ISSN No:-2456-2165

Crop Yield Prediction Analysis using Feed Forward and Recurrent Neural Network

Sivanandhini P
PG Scholar
Computer science and Engineering
PSG College of Technology

Prakash J
Assistant Professor
Computer science and Engineering
PSG College of Technology

Abstract:- **In recent days, the crop yield prediction is a major area of research, where the information about the suitable crop to cultivate will be very much useful for the farmers to cultivate. The crop yield prediction in agricultural helps the farmers to know how much yield they can expect from the cultivation. It also helps in minimizing the loss to the farmers when unfavorable condition occurs. The proposed work is to predict the yield of the crop based on the suitable crop parameters like TemperatureMin, TemperatureMax, Humidity, Windspeed, Pressure using neural network model. In this research paper, crop yields predictions were established using Feed Forward Neural Network and Recurrent Neural Network model which predict the crop yield. The performances of neural network models were evaluated using the metrics like Root Mean Square Error (RMSE) and Loss.**

Keywords:- *Crop yield prediction, Feed Forward Neural Network, Recurrent Neural Network, Root Mean Square Error.*

I. INTRODUCTION

Agricultural is the backbone of Indian economy. Every farmer is interested in knowing the yield of crop that he could expect during the harvest period, since agricultural is more reliable on weather condition, yield prediction is an important aspect for them. There are number of reasons for the loss of the crop, some of the most important factors is that the crops, fertilizers and crop productivity not properly known to the farmers. The crop yield with forecast of climate change will be more helpful for the farmers in cultivation. The major challenge of crop prediction is climatic change, weather decided the crops yield. When rainfall or humidity is weak, it is sometimes difficult to predict the crops on time and resolving food security problems to predict crop yields before harvest the crops. The yield prediction could be more beneficial and extremely challenging factor in the crop productivity. Thus, crop yield prediction needed to analyze the varies climatic parameters, to estimate the production of crop yields profits or loss. The main goal of crop yield prediction is achieving high agricultural crop productivity. The Evolution of many prediction model, make the techniques improve and increases their effectiveness. The yield prediction is achieved with the flexible methodology of artificial neural networks (ANNs) using feed forward and recurrent neural network.

II. RELATED WORK

Shruti Kulkarni in the paper Predictive Analysis to improve Crop Yield using a Neural Network Model discussed about the rice yield variability from season to season is harmful to the farmer's profits and livelihoods. Improving farmers ability to predict crop productivity under different climatic conditions will help farmers and other stakeholders make important agronomic and crop selection decisions. Neural network can be used to predict rice production and investigate factors affecting the production of rice crops for different districts of Maharashtra State India. The parameters considered were precipitation, minimum temperature, average temperature, maximum temperature and reference crop evapotranspiration, area, production and yield for the Kharif season (June to November) for the years 1998 to 2002. The dataset was processed using WEKA tool. Author used Artificial Neural Networks to forecast rice crop yields for the state of Maharashtra, India and evaluate the efficiency of the Artificial Neural network [1].

Petteri Nevavuorib in the paper Crop yield predictive with deep convolution Neural Networks discussed about improving crop yield prediction model that will be useful in agriculture, which is important sector of since it feeds the population of the country and contributes to GDP. Agricultural combination of some factors, including soil properties, climate, elevation and irrigation techniques, determines the crop yield. The agricultural basis for the model of predictive analysis about the rainfall is based on the Supervised Learning method of the Time-Series. The technology used for the final crop yield prediction is branch of Machine Learning called Recurrent Neural Networks. The author proposed an architecture that provides computational element to improve information prior to the crop sowing time regarding yield. It is made possible by a hybrid model which is driven by data.[2].

NariKim in the paper A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction discussed about evaluating machine learning methods for output prediction based on remote sensing which requires the availability of yield mapping tools that are still not very popular among farmers and increasing research into data-producing devices and sensors has been an ongoing major trend in agriculture, allowing farmers to shift towards data-driven decision-making. In this study Convolutional Neural Networks (CNNs) – a deep learning methodology showing

outstanding performance in tasks of image classification. The researchers were able to derive the field variability steps to be taken to secure and also increased the level of crop yields. This is commonly called smart farming techniques. Satellite data are predefined wavelength bands from both the spectral regions visible and Near Infrared (NIR).The spectral bands are designed to measure appropriate indices, such as the Normalized Vegetation Difference Index (NDVI).Unmanned Aerial Vehicles (UAVs), or drones, uses the better spatial resolution for data acquisition, the user can choose the data acquisition time, and the data can also be collected in cloudy conditions. The drawback is that the UAV must be operated locally, and it requires highly specialized skills to manage data and extract relevant information. The main issue with the satellite data is that no useful data will be collected if there is a cloud cover during the satellite's overflight. Remote sensing data can be used to extract information important for decision making in agriculture through machine learning. There are different tasks can be addressed depending on the characteristics, such as crop classification, weed identification or yield prediction. The function of crop yield prediction was implemented in several machine learning techniques. But, these methods are commonly referred to as deep learning techniques [3].

