Analysis of Crop Yield Prediction by using Machine Learning Algorithms

Naveen Kumar V[1], Kiran B N[2]

[1]PG Student, Dept. of Information Science and Engineering, NIE, Mysore, Karnataka, India
[2] Assistant Professor, Dept. of Information Science and Engineering, NIE, Mysore, Karnataka, India

---***---

Abstract - Farming is considered as one of the fundamental and a first culture that is drilled in India. Yet, the actual impact of environment changes and its unconventionality hugely affects cultivating. Thus crop yield forecast has become a vital advance towards an exceptionally successful harvest creation and the executives. Foreseeing the yields well in front of its reap time helps the ranchers in the development of the right harvests. This paper thus proposes a concise examination of harvest yield forecast utilizing AI calculations like K-Nearest Neighbor (KNN) Algorithm. The Experimental outcomes shows that the proposed work productively predicts the harvest yield.

*Key Words***: *Machine Learning, KNN,, crop yield , prediction*

1.INTRODUCTION

 Farming assumes a significant part in the development of the public economy. A portion of the variables on which cultivating relies upon are soil, environment, stickiness, precipitation, temperature, etc. The majority of the occasions, ranchers neglect to accomplish the normal outcome because of numerous reasons. Yield forecast is done which includes anticipating the yield of the harvest dependent on the current information. Before, forecast were basically founded on the ranchers past encounters. Yield expectation helps in distinguishing the most achievable harvests to be developed for a specific district as per given ecological conditions.

 In this paper, AI calculations , K-Nearest Neighbors and are utilized to anticipate the most reasonable yield. The harvest creation relies upon various factors which changes with each square meters and primarily relies upon the topography of the region,weather conditions ,soil type, mugginess, etc. Immense informational collections can be utilized for anticipating their impact on the significant yields of that specific region or state. AI methods have progressed significantly in the course of recent many years.

 KNN is a managed order strategy. It is additionally called as languid student strategy since it relies upon learning by similarity. It doesn't have a specific preparing stage and uses all the information for preparing while arrangement. It utilizes highlight closeness way to deal with anticipate the upsides of new information focuses which implies the new information focuses will be relegated a worth dependent on how intently it coordinates with the focuses in the preparation set.

 It comprises of highlights like area, region in square foots, temperature (in Celsius), dampness, yield, precipitation (in mm), sort of the dirt, yields, and cost. The paper is coordinated as follows: Section II presents the connected work and Section III examines about the proposed framework. The test results on farming information are examined in segment IV. At long last, Conclusion is given in the segment V.

2. RELATED WORK

[1] A Survey on Crop Recommendation Using Machine Learning by M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran.

Depiction: In this paper the creators proposed crop suggestion dependent on information mining ideas like yield and suggestion and forecast of soil and environment condition. Here they have utilized the ensembling method and a similar investigation of soil order. The proposed structure will arrange the data got from document, environment office and by applying AI estimation, Multiple Linear Regression, an assumption for most sensible yields as shown by current regular conditions is made. This outfits an agriculturist with combination of options of harvests that can be created. This investigation goes for assessment of soil dataset using data mining systems. It revolves around portrayal of soil using various estimations open. Another fundamental plan is to predict untested qualities using backslide method, and utilization of electronic soil test gathering.

Advantages:
•   The crop is suggested dependent on climatic conditions.
•   The request of soil was seen as essential to consider considering the way that depending on the productivity class of the dirt the region learning experts sorts out which yields should be gone facing that particular soil and which fertilizers should be used for the same.
Limitations:

• The primary substance relies upon soil condition and different insights were adequately not to portray the genuine yield.

Different employments of precision Agriculture are: assumption for diseases, assumption for environment expecting, course of action of soil, checking crop, yield conjecture, customized water framework structure, etc and a portion of these are not suggested in this paper.

[2] A Review on Data Mining Techniques for Fertilizer Recommendation by Jignasha M. Jethva1, Nikhil Gondaliya, Vinita Shah.

Portrayal: Agriculture assumes a critical part in the existence of an economy. It is the spine for non-industrial nations like India as over 70% of populace relies upon horticulture. To expand crop creation numerous elements are mindful like soil, climate, downpour, composts and pesticides. They have utilized soil boundaries to build crop creation since it's anything but a fundamental key factor of farming. To keep up with supplement levels in the dirt in the event of insufficiency, composts are added to soil. The normal issue existing among the Indian ranchers is that they pick inexact measure of manures and add them physically. Abundance or inadequate expansion of compost can hurt the vegetation and lessen the yield. The paper gives survey of different information mining procedures utilized on farming soil dataset for compost proposal. For the most part centered around different soil boundaries like Fe, S, Zn, Cu, N and Ph esteem and so forth In this review, creators additionally portray some Agriculture issues that can be addressed by utilizing information mining procedures like Agriculture, Soil Fertility, Fertilizer Recommendation, Data Mining, Clustering, Classification, Neural Network. Calculations utilized here are K-mean in Agriculture, K-closest neighbor in Agriculture, SVMs in Agriculture, Decision Tree in Agriculture.

