CROP YIELD PREDICTION USING MACHINELEARNING ALGORITHMS

Dr. MURALI G, BAKKA CHENNA HARI CHARAN REDDY, BHARATH V S, DHANUSH S R, DILIP S
(Assistant Professor, Dayananda Sagar Academy of Technology and Management,Bangalore)(Dayananda Sagar Academy of Technology and Management,Bangalore)

Abstract: India is an Agriculture based economy whose most of the GDP comes from agriculture and its allied sectors which accounts for 20% of total GDP. The motivation of this project comes from the increasing suicide rates in farmers which may be due to low yield in crops. Climate and other environmental changes have become a major threat in the agricultural field. Machine learning is an essential approach for achieving practical and effective solutions for this problem. Predicting yield of the crop from historical available data like weather, soil, rainfall parameters and historic crop yield. We aim to achieve this using the machine learning algorithm. We did a comparative study of various machine learning algorithms, i.e., ANN, K Nearest Neighbor, Random Forest, SVM and Linear Regression. In this project a mobile application will be developed which predicts the crop yield in general and also for a particular crop. Along with that, it will suggest to the user which crop to grow in that particular season.

I. INTRODUCTION

India is ranked 2nd in farm output.As of 2018 more than 50% of indian workforce is involved in agriculture and it contributed to country's GDP of 17-18%.The financial contribution of agriculture to india's GDP is declining.The crop yield depends on variousfactors like climate, geographical location,financial elements etc.It is difficult for farmers to choose which crop to grow as there multiple crops available and market prices are not known to them.The google statistics showed the farmer suicide rate in india which ranged from 1.6 and 1.9 per 100000 total population over a period of 10years.While 2020 saw 5098 farmers suicide and was increased by 18% from last year.

In the recent times,it has become a mandate to use technology to create awareness in farming.The seasonal climate conditions are being changed against soil,water which is leading to inconsistency in food.The crop yield rate to consumption demand is falling short and there is a need for smart system to solve this problem and reduce the inconsistency in crop yield.Therefore,we propose a system which will provide crop selection depending on economic and environmental factors to get maximum yield to farmers and to satisfy the needs of country's demand in food.The proposed system uses machine learning for prediction.The system will produce crop prediction and yield prediction based on weather attribute that is suitable for crop to get maximum yield to farmers.The system makes prediction of crops and yield based on the factors such as rainfall,temperature,land size(hectares),season, etc.

Crop yield prediction is an important aspect in farming. Every farmer always tries to know how much yield will be produced and whether it meets their expectations.In the past yield prediction was done based on farmers experience.The crop yield mainly depends on weather conditions and harvesting operation.Accurate information about history of crop yield is important factor for making decisions using machine learning which is related to agriculture risk management.

II. SCOPE OF PROJECT

Integrating farming and machine learning, we can lead to further advancements in agriculture by maximizing yield and optimizing the use of resources involved. Previous year‘s production data is an essential element for predicting the current yield.
The goal of this project is to help the farmers by combining agriculture and technology. The end result is an application that is available on mobile. The application has the following features:

i. **Login/Register :** The user can register themselves by providing a username of their choice and a password. After the registration, they can login to use the application further and view all the options that are provided to them.

ii. **Prediction :** This is one of the modules available in the application that enables the user to view the yield predictions of crops. The user is given two choices here:

- I know what to plant : This option is for those users who already have a crop in their mind that they want to grow. When chosen, the user will be given choices of crops that they must select along with other inputs .i.e., Area and the soil type. After processing the inputs, the application will return the predicted yield on the user‘s screen.
- Yet to decide the crop : This option is when the user is not sure between some crops or has no crop in mind. The user has to input the soil type and the area. The input is again processed at the back end by the modelled algorithm.

iii. **Sign Out :** The user may sign out at the end which will take them back to the login/register page.

Agriculture is one of the main sources of income in India and there is an enormous need to maintain agricultural sustainability with the increasing rate of farmer suicides. Hence it is a significant contribution towards the economic and agricultural welfare of the countries across the world.

III. DATASET DESCRIPTION

The section describes the crop dataset providing information on production from 1997 to 2014. The section provides details on the attributes, instances, missing values in the dataset.

