Predicting Agricultural Productivity through Machine Learning

Submitted by:

Shiyam Talukder	**16101243**
Habiba Jannat	**16101191**
Sukanta Saha	**15301104**
Katha Sengupta	**16101280**

Supervised by:

Dr. Muhammad Iqbal Hossain

Assistant professor

Department of computer science and engineering

Abstract

Bangladesh is an agricultural country. Its economy mostly depends and influenced by agricultural products. In Bangladesh, its agricultural productivity highly influenced by the nature like rain water and the soil condition and they are highly unpredictable. Like rainwater there are like so many other parameters which are also plays a very important role in agricultural productivity like soil parameter namely nitrogen, phosphorus, crop rotation, soil moisture, surface temperature and also humidity of the air. In recent time Bangladesh progress in agricultural field is noticeable. The government initiates so many steps in agricultural department. Now a days Bangladesh also progressing towards technical development. We can use our technical knowledge in this agricultural field to boost up the productivity so that we can take a farther step to full fill the demand of agricultural product. By technological support we can rapidly increase crop productivity resulting in better yields to the farmers. The proposed project can be a solution for smart agriculture by monitoring the agricultural field that can assist the farmers increasing the productivity to a great extent. We can get weather forecast data from BMD (Bangladesh Meteorological Department) such as temperature and rainfall and soil parameters repository gives insight into which crops are suitable cultivating in the particular area. The proposed system will integrate the data and by applying machine learning algorithm: Multiple Linear Regression and Neural Network, a prediction can be made for suitable crop in the current weather condition. This can be beneficial to the farmers by having variety of options of crops for cultivation.

.

1. Introduction

Agriculture is the largest employment sector in Bangladesh. It is agriculture that has a great and noticeable impact on its economy. About 63 percent of population in this country lives on agriculture and 19.6 percent of national GDP comes from agricultural sectors [1]. According to the World Bank, about 70.63 percent of land are being used for agriculture and agricultural irrigated land is 52.62 percent in 2015[2] in order to satisfy the need of about 164.7 million people which is not sufficient. On the other hand population is increasing day by day but agricultural field are getting decreased. Thus, a new approach must be introduced for the proper utilization of productivity. Agricultural method must be modernized.

Bangladesh most of the farmers depend on their previous experience to cultivate their land. As they are dependent highly on climate for taking decision for agricultural production so if there is change in climate pattern they do not know which crop to produce. Therefore, they cannot produce the desired the maximum amount of crop that could be produced in that area. To solve this problem we are going to develop a system which will take weather, soil quality, pH level of the soil and water and previous production into consideration and use various machine learning technique like KNN, MULTI-LINEAR REGRESSION, LOGISTIC REGRESSION, and DECISION TREE to predict the future production. This system will propose the best feasible crops according to given conditions. As previous year production is also taken into account, the prediction will be more accurate. Thus, this system will suggest profitable crops providing a choice directly to the farmer.

2. Literature Review

In [1], researcher used various techniques like MLR, logistic regression, exponential smoothing models, Markov chain model for forecasting crops per count, forewarning low or high crop yields. Here, they takes plant characters such as height, number of ear heads and effect of weather on crops, past values of production into consideration for forecasting crop production. Researchers developed a system where they put the location, soil attributes, Weather attribute, pH level of the water and soil. They also take the previous year production value to predict which crops will be produced better in that area. Using this algorithm they can predict what crop will be suitable for that Area. However, their System cannot predict the numerical amount of the production and the individual effect of the soil attributes, weather and water and sudden variation of attributes.

In [2], Authors, use techniques like K-Nearest Neighbor, Support Vector Machine, and Least Squared Support Vector Machine for providing comparative study of various algorithm. In

this paper they apply these algorithms on datasets and it shows the accuracy of each algorithm to train the datasets and also mean squared error at the cross-validation phase which shows the rate of production and these models can handle structural accident minimization. They predict the future production in categorized way. However, it requires a huge amount data for finding variance and providing a good result.

