Journal of Informatics Education and Research
ISSN: 1526-4726
Vol 4 Issue 3 (2024)

Applied Machine Learning for Predicting Crop Performance: A Supervised Learning Perspective

Somnath Mullick,

Assistant Professor, Department of Multimedia, Brainware University, Kolkata, E-Mail: somnath.office8@gmail.com

Dr. R Naveenkumar,

Associate Professor, Department of Computer Science and Engineering, Chandigarh Group of Colleges, Jhanjeri, Mohali, E-Mail: naveen.j3390@cgc.ac.in

Sandip Bhattacharjee,

HOD and Assistant Professor, Department of Multimedia, Brainware University, Kolkata, E-Mail: iamsandipin2007@gmail.com

Rahul Singha,

Assistant Professor, Department of Multimedia, Brainware University, Kolkata, E-Mail: rsingha907@gmail.com

Abstract

Agriculture is fraught with uncertainty due to climate change, rainfall, soil types, and many other factors. Crop prediction in agriculture is a major dilemma and there are huge data sets where farmers struggle to predict the right seed. In this situation of population growth, it is necessary to increase the production of crops and agricultural products at the same time in order to meet people's needs. These problems can be solved with machine learning algorithms. This white paper focuses on those solutions. Real-time environmental parameters such as soil type, precipitation, humidity, and past weather are recorded for the Tamil Nadu district, and ANN algorithms are used for crop prediction and accuracy

Keywords: Machine learning algorithms, support vector machines (SVMs), artificial neuron networks (ANNs), genetic algorithms, and K nearest Neighbour (KNN)

Introduction:

India is an agricultural country. India's financial system is driven with the aid of using the import and export of agricultural products. Agriculture is one of the most essential factors of the Indian financial system. Uncertainty in yields has led to a pointy decline in financial conditions. The most important vegetation in India is rice, wheat, legumes, and grains. Day with the aid of using day the the populace of India is developing and the vegetation productiveness wants to be increased with the aid of using the right crop grown to feed the populace [1]. One of the great approaches to predicting unknown values is using gadgets to get to know algorithms. This work intends to expand the crop prediction version of the usage of gadgets getting to know The utility intends to expect crops, so it may assist farmers to pick out great seeds for plantation [2]. Many ML algorithms may be used, and algorithms like Regression analysis, Support Vector Machine, Neural Networks, and K-Nearest Neighbour (KNN) may be utilized. his project describes ANN. The K-nearest Neighbours (KNN) set of rules is an easily supervised gadget for getting to know the set of rules that may be used to clear up each type and regression problem [3]. It's easy to enforce and understand The process of how it works according to the procedure of the standard form of the scenarios, it has the major disadvantage of being considerably slower as the scale of the information used increases. The purpose here is to apply a version wherein facts hotspots are clustered into companies to expect various instances [4]. ANN operates primarily based totally on the minimal distance from the question example to the education pattern and determines the closest neighbour [5].

Then we acquire okay nearest friends, we take the easy majority of those nearest friends to be the prediction question object. As noted earlier it is able to additionally be used for regression output, that is the item`s reward [6]. Mostly for distance calculation in the KNN set of rules the metric used is Euclidean distance Main Factors: quantity of Rainfall: The main essential for crop harvest home is water and in Bharat, most farmland depends on rainwater. an honest} amount of precipitation will facilitate in a very decent yield of a crop. thence the quantity of rainfall is taken into account as one of the foremost important factors [7]. space of land: The area of the sphere conjointly plays a vital factor. If the amount of rainfall together with the world of the field is taken into consideration, we have a tendency to get the most effective prediction concerning crop production [8]. A lesser area with a large amount of rainfall can't be favourable for agriculture.

The area-to-precipitation magnitude relation ought to be of equilibrium values for best production. form of Crop: the kind of crop required would conjointly play a vital role in deciding the quantity of production. a couple of crops want more water than the others [9]. Alternatively add this field and their drawbacks: Censuses: Agricultural censuses, are the most typical technique of knowing the production, this involves a large manpower and takes a great deal of your time to estimate the ultimate value. This method is sometimes done at a spot of forty-five years. native governance data: The local bodies report the information collected from the farmers to the authorities and therefore the estimate is calculated [10]. Even if this involves a high workforce ANN takes a great deal of your time to return to an estimate. Satellite sensing: The aerial pictures are taken and the production per space is calculated. this is often an awfully big-ticket method and needs high-quality image capturing [11].

