Crop Yield Prediction Using Naïve Bayes Algorithm

Jitendra Chavan, Nagesh Pawade, Akshay Tale, Amit Kadam , Amit Gujar

Information Technology Department, Marathwada Mitra Mandal's College of Engineering,

Pune, Maharashtra, India

ABSTRACT

Agriculture is that the backbone of India and it plays important role in economy.It Is source of production for about 58 per cent of India's population in step with government of the day's estimates, States's food production was about 291.95 MT in 2019-20; for 2020-21, the govt. had set the target up to 298.3 MT, which was two per cent more from the previous year's output.Food production must double by 2050 to match the country's population and income growth. the small and marginal farmers, therefore, have a big role within the country's food security and meeting the SDG goals.Nearly 14 per cent of the population (189.2 million) continues to be undernourished in India, in line with State of Food Security and Nutrition within the planet, 2020 report. the planet Hunger Index 2020 placed India at the 94th position among 107 countries. Achieving 'zero hunger' by 2030 is also a humungous challenge, and needs an integrated and multi-dimensional approach for overall sustainable agriculture and food systems within the country.

Keywords: KNN, Naïve Bayes,Crop prediction.

I.   INTRODUCTION

India is an agricultural country. India's economy is determined by agricultural products export and import. Agriculture is one altogether the important aspects of Indian economy. thanks to uncertainty within the crop yield there's a wonderful fall within the economic status. the foremost crops of India are Rice, Wheat, Pulses and Grains. Day by day the population of India is growing and also the crops productivity must be increased to feed the population. one in every of the foremost effective ways of predicting unknown values is by use of machine learning algorithms. This work intends to develop crop prediction model using machine learning. the appliance intends to predict crop yield so it could help farmer to choose on best seeds for plantation. There are plenty of ML algorithms which can preferably be used, algorithms like statistical procedure, Support Vector Machine, Neural Networks, K-Nearest Neighbor (K-NN) could also be utilized. during this work we discuss about K-NN.

The k-nearest neighbors (KNN) algorithm may be an easy, supervised machine learning algorithm which is able to be accustomed solve both regression and classification problems. It's easy to implement and understand, but includes a serious drawback of

becoming significantly slows because the dimensions of that data in use grows. Here objective is to use a model where information focuses are clustered during some groups so on predict the classification of another instance. K-NN works supported minimum distance from query instance to the training samples to determine the k-nearest neighbors. Then we collect knearest neighbors, we take simple majority of these k-nearest neighbors to be the prediction query object. As mentioned before it can even be used for regression-output, which is that the item's reward .Mostly for distance calculation in K-NN algorithm the metric used is Euclidean distance.

But when the farmers know the accurate information on the crop yield it will help to minimizes the loss. Machine learning, a fast-growing approach thats spreading out and helping every sector in making viable decisions to make the foremost of its applications. Most of the modern devices are facilitated by models being analyzed before deployment. the foremost concept is to increase the throughput of the agriculture sector with using the Machine Learning models. Another factor that also affects the prediction is that the number of knowledge thats being given within the training period, because the amount of parameters was higher comparatively. The core emphasis would air precision agriculture, where quality is preferred over undesirable environmental factors.

So on perform accurate prediction and stand on the variablility in temperature and rainfall ,various machine learning classifiers like Logistic Regression, Naive Bayes etc. are applied to urge a pattern. By applying the above machine learning classifiers, we came into a conclusion that Naive Bayes algorithm provides the foremost accurate value. System predicts crop prediction from the gathering of past data.

II. MOTIVATION

Motivation for this research comes from the agricultural point of view that there's lots of labor for the farmers to be done manually. So this research deals with the automated classification approach which provides help to the farmers and reduce their efforts. With the change of climate within the world, the mainstreams of the crops within the agricultural field are severely affected in performance in previous decades. The crop prediction yielding should be healthy to assist the farmers for taking suitable procedures for selling and loading/storage. Such estimations on the yielding of the crops also will assist the industries supported the farming and production of the crops for development of the logistics of their commercials. Numerous approaches of estimating and crop yielding are established within the previous times with variable success rates.

