

## Machine Learning based Crop Recommendation System for Local Farmers of Pakistan

Sayed Mazhar Ali<sup>1</sup>; Bhagwan Das<sup>2\*</sup>; Dileep Kumar<sup>3</sup>

<sup>1</sup>Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh, Pakistan.

<sup>2\*</sup>Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh, Pakistan.

<sup>2\*</sup> engr.bhagwandas@hotmail.com

<sup>3</sup>Department of Electronic Engineering, Dawood University of Engineering & Technology, Karachi, Pakistan.

### Abstract

*Farming is one of the most fundamental and generally rehearsed work in Pakistan and it plays an imperative part in fostering the country. In Pakistan, the most part of the land is used for agriculture cultivation to meet the desires of nearby people and export want as properly. Therefore, the need of increasing crop production is the significant challenge for farmers. Crop cultivation anywhere in the world depends on the climate so called seasons and soil properties, however, the enhancing the production of crops depend on various factors like mainly on temperature. In order to address the issue of increasing crop production for Pakistan, a crop recommendation system is proposed in this work. In this work, idea of ideal harvest prior to planting it, it would be of extraordinary assistance to the farmers and others required to settle on fitting choices on upgrading the creation of yields for neighborhood utilization needs and may prompt the capacity and expanded fare choice for business. Our framework utilized Machine Learning procedures with the end goal that it proposes the appropriate corps dependent on the temperature. This framework subsequently diminishes the monetary misfortunes looked by the farmers brought about by establishing the ominous harvests and furthermore it gives the information on the occasional characterization of yields what harvest is reasonable for which season. It is concluded that proposed algorithm has an average accuracy of 90% on the given dataset. The achieved accuracy is more in comparison to existing work.*

**Key-words:** Machine Learning, Linear Regression Analysis, Polynomial Regression, Prediction of Favorable Crops, Crop Production, Climate and Temperature.

## 1. Introduction

Agriculture assumes a critical part in the economies of least developed nations like Pakistan [1]. Farming establishes the biggest offer in the economy of Pakistan [2-3]. Pakistan's farming commitment to Gross Domestic Product GDP is 26% with yearly development of 2.7% [4-6]. Agriculture gives business freedoms to 44% of the workforce and 62% of the provincial populace rely on this area for their job [7-8]. Pakistan has a place with one of the world's biggest frameworks to help Agriculture creation. It is the most important source of foreign exchange revenues, accounting for around 80% of overall export earnings [9]. Farming is considered as the foundation of Pakistan's economy, which depends intensely on its significant yields [10]. Harvest creation is reliant upon topographical variables as soil synthetic, temperature, stickiness and so forth [11]. These factors plays major role in increasing the crop production. Farmers are unaware from these factors and select the crop for cultivation based on their experience of growing particular crop again and again. Farmers' crop selection accuracy is low not cost effective. It is important to practice modern methods of farming by using technology instead of practicing traditional farming method. To solve this problem, we offer a framework that will recommend crops dependent on the temperature of the general climate. Farmers can expand their yield by developing the harvest suggested by the framework.

In this study, a Machine Learning approach is utilized to recommend the best crop, and Python is used as the programming language because it is generally accepted for experimentation in the Machine Learning field. Machine Learning is an important tool for data analysis that uses learning algorithms to iteratively learn from available data [12]. Machine Learning is an area of artificial intelligence that tries to give computer methods for accumulating, changing, and updating the knowledge of intelligent systems [12]. The model is trained using linear regression analysis. The output predictions are then made by this trained system. The proposed system suggests the favorable crop for cultivation but also provide the seasonal classification of crops to the farmers, as they get maximum production and profit.

## 2. Literature Review

There are different studies have been carried out in the past in developing and designing the crop based predicting system. The each work completed and reported in the past have their own advantages and disadvantages in developing the system based on particular problem of crop production enhancement.