Image method : Image based analysis was one in all the ways that was antecedental used for detection land sort so analysis was done. Process is predicated on image analysis results aren't correct as during this technique soil conditions are not considered [12]. Image process is a long process. downside STATEMENT: In country like Bharat the assembly of crops are plagued by many factors. Factors like Humidity, temperature, rainfall, soil sort play an important role in crop prediction, and factors like these disagree by giant with regard to region. In India, Farmers majorly still have confidence ancient techniques heritable from their forefathers. These techniques would work earlier once the climate was a lot of healthier and predictable [13]. currently with factors like heating and pollution moving the atmosphere individuals have to be compelled to be good and begin utilizing fashionable techniques. it's time to research large set of information and are available up with a system that may offer enough data concerning appropriate crop to urge the nice yield. The new age methodology needs giant structured data sets and an formula capable of providing resolution victimization the provided datasets.

Proposed Work: The main goal of the system is to give farmers an overview of crops that they can take to their farms with the estimated crop yield and the market price prediction of that particular crop. The system architecture includes an input module responsible for taking input from the farmers. In that, the farmer has to provide Location, Soil type Land type, Size of land, and Water Source [14]. The farmer is also responsible for interacting with predicted results. After selecting the location parameter, the input query block is responsible for the subset selection of an attribute from crop details. It gives all possible crops to be sown at a given time stamp suitable for a given location. The prediction model is used to predict weather, pest disease based on crop and market price. Yield rates of these crops are evaluated, if the yield rate per day of these crops is fair (within tolerance) then those crops are selected for crop sequences [15].

The main feature of the system is it gives output in statistics form. To this end, the methodological approach it will follow is composed of these steps: 1) The description of the user's farm. 2) The selection of crops suitable. 3) Predictive analytics modelling for predicting total yield and market price of those crops. Methodology: Dataset collection When implementing an accurate predictive model, it may not be enough to consider one or two parameters. Data on past weather, such as precipitation, temperature, humidity, and various other factors, are collected and analysed. This analysis is sent to the predictive model. Here, we collect information from multiple sources and create a dataset. Numerous online portals such as Raithamithra, Uzhavan, and Data.gov.in are available for information.

An annual crop report for each crop is collected. Collect previous harvest history data from locations such as Madurai, Coimbatore, Theni, Tirunelveli and Thanjavur in Tamil Nadu. Collection of serial and adjacent data, such as sugar cane, coconut, cardamom, coffee, culture, ginger, tea, paddy field, peanuts, cashew nuts, and pepper, usually growing in these areas. Collect data with historical weather like rain. Humidity and other symptoms such as floors, irrigation type, operation,

location, price, year, crop, crop disease, symptoms and other symptoms. The prediction types including data mining technology are as follows: Classification: Predicted category or what class it will fall. Regression: Prediction of numerical value possessed by variables (If it is a variable that changes with time, it is called `time series' prediction).

Classification: The purpose of the classification problem is to identify the functions of each case. The understanding and prediction of existing data, and how to run a new instance is a task of classification problems. Data mining examines the classified data (case) to create a

classification model and detect predictive patterns inductively. These existing cases come from historical things. Samples across the database can come from experiments that are tested in the real world and experiments used to create classifiers.

