# The Best Accurate Crop Yield Prediction System Using Machine Learning Algorithms in Deep learning

## Dr.S.Amutha[1], S.Deepa Lakshmi[2], R.Dhivya[3], G.Subha Sri[4]

Department of Computer Science Engineering

P.S.R Engineering College,

Sivakasi.

Department of Computer Science Engineering

P.S.R.Rengasamy College Engineering for Women,

Sivakasi.

-----------------------------------------------------------------------***---------------------------------------------------------------------

**Abstract:** *This project proposes a model that will enable us to predict the agricultural outcome of a country. The prediction will include the agricultural yield for the upcoming years as well as predict the profit margin for particular crops. Although some researchers have been done in this field but most of those are in India perspective. In our country, being the oldest form of raising economy, agriculture is a field that has not been blessed by, modern technology or data analysis. This prediction is made by analyzing the dataset of certain vital parameters for rice production, such as Temperature, Humidity, Sunshine and Area using different Regression Analysis and Support Vector Machine(SVM) techniques. The dataset containing the history of Rice Production, Price Diversity and Fertilizer Usage, taken from the yearbook of Indian Agricultural Development Corporation(BADC), India Rice Research Institute(BRRI), Ministry of Agriculture(India) and by some primary data collection from Ghazipur and Patna,is also analyzed to calculate the profit per year along with comparing the accuracy of the prediction. This proposed model aims to pave the way for data science to touch the sector that keeps our economy running aiming to maximize the production of crops which will result in more profit for the farmers as well as contribute to the economy with the ancient form of revenue collection.*

**Keywords: Classification Algorithms in Agriculture, Data set, Decision tree, Machine learning, prediction of crops.**

## 1. Introduction

India, our motherland is a country that has been gifted by nature through its fertile land. Although the agricultural production is satisfying. It still fluctuates. Therefore, predicting how much crop our country can produce in any given requirement. The model we are proposing will focus on predicting the agricultural production of our country more specifically the production of rice. Given the specific data that is required by the model, the model can then analyze and form a prediction for the production. Being able to predict the agricultural outcome not only helps us to be prepared for any kind of shortage, but also it helps to maximize the profit for the farmers. Farmers ,who are the solely responsible to produce the food we eat are not compensated. It happens very often that they do not get the price they deserve just because we are not able to know what the total production might be and what the estimated price would be. Therefore, the multiple factors including humidity, minimum and maximum temperature, sunshine, year-base production of rice, price of rice in the local market and so on. If we are able to generate a well enough prediction, then we can know how the price of each unit of crop is going to be throughout the entire year. This will help the farmers to get the price they deserve and maximize their profit. Therefore, the government can take necessary steps to prevent a chaos and import rice that are necessary. We plan to achieve our goal by applying data analysis mainly. The system will require historical data to

work with. The data about the related fields are collected from various sources in order to analyze and generate prediction.

## 2. Literature Review

Prediction, or predicting the future data to be more precise is now one of the most explored topic in computer science. In data science, a major part is mining of data and predicting the future outcome of any certain field. Many algorithms have already been developed in order to achieve this. These predictions are being made by using algorithms in almost all the major fields. Companies and organizations are using these predictions to estimate their future sales or productions that they can take proper steps to maximize their profit and grow their business. Already researchers are using data analysis techniques in order to make predictions. Predicting the future outcome, these researches have also extended to the field of agriculture. Researchers have already tried to predict the future yield of crops applying the techniques of data science and machine learning.

Previously, researches have used BP Neural Network approach to achieve this same endeavor. According to the researchers this approach has three layers that need to be implemented. They are known as input layer, hidden layer

and output layer. The hidden layer itself has many multiple layers that do not have any individual connections but has only connection to adjacent layers. This approach of predicting the outcome is quite difficult to implement and researchers had found some limitations.

In perspective our country, finding a very well-defined data set is not possible. In fact, available data are very much scattered and noit of any particular format. So applying these techniques require a heavy data preprocessing which is not efficient. Our model looks for algorithms that do not require very well-structured data.

Researchers have applied decision tree model in order to make prediction about objects that are not discrete and real but structured. They calculated the expected loss or profit from their training data classification. Previous work based on information dimension system.

Researchers who applied BP neural network has found that it has shortcomings of its own. These are mainly seen when dealing with large number of multidimensional data. When the types of datasets are found, the networks structure becomes hard to design and it requires more training time as the convergence time become longer. Also, it is prone to fall in local minimum state.

