Crop Yield Prediction based on Climatic Parameters

Abhijeet Pandhe[1], Praful Nikam[2], Vijay Pagare[3], Pavan Palle[4], Prof. Dilip Dalgade[5]

[1,2,3,4]*Student, Department of Computer Engineering, Rajiv Gandhi Inst. of Technology, Mumbai, Maharashtra*
[5]*Professor, Department of Computer Engineering, Rajiv Gandhi Inst. of Technology, Mumbai, Maharashtra*

---***---

Abstract - *During the last decade the climate has become very uncertain. Due to this the farmers who were planting traditional crops are now facing problems, the yield of their traditional crops is getting reduced, therefore the rate of suicide done by farmers has increased specifically in India. In our study we found out that if farmers knew the yield of the crop that they are planting beforehand, they would choose the crop that will produce better yield based on the climate of that region. Knowing the crop yield before the harvest will also help farmers and policy makers to make important decisions like import-export, pricing, marketing, etc. The application (Smart Farm) that is developed in our research will help the user to predict the crop yield based on the climatic parameters. Several methods of predicting the crop yield have been developed in the past with different rate of success, as they don't take into account the climatic conditions. Machine Learning is an essential approach for getting practical solution of this problem. Random Forest algorithm is used to train the model so as to get accurate prediction, 5 climatic parameters were chosen to train the model. Other agro-inputs such as soil quality, pest, chemicals used, etc. were not used as they vary from field to field.*

Key Words: ***Machine Learning, Crop Yield, Prediction, Climate, Random Forest Algorithm, Application***

1. INTRODUCTION

Agriculture is backbone of India, but recently agriculture in India is wilting, almost all of the problems originating in policymaking, political structures and economics. The government encourages mostly rice and wheat, with a few others throw in, through instrumentalities like "minimum support price". Subsidizing rice and wheat, maybe a few other crops, is not a robust system. If the crops fail, the farmer has nothing to fall back on, and he goes into debt on top of what he owes in the previous crop season. An offshoot of chemically-intensive agriculture and subsidies for a few crops is mono-cropping on such a huge scale that India's diverse food cultures are on the brink of being lost. Soils are depleted, water is contaminated, farmers' incomes are forever falling, and food security is getting harder to achieve for all of India. This situation can be overcome by trying out new crops, but it is difficult as farmers are unaware of new crops and the yield they will produce. If farmer would be able to know the yield of the crop in advance, it will help him to choose which crop he should plant. The application developed in this research will provide farmers an user friendly interface to predict crop yield.
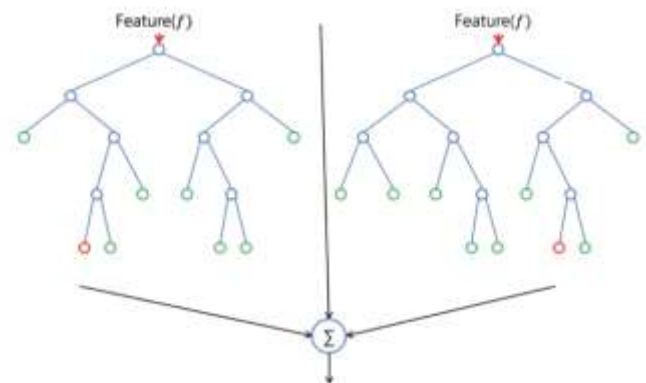
Machine Learning is a suitable approach for solving such type of problems. Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. As our system will be developed by creating a model, and training that model using the historical crop production and climate data. We will be using supervised machine learning to train the model and get accurate prediction for the future data. There are various supervised machine learning algorithms for regression, such as linear regression, Polynomial regression, Decision Tree, Random Forest, Support Vector Machine, etc. For getting high correlation between the desired and the actual outcome of prediction we are using Random Forest algorithm. In Random forest algorithm decision trees are created so that rather than selecting optimal split points, suboptimal splits are made by introducing randomness. The models created for each sample of the data are therefore more different than they otherwise would be, but still accurate in their unique and different ways. Combining their predictions results in a better estimate of the true underlying output value. As decision trees are used for making random forest over fitting does not take place in random forest algorithm.

Fig -1: Random Forest

We have used 5 climatic parameters for prediction, i.e. Precipitation, Temperature, Cloud Cover, Vapour pressure, Wet Day Frequency. 10-fold cross validation technique was used to find the accuracy of our trained model, the accuracy was found out to be 87%. It shows very high correlation between the depend and the independent variables of our model. Other agriculture related parameters such as soil quality, pest, chemicals used by farmers, etc. were not used as they vary from field to field.

