AGRIASTRO: A Software Tool for Envisaging Crop Yield

D.Maghesh Kumar[1], **Dr.K.Mohan Kumar[2]**

[1]Research Scholar, PG and Research Department of Computer Science, Rajah Serfoji Government College, Thanjavur Tamil Nadu, India
[2]Research Supervisor, Head, PG and Research Department of Computer Science, Rajah Serfoji Government College, Thanjavur Tamil Nadu, India

ABSTRACT

Agriculture is one of the main sectors that have huge impact on India`s economy. India is highly populated country and it's essential to increase the food production in order to feed the growing population. Due to the uncertainty in weather like excess rain, drought etc., farmers unable to envisage the crop yield. This makes heavy loss for the farmers and it increases the rate of suicide done by farmers in India. The research work deals with the envisaging of crop yield, by using a machine learning approach, namely a powerful optimized neural network model to increase their effectiveness. In this present study a software application tool named 'AGRIASTRO' has been developed as a user friendly and interactive in predicting the crop yields for the selected session, selected crops in the districts of Tamil Nadu. This tool uses the crop prediction methodology in envisaging the crop yield to maximize the profit and minimize the loss.

I. INTRODUCTION

Indian economy basically depends on agro industry. With the growing population every year it is necessary to increase the crop production but the crop production is a multifaceted because that is influenced by various agro input parameters like climatic, soil, water etc., These parameters various from regions to regions and the cultivation methodology as well[1]. The most regrettable truth about the agriculture is even though a most important sector in India but there is numerous factors that are hurdles in keep up high yield, including weather change, inadequate irrigation, conventional farming practices, below literacy rates and lack of affluence. Agricultural production is to accommodate the needs of the people, for that all the scientist, farmers and government join hands together by giving full effort and coin new strategies to increase the yield.

Worldwide efforts to increase the agricultural production are the necessities for the existence of the ever growing population of human and cattle. In this regard, it has become the need of the hour to effectively increase the crop production; traditional Indian agriculture was converted into an industrial system due to the adoption of modern methods and technology during green revolution [2]. This revolution changed India's status from a food-deficient country to one of the world's leading agricultural nations.

Most of the farmers are in the rural areas and they lack in knowledge about the new crops and its yield. Envisaging the crop yield well in advance before the harvest time would lend a hand to farmers in taking appropriate steps for marketing their cultivated crops. Precise forecast the stages of crop development play a vital task in crop production management system. Such aids not only to the farmers but also for the government in making policy

decisions regarding the food production. Furthermore, these forecasts could be used to make full use of the crop forecast has assumes even more significance given the daunting task before the researchers. Various approaches have been devised and developed in the earlier researches to predict crop yield had varying success [3]. An ease of use application tool developed in this research work will assist the farmers in predicting crop yield.

Neural networks are widely used for the data prediction purposes and can be used for estimating non-linear dependencies. Artificial neural network (ANN) is a powerful machine learning technique which provides diversified solutions to the cumbersome problems in agriculture. ANN has been assessed with examples that telling the how these techniques can be applied on agricultural data sets [4]. A parameter-based reengineered Artificial Neural Network (ReANN) was used in the proposed model for the estimation of district wise crop yield of Tamil Nadu [5].

The rest of this paper is planned as follows: Section II provides the background of the earlier research works related to crop yield prediction using various machine learning techniques. Section III illustrates the methodology used for to predict the crop yield by using optimized neural network algorithm, Section IV describes about the software tool used to obtain the results and Section V concludes with the discussion about the effectiveness of the developed tool and further enhancements to be made in the tool.

II. LITERATURE SURVEY

Jeeva C, and Shoba S.A., summarized the applications of data mining techniques such as k-means, bi clustering, k nearest neighbour, Neural Networks, Support Vector Machine and Naïve Bayes Classifier the field of agriculture. An efficient model for agricultural production has been developed by using the swing toolkit in Java. The developed GUI was used to predict the expected annual rainfall, annual production and productivity of major crops in Tamil Nadu[6].