III. LITERATURE SURVEY

SR.NO	PAPER TITLE	AUTHOR	DESCRIPTION
1	Crop Yield Prediction Using Machine Learning Algorithms 2019	Thomasvan K lompenburg AyalewKassa Hun Cagatay Catal	The Propose system incorporates data mining and android application technology to predict on the crop yield rate of a given year byimplementing KNNalgorithms and analyzing past days average temperature average rainfall.

2	Crop Prediction And Disease Detection	Mukul Chhanikar, Rushikesh Thakare, Aniket Tapre, Sunil Thorat	Neural Network performs better on downfall prediction and LSTM is best for temperature prediction. The researchconsiders factors like temperature, rainfall other seasonal variations, area etc. together for yield prediction. Results reveals that Random Forest is the best classier when all parameters are combined
3	Application of machine learning techniques in agricultural crop production,	Leo Brieman	Author is dealing with the accuracy and strength & correlation of random forest. Random forest creates decision trees on various data samples and then predict the data from every subset .
4	Machine Learning Algorithms in Forecasting	Di Su; Efstratios Batzelis; Bikash Pal	Creator talked about different AI strategies that can be utilized in different determining regions. In any case, their work neglects to carry out any calculations and hence can't give an unmistakable understanding into the common sense of the proposed work

IV. DATA COLLECTION

In this project, the machine learning technique is applied to create a basic web application which will predict the crop yield by applying machine learning algorithm to historical crop yield data using python libraries like Pandas, NumPy, Matplotlib, seaborn and sklearn.

The collection of data is the most important aspect of this project. There are various sources of the data on different websites which are used to train the models.

1. Data Collection:

Data selection is the first step where historical data of crop yield is gathered for the model to predict crop yield. Our dataset consists of more than 10,000 records of data related to crop.

2. Cleaning and preparing data:

In this study, the dataset is cleaned by removing the duplicate values and null values. If these values are not removed it would affect the accuracy of the model. All the collected data needed a lot of work so

after the collection of data, it is necessery to be clean and prepare according to the model requirements.

3. Analysing data: Information planning is trailed by investigating the information, uncovering the secret patterns and afterward applying different AI models.

V. PROPOSED MODEL

Machine Learning Algorithm

ML calculations are numerical model planning techniques acclimated learn or uncover basic examples implanted inside the information. ML involves a group of computational calculations which will perform design acknowledgment, arrangement, and forecast on information by gaining from existing information (preparing set). Machine learning algorithms are wont to automatically understand and realize the day-to-day problems that individuals face. the amount of hidden layers in a man-made neural network reflects within the kind of learning. The intent is to achieve knowledge by learning through datasets using customized methods. they're as follows: R-squared value, Mean Absolute Error (MAE) and Mean Squared Error (MSE). The formulas for these three parameters are as follows:

$$R^2 = 1 - \frac{\sum_n^{t=1}(yi - \hat{yi})^2}{\sum_n^{t=1}(yi - \bar{yi})^2}$$

$$MAE = \frac{1}{n}\sum_n^{t=1}|yi - \hat{yi}|$$

$$MSE = \frac{1}{n}\sum_n^{t=1}(yi - \hat{yi})^2$$

Machine learning may be supervised Learning, or non-supervised Learning. Problems in Machine Learning include-

➤ Clustering: Clustering could be a machine learning technique, which group unlabelled Dataset. It is defined as "A way of grouping the information points into different clusters, consisting of comparable data points".

➤ Regression: Regression model are accustomed predict a nonstop value. it's supervised techniques.

➤ Classification: Classification refers to a predictive modelling problem where as a category label is predicted for a given example of input file.

K-NN Algorithm:

K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm considers the commonality between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN calculation stores every one of the accessible information and groups another information point in view of the shared characteristic.

