



# OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

## Survey on Hybrid Approach for Crop Yield Prediction based on Soil Properties using Data Mining and Machine Learning Techniques

Surabhi Misar<sup>1</sup>, Prof. H. P. Channe<sup>2</sup>

PG Student, Dept. of Computer Engineering, Pune Institute of Computer Technology, Pune<sup>1</sup>  
 Assistant Professor, Dept. of Computer Engineering, Pune Institute of Computer Technology, Pune<sup>2</sup>  
 surabhimisar@gmail.com<sup>1</sup>, hpchanne@pict.edu<sup>2</sup>

**Abstract:** India is an agriculture based country. Farming and farm related business are mainly focused for earning by major population. Farmer has to make decisions manually every year for various factors such as crop to be cultivated, choosing a fertilizer quantity as per soil properties. The decisions made are not accurate as it does not consider soil properties. To overcome this problem, researchers have analysed the agricultural data for various parameters. In this dissertation work, we are going to use agricultural data which will be region specific. PAM, CLARA, modified DBSCAN data mining culturing techniques are used in literature. We are going to use machine learning techniques namely multiple linear regression, SVM, KNN and ANN for crop yield prediction. Predicting those results on parameters of soil property like NPK values, temperature and rainfall. From this analysis, farmer can get accuracy and clarity to get decision for crop to be cultivated to get better product yield while considering environmental factors, weather conditions, soil types and season.

**Keywords-** Machine learning, Data mining, Agriculture

### I INTRODUCTION

India ranks second worldwide in the farm output. Agriculture is demographically the broadest economic sector and plays significant role in the overall socio-economic fabric of India. Farmers have to make a decision every year for crop yield, choosing a fertilizer. Data in agriculture are years of data with different parameters like crop production based on rainfall, soil properties. Based on these parameters farmers have to predict the crop yield and fertilizers based on soil type. To increase crop production we have to analyse big data. Clustering is an unsupervised machine learning algorithm that groups entities, from a data set, that have high degree of similarity in the same cluster. Nowadays lots of areas are using these kinds of algorithms to separate data sets into groups in an automated way, and still have a good

quality result. Data mining clustering algorithms are used to group the data are PAM CLARA and DBSCAN. [1]

### II LITERATURE SURVEY

[1] Analysis of agriculture data using data mining techniques: application of big data

In horticulture segment where agriculturists and agribusinesses need to make incalculable choices consistently and complicated complexities includes the different elements affecting them. A fundamental issue for rural arranging aim is the precise yield estimation for the various yields engaged with the arranging. Information mining systems are vital methodology for achieving reasonable and viable answers for this issue. Horticulture has been a conspicuous focus for huge information. Natural conditions, fluctuation in soil, input levels, blends and item costs have made everything the more applicable for ranchers to utilize data and motivate help to make basic cultivating choices. This paper centres

around the examination of the agribusiness information and finding ideal parameters to boost the yield generation utilizing information mining procedures like PAM, CLARA, DBSCAN and Multiple Linear Regression. Mining the extensive measure of existing harvest, soil and climatic information, and examining new, non-exploratory information upgrades the generation and makes horticulture stronger to climatic change.

[2] Data mining techniques for predicting crop productivity

In this paper an endeavour has been made to survey the examination contemplates on utilization of information mining systems in the field of agribusiness. A portion of the systems, such as ID3 calculations, the k-implies, the k closest neighbour, fake neural systems what's more, bolster vector machines connected in the field of farming were exhibited. Information mining in application in farming is a generally new methodology for determining/foreseeing of rural product/creature the board. This article investigates the utilizations of information mining procedures in the field of farming what's more, unified sciences.

[3] Wheat yield prediction using machine learning and advanced sensing techniques.

Understanding yield restricting components requires high goals multi-layer data about elements influencing crop development and yield. In this manner, on-line proximal soil detecting for estimation of soil properties is required, because of the capacity of these sensors to gather high goals information (>1500 test per ha), and therefore lessening work and time cost of soil inspecting and examination. The point of this paper is to anticipate inside field variety in wheat yield, in light of on-line multi-layer soil information, and satellite symbolism trim development attributes. Directed self-sorting out maps equipped for dealing with existent data from various soil and yield sensors by using an unsupervised learning calculation were utilized.

[4] Detection of soil nitrogen using near infrared sensor based on soil pre-treatment and algorithm

Soil nitrogen content is one of the vital development supplement parameters of harvests. It is an essential for logical treatment to precisely get a handle on soil supplement data in exactness horticulture. The data about supplements, for example, nitrogen in the dirt can be gotten rapidly by utilizing a close infrared sensor. The information can be broke down in the location procedure, which is non-destructive and non-dirtying. So as to research the impact of soil pre-treatment on nitrogen

content by close infrared sensor, 16 nitrogen fixations were blended with soil and the dirt examples were isolated into three gatherings with various pre-treatment. The main gathering of soil tests with strict pre-treatment were dried, ground, sieved and squeezed. The second gathering of soil tests were dried and ground.