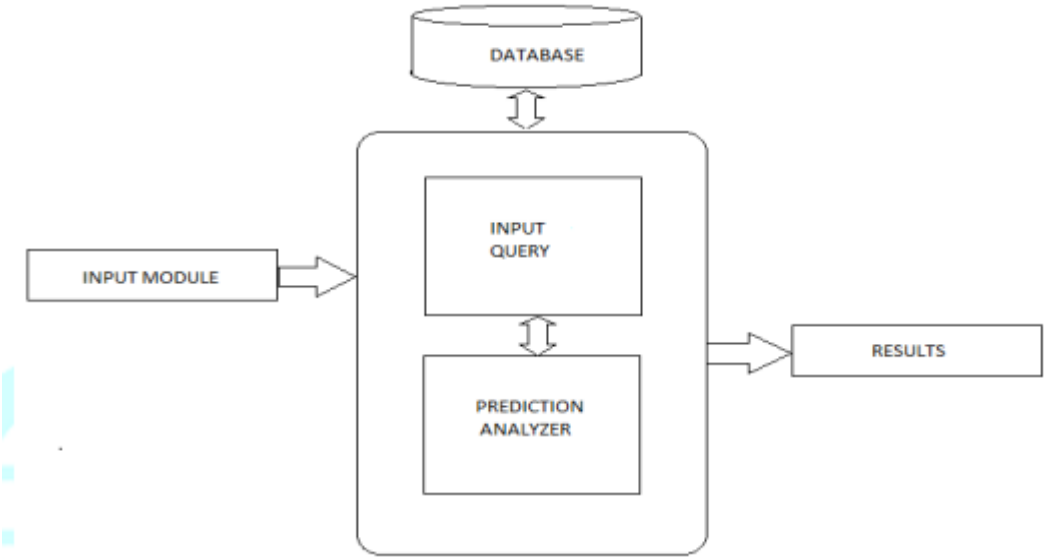

Fig 1: classification to create a model

Experts may categorize a sample of a Fig 1database and use that classification to create a model that applies to the entire database.

Regression: Predicting other values using existing values is regression. In simple cases, regression uses standard statistical techniques such as linear regression. However, many real problems are not just linear predictions of past values. For example, sales volume, stock price, and product failure rate are all very difficult to predict because they can depend on complex interactions of multiple predictors. Therefore, complicated technologies may be required to predict future values. The same model type can be used for both regression and classification. Several data mining methods such as support vector machines (SVMs), artificial neuron networks (ANNs), genetic algorithms, and K nearest Neighbour (KNN).

This system uses predictive analysis to extract knowledge from existing data used for plans, trends, and results. Predictive Analytics is to extract information from existing data sets to determine patterns and predict future results and trends. Here, prediction analysis is used to predict three terms that benefit a pretty farmer to determine the harvest he should take 1. Weather 2. Yield 3. Cultural price - Forecast model construction uses three algorithms. A Classification Algorithm for Weather Forecast, a Generalized Line Regression Model for KNN Regression Algorithm for Harvest Price Forecast.

KNN algorithm: K nearest Neighbour (KNN) method is a data mining technology that is considered in the first five technologies for data mining. In this case, considering the respective functions of the training as different dimensions of any room, the observation of this feature needs to be the coordinates of this dimension, so the number of points in the room can be obtained. You can then consider two similarities. An algorithm is to select the most common data points for new observations in a manner similar to a point from a training set to consider class selection to predict new observations. They are. For this reason, it is called the k-nearest neighbour algorithm.

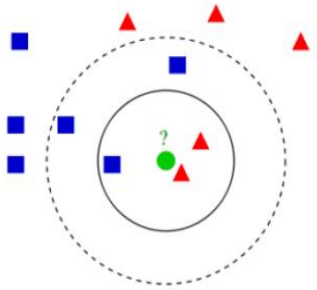

Fig 2: - K labels

The algorithm implementation can be noted as follows:

i)Load the data.

ii)Initialize K with the selected number of neighbours.

iii)For each example in the data write a query example that should be written with a current

example from the data.

iv)Add distances and indexes to the ordered collection.

v)Sorts an ordered collection of distances and indexes by distance from minimum to maximum (in ascending order).

vi)Select the first K entry from the sorted collection.

vii)Get the label of the selected K record.

viii)For regression, return the average of the K labels.

ix)When sorting, returns the mode of the k label.

Prediction of Crop Yield through KNN Here we don't forget parameters like humidity, rainfall, and red soil place. We have assigned location, place, and soil kind as enter parameters even though other parameters can also be considered.

The crop yield that is an unknown price can be anticipated the usage of the values of the closest regarded neighbours. This is possible via way of means of calculating Euclidian distance among the one's factors. Thus we can have the ability to expect crop yield for the given enter parameters.

The calculation of distance among factors in a function space, unique distance features can be used, wherein the Euclidean distance characteristic is the maximum normally 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 if a = (a1, a2) and b = (b1,b2) then the distance is given by: d(a, b) = $\sqrt{(b1 - a1)\ 2 + (b2 - a2)\ 2}$ [6]

KNN Prediction of Yield:

Here we take into account parameters such as moisture, precipitation, soil type, and area. Although other parameters can be considered, the position, area, and floor type were allocated as input parameters. The crop yield, which is an unknown value, can be predicted with the value of the following known neighbour's. It is possible by calculating the distance between these points. Therefore, we can predict crop yield for given input parameters. Calculation of the distance between points in the feature space can use different distance functions whose Euclidean distance function is most commonly used. P and q say it is represented as a feature vector. The purpose of the test was to find the operation of the KNN algorithm when

