Automated Crop Yeild Prediction Using Machine Learning

Ms.B.Anitha[1], Ms.Madhushree.V.[2], Ms.Yamini.k.[3], Ms.Pooja.H.R.[4], Mrs.Anupama P V[5], Mrs.Aruna M G.[6]
[1,2,3,4] *Student, Department of CSE, M S Engineering College, Bangalore, Karnataka, India*
[5]*Assistant Professor, Department of CSE, M S Engineering College, Bangalore, Karnataka, India*
[6]*Associate Professor, Department of CSE, M S Engineering College, Bangalore, Karnataka, India*

Abstract **- Agriculture is the pillar of the Indian economy and more than 50% of India's population are dependent on agriculture for their survival. Variations in weather, climate, and other such environmental conditions have become a major risk for the healthy existence of agriculture. Machine learning (ML) plays a significant role as it has decision support tool for Automated Crop Yield Prediction (ACYP) including supporting decisions on what crops to grow and what to do during the growing season of the crops. The present research deals with a systematic review that extracts and synthesize the features used for ACYP and furthermore, there are a variety of methods that were developed to analyze crop yield prediction using artificial intelligence techniques.**

Index Terms **- Agriculture, Crop Yield Prediction, Machine Learning Methods.**

I.INTRODUCTION

Agriculture is the most important sector of Indian Economy. Indian agriculture sector accounts for 18 percent of India's GDP and provides employment to 50% of the country's workforce. But latest studies have shown a steady decline in the contribution made by agriculture to the Indian economy.

India is a highly populated country and randomly change in the climatic conditions need to secure the world food resources. The production of agriculture is affected by several climate factors.

The main problem with agriculture in India is lack of rainfall in seasonal time.

To overcome these above issues, we need to develop a system which will be able to find the hidden facts or results, patterns and insights. The farmer can predict which crop should sow so that he/she can get more benefit. we are applying data analytics techniques on agriculture production-based datasets and find the insights so that it can help to the farmers and their decision making.

II. EXISTING SYSTEM

The purpose of crop yield prediction is to estimate production in agriculture sector for better crop management and make strategic decisions for improving crop yield in future.

The Existing model can be incorporated with a decision support system (DSS) that can be used in precision agriculture which aims at complete farm management.

Demerits

- Factors like climate and location of market and planting area is not taken into consideration
- The system doesn't take area of land being cultivated and the sowing date. The market price of the cultivated crops after harvesting is not considered.

III. PROPOSED SYSTEM

Prediction of the crop yield using the efficient algorithm and suggest how much quantity of fertilizer should be used to get the proper yield for the crop using naïve Bayesian algorithm. The data mining techniques on historical climate and crop production data several predictions are made which increase the crop productivity.

The decision support system must be implemented for the farmers to take proper decisions about soil and crop to be cultivated. They have collected the dataset with attributes of the crop season, Area and production in hectares and analyzed with various algorithms in WEKA.

IV. SYSTEM ARCHITECTURE

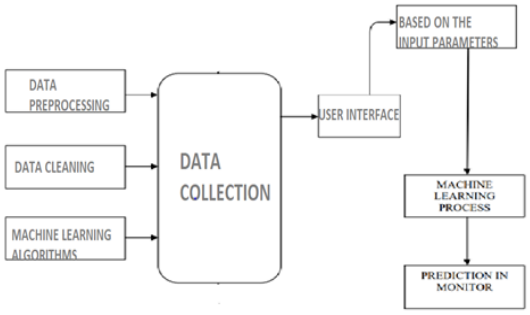

Fig.1 System Architecture

Firstly, the pre-processing of the data takes place. Then the data is collected from the dataset. Then the model is trained with the help of random forest algorithm. Then the prediction of crop yield is obtained, shown in above fig.1

B. System Modules

A. Data Collection

The below Fig.2 represents datasets are collected from various sources like GKVK, DES, agmarknet.gov.in and krishimaratavahini. The datasets include information like temperature, rainfall, price, area, production and yield of the previous 5 years (2013-2018).

