

Agricultural Crop Yield Prediction Using Deep Learning Approaches

SVM, Multiple Regression, Random forest Regression

Rohan Yadav¹, Satyam Yadav², Neha Gunjal³, Shilpa Mandal⁴

^{1, 2, 3, 4}Dept. of Computer Engineering, SKN Sinhgad Institute of Technology and Science, Lonavala, Maharashtra, India

Abstract—Manual agricultural systems are terribly advanced and agitated to use as a result of it deals with an enormous dataset and large quantity of knowledge process. several techniques and approaches are made to predict the crop yield. Deep learning is employed to predict such styles of advanced system with large quantity of dataset in it. During this survey paper a review on the utilization of the substitute Deep learning for the prediction of the crop yield is projected. This paper contains implementation of Deep learning approaches. The utilization of parameters like, Rainfall, Temperature, Crop produced in previous year is employed. Deep learning approaches are employed for prediction of the crop yield.

This paper shows the implementation of this kind of the system techniques such as multiple linear regression, Support Vector regression (SVM algorithm), Random forest regression.

Key Words: Temperature, Rainfall, Deep learning approaches: Multiple linear regression, Support vector regression, Random forest regression,

1. INTRODUCTION

To get most crop yield at minimum value is one among the first goals in agriculture. Detecting and handling troubles related with crop yield indicators in early stages of the agricultural field can give benefits in expanded yield and elevated earnings too. By reading weather styles of a specific location, massive-scale meteorological phenomena will have a completely green impact on agricultural production. The crop yield predictions can be utilized by farmers to reduce losses when negative conditions may occur. Also, predictions may be used to maximize crop prediction while there is favourable situation for farming.

Prediction of crop yield plants like jowar, bajra, wheat, rice, mung, etc. has always been an absorbing research area to agro meteorologists, because this researches are important in economics of a nation. It is an intelligent system which can predict the more accurate prediction using meteorological data. Till date, there are many yield prediction models and application, they are classified in two groups: a) Statistical Models, b) Crop Simulation Models (e.g. CERES). Lately, Artificial Intelligence (AI) based applications like Artificial

Neural Networks (ANNs) and Genetic Algorithm are more efficient in dissolving the problem than the traditional methods. The application of Artificial intelligence can make working models easier and more accurate from complex manual systems with many inputs. In this system the focus is on development of various crop yield prediction model using deep learning approaches. If an efficient AI based effective climatic factor based Crop yield predictions are done a farmer can use it very efficiently. In addition, using Artificial neural networks, a user can find the most effective factors on crop yield. To handle such a situation, a particularly versatile approach of “Deep learning” is developing rapidly.

Nowadays a part of Artificial intelligence known as deep learning is used widely in order to get more accurate results. This system does implement deep learning approaches like Multiple linear regression, random forest regression and Support vector regression to get the more accurate results. These methods have been applied for modelling and forecasting of various crop yield on the basis of various predictor variables, viz. type of Temperature, Rainfall. The proposed system also provides visualisation for better understanding and knowledge for the Client. The system also produces various pictorial representation like pie chart, bar graph and map visualisation.

2. LITERATURE SURVEY

Multiple Linear Regression has been used extensively in both agricultural farming and other areas. For example, Vinciya and A. Valarmathi performed an analysis to classify between organic farming, inorganic farming and real estate data to make a prediction of the crop yield based on the data set [1]. In this research they used Multiple Linear Regression (MLR) for the selected region. In multiple linear regressions, there was an equation for prediction and the estimation of the residuals to find the difference between actual values and the fitted values. The use of the variance was to calculate the mean-squared error.

[2] B. Ji et al [11] developed agricultural management need accurate and simple estimation techniques to predict rice yields in the planning process. The intention of this study where to: 1. observe whether artificial neural network (ANN) models could effectively predict Fujian rice yield for typical climate condition of the mountain region, 2. evaluate ANN model performance relative to variations of development parameters and 3. compare the effectiveness of multiple rectilinear regression models with ANN models. In this thesis we have described the development of deep neural network model as an alternate and more accurate technique for yielding prediction.

[3] T. Sidique et al [12] introduce a system that is focused on the climate and geographical condition of different areas of Bangladesh. It predicts cost effective crop using a prediction based algorithm. The algorithm are aimed to use is multiple linear regression with the association of some independent variable i.e. rainfall, average maximum temperature and average minimum temperature of certain location and provides prediction supported yield rate per unit area.

[4] *Bell Labs and Monmouth University Department of Electronic Engineering West Long Branch. NJ 07764
**Bell Labs + AT&T Labs: A new regression technique based on Vapnik's concept of support vectors is introduced. We compare support vector regression (SVR) with a committee regression technique (bagging) supported regression trees and ridge regression wiped out feature space. On the idea of those experiments, it's expected that SVR will have advantages in high dimensionality space because SVR optimization doesn't depend upon the dimensionality of the input space.