Ruß, Kruse, Schneider, and Wagner deliberated the previously established patterns in the field of agriculture to develop an efficient technique for modeling and by means of decision rule to transmit the raw dataset into specific statistical form. The historical datasets are very important in unleash the existing patterns and has been used for the purpose of prediction. Artificial neural network technique has been used for foreseeing the wheat yield [7].

Naheed Eja and Shabbir Abbasi used an amalgamated approach of Neural Network, Support Vector Machine and Regression to analyze and estimates the wheat yield production. They used parameters like area, yield, production, soil pH, temperature, air pressure, rainfall, water availability, humidity, pesticides and fertilizer for the prediction of wheat yield [8].

He. Y., Y. Zhang, S. Zhang and H. Fang analyzed the connection among complex attributes wheat yield by parallelise the working of a nonlinear function and used Artificial neural network this type of system. The output shows the close relationship between the wheat yield and soil related parameters such as nitrogen, potassium etc., [9].

Dahikar and Rode developed a crop yield prediction model by using Artificial Neural Network (ANN) to analyze the soil parameters and the various atmospheric conditions. The

proposed ANN model uses feed forward back propagation and was developed in Mat lab. This model more precisely shows the impact of soil and climatic conditions in crop yield and also used to approximate long-term or short-term crop production [10].

Kumar, R., M. Singh, P. Kumar and J. Singh resolved the crop selection problem by build model using machine learning technique called crop selection method (CSM). In CSM model the parameters such as climate, soil, and water crop type were used for making prediction. These constraints influence has high impact in the prediction accuracy and CSM models performance. This model demonstrates by selecting the right seasonal crop to maximize the yield minimize the loss. By using this CSM, it helps the farmers to attain the maximum profit and thus helps nation`s economy development [11].

Marinkovic et. al. studied various machine learning techniques of data mining to envisage the crop yield estimation and applied these approaches straightforwardly in the field of agriculture. Production of Soybean and maize like crops were measured on regular bases by the climate attributes. The outcome from the proposed prediction was compared with past techniques and proved that the proposed model better in envisaging crop yield [12].

Geetha surveyed on various types of data mining models used in the field of agriculture. Evaluation and assimilation made by other researchers in the agriculture related applications by using the machine learning techniques such as ANN, KNN, Decision tree, fuzzy set, Bayesian network, K-mean and J48 were studied. This consolidated work made for the purpose of getting knowledge about the current advancement technology used in agriculture [13].

Whelan, McBratney and Minasny provided a spatial prediction technique to aid the precision agriculture to enhance the production. The data has been acquired both manually and through sensors. Naturally data quantity, density and measurement quality varies but the models ability in the prediction process to individual data sets is necessary for Precision Agriculture (PA). In statistics, kriging is a method of interpolation. VESPER 1.5 (Variogram Estimation and Spatial Prediction plus Error) is a shareware software that gives the global or local kriging by using global or local semivariograms[14].

Abhijeet Pandhe et. al. applied machine learning technique to solve the problem of predicting crop yield using climatic parameters to help cultivators and policy makers to make important decisions such as pricing, marketing etc., A web application 'Smart Farm' has been developed by using the Random Forest algorithm, to assist the farmers to get precise prediction. The backend of the application was developed using python and the obtained results were provided to the web page by using Flask. The acquired result indicates that there is strong relationship between the climate and the crop yield [15]

III. MATERIALS AND METHODS

The main intention behind this research work is to develop a software application tool using optimized artificial neural networks to achieve the best possible results by envisaging crop yields in Tamil Nadu. The usefulness of neural networks in prediction has been tested several times by various researchers. The main stress is in optimizing the neural networks, to get accuracy of the results. The ReANN algorithm comes handy to solve this issue[5]. The AGRIASTRO tool trains neural network and produce a precise result, when predicting crops yield and was developed by using Python.

