

CROP PREDICTION USING KNN ALGORITHM

B. LAKSHMI PRAVEENA¹, A. BHUMIKA², A. SINDHUJA³ B. TEJASWINI
MALLIKA⁴

1: Assistant Professor, Department of Information Technology

2,3,4: Student, Department of Information Technology
Vasireddy Venkatadri Institute of Technology, Nambur, India

Abstract- *India is generally an agricultural country. Now a days the most important emerging field in the real world is agriculture and it is main occupation and backbone to our country. Recent developments in Information Technology for agriculture field has become an interesting research area to predict the crop yield. Crop yield prediction is the methodology to predict yield of the crops using different parameters like PH, temperature, fertilizers and other atmospheric conditions are parameters. We have used combined datasets from different official government websites which includes features like crop name, area, production, temperature, rainfall, humidity and wind speed. After successfully extracting the data set, six machine learning algorithms (logistic regression, random forest, support vector machine, k-nearest neighbor, naive bayes classifier and feed forward neural network) were used to establish models. Among these machine learning models, we used K-Nearest Neighbor(K-NN) algorithm for brief analysis of crop yield. The developed model can have the accuracy of predictions more than 80%. We can further recommend which crop has to be yielded based on the season, climatic conditions and PH of the soil.*

Keywords: *Agriculture, Crop-Prediction, K-NearestNeighbor*

I. INTRODUCTION

India is an agricultural country. India's economy is determined by agricultural products export and import. Agriculture is one of the important aspects of Indian economy. Due to uncertainty in the crop yield there is a great fall in the economic status. The major crops of India are Rice, Wheat, Pulses and Grains. Day by day the population of India is growing and the crops productivity need to be increased to feed the population. One of the best ways of predicting unknown values is by use of machine learning algorithms. This work intends to develop crop prediction model using machine learning. The application intends to predict crop yield so it could help farmer to choose best seeds for plantation. There are plenty of ML algorithms which could be used, algorithms like Regression analysis, Support Vector Machine, Neural Networks, K-Nearest Neighbor (K-NN) can be utilized. In this work we discuss about K-NN. The k-nearest neighbors (KNN) algorithm is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It's easy to implement and understand, but has a major drawback of becoming significantly slower as the size of that data in use grows. Here objective is to use a model where information focuses are clustered in a few groups in order to predict the classification of another instance. K-NN works based on minimum distance from query instance to the training samples to determine the k-nearest neighbors. Then we collect k-nearest neighbors, we take simple majority of these k-nearest

neighbors to be the prediction query object. As mentioned before it can also be used for regression-output, which is the item's reward. Mostly for distance calculation in K-NN algorithm the metric used is Euclidean distance.

II. PROBLEM STATEMENT

In country like India the production of crops are affected by several factors. Factors like Humidity, temperature, rainfall, soil type play a vital role in crop prediction, and factors like these differ by large with respect to region. In India farmers majorly still rely on traditional techniques inherited from their forefathers. These techniques would work earlier when the climate was much healthier and predictable. Now with factors like global warming and pollution affecting the environment people have to be smart and start utilizing modern techniques. It is time to analyze large set of data and come up with a system that can provide sufficient information regarding crop yield. The new age methodology requires large structured data sets and an algorithm capable of providing solution using the provided

datasets.

III. METHODOLOGY

A. Dataset Collection

When implementing an accurate prediction model it might not be sufficient to just consider one or two parameters. Data about temperature, PH and various other factors are collected and analyzed. This analysis will be fed to the prediction model.

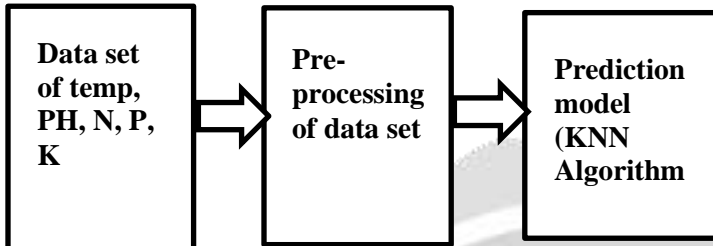


Fig 1:- Flow graph of the methodology Data Collection Here we gather information from several sources

and

construct datasets. Plenty of online portals like kaggle, and Data.gov.in are available for information collection. Annual crop report of each crop is collected Collecting previous crop history data. Collecting data related to crops like sugarcane, Onions, Oranges, Garlic, Ginger, Peas, Paddy, Ground nut, Potatao, Tomato are the crops which are commonlygrown in the regions. We also collect data related to Rainfall. Humidity,

Soil type, Irrigation type, Previous Yields, Location, Price, Year, type of crop, Crop diseases and its symptoms