This implies when new information shows up then it tends to be effectively arranged into a well suite class by utilizing K-NN calculation. K-NN calculation can be utilized for Regression as well with respect to Classification however generally it is utilized for the Classification issues. K-NN is a **non-parametric algorithm**, which means it does not make any assumption on underlying data. It is also called a **lazy learner algorithm** because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data..

Naïve Bayes Algorithm:

Naïve Bayes is one among the foremost efficient and effective induction learning algorithms for machine learning and data processing. It's competitive performance in classification is surprising because the conditional independence assumption on which it's based isn't true in world applications.

A. Architecture

The system will follow Naive Bayes classifier, the supervised learning algorithm contains the four levels to calculate and predict the crop for the acceptable climate in phenomenon like, the architecture is shown inthe Fig. 4.1..

System Architecture
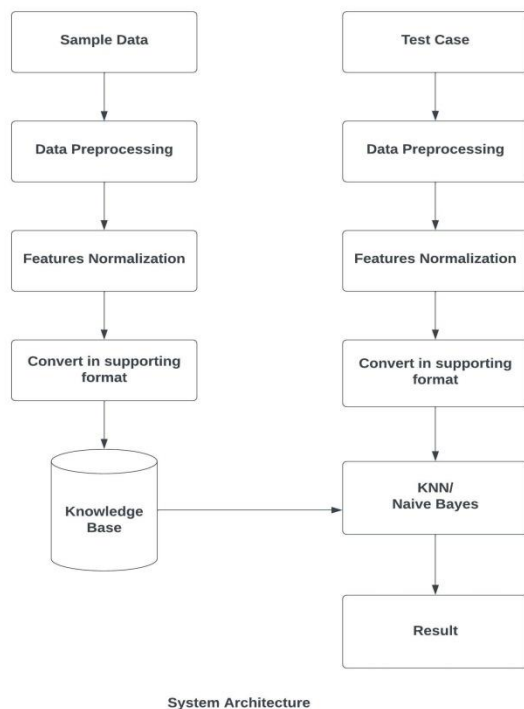
a. Data collection :

Data consists from a unique source and optimized for data sets. and also the data are going to be use to judge descriptively. Several abstract online outlets, like Kaggle, Google weather forestation and data government, provide the information for up to 10years serial. the info sets like soil nature, climate and seed data are used for the crop prediction and better crop yields.

b. Processing step :

Preprocessing the information is taken into account as significant step a machine learning phase.

Preprocessing involves adding the missing values, the right set of information, and extracting the functionality. Data set form is very important to the method of study. the info collected during this step will induced in Google Colab platform within the type of python programming so as to induce the required output.

c. Feature Extraction:

Extracting the features would shrink the data size involved to characterize a wide collection of data. The features of soil, crop and rainfall collected from the pretreatment process establish the final training data collection. This approach selects the features grounded on the correlation matrix i.e. the features that has further correlation value is named as an important prophetic function for yield.

d. Data Prediction

In advance to this step there need to resolve the data into train dataset and test dataset. By applying the Naïve Bayes Gaussian classifier the data is trained with available input and affair data. In the test phase, the data are tested if the delicacy of e model is satisfied. Also the new data is prognosticated by machine literacy module.

In Naïve Bayes system the delicacy of the model is 97. In order to ameliorate the delicacy boosting system is used. In boosting system the weak rule has to find and combine these rules to make it stronger. The weak rule has been plant by iterative process. The weak rule was plant by distribution of Naïve Bayes algorithm.

For right distribution

1. The distribution taken by base learner and lot same weight to all observation.
2. Still, also further attention has to pay, If any error passed by vatication of Ist Naïve Bayes base

algorithm. Also coming NaïveBayes base algorithm has to apply.