Adithya Sharma in the paper Crop Yield Prediction using Convolution Neural Network discussed about the artificial intelligence (AI) models and compare with other models, in order to develop the best crop prediction model using Midwestern United States. The accurate crop yield estimates are important aspects such as agronomic problems, including farm management, national food policies and foreign crop trade and used the optimization processes to ensure the best layer structure, cost function, optimizer, activation function, and drop-out ratio configurations, especially for the deep neural network (DNN) model. The mean absolute error (MAE) in DNN model with JA database was approximately 21–33 percent and 17–22 percent more accurate for maize and soyabean yields than the other five AI models, respectively.AI could be an alternative to statistical modeling , as it can address non-linearity and complexity problems effectively.AI includes machine learning models, such as the random forest (RF) and support vector machine (SVM), and neural network models, such as the artificial neural network (ANN) and deep neural network (DNN).RF produces a large number of decision trees with slightly different characteristics.ANN is a network model that can produce an approximation by optimizing a weight and bias set for a node-link structure consisting of input, hidden, and output layers.DNN can overcome the local minima problem found in traditional ANN approaches, as well as the problem of overfitting seen in traditional machine learning models. Because of DNN, which can optimize a deep network structure through the back-propagation algorithm, which combines forward and backward optimization processes, and through the activation function, which avoids the issue of loss-function gradients in the back-propagation process.[4].

III. DATASET DESCRIPTION

The data set used for this work is the dataset obtained from forecasting of crop yield prediction year 2013 for Maharashtra state of India. The dataset consists of different predictor variables including Humidity, Temperature Max, Temperature Min, Wind Speed and yield. The description of the dataset is described below.

Attributes	Description
Humidity	The air is totally saturated with water vapor.
Windspeed	The rate at which air is moving in a particular area.
Temperature Maximum	The largest number or highest number of something that is permissible/ possible.
Temperature Minimum	The lowest temperature recorded—diurnally, monthly, seasonally, or annually, or the lowest temperature of the entire record.
Pressure	The effect on the amount of water required by the crop to maintain optimal growth
Yield	Yield prediction is the outcome of agricultural output.

Table-1: Dataset Description

IV. CROP YIELD PREDICTION

The crop yield predication model uses the WHEAT 2013 dataset as mentioned above. The complete flow of crop yield prediction is shown in the flowchart below.

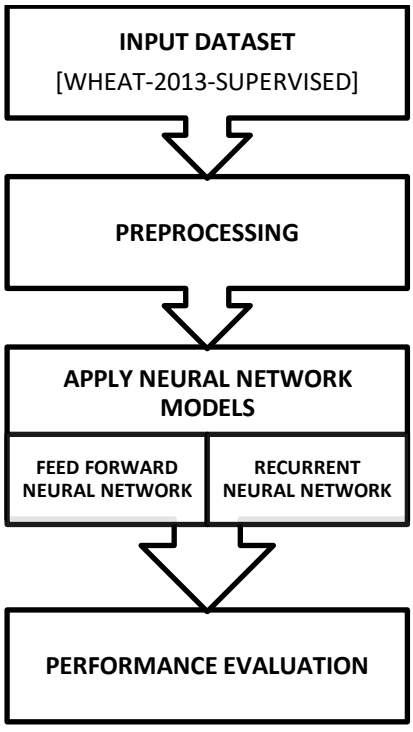

Fig- 1: Flow of Crop Yield Prediction

❖ *Preprocessing:*

The dataset contains missing values so these missing values need to be properly handled in order to apply models. The preprocessing technique of backward filling is used to check and remove the null values. Once the null values are removed it can be fed into the model to predict the crop yield.

A. Feed Forward Neural Network

➢ *Preprocess and Load Data :*

The working of neural network and we need to process it before feeding to the neural network then will also useful for virtualize the data. Preprocessing applied before feeding it to algorithm. Data preprocessing technique to remove the null values so Backfilling is used.