In the paper [1] published by Naveen Kumar P R et al., proposed the prediction for crop using neural network. The system is developed based on the concept of linear predictive model to predict yield and best crop suitable to grow. The model has trained to produce the best answer using the train and test datasets. The system processed the information and generated a forecasted lucrative crop, as well as a list of required fertilizers and an overall yield per hectare. The system only predicated the yield of the selected crop with an accuracy of 85%. Kevin Tom Thomas et al. [2] proposed the yield expectation utilizing conventional neural systems administration dependent on the dirt supplement esteems and pH given as information. The model is trained using the KNN algorithm for the crop prediction. The system recommends crop using soil properties only with accuracy of 88%. In the paper [3], Rohit Kumar Rajak et al., proposed the crop suggestive system. The system is modeled using SVM and ANN to recommend the crop for site using soil parameters. The system is valid for only Maharashtra state, India. The only soil properties are used as input. Shashikala et al, in the paper [4] designed a rain prediction using Polynomial Regression for the Field of Agriculture Prediction for Karnatakka. The best match of rain and crop category is out final output. The system recommends crop using only rain prediction as input. The proposed system is valid for only Karnatakka state, India. N.L Chourasiya et al., [5] proposed the crop prediction system using ANN machine learning tool to predict seed sowing for farmers which ends up in high crop production. The system takes seed and soil data as input then predict the crop for planting. Devdatta A. Bondre et al., [6] proposed the prediction of crop using the Random Forest and SVM for crop yield prediction and fertilizer recommendation. In [7] by P. Surya, the system based on crop yield using Data Mining Predictive Analytical Techniques. The analysis was conducted using data from the state of Tamilnadu. The predictor model predict the yield of crops and compare the yield of different crops that which crop is most useful for the Tamilnadu state. The proposed system only compares the yield prediction of different crops using previous year data for a Tamilnadu state. In [8] by Mannem Ganesh Reddy proposed the system of classification KNN algorithm for prediction of crop. The system takes season, temperature, rainfall and soil characteristics as input to predict the yield of crops. The proposed system only predict yield per hectare and valid for only Telagana state, India.

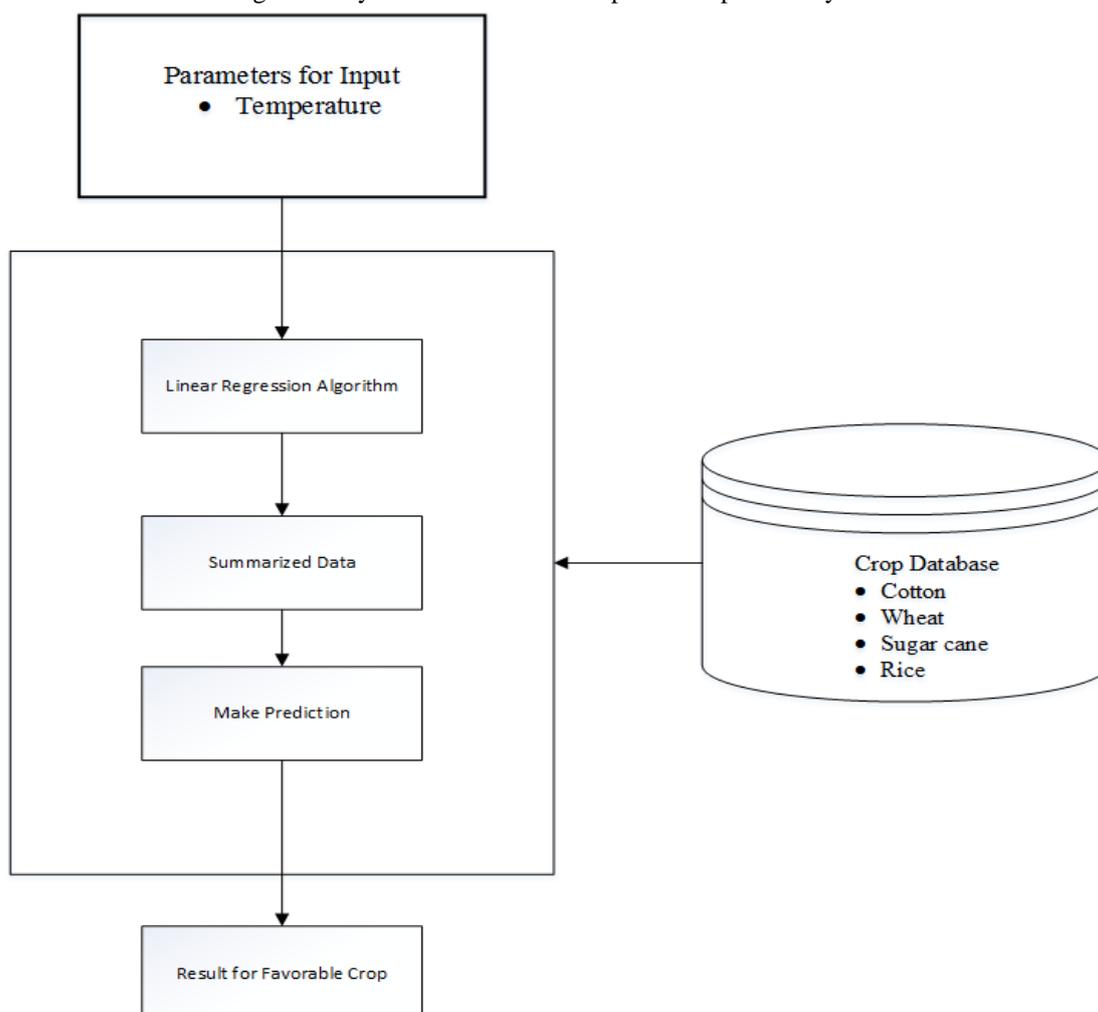
The existing systems using Machine Learning different algorithms are valuable with good accuracy used for their own regions. Mainly the existing systems use the soil parameters as input for the crop prediction. The main focuses of these systems were based on yield prediction, not for the crop suggestion. In the proposed system the input parameters are climatic conditions and soil properties to suggest the favorable crop to get maximum yield for enhancing the crop production for local farmer of rural of area of Sindh, Pakistan. Furthermore, the proposed system for the classification of knowledge

of crops to the farmers. In addition to that the accuracy of existing systems is low as compared to our work. In comparison to that proposed system offer better accuracy. The one of the limitation of the proposed work is that dataset in the system is particularly used for Nawabshah, Sindh, Pakistan region only. However, the similar system can be implemented for any region across the world using proposed methodology. In the next, methodology of the proposed system is discussed.