Number of attributes: 10
Number of instances: 246092

State_Name - The attributes contain the information about the State name of harvest
District_Name - The attributes contain the information about the District name of harvest
Crop_Year - The attributes contain the information about the Year of harvest
Season - The attributes contain the information about the Season of harvest
Crop - The attributes contain the information about the Crop grown
Area - The attributes contain the information about the Area of cultivation
Temperature - The attributes contain the information about the Temperature
Humidity - The attributes contain the information about the Humidity
Rainfall - The attributes contain the information about the Rainfall
Production - The attributes contain the information about the Production values

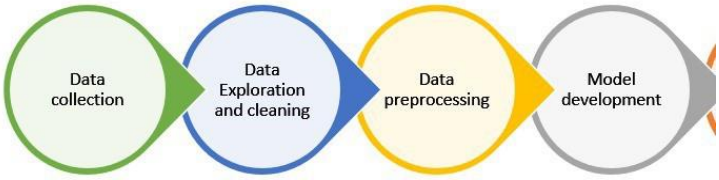
IV. METHODOLOGY

Data Pre-Processing
Data Preprocessing is a method that is used to convert the raw data into a clean data set. The data are gathered from different sources, it is collected in raw format which is not feasible for the analysis. By applying different techniques like replacing missing values and null values, we can transform data into an understandable format. The final step on data preprocessing is the splitting of training and testing data.

Data is a very important part of any Machine Learning System. As the climate changes from place to place, it is 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 websites. The data about the crops of each districts of India was gathered from www.data.gov.in and the data about the climate was gathered from www.imd.gov.in. The climatic parameters which affect the crop the most are precipitation, temperature, cloud cover, vapour pressure, wet day frequency. So, the data about these climatic parameters was gathered at a monthly level.

Dataset Collection: In this phase, we collect data from various sources and prepare datasets. And the provided dataset is in the use of analytics (descriptive and diagnostic). There are several online abstracts sources such as Data.gov.in and indiastat.org. For at least ten years the yearly abstracts of a crop will be used. These datasets usually accept behaviour of anarchic time series. Combine the primary and necessary abstracts. Random Forests for Global and Regional Crop Yield Predictions.

Data Partitioning: The Entire dataset is partitioned into 2 parts: for example, say, 75% of the dataset is used for training the model and 25% of the data is set aside to test the model. To predict future events Machine Learning

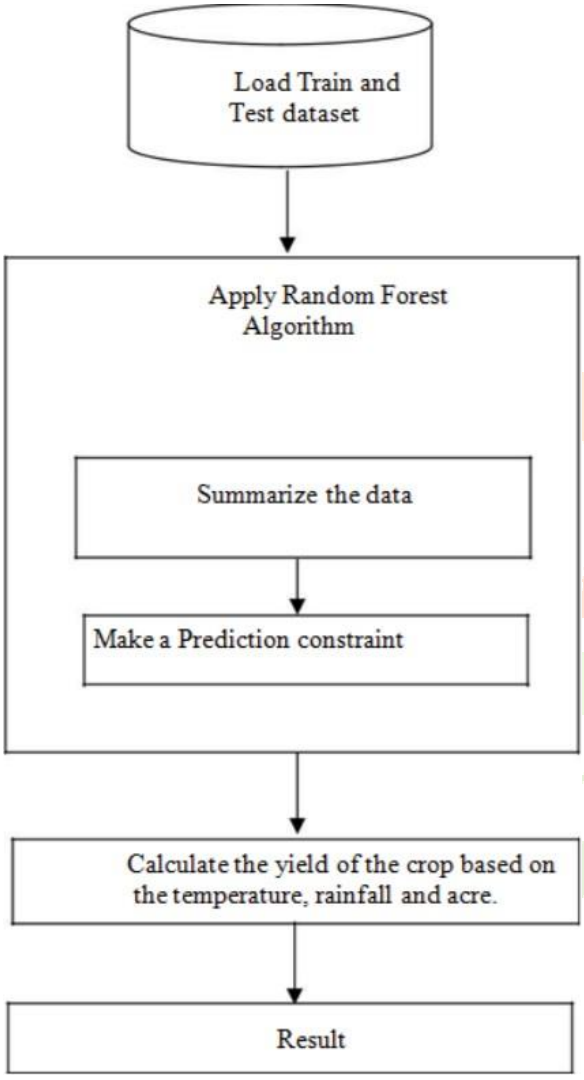


Algorithms:

Supervised learning: Supervised machine learning algorithms can apply what has been learned in the past to new data using labelled examples. After Sufficient training the system can provide targets for any new input. IN order to change the model accordingly the learning algorithm can also differentiate its results with the correct, intended output and find errors.