A) Data source:

The present study was carried out in the Tamil Nadu state, India. Online sources providing information about the state`s agricultural statistical data was mainly gathered from Indian Statistics and Agriculture department`s web sites tnagrisnet.tn.gov.in and data.gov.in.

B) Input variable selection:

Inputs play a significant role in the manipulation of output. Improper input leads to inconsistent results and this paves the way for the uncertainty in the prediction. The selection of input variable is very crucial and these have significant impact on the output. It is awfully essential to discover an effectual method for the input variable selection. All input variables are evaluated by using the correlation measure. There are two kinds of input variable one is independent variable and other is dependent variable and very careful in classifying these variables. The dependable variable can be difficult to find. The variable with utmost importance with respect to the dependent variable is selected and grouped with the second utmost implication variable and vice versa [16]. Selected parameters for the crop yield are depicted in the Table 1.

Table 1 : Crop yield Parameters

Parameters	Description
District_Name	District name of the state
Crop_Year	Cultivated Year of the crop.
Season	Season in which the crop cultivated
Crop	Cultivated crop name
Area	Cultivated Area / Hectare
Production	Crop production in Tones/ Hectare
Yield	Crop yield in Kg/ Hectare

C) Data Preprocessing:

The purpose of data pre-processing is used to transform the raw data in a useful and efficient format, such that it becomes machine-readable in any Machine Learning process. After pre-process the data it helps the machine to easily parse and the features of the data can now be easily interpreted by the algorithm. A feature is an individual measurable property is often called as variables, characteristics, fields, attributes, or dimensions.

The first step in the data pre-processing is data cleaning, to remove the irrelevant data, handles the missing and noisy data. The second step data Transformation is taken in order to transform the data in appropriate forms suitable for further process. The final step is data reduction, to maximize the storage efficiency and minimize data storage and analysis costs.

Clear identification of the dependent and independent factors in the pre-processed dataset is essential. From Table 1 production and yield are dependent factors because it changes with the independent factors season, crop and area. After cleaning, the last 17 (2002-

2018) year's agricultural data from various districts of Tamil Nadu, India. The collected data contains the crop details are about 13,547 rows

D) Dataset splitting:

It's necessary for every machine learning model to learn because to know the relations between attributes and can envisage for new interpretations. This learning activity requires a training dataset to train the model. Usually a single mammoth dataset is used and it is mandatory to split the dataset into 80:20 ratio of subsets namely a training dataset and a test dataset. For the cross validation purpose the training dataset is further split in the ratio of 60:20. After splitting the data, training, validation and test in the ratio of 60:20:20

E) Reengineered Artificial Neural Network(ReANN):

In this software tool for the purpose of envisaging, a novel optimized machine learning technique called ReANN is used. In the ReANN algorithm hyper-parameter optimization was used i.e. it Optimize the values of Learning Rate, Number of Epochs, Hidden Layers, Hidden Units, momentum and Activations Functions.

The ReANN model was trained using damped-least squares method, it uses Leaky ReLU activation function during back propagation in the hidden layers and softmax function was used for output layer [5].

In order to make accurate prediction, the historical data is very much useful which is good for presumption but is extremely susceptible to error and make our judgement sometime wrong. To reduce the error, lot of models were developed by various researchers. A novel approach was followed in the development of new Crop Yield prediction model is shown in fig.1 and the prediction process will be incessantly repeated until attaining the precision in the predicted yield.
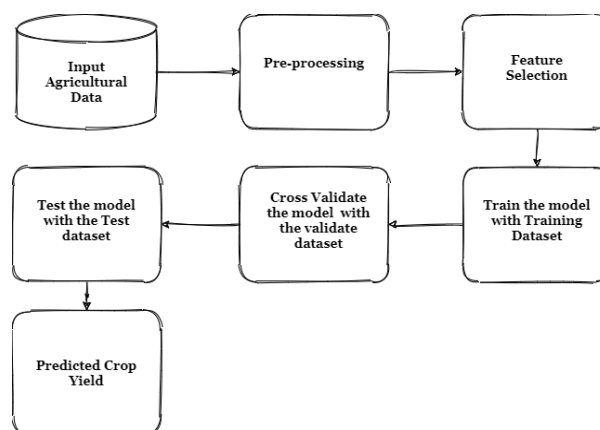
Fig: 1 Crop yield Prediction model

