistan: A quality-quantity assessment. *International Journal of Agricultural Extension, 9*(1), 69-78.

Tushaar, S., Singh, O. P., & Mukherji, A. (2006). Some aspects of South Asia's groundwater irrigation economy: analyses from a survey in India, Pakistan, Nepal Terai and Bangladesh. *Hydrogeology Journal, 14*, 286-309.

Ullah, G., Khan, E. A., Shah, S. H., Kiran, M., & Haq, F. (2023). Unlocking the secrets of wheat production: Evaluating phosphorus levels and seeding rates in alluvial soil conditions. *Trends in Horticulture, 6*(1), 34-45.

Wang, P., Zhang, W., Li, M., & Han, Y. (2019). Does fertilizer education program increase the technical efficiency of chemical fertilizer use? Evidence from wheat production in China. *Sustainability, 11*(2), 1-14.

Zhang, Q., Dong, W., Wen, C., & Li, T. (2020). Study on factors affecting corn yield based on the Cobb-Douglas production function. *Agricultural Water Management, 228*, 1-11.