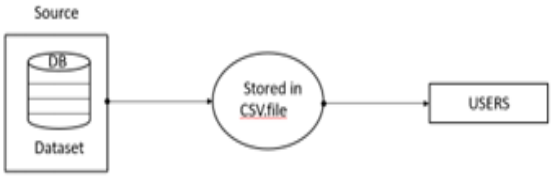

Fig.2 Block Diagram of Data Collection

B. Data Pre-processing

The below Fig.3, Data pre-processing includes removing of the unwanted attributes from our datasets. Feature extraction is done in order to extract only the attributes that affect the price and yield of a crop like rainfall, temperature, location, area, production and yield.

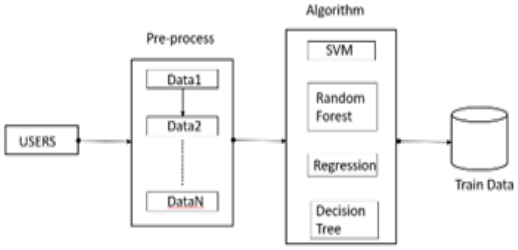

Fig.3 Block Diagram of Data Pre-processing

C. Data Analysis and Prediction

The below Fig.4, shows, the patterns are recognized, percentage correlation between various factors affecting crop yield and price are determined. Various data visualization techniques are used to study the patterns in data and factors causing change. The accuracy of these algorithms is compared using mean absolute percentage error thus helping us determine the most suitable approach for prediction.

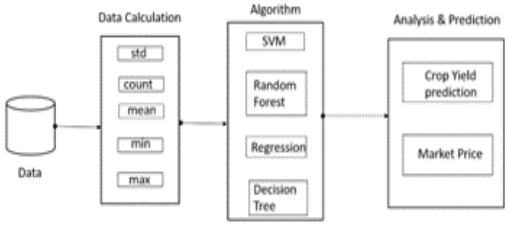

Fig.4 Block Diagram of Analysis and Prediction

V.METHODOLOGY

RANDOM FOREST

Random Forest is an excellent supervised learning algorithm that can train a model to predict which classification results in a certain sample type belong to base on a given dataset's characteristic attributes and classification results. Random Forest is based on a decision tree and adopts the Bagging (Bootstrap aggregating) method.

SUPPORT VECTOR MACHINE

SVM is one of the most widely used computer supervised learning model to perform prediction and classification. An SVM model interprets the training data points as points within the function domain, distributed in a way that distinguishes as broadly as possible the points belonging to different classes the same area, the test data points are then drawn and graded by which side of the threshold they fall.

DECISION TREE

Decision Tree is one of the Supervised Learning algorithms. Classification issues are mostly dealt with by the use of a decision tree. Based on important predictors, the population divides into two or more related DT sets. The DT's first step is to calculate entropy for each and every attribute.

LINEAR REGRESSION

Linear Regression is a machine learning algorithm based on supervised learning. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on – the kind of relationship between dependent and independent variables they are considering, and the number of independent variables getting used.

VI.IMPLEMENTATION

A.Data Collection

In this module datasets are collected from various sources like GKVK, DES, agmarknet.gov.in and krishimaratavahini. The datasets include information like temperature, rainfall, price, area, production and yield.

Pseudocode for reading data from the csv file

```
Procedure: ReadData()
//Input: Data from csv file
//Output: Dataset
Begin
STEP 1: Read data from csv file
STEP 2: Store values in dataset
STEP 3: Return raw dataset
End
```

B. Data Pre-processing

Data pre-processing includes removing of the unwanted attributes from our datasets. In this preprocessing step we are removing null valued and duplicate values in datasets.

Pseudocode to drop missing values

```
Procedure: Data Preprocess()
//Input: Raw Dataset
//Output: Cleaned dataset
Begin
STEP 1: Read dataset
STEP 2: Find any missing values
STEP 3: IF any missing values THEN
        Remove the row by using dropna()
STEP 4: Return raw dataset with no null values
End
```

C. Data Analysis and Prediction

patterns in data are recognized, percentage correlation between various factors affecting crop yield and price are determined. Various data visualization techniques are used to study the patterns in data and factors causing change. Algorithms like Multiple Linear Regression and Random Forest are used to predict crop yield and price. The accuracy of these algorithms is compared using mean absolute percentage error thus helping us determine the most suitable approach for prediction.