B. K-NN Algorithm

The k-nearest neighbor (k-NN) method is a data mining technique considered to be among the top five techniques for data mining. In this, we consider each of the characteristics in our training set as a different dimension in some space, and take the value an observation has for this characteristic to be its coordinate in that dimension, so getting a set of points in space. We can then consider the similarity of two points to be the distance between them in this space under some appropriate metric. The way in which the algorithm decides which of the points from the training set are similar enough to be considered when choosing the class to predict for a new observation is to pick the k closest data points to the new observation, and to take the most common class among these. This is why it is called the k Nearest Neighbors algorithm. The implementation of algorithm can be noted as below :

1. Load the data
2. Initialize K to your chosen number of neighbors
3. For each example in the data
 - ❖ Calculate the distance between the query example and the current example from the data.
 - ❖ Add the distance and the index of to an ordered collection.
4. Sort the ordered collection of distances and indices from smallest to largest (in ascending order) by the distances
5. Pick the first K entries from the sorted collection
6. Get the labels of the selected K entries
7. If regression, return the mean of the K labels
8. If classification, return the mode of the K labels

C. Prediction of Crop Yield through KNN

Here we consider parameters like soil type, PH, Nitrogen, Phosphorous, Potassium etc. We have assigned N, P, K, soil type as input parameters although other parameters may also be considered. The crop yield which is an unknown value can be predicted using the values of the nearest known neighbors. This is possible by calculating Euclidian distance between those points. Thus we will be able to predict crop yield for the given input parameters.

The calculation of distance between points in a feature space, different distance functions could be used, in which the Euclidean distance function is the most commonly used one. Say p and q are represented as feature vectors. To measure the distance between p and q, the Euclidean metric is generally used by if a = (a1, a2) and b = (b1,b2) then the distance is given by:

$$d(a, b) = \sqrt{(b1 - a1)^2 + (b2 - a2)^2}$$

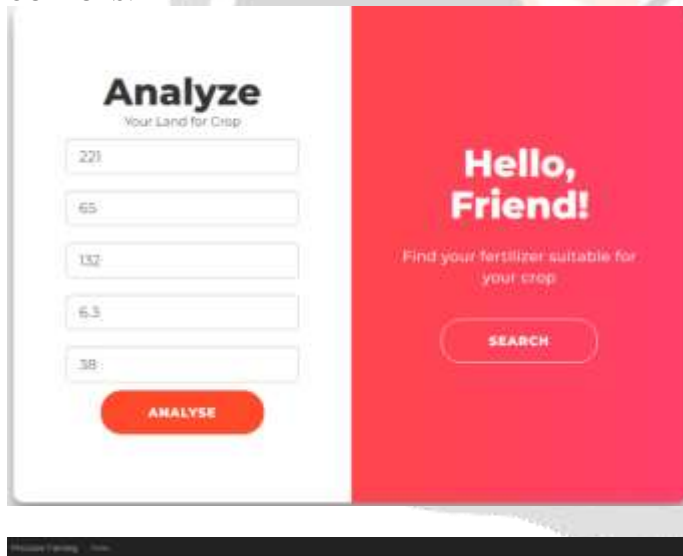
IV. PROPOSED SYSTEM

Prediction of CROP is done using KNN classification. The system can compute certain crop by using the value of humidity, temperature and soil properties. By using this system farmer will be able to increase agriculture productivity. Hence, this will have an overwhelming impact on poverty alleviation, boosting employment rate, human resource development and food security. The data set that we used to train our system was collected from the official website of Indian Agricultural Statistics and Indian Remote Sensing.

V. RESULTS

To predict the crop yield rate a application is created. This application includes three parts. First is managing datasets, second is testing datasets and third is analyzing datasets. In managing datasets we can get the datasets of previous years and they can also be converted into supporting format.

OUTPUTS:



Onions Revenue/Hectare: Rs440000

Planting

- 1. Plant onions in rows or beds in the field, leaving 10-12 cm between plants and 30-40 cm between rows.
- 2. Water the plants regularly and keep the soil moist.
- 3. Weeding should be done regularly to remove weeds.
- 4. Fertilizer should be applied regularly to the plants.
- 5. Harvest the onions when they are ready to be harvested.