IV. RESULTS AND DISCUSSION

The ultimate objective of the research work was to develop a user friendly software application that will help the agriculturalist and other policy makers to predict the crop yield based archived agricultural data from the year 1998. A software application 'AGRIASTRO' has been developed for this purpose.

The software has been made using python because it supports handful of libraries making implementation of machine learning easier for developer. The following python libraries has been used in the software are NumPy is an open-source numerical Python library used for working with N-dimensional array objects, pandas providing high-performance, easy-to-use data structures and data analysis tools, Matplotlib is a comprehensive library for creating static, animated, and interactive visualizations in Python., Scikit-learn used to split the data into training set and testing set and Tkinter is Python's de-facto standard Graphical User Interface package used.

The main menu of AGRIASTRO software tool used for crop yield prediction is shown in fig.2.
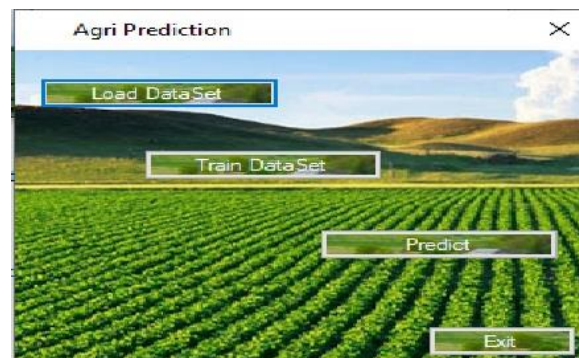
Fig 2. Main Form

The user has to load the dataset first for the processing. The loaded dataset is depicted in the fig. 3

Fig 3. Load dataset

After loading the dataset, it is mandatory to clean the data in order to avoid anomalies in result. The preprocessed dataset is shown in fig.4.

Fig 4. Preprocessing the Dataset.

The dataset contains the all the details of crops cultivated in Tamil Nadu. The user has the option of selecting any district in Tamil Nadu, any season such as Karif, Rabi etc., and any type of crop and selects the year for making prediction. After entering the details user will press 'Predict'. This will make the prediction by using the model which is trained by using the ReANN algorithm. The user can see in Fig. 5, the result of the prediction of the crop yield. For making future prediction the details are to be loaded in the test data.
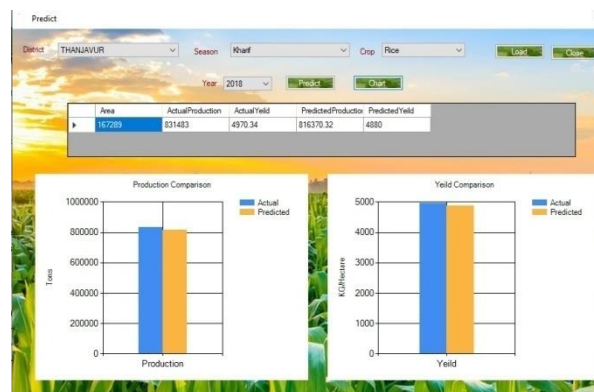

Fig 5. Crop Yield prediction.

Henceforth, using the AGRIASTRO tool, the user can predict the crop yields of their choice by selecting district, season and crop type and crop yield values of previous known years to arrive at an expected value.