### 3. Research Methodology

The goal of this study is to demonstrate the impact of meteorological variables on agricultural production in order to improve crop yields, which will help farmers. As shown in Fig. 1, the linear regression system model is created in Python. Other design steps will be explored in detail in the following sections.

Figure 1 - System Model of the Proposed Crop Favor System



## A. Data Collection

Statistical data for the study is gathered from a variety of sources, including the agricultural department, websites, agricultural research institutions, and so on. The data was gathered in the Nawabshah area. Our information includes the temperate as an attribute, as well as the outcomes of crops that may be cultivated in that soil type. Dataset consists of few major crops which are mostly cultivated as wheat, sugarcane, rice and cotton.

## B. Data Preprocessing

Data preprocessing is a data mining approach that entails converting raw data into a format that can be understood. Because the original dataset may have a large number of missing values, all of them should be eliminated at first. Missing values are represented by a dot in the dataset, and their existence can degrade the overall value of the data as well as impair performance. As a result, we replace these numbers with the mean values to fix this problem. The second step is to create the class labels. Because we're utilizing supervised learning, there should be a class label for each entry in the dataset, which is produced during the preprocessing phase.

## C. Regression Analysis

Regression Analysis is a predictive modelling technique that examines the relationship between a dependent or target variable and an independent or predictor variable. It covers linear, multiple linear, and non-linear regression models, among others. Simple linear regression is the most used model.

Polynomial regression is a type of regression method in which the link between the independent variable  $x$  and the dependent variable  $y$  is described as an  $n$ th degree polynomial in  $x$ . polynomial regression fits a nonlinear relationship between the value of  $x$  and the associated conditional mean of  $y$ , denoted by  $E(y | x)$ . Despite the fact that polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear in the sense that in the unknown parameters inferred from the data, the regression function  $E(y | x)$  is linear. As a result, polynomial regression is regarded as a subset of multiple linear regressions. The predicted value of  $y$  may be modelled as an  $n$ th degree polynomial in general, producing the generic polynomial regression model given in (1):

$$y = b_0 + b_1x_1 + b_2x_2^2 + \dots + b_kx_k^k \quad (1)$$

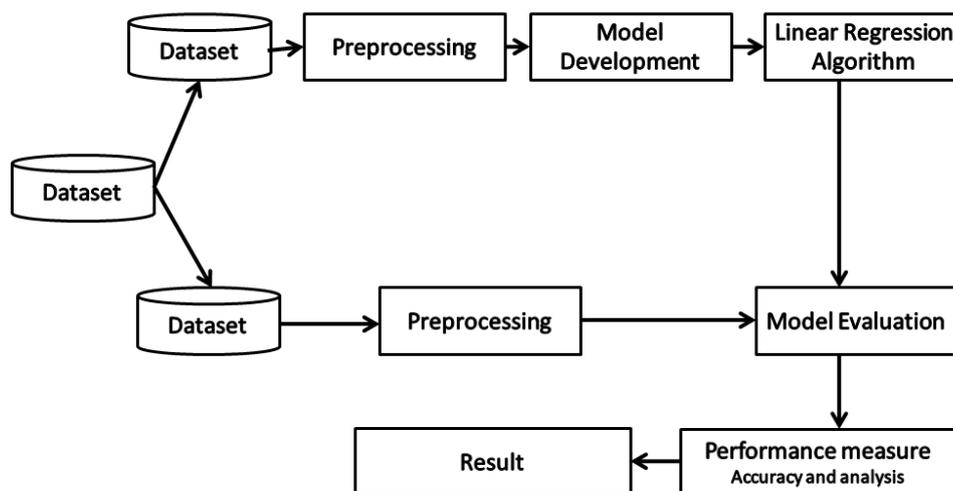
Where  $y$  is the value predicted by the polynomial model with regression coefficients  $b_1$  to  $k$  for each degree and a  $y$  intercept  $b_0$ .

#### D. Trained Model

Prepared model is gotten subsequent to applying the dataset to the Machine Learning algorithms. Our work gives a crop proposing framework which depends on the Polynomial Regression algorithm dependent on the temperature of climate are given as an input to the model. The algorithm will seek for a crop that is the most similar in value to the provided values. The output will be the results of crops, crop with high result are suitable for the inputted values.