Researchers also tired to apply time series analysis in order to predict the price. Researchers were conducted to predict and analyze the fluctuation price of vegetables and crops. Their study aimed to identify trends and seasonality of individual vegetables and investigate factors of supply and demand.

Furthermore, many researchers have also proposed model that utilized APC and CV in order to determine indicators that showed the fluctuating prices. ARIMA and GARCH has also been applied by researchers as time series models. Researchers also used satellite data in order to predict the temperature in the areas where crops are produced. But the major limitation for this approach is that this model will not be accessible to many because it requires turned out to not satisfactory.

In this project, we propose a model or a system that will make prediction about the future production of crops and their price and profit margin. In order to do so, we implement the regression and Support Vector Machine (SVM) algorithm. This model implements multiple methods of regression analysis to determine the accurate outcome. Also this approach will give us overall idea about the most suitable algorithm for this use case as we will be implementing multiple algorithm on the same data set. Therefore, we can compare and determine which algorithm delivers the most accurate result and in a more

efficient manner. These method scan work with dataset that are structured and well defined. It is not possible to find such dataset in our country. So we have manually collected all data ourselves and processed them in order to make them suitable for our algorithms.

## 3. Existing System

The prediction is based on linear regression is discussed in many works. It is a statistical method applied over linear systems. By using this, the relationship between dependent and independent variables can be measured. If independent variable is having more than one input attribute, multiple regressions can be applied. Regression based models are used in prediction as this technique shows consistent results during standard tests.

## 4. Proposed Model

In order to predict the outcome, we initially wanted to use a couple of algorithms. However, some of them had taken out because of their lower efficiency and accuracy. During the implementation, the efficiency and accuracy of linear regression well below the acceptable of the error. However, Random Forest and Support Vector Machine(SVM) has performed well and had very good accuracy.

### 4.1 Workflow

The research work has been divided into a few major parts. All these parts together represent the research model. The procedure have taken areas as follows:

1.Data collection

2.Data processing

3.Feature extraction

4.Feature scaling


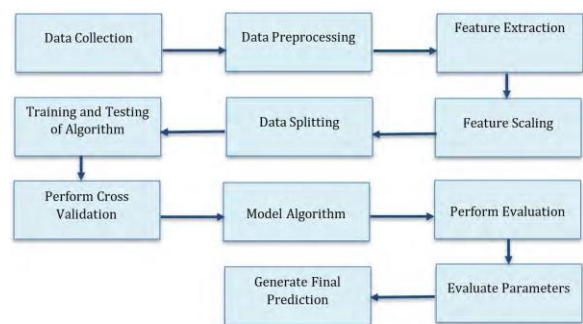
Fig.4.1 Work Flow of the Research

5.Data splitting

6.Training and testing algorithms

7.Perform cross validation

8.Perform evaluation of algorithms

9.Predict parameters

10.Generate final prediction

## 5. Algorithms

### 5.1Regression Analysis

Regression analysis is a method in which observations of events are used to determine and quantify relationship between one variable and some other independent variable, if any. Researchers often have used regression analysis in order to generate a model for analyzed data. For the goal of making predictions based on multiple factors, there are many types of regression types that are available and widely used. However, all the regression techniques are based on mainly three main features or parameters[8].

The three parameters are:

1.Total number of independent variables

2.Types of all the different variables

3.Shape of the regression line.

### 5.2 Linear Regression

To make any kind of model for data prediction, one of the most widely used method is linear regression. The dependent variable, which is one of the key factors of regression method, is usually continuous in nature. However, the independent variable can either be both of continuous or discrete. The regression line that is produced is linear by nature, hence the name linear regression.

### 5.3 Support Vector Machine

Support Vector Machine (SVM) is a supervised machine learning technique. This is used for both regression and classification. The portion of this which is used for regression is known as Support Vector Regression(SVR).There are two types of SVR. They are,

1.Linear Support Vector Regression

2.Non-Linear Support Vector Regression.

### 5.4 Random Forest

Random Decision Forest or Random Forest is one of the popular methods that are used by researchers in order to generate predictive models. This is a very good machine learning algorithm which predicts outcomes with higher accuracy.