Benefits:
• Recommendation could be given to ranchers utilizing past agrarian information with assistance of information mining ideas and they can get enhanced outcomes from proposal.
• The inspiration driving this work by the creators is to investigate information mining methods, which are reasonable for tackling complex rural issues.Limitations:
  • The survey is based on past data and if data with new entry is not recommended by the system.
    • Soil parameter is the only main thing they have used in this paper to recommend the crop.

[3] Crop Recommendation System for Precesion agriculture by S.Pudumalar, E.Ramanujam, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan.
Description: Data mining is the practice of examining and deriving purposeful information from the data. Data mining finds its application in various fields like finance,

retail, medicine, agriculture etc. Data mining in agriculture is used for analyzing the various biotic and abiotic factors. Agriculture in India plays a predominant role in economy and employment. The common problem existing among the Indian farmers are they don't choose the right crop based on their soil requirements. Due to this they face a serious setback in productivity. This problem of the farmers has been addressed through precision agriculture. Precision agriculture is a modern farming technique that uses research data of soil characteristics, soil types, crop yield data collection and suggests the farmers the right crop based on their site specific parameters. This reduces the wrong choice on a crop and increase in productivity. In this paper, the problem has been solved by proposing a recommendation system through an ensemble model with majority voting technique using Random tree, CHAID, K-Nearest Neighbor and Naive Bayes as learners to recommend a crop for the site specific parameters with high accuracy and efficiency.

Benefits:
• One can get the harvest suggestion dependent on his site boundaries.
• Farmers can get the anticipated yield by exactness farming strategies utilized in this framework and urged to utilize this framework.
• The fundamental strategy utilized in this framework were notable and one can trust on this framework and utilize this.
Impediments:
• More information is required and sets aside more effort for forecast.
• Suitable for crop idea and not for crop yield expectation.
[4] Crop Yield Prediction Using Deep Neural Networks by Saeed Khaki and Lizhi Wang.

Portrayal: Crop yield is a profoundly mind boggling characteristic dictated by numerous variables like genotype, climate, and their communications. Precise yield expectation requires crucial comprehension of the utilitarian connection among yield and these intelligent variables, and to uncover such relationship requires both far reaching datasets and amazing calculations. A profound neural organization (KNN) approach that exploited cutting edge demonstrating and arrangement methods. Their model was found to have an unrivaled expectation precision, with a root-mean-square-mistake (RMSE) being 12% of the normal yield and half of the standard deviation for the approval dataset utilizing anticipated climate information. With amazing climate information, the RMSE would be diminished to 11% of the normal yield and 46% of the standard deviation. They likewise performed include choice dependent on the prepared DNN model, which effectively diminished the component of the info space without critical drop in the expectation precision. Their computational outcomes

proposed that this model altogether beat other mainstream strategies like Lasso, shallow neural organizations (SNN), and relapse tree (RT). The outcomes likewise uncovered that natural elements greatly affected the harvest yield than genotype.

Benefits:

• They introduced an AI approach for crop yield forecast, which showed prevalent execution in the 2018 Syngenta Crop Challenge utilizing enormous datasets of corn cross breeds.

• The approach utilized profound neural organizations to make yield forecasts (counting yield, check yield, and yield distinction) in light of genotype and climate information. Limitations:

- Genotype and environment data are often represented by many variables, which do not have equal effect or importance in yield prediction.
- It is vital to find important variables and omit the other redundant ones which might decrease the accuracy of predictive models.
- Solar radiation and temperature have considerable effects on the variation in corn yield across different environments.

[5] Analysis of Crop Yield Prediction utilizing Data Mining Techniques by D Ramesh and B Vishnu Vardhan. Portrayal: Agrarian area in India is dealing with thorough issue to amplify the harvest efficiency. In excess of 60% of the harvest actually relies upon storm precipitation. Ongoing advancements in Information Technology for farming field has become an intriguing examination region to foresee the harvest yield. The issue of yield forecast is a significant issue that still needs to be addressed dependent on accessible information. Information Mining procedures are the better decisions for this reason. Distinctive Data Mining methods are utilized and assessed in agribusiness for assessing what's to come year's harvest creation. This paper presents a concise examination of harvest yield expectation utilizing Multiple Linear Regression (MLR) strategy and Density based grouping method for the chose district for example East Godavari area of Andhra Pradesh.

Benefits:

• In this paper the primary point of creators is to make an easy to use interface for ranchers, which gives the investigation of rice creation dependent on accessible information.

• Different Data mining methods were utilized to anticipate the harvest yield for amplifying the yield efficiency.