[5] Agricultural Crop Yield Prediction Using Artificial Neural Network Approach

By considering different circumstances of climatologically marvels influencing neighbourhood climate conditions in different parts of the world.

These climate conditions directly affect edit yield. Different investigates have been done investigating the associations between huge scale climatologically wonders and product yield. Counterfeit neural systems have been shown to be incredible assets for demonstrating and expectation, to build their viability.

Yield expectation system is utilized to anticipate the reasonable product by detecting different parameter of soil and furthermore parameter identified with air. Parameters like kind of Soil, PH, Nitrogen, Phosphate, Potassium, Natural Carbon, Calcium, Magnesium, Sulphur, Manganese, Copper, Press, Profundity, Temperature, Precipitation, and Dampness. For that reason we are used Artificial Neural Network (ANN).

### III SYSTEM DESIGN

The dataset in agriculture content NPK values and type of soil ,apply clustering algorithms to form cluster .data mining clustering algorithm are divide the data and apply linear regression and support vector machine. According to accuracy of suitable parameter we can predict the next crop yield and suitable fertilizer.

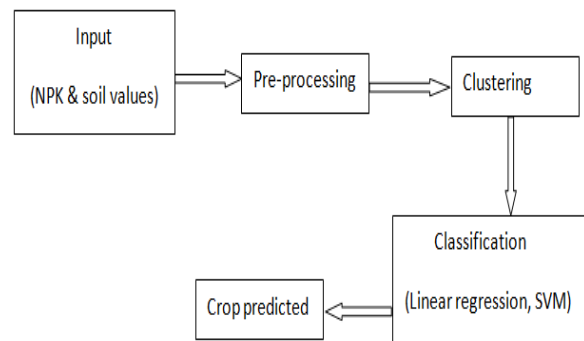


Figure 1: Block Diagram of Proposed System

a) **Input-** The input content Nitrogen, Phosphorous and potassium (NPK) values and type of soil. The proper combination of NPK content in a soil is very useful for the crop growth. There are eight different type of soil found. Among them four types of soil are good for crop

cultivation. Rest are found in deserts, mountains, forest and alkaline. Different variety of crops content different NPK values and different soil type as well. For growth of crop combination of both is important.

**b) Pre-processing-** In pre-processing the invalid values, noise and errors can be removed from the dataset.

**c) Cluster-** There is three data mining algorithm are used for clustering the data. PAM, CLARA, DBSCAN are used to divide the data into groups.

#### IV ALGORITHM

##### DBSCAN Algorithm

DBSCAN is a thickness based bunching calculation. It contains vast measure of information which has commotion and anomalies DBSCAN have two parameters-chiefly Eps and MinPts. In conventional DBSCAN, It can't deliver ideal Eps esteem consequently. In Modified methodology, we can and out least focuses and epsilon consequently. Along these lines KNN plot is utilized to discover epsilon esteem and the K esteem is client characterized. To dodge client depended K esteem, we apply Bachelors Wilkins bunching calculation is connected on the database and K esteem is acquired alongside its separate grouping focus.

##### Procedure Steps:

1. Calculate Eps and MinPts for further calculations.
2. Calculate minimum points and epsilon automatically.
3. Apply Bachelors Wilkins clustering algorithm.
4. Obtain k value from Bachelors Wilkins clustering algorithm.
5. Use k value to plot KNN Distance Graph.
6. Obtain Eps value from KNN plot.

#### V CONCLUSION

To analyse the accuracy of data, we can predict the next crop yield and fertilizer required for the crop to be cultivated.

#### VI FUTURE WORK

In this system we analyses the soil quality from different regions to predict suitable crop. In future work we want to study all parameters like soil, rain fall, weather conditions for more accurate prediction of crops.

#### REFERENCES

- [1] Jharna Majumdar, Sneha Naraseeyappa and Shilpa Ankalaki, "Analysis of Agriculture Data Using Data Mining Techniques: Application of Big Data", Journal of Big Data 2017.
- [2] Veenadhari S, Mishra B, Singh CD., "Data Mining Techniques for Predicting Crop Productivity- A Review Article", 2011.

[3] Pantazi XE, Moshou D, Alexandridis T, Mouazen AM., "Wheat Yield Prediction using Machine Learning and Advanced Sensing Techniques", Comput Electron Agric. 2016.

[4] Veenadhari S, Misra B, Singh D., "Machine Learning Approach for Forecasting Crop Yield based on Climatic Parameters", In Paper presented at International Conference on Computer Communication and Informatics (ICCCI-2014), Coimbatore 2014.

[5] Ester M, Kriegel HP, Sander J, Xu X., "A Density-based Algorithm for Discovering Clusters in Large Spatial Databases with Noise", In Paper Presented at International Conference on Knowledge Discovery and Data Mining, 1996.