Taking into this account, this algorithm search for the similarities in the whole dataset. The outcome is determined dependent on the most practically identical or nearest values. Depending on the amount of records in the dataset, the complete dataset is split into a number of classes or possible outcomes in this method, as illustrated in Fig.2.

Figure 2 - Diagram of Model Training



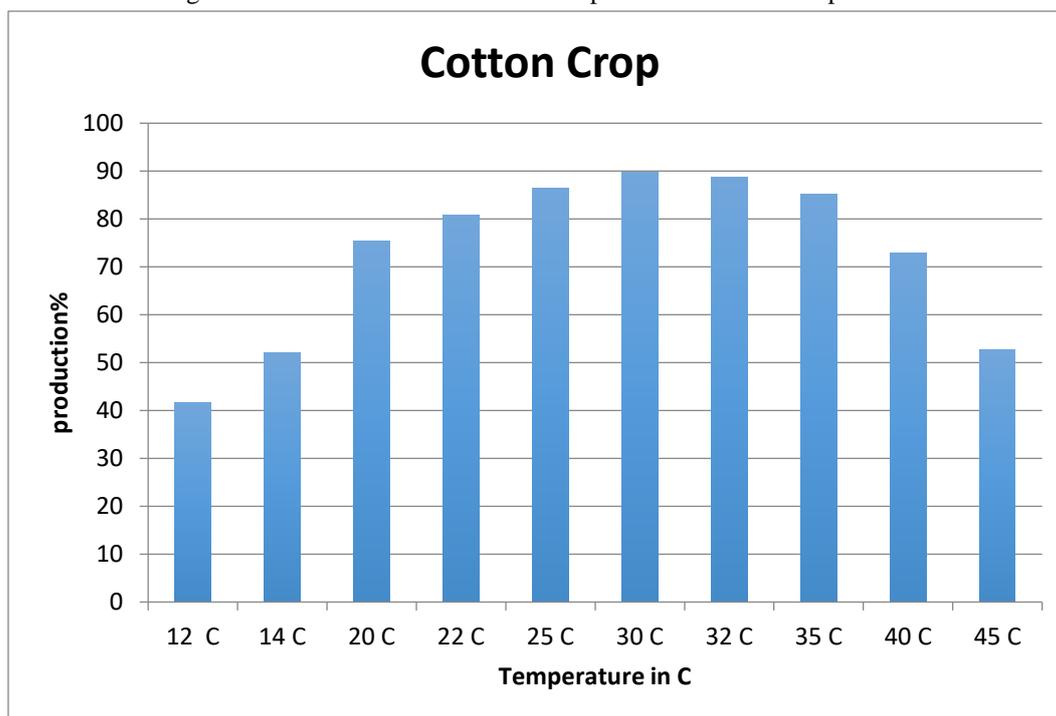
#### E. Programming Language

In our proposed work Python programming language is used. Python is a programming language that runs on a variety of platforms and has a syntax that is close to that of the English language due to which it provides fewer lines to write a program than other programming language. Python code is justifiable by people, which makes it simpler to assemble models for machine learning.

#### 4. Results and Discussions

In order to address the problem of increasing crop production for Pakistan, the above machine learning system is proposed for crop recommendation that suggests of favorable crop prior to planting it, it would be extremely beneficial to farmers and other stakeholders in making informed decisions on how to increase output. The crop production is highly affected by the environmental and soil conditions. The production is reduced when cultivated crop has not provided the favorable environmental and soil conditions. It is necessary to cultivate the favorable crop according to the environmental and soil parameters to get maximum profit. Based on the trained model discussed in Fig. 2. The analysis of crops production based on different temperature is attained using the proposed machine learning model is depicted in Fig. 3. It has been analyzed from the data that the minimum production i.e. 40% of the cotton will be yield at temperature 12°C and maximum yield of cotton production i.e. more than 90% will yield at temperature 25-35°C. Based on this information, it has been extracted that for temperature ranging from 25-35°C will be recommended to the farmer using proposed system.

Figure 3 - Relation between Cotton Crop Production and Temperature



Similarly, the analysis is carried out for wheat crop yield on different temperature is defined using the proposed machine learning model as shown in Fig. 4. It is shown in Fig. 4 that the maximum wheat crop production i.e. more than 90% will be achieved at range of temperature from 12°C to 22°C

and as the temperature raises the wheat production is decreased. These are very good recommendation proposed by designed system for the farmers to enhance the wheat production, if temperature is in between 12°C to 22°C.