• The approach utilized here was the factual strategy specifically Multiple Linear Regression procedure and Data Mining strategy in particular Density-based

grouping method were taken up for the assessment of harvest yield investigation.

Constraints:

• A enormous arrangement of information should be accessible.

• It is just appropriate for one harvest and not for different yields.

• It is especially kept up with for a chose locale in one state

3. PROPESED SYSTEM

In the proposed framework, directed learning calculations are utilized to shape a model which will help us in giving decisions of the most achievable harvests that can be developed around there alongside its assessed yield. Two of the calculations utilized here is K-Nearest Neighbor and. The principle stages associated with the interaction are dataset assortment, pre-preparing the information, include extraction and characterization. *Dataset collection*

A. The dataset utilized for this task is gathered from different online sources like Kaggle.com and data.govt.in. We have taken the farming information of four areas to be specific, Mangalore, Hasana, Kodagu and Kasaragod. Some significant highlights or the boundaries which highestly affects the horticultural yield considered in the undertaking are recorded underneath. Rainfall (in mm)

- Humidity
- Temperature
- Area
- Yield
- Type of the soil
- Location
- Price

C. Pre-processing the data

D. After the choice of the dataset, it must be pre-handled into a structure that you can work with. A portion of the means are organizing, cleaning and testing. At first the information you have chosen is changed over into the configuration appropriate for you to work with. Cleaning information is the evacuation or fixing of the blended information. Inspecting is taking a little delegate test of the chose information that might be a lot quicker for investigating the arrangements than choosing the opening dataset. *Transforming the data* The last advance is changing the chose information. The pre processed information here is then changed into information that is prepared for AI calculations by utilizing different designing highlights like scaling, include total, etc. There might be a few highlights that can be joined into a solitary element which would be more significant to the

issue you are attempting to tackle. Figure 1 underneath demonstrates the last information to be utilized by the classifiers. Figures 2 and 3 shows the framework plan of the proposed framework
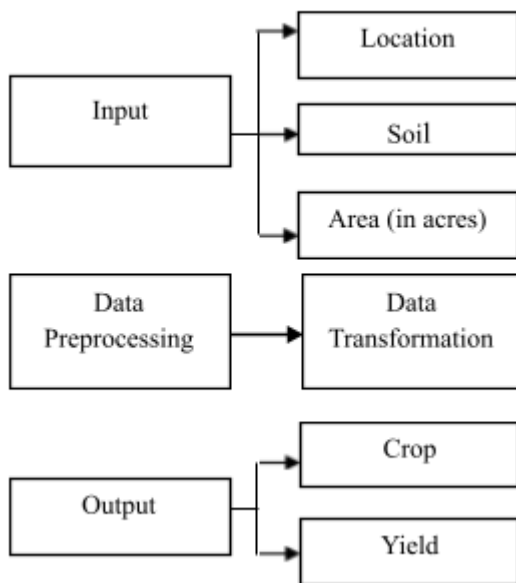
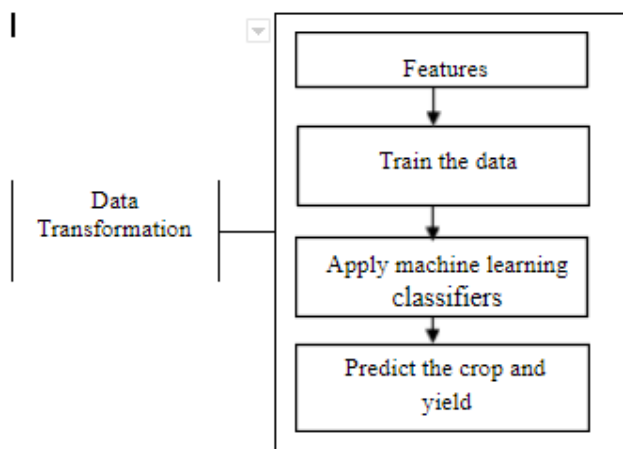**Fig 1:** Final dataset

Fig 2: System design(a)

Fig3: system design (b)

4. ALGORITHMS USED FOR PROPESED SYSTEM

The changed information is then parted into two sets specifically, preparing sets and testing sets prior to applying the AI classifiers. The two AI classifiers utilized here are K-Nearest Neighbors. When the model is prepared effectively it is tried on the testing dataset which is not the same as the preparation information .A. KNN algorithm

KNN is an administered AI calculation. It learns by similarity. It's anything but a straightforward yet an incredible methodology for making forecasts. In the venture, as per the information given, the dataset is preprocessed to get the separated dataset whichis our preparation set. Test information is chosen arbitrarily from this training set. K-most comparative records to the test record is determined. Euclidian distance is determined for discovering the likeness between the records. When k neighbors are found , a summed up expectations are made by returning the most well-known result. $( ) \sqrt{\sum} ( ) (1)$