In [3], Authors propose an android base application which predicts most suitable crops according to current environmental situation by using multi-linear regression on the weather, soil and past production data. This app uses the location of a place and recommends the best suitable crops from the past data collected from the weather station. According to their model they try to predict the production from soil and rainfall using regression. Then try to incorporate previous production using a multi-linear regression for better accuracy. However, as they uses only regression and multi-linear regression they cannot compare with other algorithm to check if they can improve their accuracy.

In [4], S. S. Dahikar and S. V. Rode implemented the ANN algorithm to determine the suitable crop for the specific soil type. As the soil have a lot of parameter such as pH, nitrogen, temperature and the effect of the climate on the soil, they have predicted one particular crop to be yielded in one particular type of soil depending on the parameters. Furthermore, they have claimed that Artificial Neural Networks (ANNs) works better on this type problem than statistical models as complex neural systems with many inputs has shown better accuracy.

In [5], Mr. V. Lamba and Dr. V. S. Dhaka discussed the techniques of wheat yield prediction comparing Artificial Neural Network (ANN) with other models such as Multiple Linear Regression (MLR), Logistic regression, Time series etc. With their research they have claimed the accuracy and efficiency of ANN better than others. They have used different types of models on different aspects. For Multiple Linear regression and logistic regression model they predicted crop yields through plant characters and pest count. Moreover, amongst the Time Series Models they used exponential smoothing models and auto-regressive integrated moving average models for predicting the area/production of the crops. Finally, they have implemented the probabilistic model Markov chain to forecast the techniques in agriculture. In this way with the comparison amongst the techniques they have predicted why Artificial Neural Network (ANN) is the best.

In [6], authors have compared different AI models for the best prediction model for the Midwestern US. Input dataset was from satellite-based vegetation indices and meteorological and hydrological variables. They have examined the effect of phenology in three periods, and one database was selected as the best months to predict crop and soybean yields. Using the DNN model they have performed an optimization process for the accuracy and they have

found it has outperformed rest five of them with prediction error around 7.6% for corn and 7.8% for soybean which is much less than other methods.

