ISSN No:-2456-2165

Prediction of Agricultural Crops using KNN Algorithm

H. K. Karthikeya[1*], K. Sudarshan[2], Disha S. Shetty[3]
[1,3]Student, Department of Computer Science & Engineering, Srinivas Institute of Technology, Mangaluru, India
[2]Associate Professor, Department of Computer Science & Engineering, Srinivas Institute of Technology, Mangaluru, India

Abstract- Agriculture is full of uncertainty due to climate change, rainfall, soil type and numerous other factors. Crop prediction in agriculture is a very big dilemma and there is huge dataset where farmers find difficult to predict the yield and seed selection. In this current situation due to rise in the population the production of crops and agricultural products needs to be increased simultaneously to meet the demands of the people. These problems could be solved using machine learning algorithms and this paper focuses on these solutions. The real time environmental parameters like soil type, rainfall, humidity etc of Mangalore, Kodagu, Kasaragod and some other districts of Karnataka state are collected and crop prediction is done along with the accuracy for the crops is done with the help of K-NN algorithm.

Keywords:- *Agriculture, Crop-Prediction, K-Nearest Neighbor .*

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 slows 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 Rainfall, temperature, humidity and various other factors are collected and analyzed. This analysis will be fed to the prediction model.

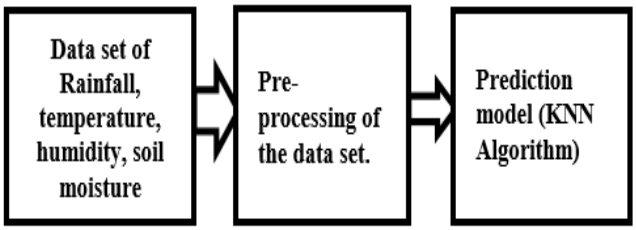

Fig 1:- Flow graph of the methodology Data Collection

Here we gather information from several sources and construct datasets. Plenty of online portals like Raitha-mithra, karnataka.gov.in and Data.gov.in [1] are available for information collection. Annual crop report of each crop is collected Collecting previous crop history data from places like Mangalore, Kodagu, Kasaragod, Mysore, Davangere, Hassan, Shivamogga, Chikkamagalur which belongs to Karnataka State.

ISSN No:-2456-2165

Collecting data related to crops like Coconut, Cardamom, Coffee, Areca nut, Ginger, Tea, Paddy, Ground nut, Black gram, Cashew, Pepper are the crops which are commonly grown in these 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 humidity, rainfall, soil type, area etc. We have assigned location, area, 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 calculation 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. TESTING AND ANALYSIS

The purpose of the test was to find the workings of K-NN algorithm and how will it predict the yield when three parameters were given as input. Input data is given as follows Location: Mangalore, Soil-Type: Coastal alluvial and Area: 1395 cents. The system predicted Coconut and Cocoa as two potential crops with the accuracy of 63.63%. While testing for Kodagu district the where soil type is Laterite soil and the area given was 1395 cents the system predicted Cardamom and Pepper as two potential crops and noted the accuracy of 56.66%

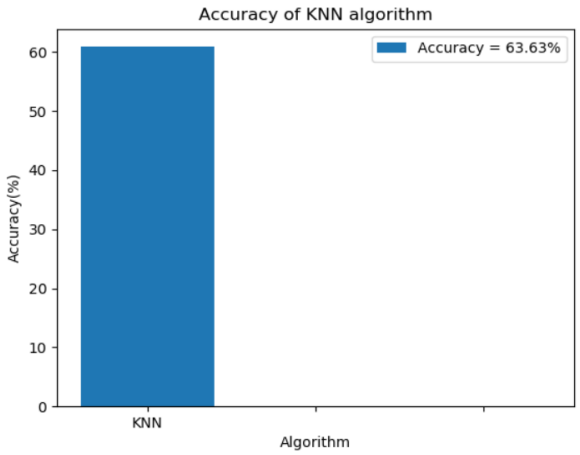

Fig 2:- Accuracy of KNN Algorithm for Mangalore Region

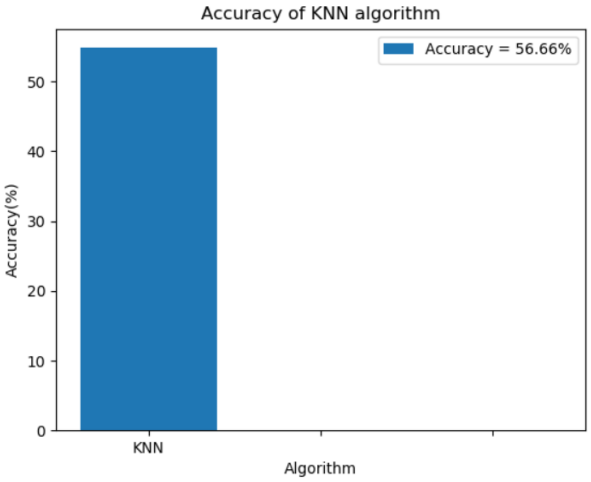

Fig 3:- Accuracy of KNN Algorithm for Kodagu Region

The system helps in avoiding the use of sensors and reduces unnecessary cost. This system results in efficient usage of time and cost. A key aspect of Crop Prediction is to identify a suitable crop quickly and suggest the farmer as to which crop to grow. Our system helps in gathering all necessary information and giving a model of output which not only increases current economical gain but also safeguards future profitability. The accuracy part of the system is noted as

decent but can be made more accurate with increase in the efficiency.

V. CONCLUSION

The implementation of the system was to learn about crops and agriculture and find an efficient way of harvesting. The study focuses on the agricultural datasets obtained from various portals belonging to some districts of Karnataka State. Datasets ordered in well structured manner. K-NN algorithm is used for the prediction model and crop yield prediction and its accuracy is obtained. The future is bright for the implementation of machine learning algorithms in the field of crop production and we hope to implement more advanced algorithms so that the system becomes more efficient, we hope to make system prediction more stable and obtain high accuracy with the help of more datasets and advanced algorithms.

REFRENCES

[1]. P. Vinciya "Agriculture Analysis for Next Generation High Tech Farming in Data Mining", Anna University, Trichy, Tamilnadu, India, 5 May,2016.
[2]. M. A. Jayaram and Netra Marad," Fuzzy Inference Systems for Crop Yield Prediction", Journal of Intelligent Systems, 2012,21(4),pp.363-372.
[3]. Sk Al Zaminur Rahman, Kaushik Chandra Mitra, Soil Classification using Machine Learning Methods and Crop Suggestion Based on Soil Series,2018 21st *International Conference of Computer and Information Technology (ICCIT), 21-23* December 2018.
[4]. *Hart, Peter E.* (1968). "The Condensed Nearest Neighbor Rule". IEEE Transactions on Information Theory. **18**: 515–516. *doi*:*10.1109/TIT.1968.1054155*
[5]. http://agricoop.nic.in/sites/default/files/Annual_rpt_2016 17_E.pdf.