2. METHODOLOGY

Data is very important part of any Machine Learning System. To implement the system we decided to focus on Maharashtra State in India. As the climate changes from place to place, it was necessary to get data at district level. Historical data about the crop and the climate of a particular region was needed to implement the system. This data was gathered from different government web sites. The data about the crops of each district of Maharashtra was gathered from www.data.gov.in and the data about the climate was gathered from www.imd.gov.in. The climatic parameters which affects the crop the most are precipitation, temperature, cloud cover, vapour pressure, wet day frequency. So the data about this 5 climatic parameters was gathered at a monthly level.

2.1 Cropping Seasons in India

Crops are grown over the year in India. But the crops can be classified based on the season they grow. There are 2 main cropping seasons in India: Kharif and Rabi. The kharif cropping season is from July - October during the south - west monsoon and the rabi cropping season is from October - March (winter). The crops grown between March and June are summer crops, and the crops which grow over the year are whole year crops. As the data of climate which was gathered was at a monthly level, dimensionality reduction was done by combining the columns based on the cropping season of each feature. The data of crop and the climate data was then combined into 1 dataset. The historical information of the crop such as district in which the crop is grown, name of the crop, the season of that crop and the 5 most influencing climatic parameters became the independent variables. And the yield of the crop will become the dependent variable. This is the dataset that will be used for training the model by using random forest regression.

2.2 Implementation

In this research we have developed an user friendly web application, the farmer will interact with the user interface of this system. There are three main entities of the system: User, Server, Climate Forecast API as shown in Fig-2. User will select the name of the district, name of the crop and the area of his field then he will submit this information to the server. The processing of this information will be done at the server side. The model is trained at server side using the random forest algorithm and the dataset.

Steps for training the system:

1. Importing the Required Libraries.
2. Importing the dataset and Feature Selection.
3. Taking care of missing data.
4. Encoding categorical data or the Independent Variable.
5. Splitting the dataset into the Training set and Test set.
6. Feature Scaling.

7. Fit Random Forest Regressor to the Training set.
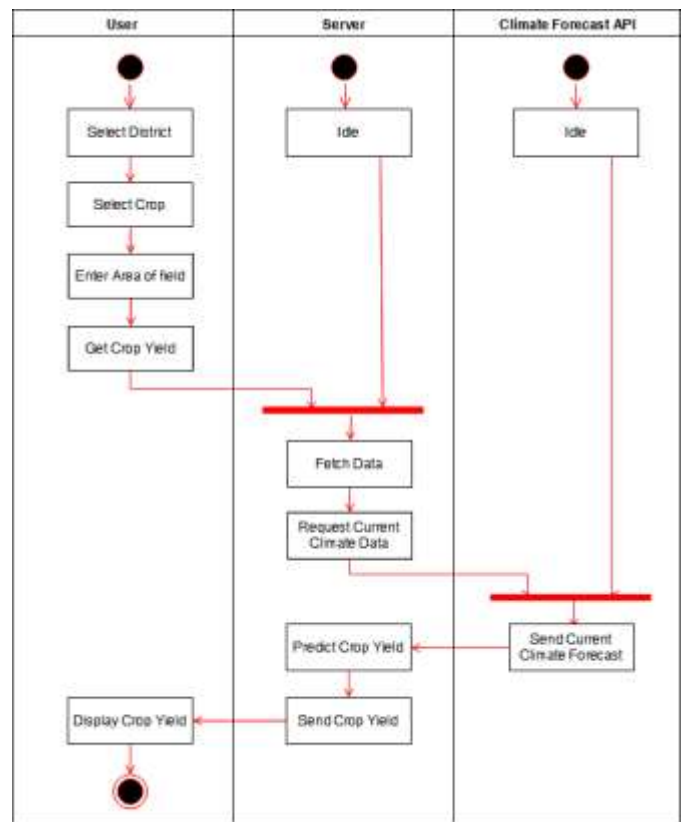8. Predicting a new result.

Fig -2: Activity Diagram

As random forest algorithm was used to train the model. The Pseudo code of random forest algorithm:

1. Randomly select "k" features from total "m" features. (Where k << m)
2. Among the "k" features, calculate the node "d" using the best split point.
3. Split the node into daughter nodes using the best split.
4. Repeat 1 to 3 steps until "l" number of nodes has been reached.
5. Build forest by repeating steps 1 to 4 for "n" number times to create "n" number of trees.
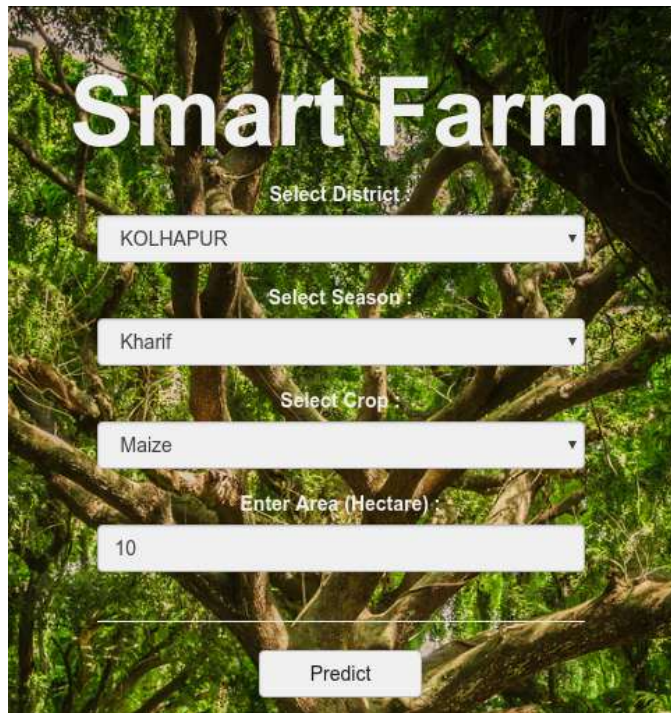
The model is trained by considering 200 decision trees to construct the random forest. 10-fold cross validation technique was used to find the accuracy of the model, Cross-validation is a re-sampling procedure used to evaluate machine learning models on a limited data sample. The accuracy of model was found out to be 87 %

Server will receive the information given by the user, for making a new prediction the model will also need the climate information about the season of that crop. Server will get the information from a climate forecast API, the server will send request for the climate information of that season and API will respond with the climate information of that season.

Then the server will make new prediction by using the trained model. This predicted value of crop yield will then be sent to the user and it will be displayed on the user interface.
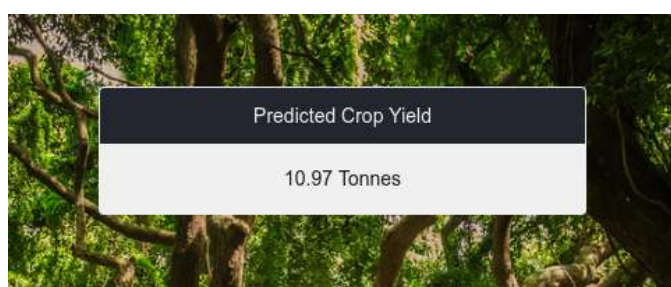
2.3 Results

The aim of the research was to construct an user friendly web application that will help the farmers and other policy makers to predict the crop yield based on the climatic parameters. A web application 'Smart Farm' was thus developed.

Fig -3: Smart Farm, Home Page

Fig -4: Smart Farm, Result Page

The backend of the system was made using python and the results of the application were rendered into a web page by using Flask. As you can see in Fig-3, it is the home page of the application where the farmer or the user of the application will enter the details such as name of district, season, name of crop and area of his field in hectare. After entering the details user will press 'Predict'. Which will send request to the server, and here the prediction will be made using the model which is trained by using the random forest algorithm. As you can see in Fig-4, the result of the prediction

of the crop yield will be send to the user and it will be displayed in tonnes.

3. CONCLUSION

The accuracy of the model is 87% which was found using 10-fold cross validation technique, it indicates that there is high correlation between the climate and the crop yield. Other agro-inputs such as soil quality, pest, chemicals used, etc. were not used as it changes from field to field. As the climate has become very uncertain, such application will help farmers to make better decision on which crop to plant based on the climate of that season. It will also help policy makers to make important decisions like import-export, pricing, marketing, etc. even before the crop is harvested.

REFERENCES

[1]   P. Priya, U. Muthaiah and M. Balamurugan, "Predicting yield of the crop using machine learning algorithm," International Journal of Engineering Science & Research Technology, vol. 7(4), April 2013

[2]   Jig Han Jeong and team, "Random forests for global and regional crop yield productions," PLOS ONE, vol. 11(6), June 2016

[3]   S. Veenadhari, Dr. Bharat Misra and Dr. CD Singh, "Machine learning approach for forecasting crop yield based on climate," International Conference on Computer Communication and Informatics, Conference Paper, January 2014

[4]   Subhadra Mishra, Debahuti Mishra and Gour Hari Santra, "Application of machine learning techniques in agricultural crop production," Indian Journal of Science and Technology, vol. 9(38), October 2016