3. Work Plan and Our Proposed Approach

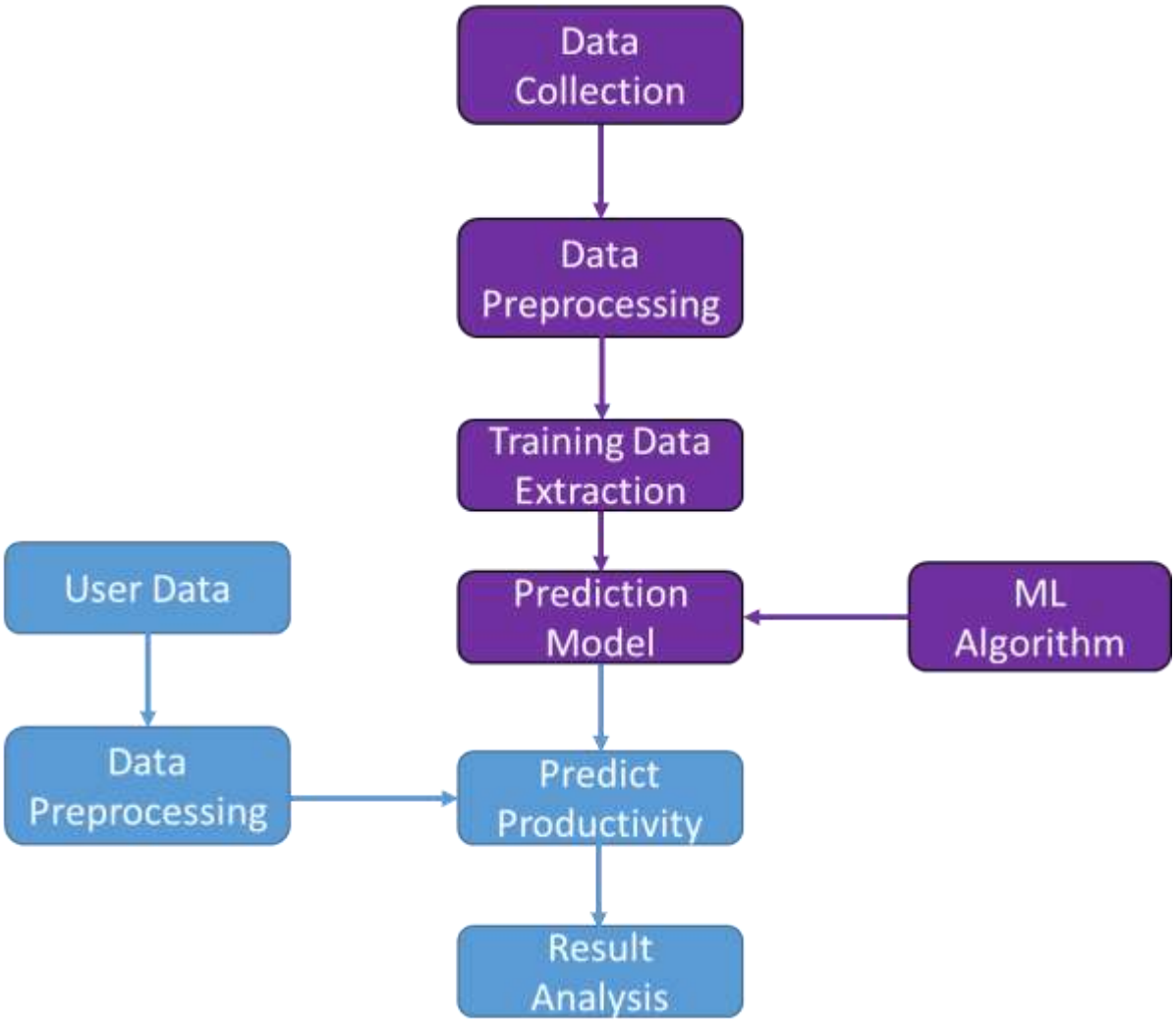

Figure: Workflow of Our Proposed Approach

Data is collected from an Agricultural Institute and many available data sets are collected for the analysis and result of the agricultural productivity. Data set is pre-processed which means the actual data (raw) collected will be converted to the theoretical data set that can be easily understand to us. It is because raw data has many errors or lacking in it. In other word, it is called Data Mining. After that, the data is extracted to a few numbers of sets so that it can be easily

used for the purpose of analysis. The extracted data is send to training and on the other hand different Machine Learning Algorithm are being tested with these trained extracted data. After being tested, Algorithm with the best result is selected to make a model for our system to predict the productivity. Meanwhile, in our system users will input their data and these data would be pre-processed so that it separates the incomplete raw data and our model system would take these pre-processed data to predict their productivity whether it is suitable to the environmental conditions or not and how will be the productivity increase. After all these, overall result will be analyzed and tested to see how our system working.

References:

[1] D.S. Zingade et al., "Crop Prediction System using Machine Learning," International Journal of Advance Engineering and Research Development, vol. 4, 2017.

[2] A. Kumar, N. Kumar and V. Vats, "EFFICIENT CROP YIELD PREDICTION USING MACHINE LEARNING ALGORITHMS," International Research Journal of Engineering and Technology(IRJET), vol. 5, no. 6, June 2018.

[3] D.S. Zingade et al., "Machine Learning based Crop Prediction System Using Multi-Linear Regression," International Journal Of Emerging Technology and Computer Science, vol.3, no. 2, April 2018.

[4] S. S. Dahikar and S. V. Rode, "Agricultural Crop Yield Prediction Using Artificial Neural Network Approach", International Journal Of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, Vol. 2, no. 2, January 2014.

[5] V. Lamba and V. S. Dhaka, "Wheat Yield Prediction Using Artificial Neural Networkand Crop Prediction Techniques (A Survey)," International Journal For Research In Applied Science and Engineering Technology(IJRASET), Vol. 2, no. 9, September 2014

[6] N. Kim, K. J. Ha, N. W. Park, J. Cho, S. Hong and Y.W. Lee, "A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction: Case Study of the Midwestern United States, 2006–2015," ISPRS International Journal of Geo-Information, vol. 8, no. 5, p. 240, 2019.