➢ *Splitting Data and Normalization :*

The dataset is processed and ready to feed in the neural network. Generally, it is better to split data into training and testing data. Training data is data on which train the neural network. Test data is used to check the trained network. Training dataset will have the 70% and test data samples data 30%. The python module will make two array values X and Y. X contains the features and Y contains the classes. Normalization is a technique used to change the values of the array of the scaled method, with using the ranges values (0,1). Because of using this check the difference of accuracies on this dataset. so if feed the unnormalized to the neural network, the gradients changes for every column. This data is totally new for the neural network model and if the neural network performs well on the dataset, it shows that there is no overfitting.

➢ *Define Model :*

The model needs to specify the number of hidden layers in the neural network and their size, the input and output size. sequential specifies to keras that creating model sequentially and the output of each layer add to the input for next layer. Dense is used to specify the fully connected layer. The Dense are input dimension which are5 and output dimension which are 1 and the activation function to be used relu function. the second layer defined the model to be sequential so keras will automatically consider the input dimension. finally, the output layer takes that the different classes of function like relu and output will be predicted based on the given dataset.

➢ *Loss and Optimizer :*

The Loss function according to given task and specify the optimizer to use with the learning rate which values of 0.3 and the Loss function is RMSE metrics to judge the performance of neural network. The optimizer is used to be Adam.

➢ *Fit Model :*

The model input data class which are used to be X->train and the labels are Y->train, the epoch iteration and batch size are also considered. It returns the history of model training and history consists of losses and after each epoch will be visualized. Here the dataset is very big size so have to be allocated the batch size number. The divides the data size are equal to batch number and then only uploaded the number of sample data in memory and processed. It will take a round a minute to train and after 50 epochs neural network are trained. so far the loss function RMSE value trained till around 0.234 value.

B. Recurrent Neural Network

➢ *Pre-processing and Data Splitting :*

Backfilling is used to handle the null values and a simple method is used to split the dataset by separating into the training datasets with 67% of the observations that can use to train our model, leaving the remaining 33% for testing the model.

➢ *Define Model :*

The network has a visible layer with 1 input, a hidden layer with 4 LSTM blocks or neurons, and an output layer that makes a single value prediction. The Long Short-Term Memory network, or LSTM network, is a recurrent neural network that is trained using Back propagation. Instead of neurons, LSTM networks have memory blocks that are connected through layers. A block has components that make it smarter than a classical neuron and a memory for recent sequences. A block contains gates that manage the block's state and output and operates upon an input sequence and each gate within a block uses the sigmoid activation units to control whether they are triggered or not, making the change of state and addition of information flowing through the block conditional. Once the model is fit, we can estimate the performance of the model on the train and test datasets. This will give us a point of comparison for new models. The network is trained for 50 epochs and a batch size of 1 is used. The predictions before calculating error scores to ensure that performance is reported in the same units as the original data.

➢ *Loss and Optimizer :*

The Loss function according to given task and specify the optimizer to use with the learning rate which values of 0.3 and the Loss function is RMSE metrics to judge the performance of neural network and Adam. Optimizer is used.

➢ *Fit Model :*

The model input data class which are used to be X->train and the labels are Y->train, the epoch iteration and batch size are also considered. It returns the history of model training and history consists of losses and after each epoch will be visualized. Here the dataset are very big size so have to be allocated the batch size number. It divides the data of batch size are equal to the batch number and then only uploaded the number of sample of data in memory and processed. It will take a round a minute to train and after 100 epochs the neural network will be trained. so far the loss function RMSE now trained till around the 0.010 value.

ISSN No:-2456-2165

V. PERFRMANCE EVALUATION

A. Dataset virtualization :

The visual representation of the various parameters like humidity, pressure, temperatureMax, temperatureMin, windspeed, yield of the dataset is shown in the below figure-2.