Unsupervised learning: unsupervised machine learning algorithms are used when the information used to train is neither labelled nor classified. Unsupervised learning does analysis of how systems can infer a function to describe a hidden structure from unlabelled data. In order to describe hidden structures from unlabelled data the system doesn't figure out the right output, but it examines the data and can draw inferences from datasets.

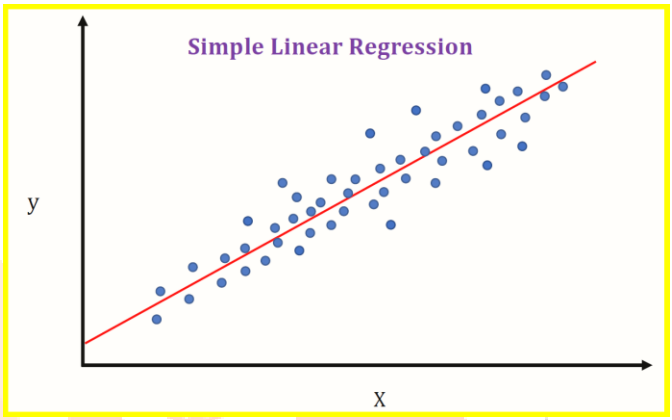
Random Forest Classifier: Random forest is the most popular and powerful supervised machine learning algorithm capable of performing both classification and regression tasks, that operate by constructing a multitude of decision trees at the time of training and generating outputs of the class that is the mode of the classes (classification) or mean prediction (regression) of the individual trees. The more trees in a forest the more robust the prediction.



Naive Bayes:- Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. The Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Linear Regression:- Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression models a target prediction value based on independent variables. 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 being used.

Linear regression performs the task to predict a dependent variable value (y) based on a given independent variable (x). So, this regression technique finds out a linear relationship between x (input) and y(output). Hence, the name is Linear Regression.



V. LITERATURE SURVEY

[1] Smart Farm Web Application Using Machine Learning Algorithm by Indhra Priyadharshini, Abhinaya, Indhuja, Keerthana Shree.

E-Farming is a web application developed for farmers. The app supports village farmers who want to use the facility and find out how it is possible and how to use e-farming. If farmers have computer knowledge, they can register directly on the site and sell their products. The project was developed to create a website that uses ML algorithms to estimate crops and their prices, and this farm information on crops, statistics and new trends.

[2] Crop Yield Prediction using Machine Learning Algorithms by Anakha, Aparna, Jisu Mani.

The core emphasis would be on precision agriculture, where quality is ensured over undesirable environmental factors. So as to perform accurate prediction and stand on the inconsistent trends in temperature and rainfall various machine learning classifiers like Logistic Regression, Naïve Bayes, Random Forest etc. are applied to urge a pattern. By applying the above machine learning classifiers, we came

into a conclusion that the Random Forest algorithm provides the foremost accurate value. System predicts crop prediction from the gathering of past data. Using past information on weather, temperature and a number of other factors the information is given. The Application which we developed, runs the algorithm and shows the list of crops suitable for entered data with predicted yield value.

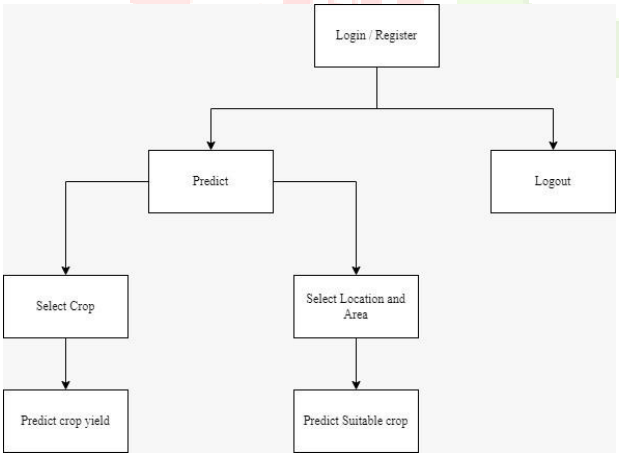
[3] Prediction of Crop Yield using Machine Learning Techniques by Prajwal G.S, Manoj G.S, Anitha, Prashanth.