Precision Farming [View](#)

Care

- Yellow Sweet Shiner will rot due to fungal diseases. Control fungicides when the plants push the soil away and the tubling process has started. Control soil-borne fungi around the roots. The fish needs knowledge about the soil.
- Generally, when plants do not have enough water, it is not a cost. About 100 ml of water per week (including rain water) is sufficient. If you want sweet shiners, water more.
- Plants will look healthy when they are healthy. Be sure to water during drought conditions.
- Make sure water has enough nutrients. Make it up with extra nutrients and other water.
- Use a quality system that can help you know what the plants have "drinking" and how to

Pests/Diseases

- **Thrips:** To control thrips, try to keep about 10% of a healthy thrips. Use a weak power of paper into the garden and avoid the roots (especially). If thrips are present, you will spot them (avoided) holes on the paper. A couple of treatments will have a good effect on them. Follow the package directions. Spray the plants before, three days after, and then three days after, if possible.
- **Green Midge:** Green midges are common and will cause damage. Get it by mounding soil around the roots. The green midges are to lay its eggs at the base of plants, so the mounding should prevent that. You should also keep the plants dry because the insects are a water lover. And make sure you completely harvest your plants at the harvest process. Green midges are usually a problem in very dry periods, so the recommendation may be necessary if you have a dry season.

Harvest/Storage

- When the plants start to mature, the tops become yellow and begin to fall over. At that point, test the condition or even stop or slow to speed the fish feeding process.
- Leave the soil for knowledge drying, and after a few days turn them up and in the container they grow. Always handle them very carefully—the slightest touch will encourage them to rot.
- After the plants are harvested, put them in a bag.
- Be sure to harvest at the right time. Before you harvest, the plants should be dry for 10 days.
- Make sure you dry the plants before you store them in a cool place or any other storage area. Spread them out on a open screen of the ground is dry.
- Store at 40 to 55 degrees F (4 to 10 degrees C) in bags or with the stems broken off.
- Make sure you harvest them in a cool and dry.
- Don't store them with apples or pears, as the ethylene gas produced by the fruit will ripen the plants. Control the storage of these fruits as well as possible.

Recommended Varieties

Other varieties are classified into two categories: complete, best in the North, and strong, best in the South.

- Yellow Sweet Shiner: long, yellow, large, round shape, yellowish.

Precision Farming [View](#)

• Control the soil to encourage drying, and after a few days turn them up and in the container they grow. Always handle them very carefully—the slightest touch will encourage them to rot.

• After the plants are harvested, put them in a bag.

• Be sure to harvest at the right time. Before you harvest, the plants should be dry for 10 days.

• Make sure you dry the plants before you store them in a cool place or any other storage area. Spread them out on a open screen of the ground is dry.

• Store at 40 to 55 degrees F (4 to 10 degrees C) in bags or with the stems broken off.

• Make sure you harvest them in a cool and dry.

• Don't store them with apples or pears, as the ethylene gas produced by the fruit will ripen the plants. Control the storage of these fruits as well as possible.

Recommended Varieties

Other varieties are classified into two categories: long, best in the North, and strong, best in the South.

- Yellow Sweet Shiner: long, yellow, large, round shape, yellowish.
- Shiner: long, yellow, large, round shape, yellowish.

2. GARLIC Revenue/Hectare: Rs980000

3. ORANGE Revenue/Hectare: Rs4100000

[Predict Bio-Fertilizer](#)

VI. CONCLUSION

This project presents a Machine Learning framework for the task of crop yield prediction, based on inexpensive remote sensing data. It allows for real time forecasting throughout the year and is applicable world-wide, especially for developing countries where field surveys are hard to conduct. We use modern representation learning ideas for crop yield prediction and successfully learns much more effective features from raw data compared with the hand-crafted features that are typically used. We propose a dimensionally reduction approach based on KNN classifier algorithm for model provides us with the state-of-art prediction accuracy and will have great impact in sustainable agriculture and food security.

VII. FUTURE ENCHANCEMENTS

We believe the proposed system will be able to help farmers to take the right decision of cultivating the right crop. A farmer can plant different crop in different districts based on the system recommendations. So, every farmer will get the chance of maximizing their yield and profit by using the system. Our main goal is to produce more with less as even being a developing country; we are almost using all our resources to keep up

to data with the rest of the world. In addition, any sort of contribution to the agriculture can be beneficial for the country as well to its people.

VII. REFERENCES

- [1] Karan Deep Kauri, 2019, "Machine Learning: Application in Indian Agriculture", International Journal of Advanced Research in Computer and Communication Engineering, 46(10), pp. 1701-1711.
- [2] S. Djodiltachoumy, 2018, "A Model for Prediction of Crop Yield", International Journal of Computational Intelligence and Informatics, pp. 1-11.
- [3] Nishit Jain, Amit Kumar. At.al, 2017, "Crop Selection Method Based on Various Environmental Factors Using Machine Learning", pp. 1584.
- [4] Raorane A.A, Dr. Kulkarni R.V, 2016, "Application of Datamining Tool To Crop Management System", 15(1), pp. 61-69.
- [5] Ashwani Kumar Kushwaha, Sweta Bhattacharya, 2016, "Crop yield prediction using Argo Algorithm in Hadoop", 14(3), pp. 106-114.