Pseudocode for using algorithms to build prediction Model

```
Procedure: DesignModel ()
//Input: Cleaned dataset
//Output: Prediction model
Begin
STEP 1: Read dataset T
STEP 2: Read n //to determine algorithm
STEP 3: IF(n==1) {
Initialize
RF=RandomForestAlgorithm ()
Call RF. Fit (X, y)
Store RF. Predict () in predictions}
STEP 4: ELSE IF(n==2) {
Initialize
LR=LinearRegressionAlgorithm()
Call LR. fit (X, y)
Store LR. Predict () in predictions}
STEP 5: ELSE IF(n==3) {
Initialize  SVM=SupportVectorMachineAlgorithm()
Call SVM. Fit (X, y)
Store SVM.predict() in predictions}
STEP 6: ELSE IF(n==4) {
        Initialize
DT=DecisionTreeAlgorithm ()
        Call DT. Fit (X, y)
        Store DT. Predict () in predictions}
END
```

VII. EXPERIMENTAL RESULT

In our tests, we looked at the performance of training set that had different features in it. The proposed system requires the operating system of windows 10 and above. It requires 8Gb of RAM and 40Gb of storage disk. The proposed system is coded using python programming language along with anaconda tool and some library functions.

The fig.5 shows, the attributes of yield to calculate the most suitable crop to grow in your Farm.

Fig.5 The attributes of crop yield

The fig.6 Shows, the Output of the Crop yield Prediction with different algorithms.

Fig.6 The output of the crop yield prediction

The fig.7 Shows, the attributes of the Market Price

Fig.7 The attributes of market price

The fig.8 Shows, the Final Output of the Market Price with prediction to Random Forests algorithm.

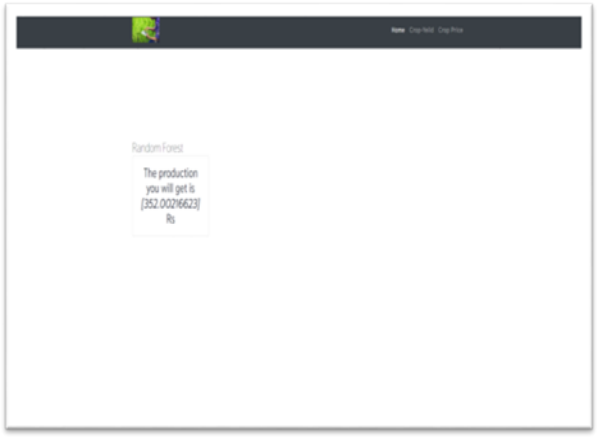

Fig.8. The output of the market price

VIII.PERFORMANCE ANALYSIS

The performance evaluation of the crop yield prediction system explained by various line and bar graphs in Fig.9&10.The line Graph is plotted between year and prediction to estimate production. The bar Graph is plotted between Season and production to estimate prediction.

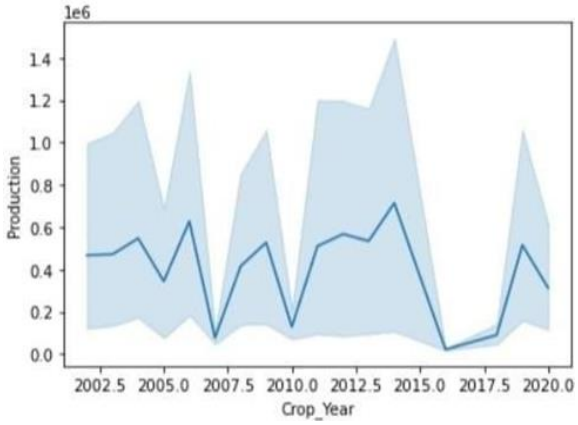

Fig.9 Line graph for year and prediction

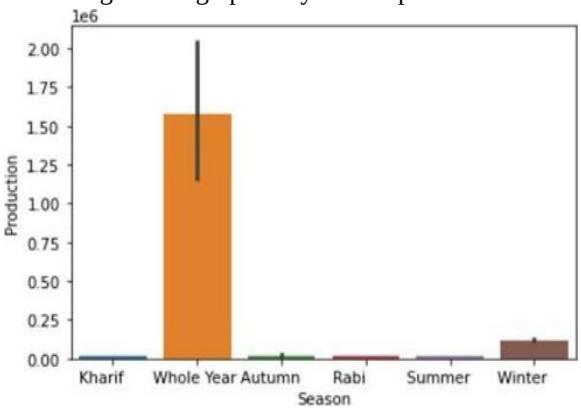

Fig.10 Bar graph season and production

IX.CONCLUSION

Agriculture is the backbone of counties like India. However, the usage of technology towards agriculture is to be given paramount importance towards preclusion agriculture.