This model is going to help farmers to make better decisions as to decide which crop to plant. Based on the season's climate it will help farmers to make important decisions, such as import, export, pricing, marketing before the crop is harvested .Crop production was influenced by the various economy, season and Agriculture in this country assumes a significant job in economy and work. This technique is described by a dirt database gathered from the ranch, crop gave by farming specialists, accomplishment of parameters, for example, soil by soil testing lab dataset.

[4] Analysis of Crop Yield Prediction using Data Mining Techniques.

Data Mining is the process of extracting helpful and significant information from huge sets of data. Data Mining in the agriculture field is a comparatively novel research field. Yield prediction is a very important agricultural problem. Any farmer is interested in knowing how much yield he is concerned to be expecting. In the earlier period, yield prediction was performed by considering farmer's experience on a particular field and crop. In any of Data Mining actions the training data is to be collected from past data and the gathered data is used in terms of training which has to be exploited to study how to categorize future yield predictions.

V. SYSTEM ARCHITECTURE



System architecture mainly consists of weather API where we fetch the data such as temperature, humidity, rainfall etc. The data fetched from the API are sent to the server module. The data gets stored on to the database on the server. Using the mobile application, the user can provide details like location, area, etc. The user can create an account on the mobile app by one-time registration and all this entered data is sent to the server. The trained Random forest model deployed on the server uses all the fetched and input data for crop yield prediction, finds the yield of predicted crop with its name in the particular area

VI. SYSTEM ANALYSIS

Python 3.8.5(Jupyter Notebook):

Python is the coding language used as the platform for machine learning analysis. Jupyter Notebooks illustrates the analysis process and gives out the needed result.

Weather API (Open Weather Map):

Weather API is an application programming interface used to access the current weather details of a location. The generated API key illustrates current weather forecast needed for crop prediction.

Android Studio (Version 3.4.1):

Android Studio is the official integrated development environment (IDE) for Android application development. This paper uses java as the framework for frontend designing. USB debugging method is used for the connection of IDE and app.

Python Flask Framework (Version 2.0.1):

Flask is a micro framework in python. Flask is based on WSGI. (Web Server Gateway Interface) toolkit and Jinja2 template engine. In this paper flask is used as the back-end framework for building the application. It is the collection of modules and libraries that helps the developer to write applications without writing the low-level codes such as protocols, thread management, etc.

VII. CONCLUSION

This paper focuses on predicting the crop yield and suitable crop with the help of machine learning algorithms. This system is proposed to deal with the increasing rate of farmer suicides and to help them to grow financially stronger. Appropriate datasets are to be collected, studied and trained using machine learning tools. This system tracks the user's location and fetches needed information from the backend based on the location. Thus, the user needs to provide a few details like location and area of land. This system provides the yield of the crop and which crop to grow in a particular season

VIII. FUTURE WORK

The Future Work targets a fully automated system that will do the same. Another functionality that we are trying to implement is to provide the correct fertilizer for the given crop and location. To implement this through study of fertilizers and their relationship with soil and climate is required. We are also aiming to predict the crisis situation in advance like the recent hike of tomato prices.

XI. REFERENCES

[1] Smart Farm Web Application Using Machine Learning Algorithm by Indhra Priyadharshini, Abhinaya, Indhuja, Keerthana Shree.
[2] Crop Yield Prediction using Machine Learning Algorithms by Anakha, Aparna, Jisu Mani.
[3] Prediction of Crop Yield using Machine Learning Techniques by Prajwal G.S, Manoj G.S, Anitha, Prashanth.
[4] Analysis of Crop Yield Prediction using Data Mining Techniques.by Ms. Fathima, Ms. Sowmya K, Ms. Sunita Barker, Dr. Sanjeev Kulkarni
[5] Aruvansh Nigam, Saksham Garg, Archit Agrawal "Crop Yield Prediction using ML Algorithms ",2019
[6] Dr.Y Jeevan Kumar,"Supervised Learning Approach for Crop Production",2020
[7] Ramesh Medar,Vijay S, Shweta, "Crop Yield Prediction using Machine Learning Techniques", 2019
[8] Sangeeta, Shruthi G, "Design And Implementation Of Crop Yield Prediction Model In Agriculture",2020