Condition 1 shows the euclidian distance recipe. It is utilized to ascertain the distance between the two information focuses in a plane. KNN don't have a learning stage thusly. It simply computes the distance between the test set and each line of the preparation set and returns the most comparable ones as its neighbors. The lesser the distance, more comparative the records are. Consequently it is called as sluggish student method or learning by similarity. Yet it is considered as on of the most powerful algorithms.
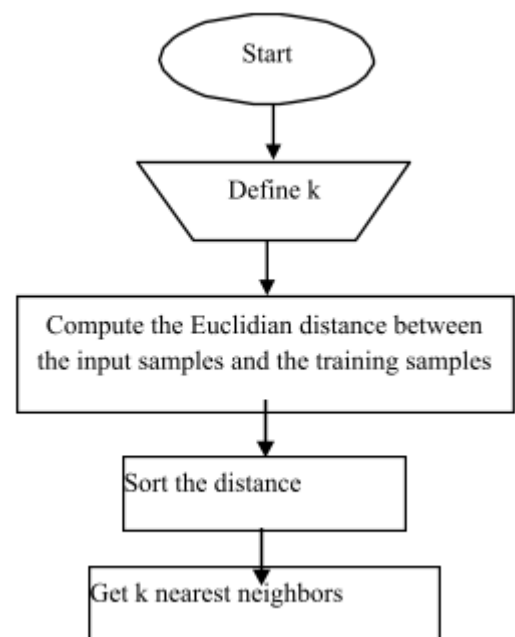

Fig 4: KNN Classifier

Figure 4 shows the steps involved in the KNN Algorithm.

5. METHODOLOGY AND RESULTS

In the proposed framework, at first it assembles fundamental client data like the Name, Email id, Password and telephone number from the client. It is then put away into worker's information base. Once enlisted, Username and the secret word gets confirmed from the worker side each time the client attempts to sign in to the site. The client would then be able to choose the choice to foresee the yield which is appropriate for the specific area and the dirt kind as displayed in the figure 6 underneath. He will likewise get the assessed yield of the anticipated harvest as per the quantity of sections of land of land he has given as the info
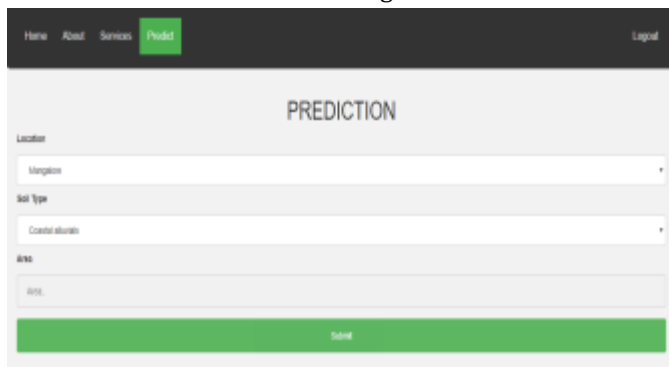

Fig 5: Crop yield prediction

As displayed in the figure 7 beneath, From the given information sources and by applying the AI classifiers, two decisions of the harvests are anticipated alongside its assessed yield for the given number of sections of land of land.

Fig 6: Prediction result

The execution of KNN is done in Python, Execution time in seconds from epoche, and as far as their Precision and Recall scores. Order Report is a report that is utilized to quantify the nature of expectations from a characterization calculation. Exactness, Recall, F1score are a portion of the measurements given in the Classification report. Accuracy is the capacity of a classifier not to name an example positive that is really negative. Review is the capacity of a classifier to track down every certain occurrence.

6. CONCLUSION

This project provides an easy to access and an efficient system to predict the crops and the estimation of its yield under the given conditions for a particular region. Farmers are still not connected with the modern technologies. It efficiently bridges the gap between the rural farmers and the modern technologies. Machine learning algorithms have proved very effective in predicting the crops and its yield. From the comparison analysis of KNN. In the future, Farming can be taken to next levels by connecting all the farming devices to the internet using IOT.

REFERENCES

[1] A Survey on Crop Recommendation Using Machine Learning by M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran.
[2] A Review on Data Mining Techniques for Fertilizer Recommendation by Jignasha M. Jethva1, Nikhil Gondaliya, Vinita Shah.
[3] Crop Recommendation System for Precision Agriculture by S.Pudumalar, E.Ramanujam, R.Harine Rajashreen, C.Kavyan, T.Kiruthikan, J.Nishan.
[4] Crop Yield Prediction Using Deep Neural Networks by Saeed Khaki and Lizhi Wang.
[5] Analysis of Crop Yield Prediction using Data Mining Techniques by D Ramesh and B Vishnu Vardhan.
[6] agrifarming.in website
[7] kaggle.com website