3. PROPOSED SYSTEM

The system prepared predicts about the 10 major crops yield in a particular district in Maharashtra. The client on their first login has to register themselves. Once the user logs into the system he gets all the access for predicting crop yield and using it for betterment.

This system in the backend uses various deep learning algorithms that helps the system give more accurate answers. The proposed system is based on following deep learning algorithms:

(1) Multiple regression-

Multiple regression is used when we want to predict the value of a variable based on the value of two or more other variables. The variable we would like to predict is named the variable (or sometimes, the result, target or criterion variable). Here, in this case the dependent variables is the crop which is to be predicted and the independent variables on the basis of which crops are predicted are temperature and rainfall.

(2) Random forest regression:

Random forest regression is an ensemble machine learning technique. Ensemble methods are supervised learning models which combine the predictions of multiple smaller models to improve predictive power and generalization.

(3) Support Vector regression:

Support Vector Regression (SVR) uses the Support Vector Machine (SVM) algorithm to predict a continuous variable. Support Vector Regression tries to fit the best line within a predefined or threshold error value.

Algorithm details:

(1) Multiple regression algorithm-

Step 1: Data Pre Processing

a. Importing The Libraries.:

b. Importing the Data Set.

c. Encoding the Categorical Data.

d. Avoiding the Dummy Variable Trap.

e. Splitting the Data set into Training Set and Test Set.

Step 2: Fitting Multiple rectilinear regression to the Training set

Step 3: Predicting the Test set results.

The Formula for Multiple Linear Regression Is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \epsilon$$

where, for $i=n$ observations:

y_i =dependent variable

x_i = explanatory variables

β_0 =y-intercept (constant term)

β_p =slope coefficients for each explanatory variable

ϵ =the model's error term (also known as the residuals)

(2)Random forest regression algorithm-

Step 1 : Import the required libraries.

Step 2 : Import and print the dataset.

Step 3 : Select all rows and column 1 from dataset to x and all rows to y.

Step 4 : Fit Random forest regressor to the dataset.

Step 5 : Predicting a new result.

Step 6 : Visualising the result.

(RF) construct many individual decision trees during training. Predictions from all trees are pooled to form the ultimate prediction; the mode of the classes for classification or the mean prediction for regression. RF calculates a nodes importance using Gini Importance, assuming only two child nodes (binary tree): Random forests (RF) construct many individual decision trees at training. Predictions from all trees are pooled to make the final prediction; the mode of the classes for classification or the mean prediction for regression. RF calculates a nodes importance using Gini Importance, assuming only two child nodes (binary tree):

$$ni_j = w_j C_j - w_{left(j)} C_{left(j)} - w_{right(j)} C_{right(j)}$$

$ni_{sub(j)}$ = the importance of node j

$weight_{sub(j)}$ = weighted number of samples reaching node j

$C_{sub(j)}$ = the impurity value of node j

$left(j)$ = child node from left split on node j

$right(j)$ = child node from right split on node j

(3)Support Vector regression algorithm-

Step1 :Importing the libraries.

Step2: Importing the dataset.

Step3: Feature Scaling.

Step4: Fitting the Support Vector Regression Model to the dataset.

Step5: Predicting a new result.

Step6: Visualising the Support Vector Regression results.

In contrast to OLS, the objective function of SVR is to minimize the coefficients — more specifically, the l_2 -norm of the coefficient vector — not the squared error. The error term is instead handled in the constraints, where we set the absolute error less than or equal to a specified margin, called the maximum error, ϵ (epsilon). We can tune epsilon to gain the desired accuracy of our model. Our new objective function and constraints are as follows:

Minimize:

$$MIN \frac{1}{2} ||w||^2$$

Constraints:

$$|y_i - w_i x_i| \leq \epsilon$$

4.RESULTS AND SCREENSHOTS:

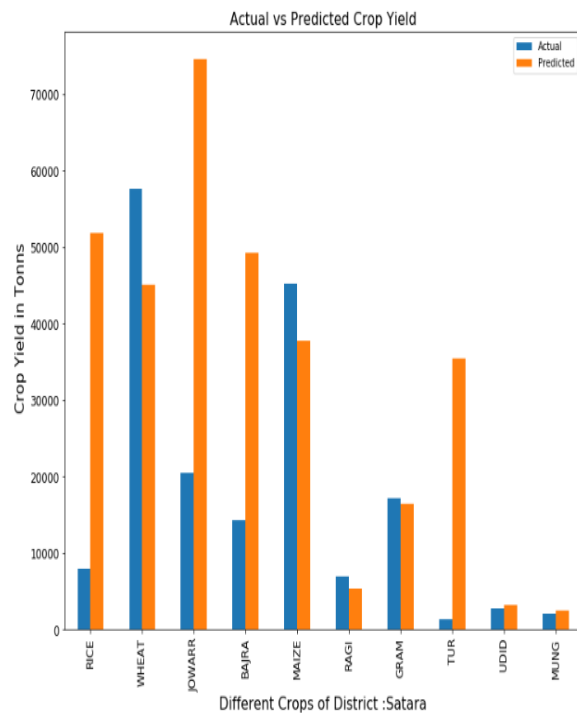


Figure 1:Actual vs Predicted crop yield

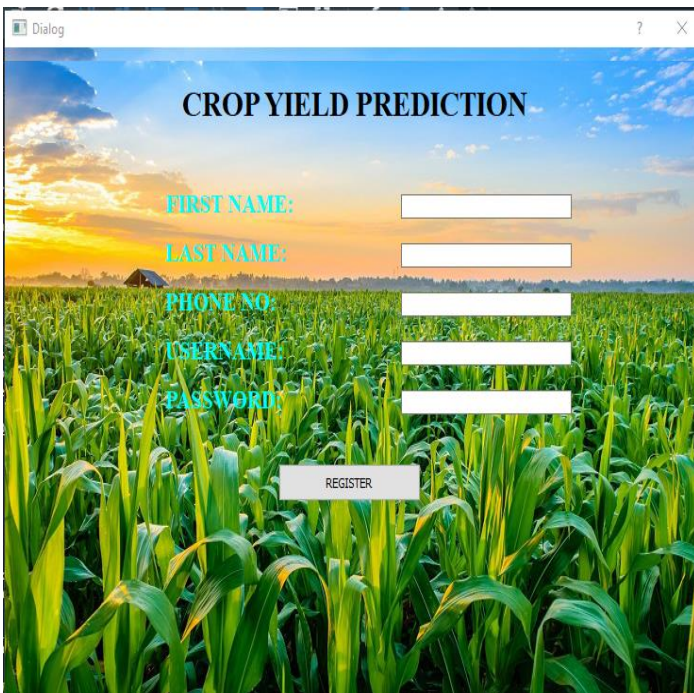


Figure 6: Signup page

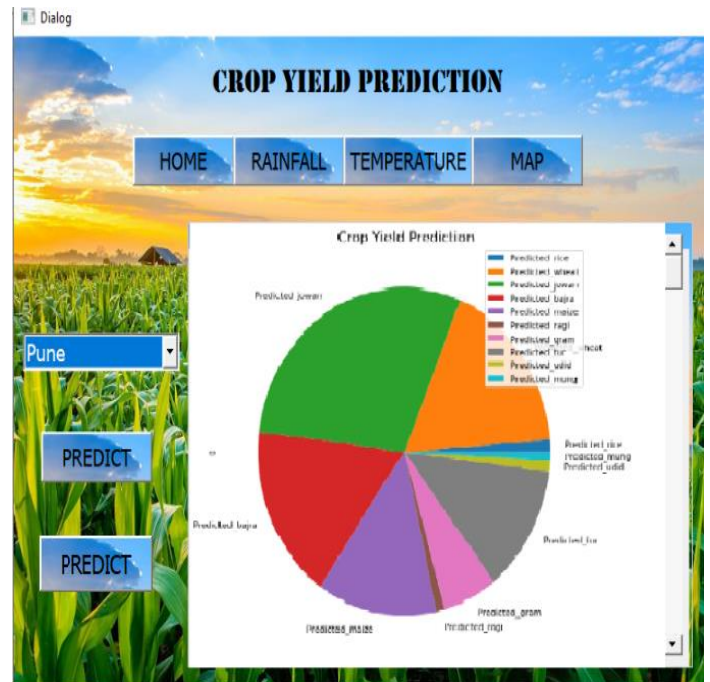


Figure 8: Output in pie representation

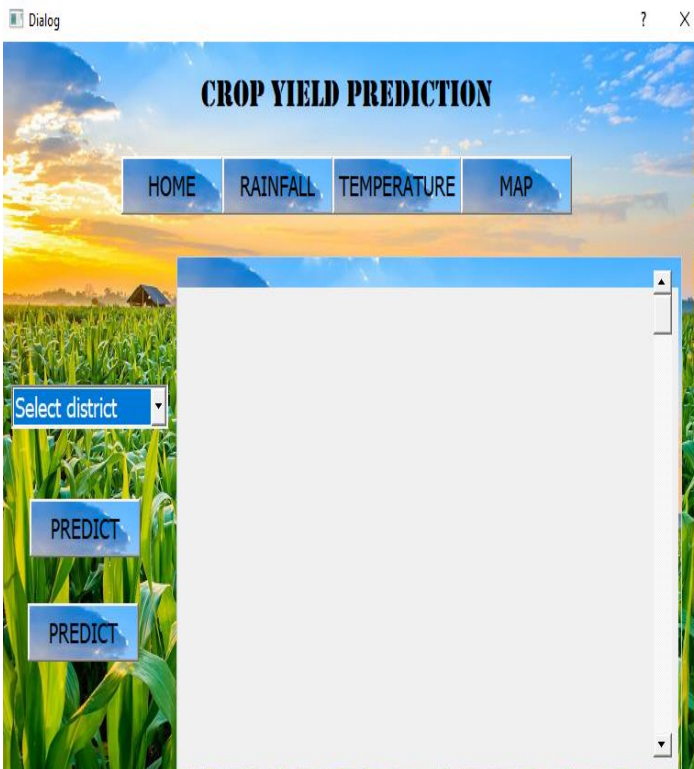


Figure 7: Output page

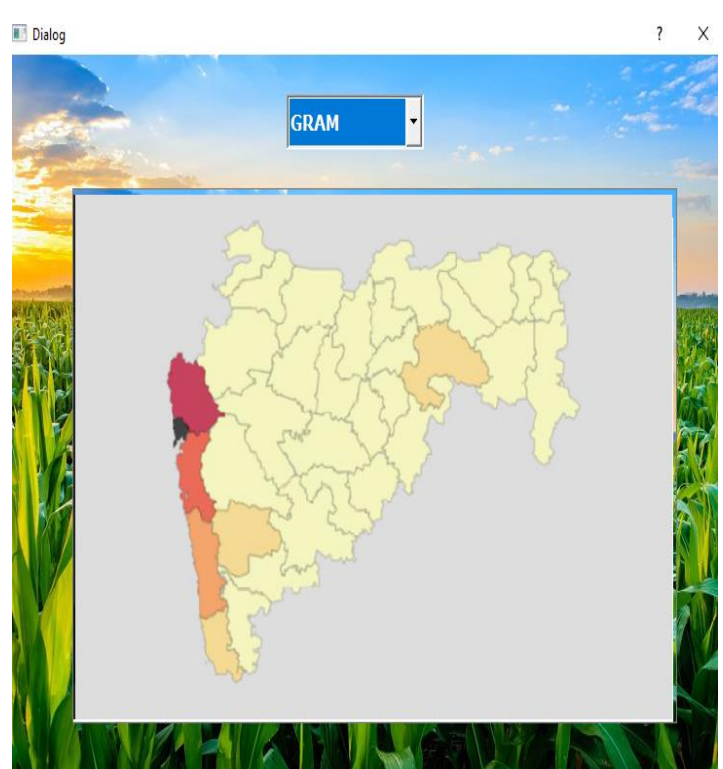


Figure 9: Map representation of crop yield

5.CONCLUSION

We have successfully developed a system that would help in the agricultural sector by using the 21st century's modern Deep learning Approaches. Our study tried to give a prediction on how a simple machine learning algorithm can change our country's agriculture image. Being dependent on the agricultural side for a long time our country has not meet much between agriculture and technology so far. Right now, the people of our generation are in a position where everyone is in touch with modern things. So it is high time we should aim at a future to live a better lifestyle. Our government has already taken so many good initiatives in the agricultural sector. It is high time to precede digitally in this sector so that; not only the government but also stockholder and society might get benefitted out from it. Our one little step are going to be enough to introduce digital agriculture system for best crop selection and yield prediction.

Deep learning is a beneficial tool for crop prediction.

5.REFERENCES:

- [1] Nikhil,Dr.Kanwal, "Exploiting Data Mining Technique for Rainfall Prediction," International Journal of Computer Science and Information Technologies (IJCSIT),vol. 5, no. 3,pp. 3982-3984, 2014.
- [2] Manpreet, Heena, Harish, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications," International Journal of Computer Applications, vol. 99, no. 12, pp.1-3, August.2014.
- [3] Shastry, K., Sanjay, H. &Deshmukh, A. (2016). A Parameter Based Customized Artificial Neural Network Model for Crop Yield Prediction. Journal of Artificial Intelligence,9, 23-32 doi: 10.3923/jai.2016.23.32
- [4] Tanhim Islam, Tanjir Alam Chisty, BRAC University, A Deep Neural Network Approach for Crop Selection and Yield Prediction in Bangladesh(2018).
- [5] Talha Siddique, Dipro Barua, Zannatul Ferdous, Amitabha Chakrabarty Department of Computer Science and Engineering BRAC University, Automated Farming Prediction(2017).