This paper proposes a system which will help farmers to have an idea of yield estimates based on weather parameters and area under cultivation Using this farmer can make decisions on whether to grow that particular crop or go for alternate crop in case yield predictions are unfavorable.

This research work can be enhancing to the next level. We can build a recommender system of agriculture production and distribution for farmer. By which farmers can make decision in which season which crop should sow so that they can get more benefit. This system is work for structured dataset. The accuracy estimated from the project is 83%.

X.FUTURE ENHANCEMENT

In future, this system can be enhanced by considering the live weather conditions to make the accurate crop yield prediction

REFERNECE

[1] Tishampati Dha, Doug Gray, Carl Menges (2009). Multiple Crop Yield prediction using dual- polarimetric TerraSAR-X strip map imagery.

[2] Zengnan Li, Zhongxin Chen (2011). Remote sensing indicators for crop growth monitoring at different scales.

[3] Julian Ramiaz Villegas, Andy Challinor (2012). Assessing relevant climate data for agricultural applications.

[4] Grajales, D. F. P., Mejia, F., Mosquera, G. J. A., Piedrahita, L. C., Basurto, C. (2015). Crop-planning, making smarter agriculture with climate data. 2015 Fourth International Conference on Agro-Geoinformatics (Agro-Geoinformatics).

[5] Niketa Gandhi, Leisa J. Armstrong, Owaiz Petkar, Amiya Kumar Tripathy (2016). Rice crop yield prediction in India using support vector machine.

[6] Shakoor, M. T., Rahman, K., Rayta, S. N., Chakrabarty, A. (2017). Agricultural production output prediction using Supervised Machine Learning techniques.2017 1st International Conference on Next Generation Computing Applications (NextComp).

[7] Sekhar, C. C., Sekhar, C. (2017). Productivity improvement in agriculture sector using big data tools. 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC).

[8] Garg, A., Garg, B. (2017). A robust and novel regression based fuzzy time series algorithm for prediction of rice yield. 2017 International Conference on Intelligent Communication and Computational Techniques (ICCT).

[9] Jheng, T.-Z., Li, T.-H., Lee, C.-P. (2018). Using hybrid support vector regression to predict agricultural output. 2018 27th Wireless and Optical Communication Conference (WOCC).

[10] Manjunath, M., Parkavi, A. (2018). Estimation of Arecanut Yield in Various Climatic Zones of Karnataka using Data Mining Technique: A Survey. 2018 International Conference on Current Trends Towards Con- verging Technologies (ICCTCT).

[11] Mr. A Suresh, Dr. P. Ganesh Kumar, Dr. M. Ramalatha (2018). Prediction of major crop yields of Tamil Nadu using K-means and Modified KNN.

[12] Xinran Gao, Jianxi Huang, Hongyuan Ma, Wen Zhuo, Dehai Zhu (2018). Regional Winter Wheat Maturity Date Prediction Using Remote Sensing-Crop Model Data Assimilation and Numerical Weather Prediction.

[13] Feilong Wang, Fumin Wang, Yao Zhang, Jinghui Hu, Jingfeng Huang, Lili Xie, Jingkai Xie (2019). Rice yield estimation at pixel scale using relative vegetation indices from unmanned aerial systems.

[14] Potnuru Sai Nishant, Pinapa Sai Venkat, Bollu Lakshmi Avinash, B. Jabber (2020). Crop yield prediction based on Indian Agriculture using Machine Learning.

[15] D. Jayanarayana Reddy, Dr. M. Rudra Kumar (2021). Crop Yield Prediction using Machine Learning Algorithm.