three parameters were given as input. The input data is given as follows. Location: Madurai, BODTYPE: Floor. The system predicted paddy fields and sugar cane as two potential cultures with curacy. During the test of the heavenly area where the floor type is the floating bottom and dizziness floor, the system is determined by the accuracy of the cardamom and turmeric as two possible cultures. When testing the Theni area, where the oil types are laterite soil and alluvial soil, the system predicted and accurately recorded cardamom and turmeric as two potential plants. Testing the Thirunelveli area, where the oil types are laterite soil and alluvial soil, the system predicted and accurately recorded rice as a potential crop. When testing the Coimbatore area, where the soil types are black soil and alluvial soil, the system predicted and accurately recorded cotton and coconut as two potential crops. This system helps to avoid the use of sensors and reduces unnecessary costs. This system allows you to use your time and money efficiently. An important aspect of crop prediction is to quickly identify suitable crops and suggest to farmers which crops to plant to produce effective yields. Our system collects all the information you need and helps you create a performance model that not only increases your current economic benefits, but also guarantees your future profitability. The accuracy part of the system is noted, but it can before accurate as efficiency increases. K-Nearest Neighbors (KNN) is a simple, instance-based learning algorithm used for classification and regression. To illustrate how KNN can be used for predicting yield, let's walk through a step-by-step example with sample data.

E.g. Data: Suppose we have a dataset of different crops with the following features:

1. Temperature (in degrees Celsius)
2. Rainfall (in millimetres)
3. Fertilizer Amount (in kilograms per hectare)
4. Yield (in tons per hectare) - this is what we want to predict

A sample table with 6 data points:

Crop	Temperature (°C)	Rainfall (mm)	Fertilizer Amount (kg ha)	Yield (tons ha)
A	25	100	50	10
B	30	80	60	12
C	22	120	40	8
D	28	90	70	11
E	24	110	55	9
F	27	95	65	10.5

Table 1: Relationship between environmental and agricultural factors

The data presented provides insights into the relationship between environmental and agricultural factors and crop yield for six different crops. Crop A, cultivated at a moderate temperature of 25°C with 100 mm of rainfall and 50 kg ha of fertilizer, yields 10 tons ha. Crop B, growing in warmer conditions at 30°C with less rainfall at 80 mm and higher fertilizer application of 60 kg ha, achieves a yield of 12 tons ha. In contrast, Crop C, which thrives at a cooler temperature of 22°C and receives ample rainfall of 120 mm with 40 kg ha of fertilizer, results in a lower yield of 8 tons ha. Crop D, cultivated at 28°C with 90 mm of rainfall and 70 kg ha of fertilizer, yields 11 tons ha. Crop E, growing at 24°C with significant

rainfall of 110 mm and 55 kg ha of fertilizer, produces a yield of 9 tons ha. Lastly, Crop F, with a temperature of 27°C, 95 mm of rainfall, and 65 kg ha of fertilizer, achieves a yield of 10.5 tons ha. These variations in yield across different crops highlight how temperature, rainfall, and fertilizer levels interact to influence agricultural productivity.