## 6. Results and Analysis

We have applied multiple algorithms in the data set in order to determine and compare the performance level of the algorithm.

| Item | Linear Regression | Random Forest | Support Vector Regression (SVR) |
|---|---|---|---|
| Minimum Temperature | 7.59 | 40.79 | 10.24 |
| Maximum Temperature | 16.55 | 24.21 | 19.07 |
| Yearly Rainfall | 11.96 | 91.82 | 16.07 |
| Yearly Humidity | 43.85 | 62.91 | 52.53 |
| Rice production | N/A | 68.49 | 59.78 |

Table6.1 Algorithm Accuracy Comparison

Confusion matrix
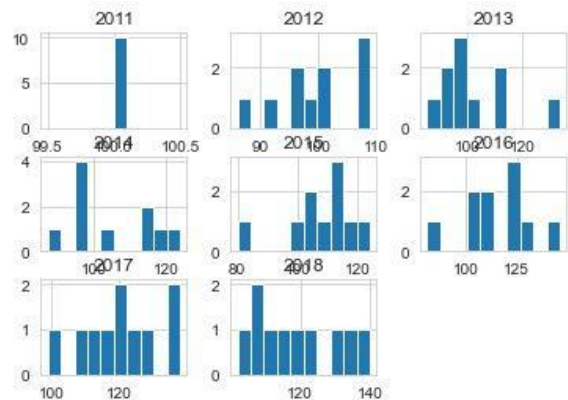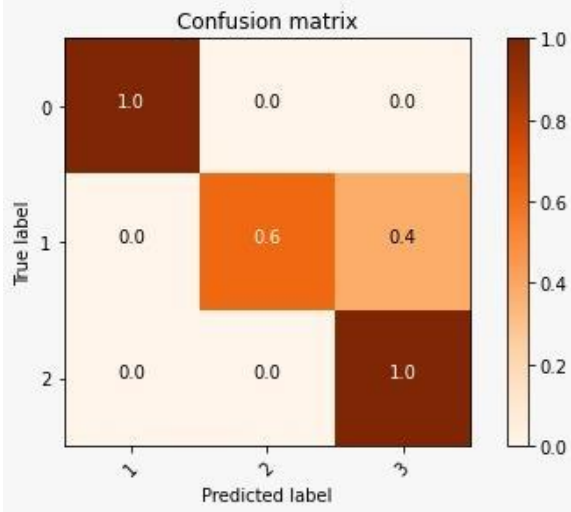
## 7. Conclusion

In this research, the model proposed analyses some real time data of Aush and Aman rice, two major types of rice and predict the yield, by which the profit intended for future production can be calculated. Linear Regression, Random Forest and Supervised Machine Regression algorithms has been used to demonstrate the prediction. Throughout this research and field work, it is found that the unavailability of expected data for different agricultural crops is the major problem for such analytically implementation. In the future, this model can be implemented as web and mobile applications so that the agriculture department of the country and field workers can use the predicted result before planning of a production and have a convenient profit margin.

## REFERENCES

[1] J.Ming, L.Zhang, J.Sun, and Y.Zhang, "Analysis models of technical and economic data of mining enterprises based on big data analysis",in 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis(ICCCBDA),pp.224-227,April 2018.

[2] W.Minghua,Z.Quaolin,Y.Zhijian,and Z.Jingui," Prediction model of agricultural product's price based on the improved bp neural network", in 2012 7th International Conference on Computer Science Education(ICCSE0,pp.613-617,July 2012.

[3]M.T.Shakoor,K.Rahman, S.N.Rayta,and A.Chakrabarty," Agriclutural production output prediction using supervised machine learning techniques",in 2017 1st International Conference on Next Generation Computing Applications (NextComp),pp.182-187,July 2017.

[4] P.Vinciya and D.A.Valarmathi,"Agricluture analysis for next generation high tech farming in data mining",International Journal of Advanced Research in Computer Science and Software Engineering(IJARCSSE),vol.6,no.5,pp.481-488,2016.

[5] W.Lin,G.Dean, and C.Moore,"An empirical test of utility vs.profit maximization in agricultural production", American Journal of Agricultural Economics, vol.56,no.3,pp.497-508,1974.

[6].H.Lee and A.Moon,"Development of yield prediction system based on real-time agricultural meterological information", in AdvancedCommunication Technology(ICACT),2014.16th International Conference on ,pp.1292-1295,IEEE,2014.

[7].S.Bagis, B.B.Ustundag, and E.Ozelkan,"An adaptive spatiotemporal agricultural cropland temperature prediction system based on ground and satellite measurements", in 2012 first International Conference on Agro-Geoinformatics,pp.1-6,Aug 2012.