Figure 4 - Relation between Wheat Crop Production and Temperature

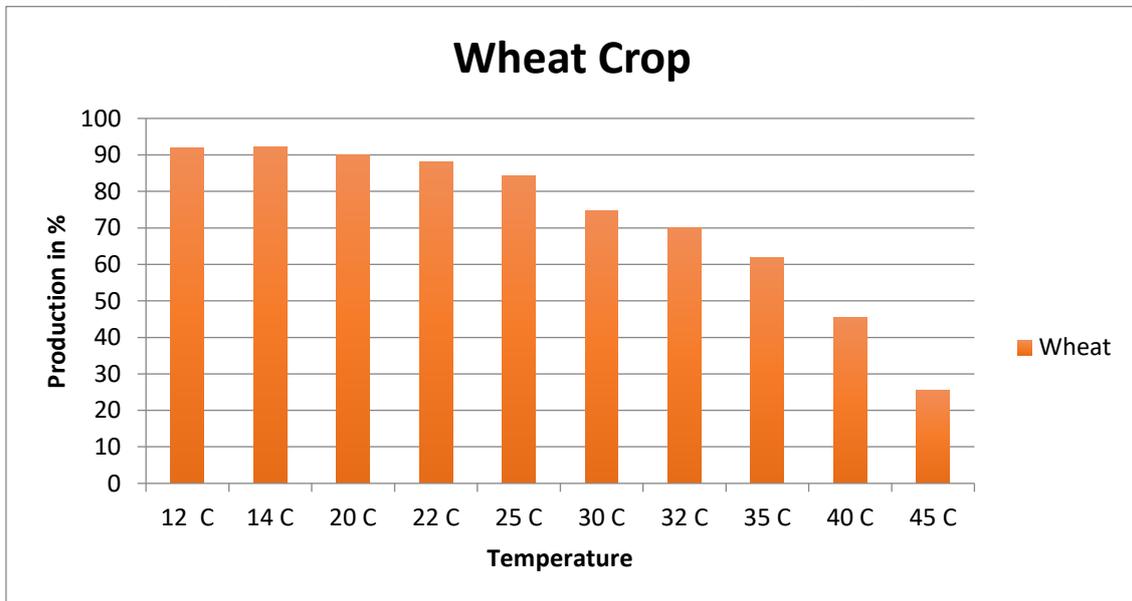


Fig. 5 discussed the Sugarcane yield on different temperature. It is shown in Fig. 5 that the maximum wheat crop production i.e. more than 80% will be achieved at range of temperature from 20°C to 32°C and as the temperature raises the sugarcane production is decreased.