V. CONCLUSION AND FUTURE ENHANCEMENTS

The tool mainly developed for the purpose of predicting, the crop yield from large agriculture dataset. This application will assist farmers to take better decision on which crop to be planted based on yield of that season and also helps both the formers and policy makers in making important decisions like import-export, pricing, marketing, etc. even before the crop is being harvested. Currently this tool used to give, crop yield details of Tamil Nadu state alone but having provisions for extending to other states of India

REFERENCES

[1] Pandhe, Abhijeet, Praful Nikam, Vijay Pagare, Pavan Palle, and Dilip Dalgade. (2019) "Crop Yield Prediction based on Climatic Parameters." *METHODOLOGY* 6, no. 03.

[2]   Fujita, Koichi. (2012), "Green revolution in India and its significance in economic development: Implications for sub-saharan africa." *JICA Research Institute* 15: 5-6.

[3]   Yuvaraj, N., and K. Menaka. (2016), "A survey on crop yield prediction models." *Int. J. Innov. Dev* 5: 12.

[4]   Ejaz, Naheed, and Shabbir Abbasi., (2020), "Wheat Yield Prediction Using Neural Network and Integrated SVM-NN with Regression." *Pakistan Journal of Engineering, Technology & Science* 8, no. 2.

[5]   Maghesh Kumar D, and Mohan Kumar K. (2020) "Decision Support System For Prediction Of Crop Yield Using Re-Engineered Artificial Neural Network." *Journal of Critical Reviews* 7, no. 12 : 2290-2297.

[6]   Jeeva C, Shoba S.A., (2016), "An Efficient Modelling Agricultural Production Using Artificial Neural Network (ANN)", *International Research Journal of Engineering and Technology (IRJET)*, Volume: 03 Issue: 05.

[7]   Ruß, G., R. Kruse, M. Schneider, and P. Wagner,(2008), "Data mining with neural networks for wheat yield prediction." *Advances in Data Mining, Medical Applications, E-Commerce, Marketing, and Theoretical Aspects*, 47-56.

[8]   Naheed Ejaz, Shabbir Abbasi,(2018) "Wheat Yield Prediction Using Neural Network and Integrated SVM-NN with Regression", *Pakistan Journal of Engineering Technology and Science (PJETS)* Volume 8, No 2.

[9]   He, Y., Y. Zhang, S. Zhang and H. Fang, (2006), "Application of artificial neural network on relationship analysis between wheat yield and soil nutrients.", *Engineering in Medicine and Biology Society, 27th Annual International Conference of the, IEEE.*

[10] Dahikar, S. S., & Rode, S. V. (2014). "Agricultural crop yield prediction using artificial neural network approach.", *International journal of innovative research in electrical, electronics, instrumentation and control engineering*, 2(1), 683-686.

[11] Kumar, R., M. Singh, P. Kumar and J. Singh, (2015) "Crop selection method to maximize crop yield rate using machine learning technique", *In Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), International Conference* on (138-145).

[12] Marinkovic, B., J. Crnobarac, S. Brdar, B. Antic, G. Jacimovic and V. Crnojevic, (2009) "Data mining approach for predictive modeling of agricultural yield data", *Proc. First Int Workshop on Sensing Technologies in Agriculture, Forestry and Environment Novi Sad, Serbia.*

[13] Geetha, M., (2015), "A survey on data mining techniques in agriculture". *International journal of innovative research in computer and communication engineering,* 3(2): 887-892.

[14] B.M. Whelan, A.B. McBratney and B. Minasny, (2002)"VESPER 1.5 – spatial prediction software for precision agriculture", *Proceedings of the 6th International Conference on Precision agriculture, ASA/CSSA/SSSA, Madison, Wisconsin*, 14p.

[15] Abhijeet Pandhe, Praful Nikam, Vijay Pagare, Pavan Palle, Prof. Dilip Dalgade, (2019) "Crop Yield Prediction based on Climatic Parameters", *International Research Journal of Engineering and Technology (IRJET)* Volume: 06 Issue: 03.

[16] Mundry, R. and Nunn, C. L.(2009)"Stepwise model fitting and statistical inference: turning noise into signal pollution", *The American Naturalist*, vol. 173, no. 1, pp. 119–123.