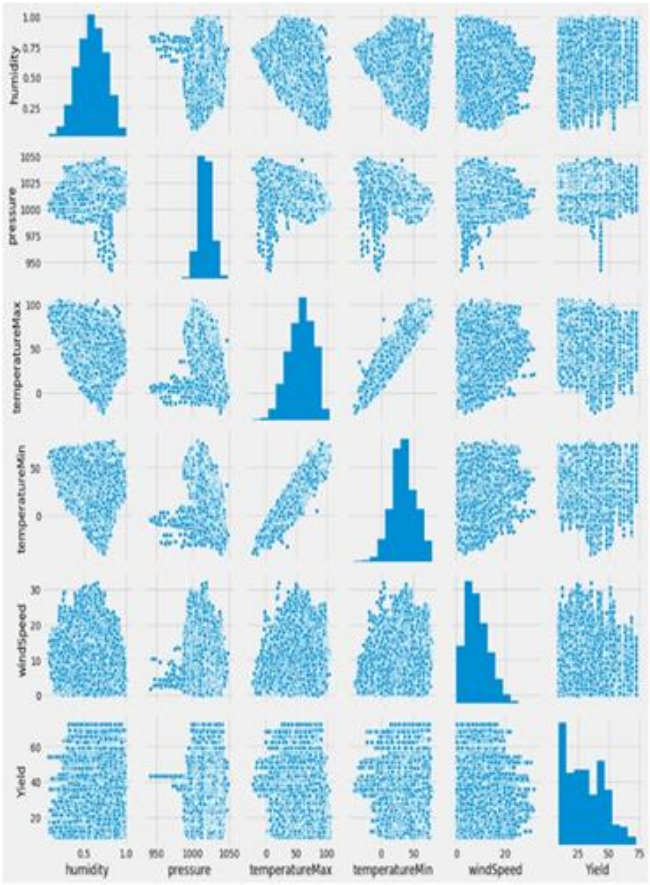

Fig- 2: Visualization of Data in The Dataset

The model is made run for the 50 EPOC and the evaluation value determined in the models are represented in the below table – 2. The results show that the Recurrent Neural Network outperforms the feed forward neural network model.

Model	No of EPOC	RMSE	LOSS
Feed-Forward Neural Network	50	0.232	0.1141
Recurrent Neural Network	50	0.00011	0.0103

Table 2:- Evaluation of Neural Network Models

B. Metrics for Evaluation :

The performance of the neural network models was evaluated by determining the Root Mean Square Error. The result obtained for Feed forward Neural Network and Recurrent Neural Network models are shown in the below fig-3 and fig-4 respectively.

Fig- 3: Evalution of Rmse Using Feed Forward Neural Network

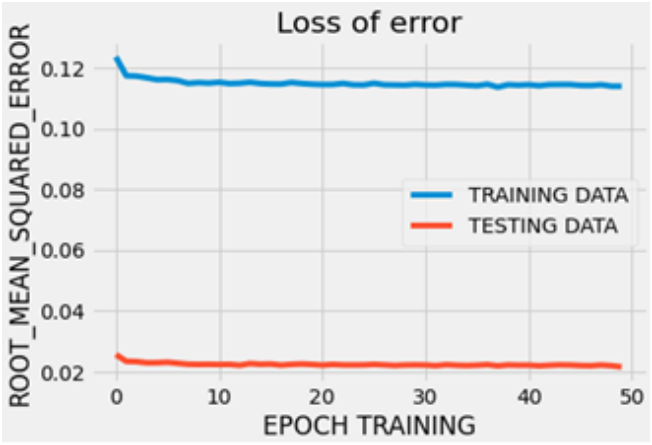

Fig- 4: Evalution of Rmse Using Recurrent Neural Network

VI. CONCLUSION & FUTURE ENHANCEMENT

The work to predict the yield of the crop based on the suitable crop parameters like TemperatureMin, TemperatureMax, Humidity, Windspeed, Pressure using neural network model like Feed forward Neural Network and Recurrent Neural Network. The performance of neural network model was evaluated using the metrics like Root Mean Square Error (RMSE) and Loss. Comparing the FNN and RNN based on loss of error RNN has low error rate at the same it is better for crop yield prediction. This makes the farmers to take decision right for right crop to cultivate the agriculture sector. The future study can be made to enhance the type of diseases that affect the crop and also suggesting to use a type of pesticide in order to overcome the diseases.

REFERENCES

[1]. Shruti Kulkarni , Shah Nawaz , G Srivatsa Sharma "Predictive Analysis to Improve Crop Yield using neural network" -2018, vol pp: 1- 15,1ssn No : 1875-1890.
[2]. Petteri Nevavuorib , Nathaniel Narraa , Tarmo Lippinga "Crop yield prediction with deep convolutional neural networks" – Elsevier Journal 2018.-Issn No : 3456-3477.
[3]. Nari Kim , Kyung-Ja Ha "A Comparison Between Major Artificial Intelligence Models for Crop Yield Prediction"-2018- Issn No : 434-440. Prakash J, "Enhanced Mass Vehicle Surveillance System", J4R, Volume 04, Issue 04 ,002, 5-9, June 2018.
[4]. Adthya sharma " Crop yield prediction using convolution neural network -2018- Issn No: 489-497.
[5]. Prakash J, Bharathi A, "Predicting Flight Delay using ANN with Multicore Map reduce framework, Communication and Power Engineering", De Gruyter 2016, 280–287.
[6]. Sandhya,J.Prakash,B.VinothKumar,"Comparative Analysis of Clustering Techniques in Anomaly Detection Wind Turbine Data", Journal of xi'an University of tech.