Figure 5 - Relation between Sugarcane Crop Production and Temperature

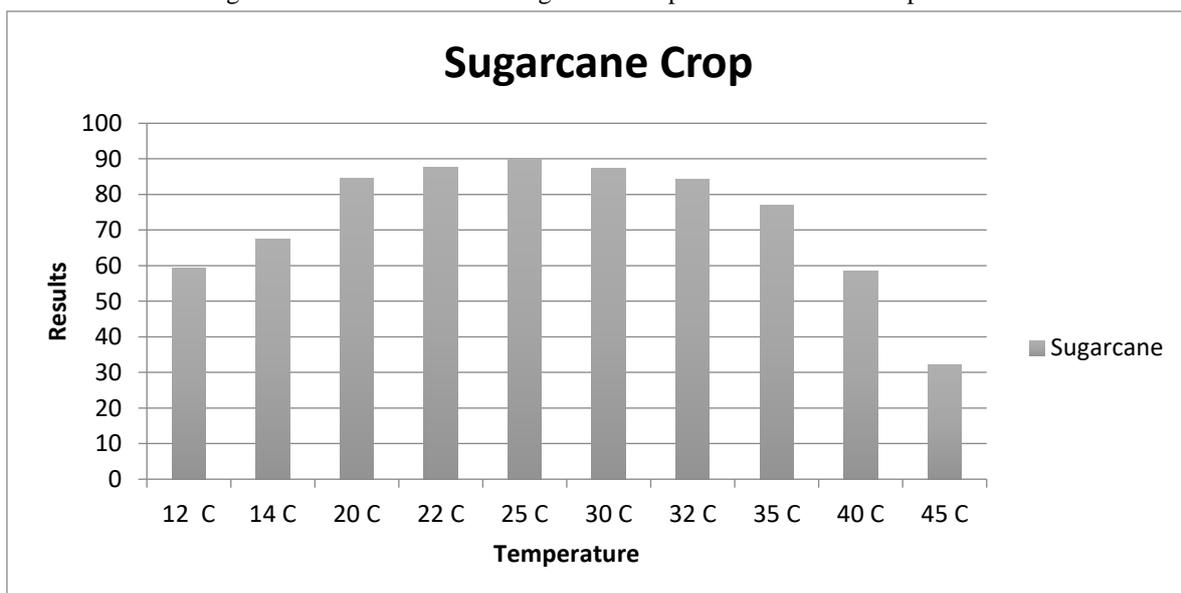
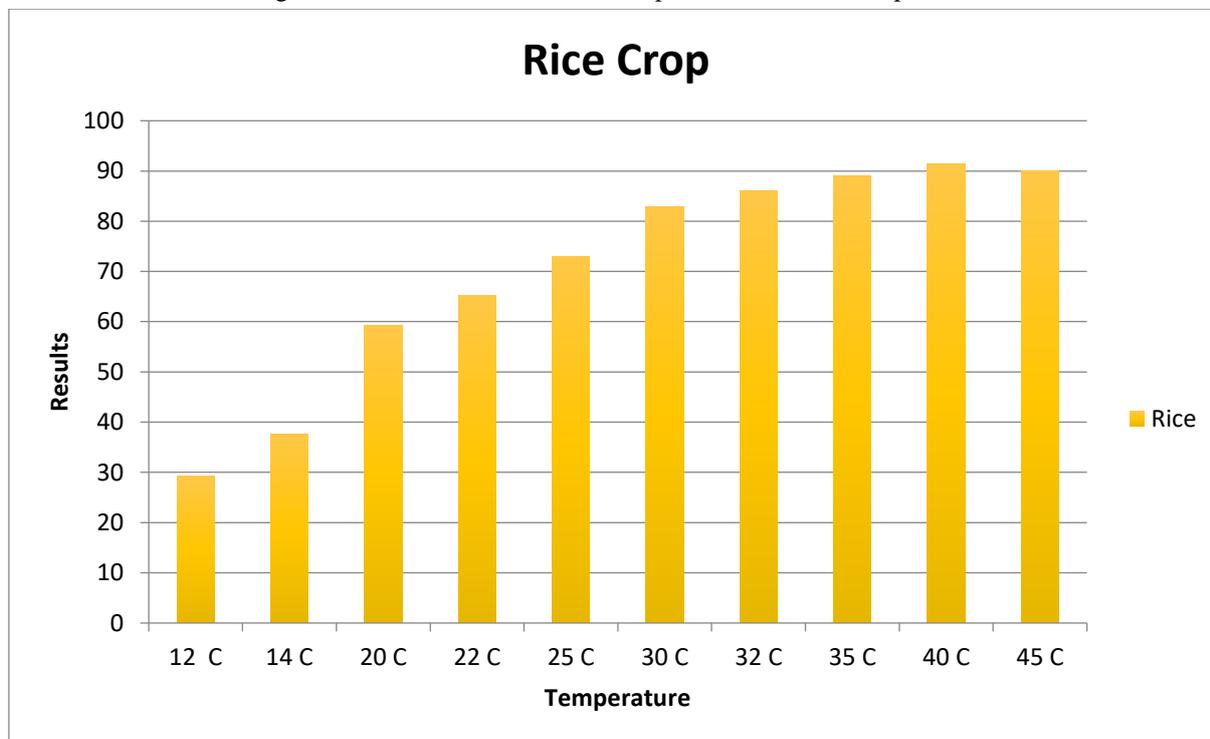


Fig. 6 discussed the rice yield on different temperature. It is shown in Fig. 6 that the maximum wheat crop production i.e. more than 80% will be achieved at range of temperature from 30°C to 45°C and as the temperature is recued to 12°C the production is affected drastically.

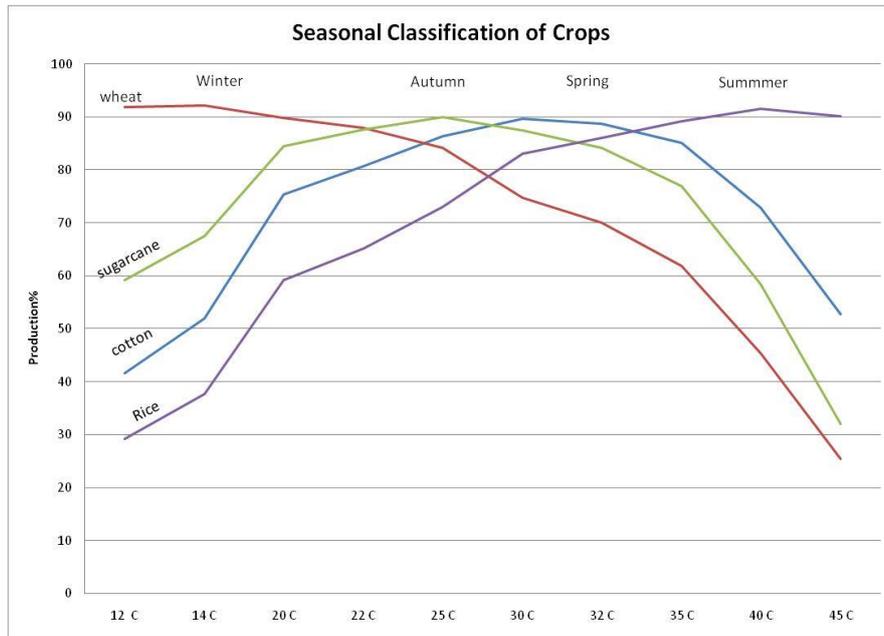
Figure 6 - Relation between Rice Crop Production and Temperature



The proposed system provides the solution for selecting the favorable the crop based on Temperature. The values are given as input to the system and system analyzed the data gives results of crops as output. The system suggested crop grown efficiently and farmers get the maximum production and profit. Our system also gives the knowledge of seasonal classification of different crops, that the crops grow efficiently and healthy in which environment and soil. Our algorithm has an average accuracy of 90% on the given dataset.