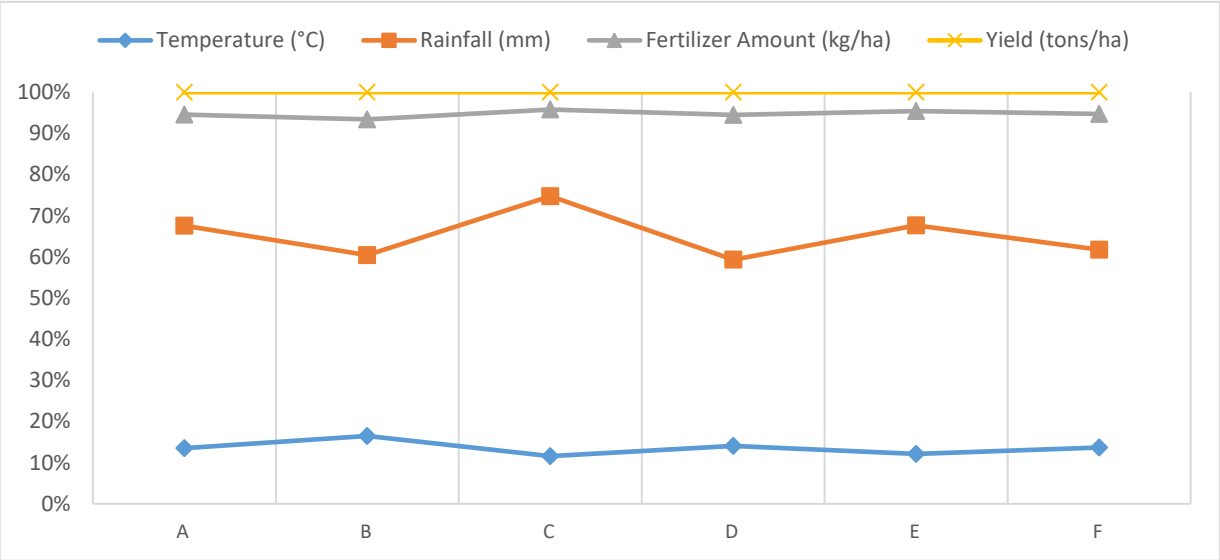

Fig 3: Environmental and agricultural factors in Chart comparison

KNN Prediction

1. Choose the value of K: This is the number of nearest neighbours to consider for making the prediction. We'll use K = 3 for this example.
2. Calculate the distance: For simplicity, we'll use Euclidean distance. Let's predict the yield for a new crop with the following features:
    - Temperature: 26°C
    - Rainfall: 100 mm
    - Fertilizer Amount: 60 kg ha
3. Find the K nearest neighbours: Compute the Euclidean distance from the new crop to all the existing crops.
4. Average the yields of the nearest neighbours: Use the yields of these K nearest neighbours to predict the yield for the new crop.

Detailed Calculation: Calculate Euclidean Distance

For each crop in the dataset, calculate the Euclidean distance to the new crop (26°C, 100 mm, 60 kg ha):

Distance=(T−Tnew)2+(R−Rnew)2+(F−Fnew)2\text{Distance} = \sqrt{(T - T_{\text{new}})^2 + (R - R_{\text{new}})^2 + (F - F_{\text{new}})^2}Distance=(T−Tnew)2+(R−Rnew)2+(F−Fnew)2

Sort the Distances: Sort the crops by their distances from the new crop:

Crop	Distance
F	7.14
A	10.05
E	11.36
D	14.28
B	20.39
C	28.56

Table 2: the distances associated with various crops

The data presents the distances associated with various crops in Table 2, likely reflecting their distance from a reference point or their spatial arrangement within a study. Crop F is closest, situated at a distance of 7.14 units, indicating it is relatively nearer compared to other crops. Crop A follows, with a distance of 10.05 units. Crop E is slightly farther away at 11.36 units. Crop D is positioned at a greater distance of 14.28 units, while Crop B is even further at 20.39 units. The most distant crop is Crop C, located 28.56 units away. This distance gradient highlights the varying spatial placements or arrangements of these crops, which may influence their management or study outcomes.

Select , For K = 3, the nearest neighbours are F, A, and E. Average the Yields of the nearest neighbours are:

- Crop F: 10.5 tons ha
- Crop A: 10 tons ha
- Crop E: 9 tons ha

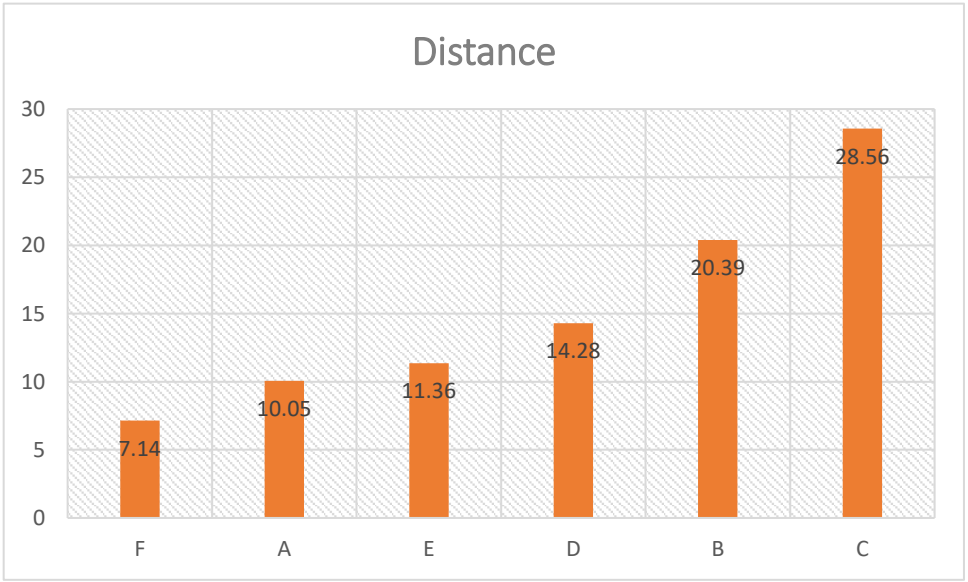

Fig 4: Average yield of K Nearest Neighbours

Average yield :

Average Yield=10.5+10+93=29.53≈9.83 tons ha text{Average Yield} = frac{10.5 + 10 + 9}{3} = frac{29.5}{3} approx 9.83 text{ tons ha}Average Yield=310.5+10+9=329.5≈9.83 tons ha.