3. Repeat the step 2 untill it gives advanced delicacy.

VI. GUI

Shown below are the screenshots of the various activities from the web application along with their description:-
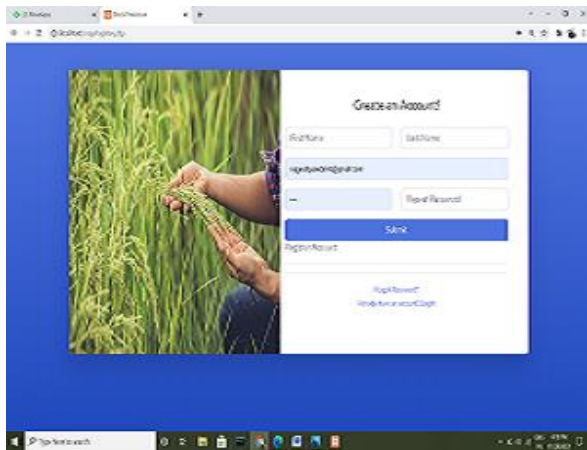
Description:

This is the Sign-up page of our website. This page offers the all necessary tabs which are useful for the user to create account for crop yield.
This website includes the following tabs:
1. First Name
2. Last Name
3. E-mail Id
4. Password
5. Reset Password
6. Submit

This is the Sign-up page of our website. This page offers the all necessary tabs which are useful for the user to create account for crop yield.
This website includes the following tabs:
1. First Name
2. Last Name
3. E-mail Id
4. Password

5. Reset Password
6. Submit

Login page as Follows:
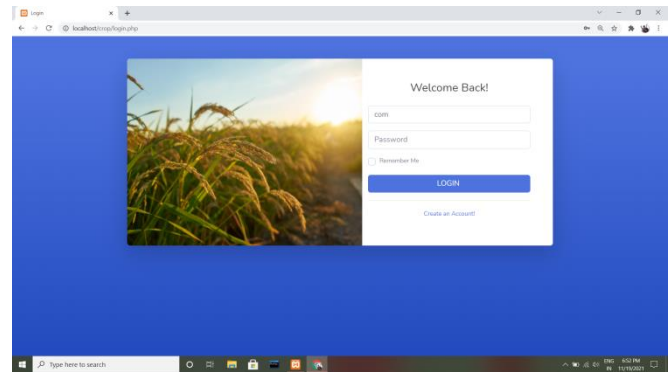
Description:

After clicking on Sign-In tab on the Home Page, the user is able to log in our website. If the user is already registered on our website then he/she directly log into the website, but if he/she is not registered then he/she wants to register and then log into the website.

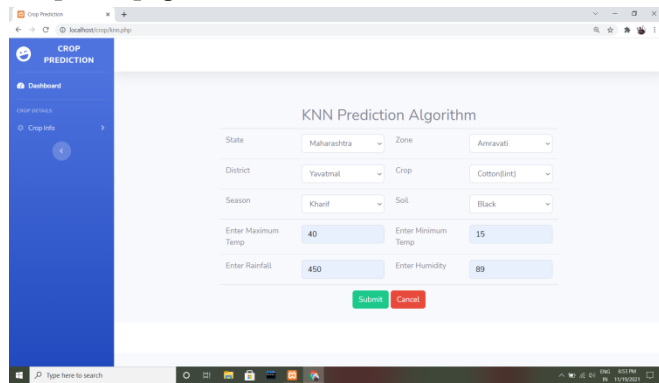This website includes the following tabs:

1. E-mail Id
2. Password

Dashboard is as fallows:
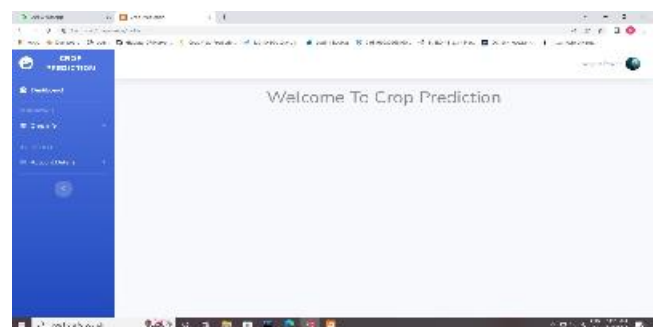
Description:
After user get login into the system then user will redirect to the dashboard page. In that page you will see crop details option after clicking on it will redirect to crop details in that this page contain crop info and crop prediction.
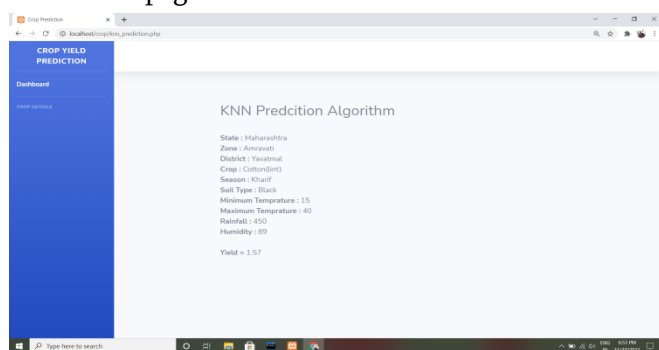
Crop info page is as fallows:

Description:

After clicking on crop info page the user will able to fill the details for prediction.

Details are as fallows:

1.State

2.District

3.Season

4.Enter maximum temperature

5.Rainfall

6.Zone

7.Crop

8.Soil

9.Enter minimum temperature

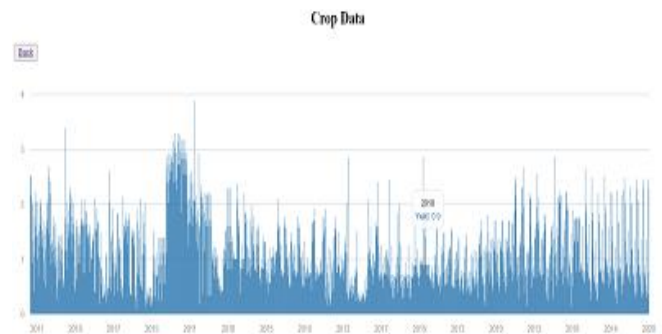10.Humadity

Prediction page is as fallows:

Description:

In that system will predict crop yield based on KNN algorithm which we have used in our proposed system.

Result page is as fallows:

Following are the result snapshot of the proposed system :

VII. CONCLUSION

The application on successful implementation will helps farmers in deciding of which crop to cultivate within the field. This work is used to look out the gain knowledge about the crop which will be deployed to create an efficient and useful harvesting.This improves our Indian economy by maximizing the yield rate of crop production. This paper focuses on the prediction of crop and calculation of its yield with the assistance of machine learning techniques. Several machine learning methodologies used for the calculation of accuracy. Implemented a system to crop prediction from the gathering of past data. The proposed technique helps farmers in deciding of which crop to cultivate within the field. This work is used to look out the gain knowledge about the crop which will be deployed to create an efficient and useful harvesting.

VIII. REFERENCES

[1]. Smart Crop Prediction using IoT and Machine Learning International Journal of Engineering Research Technology (IJERT) ISSN: 2278-0181 Published by, www.ijert.org NTASU – 2020 Con- ference Proceedings.

[2]. Crop Prediction and Disease Detection .International Research Journal of Engineering

and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 11 — Nov 2019.

[3]. Smart Agricultural Crop Prediction Using Machine Learning Journal of Xi'an University of Ar- chitecture Technology Volume XII, Issue V, 2020 ISSN No : 1006-7930.

[4]. Crop Yield Prediction Using Machine Learning Algorithms 2019 Fifth International Confer- ence on Image Information Processing.

[5]. Balamurugan have implemented crop yield prediction by using only the random forest classi- fier. Various features like rainfall, temperature and season were taken into account to predict the crop yield. Other machine learning algorithms were not applied to the datasets. With the absence of other algorithms, comparison and quantification were missing thus unable to provide algorithm.

[6]. Aruvansh Nigam, Saksham Garg, Archit Agrawal conducted experiments on Indian govern- ment dataset and its been established that Random Forest machine learning algorithm gives the best yield prediction accuracy. Sequential model thats Simple RNN works for rainfall prediction while LSTM is good for temperature . The paper puts several factors like rainfall, temperature, season, area etc. together for yield prediction. Results shows that Random Forest is best classier algorithm when all parameters are combined.