Fig. 7 is the outcome of our proposed algorithm that using our proposed system, the farmers may get the information that in which month based on temperate the crops can be cultivated so that can be given maximum production. It can be seen in Fig. 7 that the wheat crop may yield maximum production in the late winter season possess in Pakistan. The sugarcane may yield the maximum production in the season of autumn.

Figure 7 - Seasonal Classification of Crops



Furthermore, Fig. 7 demonstrates that in order to get optimum Cotton output, the month of spring is most feasible as per season of Pakistan. For the rice production, the season of summer is appropriate to cultivate in order to achieve the maximum production. In the next, the accuracy of all crop using proposed algorithm is depicted using Fig. 8.

Figure 8 - Comparison between Predict and Test Values of Major Four Crops

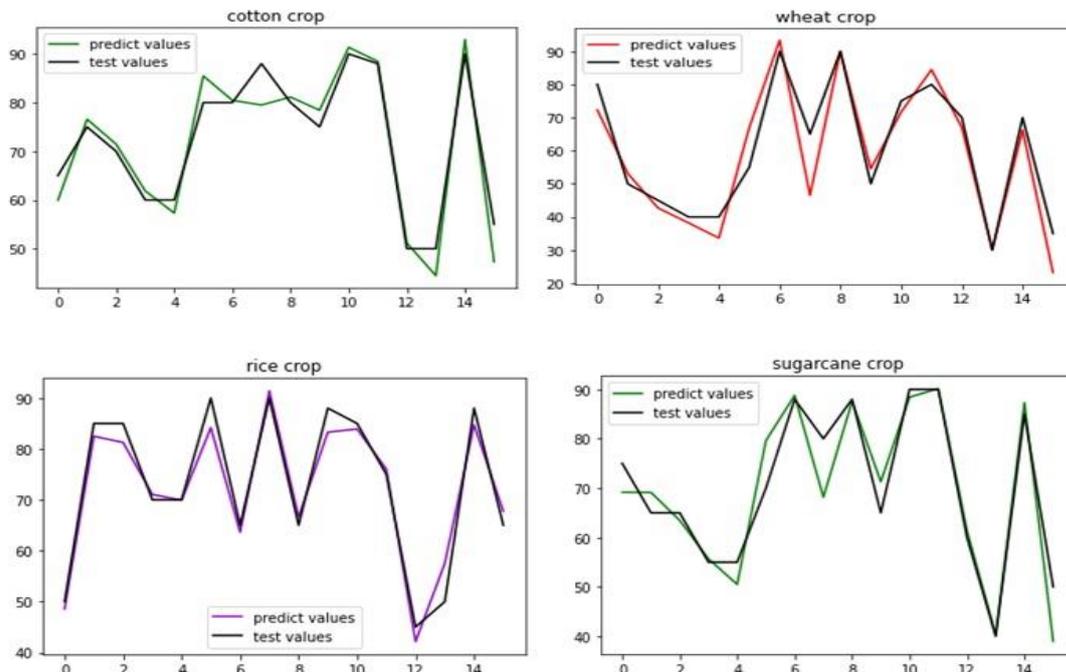


Fig. 8 demonstrate the accuracy of proposed machine learning algorithm for crop suggestion. It has been analyzed that for all the crops, Rice, Cotton, Sugarcane, and Wheat the test values and predict value almost follow same pattern and based on that analysis of the algorithm the 90% accuracy is attained using the proposed system for crop suggestion for Pakistani famers in order to get the maximum accurate information for enhancing the crop yield.

## 5. Conclusion and Future Work

Agriculture assume an indispensable part in endurance for everybody. Farmers confront several challenges as a result of a variety of uncontrollable factors. As a result, we employ prediction models to overcome the unpredictability of favorable crop or other agriculture-related problems. To forecast the favourable crop, the regression model is employed as a prediction tool. Thus, in this research, linear regression analysis is utilised to create a link between the above-mentioned independent factors and their impacts on crop output, with the goal of increasing agricultural productivity by utilizing the model's favourable crop forecast. The dependence of crop production on temperature is also examined in this paper, and a crop prediction is made based on the results. The crop recommendation system is capable of providing a lasting solution to the problem encountered by farmers.

There are some aspects of this study that require additional investigation in order to learn more about crop recommendation prediction using regression analysis. Rainfall, minerals accessible in soil such as potassium, nitrogen, and phosphorous, rainfall, and crop market value may all be included to the input parameters, making the model forecast more accurate and lucrative. Other approaches, such as fuzzy logic and neural networks, might be employed in future studies.