Conclusion :

The implementation of this system consisted of learning about crops and agriculture and finding efficient harvesting methods. This study focuses on agricultural datasets obtained from various portals in several districts of Tamil Nadu. A well-structured and ordered dataset. The KNN algorithm is used for predictive modelling and yield prediction, and its accuracy is maintained. The predicted yield for the new crop is approximately 9.83 tons per hectare based on the KNN algorithm with K = 3. The future of implementing machine learning algorithms in the field of crop production is bright. We want to implement more sophisticated algorithms to make our system more efficient. We want to make system predictions more stable and achieve high.

References:

[1] Dr R. Naveen Kumar, Amit Kumar Bhore, Sourav Sadhukhan, Dr G. Manivasagam, Rubi Sarkar "Self-Monitoring System for Vision-Based Application Using Machine Learning Algorithms" DOI: 10.5281/zenodo.10547803, Vol 18 No 12 (2023), Page No 1958 – 1965, Published on 31-12-2023.

[2] Balachandran.S , Ashokan R. and Sridharan S (2006). "Global surface temperature about northeast monsoon rainfall over Tramlined", Journal of Earth System Science, vol. 115, no.3, pp. 349-362.

[3] Mohapatra S.K, Upadhyay. A, and Gola C. "Rainfall prediction based on 100 years of meteorological data." 2017 International Conference on Computing and Communication Technologies for Smart Nation (IC3TSN), pp. 162-166. IEEE, 2017.

[4] Renuka, Sujata Terdal." Evaluation of Machine Learning Algorithms for Crop Yield Prediction". International Journal of Engineering and Advanced Technology (IJEAT) Volume-8 Issue 6, August 2019.

[5] Shivam Bang, Rajat Bishoni, Ankit Singh Chauhan, Akshay Kumar, Indu Chawla." Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models" 2019 Twelfth International Conference on Contemporary Computing (IC3).IEEE, 2019.

[6] Sundaram, Meenakshi S, and Lakshmi M. "Rainfall prediction using seasonal autoregressive integrated moving average model." Computer science 3, no. 4 (2014), pp. 58-60, 2014.

[7] Veenadhari S, Misra B, and Singh C.D. "Machine learning approach for forecasting crop yield based on climatic parameters." 2014 International Conference on Computer Communication and Informatics, pp. 1-5. IEEE, 2014

[8] RN Kumar, MA Kumar Medical data mining techniques for health care systems International Journal of Engineering Science 3498, 2016.

[9] Yan M et al (2018) Research on precision management of farming season based on big data EURASIP J Wireless Commune Networking 2018(1):143

[10] Priya R, Ramesh D, Khosla E (2018) Crop prediction on the region belts of India: a Naïve Bayes MapReduce precision agricultural model. In: 2018 international conference on advances in computing, communications and informatics (ICACCI). IEEE

[11] Pawar M, Chillarge G (2018) Soil toxicity prediction and recommendation system using data mining in precision agriculture. In: 2018 3rd international conference for convergence in technology (I2CT). IEEE

[12] Sk Al Zaminur Rahman, Kaushik Chandra Mitra, Soil Classification and Soil Series-based Crop Proposals Using Machine Learning Techniques, 2018 21st International Conference on Computers and Information Technology (ICCIT), December 2, 2018.

[13] Hart, Peter E. (1968). "Condensed Nearest Neighbour Rule". IEEE transactions related to information theory. 18: 515-516. Doi: 10.1109 TIT.1968.1054155

[14] H. K. Karthikeya1*, K. Sudarshan2, Disha S. Shetty3 "Prediction of Agricultural Crops using KNN Algorithm" Volume 5, Issue 5, May – 2020 IJISRT May – 2020

[15]. Sandip Bhattacharjee, Dr R Naveenkumar, Rahul Singha, Somnath Mullick, Rubi Sarkar The Impact of Machine Learning on Enhancing Diversity and Inclusion through Advanced Recommence Screening Techniques Journal of Informatics Education and Research - ABDC 4 (2), 3141-3159, 2024.