## Acknowledgement

The authors would like to thanks the Directorate of Postgraduate Studies and Department of Electronic Engineering, Quaid-e-Awam University of Engineering, Science and Technology, Nawabshah, Sindh, Pakistan for their support in conducting this research.

## References

Naveen Kumar PR, Manikanta KB, Venkatesh BY, Naveen Kumar R, Amith Mali Patil, “Smart Agricultural Crop Prediction Using Machine Learning”, *Journal of Xi'an University of Architecture & Technology* Volume XII, Issue V, 2020, ISSN No: 1006-7930.

Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese, Er. Jinu Thomas, “Crop Prediction Using Machine Learning”, *International Journal of Future Generation Communication and Networking*, Volume 13, No. 3, (2020), pp. 1896-1901, ISSN No: 2233-7857 IJFGCN.

Rohit Kumar Rajak, Ankit Pawar, Mitalee Pendke, Pooja Shinde, Suresh Rathod, Avinash Devare, “Crop Recommendation System to Maximize Crop Yield using Machine Learning Technique”, *International Research Journal of Engineering and Technology*, Volume 04, Issue 12, Dec-2017, e-ISSN: 2395-0056, p-ISSN: 2395-0072.

Shashikala, Bikash Kumar Shah, Sundeep Thapa Chettri, “Rain Prediction using Polynomial Regression for the Field of Agriculture Prediction for Karnatakka”, *International Journal of Advances in Engineering and Management*, Volume 2, Issue 3, pp: 62-66, ISSN: 2395-5252.

Devdatta A. Bondre, Mr. Santosh Mahagaonkar, “Prediction of Crop Yield and Fertilizer Recommendation using Machine Learning Algorithms”, Volume 4, Issue 5, pages 371-376, Sept-2019, ISSN: 2455-2143.

P. Surya, Dr. I.Laurence Aroquiaraj, “Crop Yield Prediction in Agriculture using Data Mining Predictive Analytic Techniques”, *International Journal of Research and Analytical Reviews*, Volume 5, Issue 4, 2018, E-ISSN: 2348-1269, P-ISSN: 2349-5139.

Mannem Ganesh Reddy, Goli Vineeth, Chapala Rithik Reddy, Dr. P. Indira Priyadarsini, “Prediction of Suitable Crop Using Machine Learning”, *International Journal of Scientific Development and Research*, Volume 5, Issue 5, May-2020, ISSN: 2455-2631.

Anam Azam, Muhammad Shafique, “Agriculture in Pakistan and its Impact on Economy – A Review”, *International Journal of Advanced Science and Technology*, Volume 103, 2017, pp. 47-60, ISSN: 2005-4238 IJAST.

Waqar Ahmed Pahore, Abdul Saboor Soomro, Nisar Ahmed Pahore, “Climate Change Impacts on Soil Resources and Crop Productivity: A case study of District Jacobabad, Sindh Pakistan”, *International Journal of Scientific & Technology Research*, Volume 5, Issue 10, Oct-2016, ISSN: 2277-8616.

Abdul Rehman, Luan Jingdon, Babar Shahzad, Abbas Ali Chandio, Imran Hussain, Ghulam Nabi, Muhammad Shahid Iqbal, “Economic perspectives of major field crops of Pakistan: An empirical study”, *Journal Pacific Science Review B: Humanities and Social Sciences*, Volume 1, Issue 3, 2015, pages 145-158.

Nishit Jain, Amit Kumar, Sahil Garud, Vishal Pradhan, Prajakta Kulkarni, “Crop Selection Method Based on Various Environmental Factors using Machine Learning”, *International Research Journal of Engineering and Technology*, Volume 4, Issue 2, Feb-2017, e-ISSN: 2395-0056, p-ISSN: 2395-0072.

Shubham Gupta, Vishal Bharti, Anil Kumar, “A Survey on various Machine Learning Algorithms for Disease Prediction”, *International Journal of Recent Technology and Engineering IJRTE*, Volume 7, Issue 6C, April 2019, ISSN: 2277-3878.

Sangeeta, Shruthi G, “Design and Implementation of Crop Yield Prediction Model in Agriculture”, *International Journal of Scientific & Technology Research*, Volume 8, Issue 01, January 2020, ISSN: 2277-8616.

Ayon Dey, “Machine Learning Algorithms: A Review”, *International Journal of Computer Science and Information Technologies*, Volume 7(3), 2016, ISSN: 1174-1179.

Osisanwo F.Y, Akinsola J.E.T, Awodele O, Hinmikaiye J.O, Olakanmi O, Akinjobi J, “Supervised Machine Learning Algorithms: Classification and Comparison”, *International Journal of Computer Trends and Technology*, volume 45 Number 3, June 2017, ISSN: 2231-2803.