A METHODOLOGY TO PREDICT CROPPING PATTERN USING SOIL PROPERTIES

[1]Vanamala Navya sri, [2]Bijjam Sambhavi, [3]Kode Girija Sankar, [4]Vuyyuru Sai Kiran Reddy, [5]N. Md. Jubair Basha

[1,2,3,4] B. Tech Student, CSE Department, Kallam Haranadhareddy Institute of Technology, Guntur, India

[5]Assosiate Professor, CSE Department, Kallam Haranadhareddy Institute of Technology, Guntur, India

[1]navyasrivanamala987@gmail.com, [2]bijjamsambhavi@gmail.com, [3]girijasankarkode@gmail.com,

[4]vuyyurisaikiranreddy@gmail.com, [5]jubairbasha@gmail.com

Abstract : India is generally an agricultural country. Now a days the most important emerging field in the real world is agriculture and it is the main occupation and backbone of our country. Around 60 percent of the whole land within the country is used for agriculture to satisfy the wants of 1.2 billion people, so improving crop production is therefore seen as a big aspect of agriculture. Production of crops could also be a difficult task since it involves various factors like soil type, temperature, humidity etc. If it's possible to hunt out the crop Production before sowing it, it'd be of great help to the farmers. Predicting the Cropping pattern beforehand of its harvest would help the farmers take appropriate measures for better results. Recent developments in In-formation Technology for agriculture field have become an interesting research area to predict the crop. Crop prediction is the methodology to predict the crops using different parameters like PH level of soil, Temperature, Fertilizers, Potassium(P), Phosphorous(K), Nitrogen(N) and other atmospheric conditions are parameters. Data mining techniques is very popular in the area of agriculture. Machine learning techniques are used and evaluated in agriculture for estimating the future years crop production. This paper presents a brief analysis of crop prediction using K-Nearest Neighbour (K-NN). This Paper may be a recommendation system which makes use of various data processing techniques such as recommending the acceptable crops that support the input soil parameters. The proposed technique thus reduces the financial losses faced by the farmers caused by planting the wrong crops and also it helps the farmers to hunt out new kinds of crops which will be cultivated in their area.

IndexTerms - Agriculture, K-Nearest neighbor, Machine Learning, Crop production.

I. INTRODUCTION

It is estimated that 795 million people still live without an adequate food supply, and that by 2050 there will be two billion more people to feed. A central challenge to address food security issues is yield estimation, namely being able to predict crop well before harvesting. Agricultural monitoring, in particular in developing countries, can improve food production and support humanitarian efforts in light of climate change and droughts. Existing approaches rely on survey data and other variables related to crop growth (such as weather and soil properties) to model crop yield. The aim is to use a model where information focuses are clustered in a few groups in order to predict the classification of another instance. This application's evaluations help farmers predict cropping pattern. Support Vector machine (SVM), Regression analysis, K-Nearest Neighbour (KNN), Clustering and various types of techniques are used for prediction. The Proposed system worked with agriculture annual data that was collected by the IBS form annual or seasonal conducted surveys. The dataset contains the humidity, temperature, and rainfall, estimate of acreage, production, and yield, crop losses caused by floods, annual land utilization, and annual weather and crop reports by districts.

Predicting the cropping pattern in advance of its harvest would help the policy makers and farmers for taking appropriate measures for marketing and storage. This paper will help the farmers to know the cropping pattern based on soil conditions before cultivating onto the agricultural field and thus help them to make the appropriate decisions. It attempts to solve the issue by building a prototype of an interactive prediction system. Thus, for such kinds of data analytics in crop prediction, there are different techniques or algorithms, and with the help of those Machine Learning algorithms this paper can predict crop using KNN.

In India, there are several ways to increase the economic growth in the field of agriculture. There are multiple ways to increase and improve the crop production and the quality of the crops. Data mining is also useful for predicting cropping pattern. The main objectives is to use machine learning techniques to predict crop, to increase the accuracy of crop production.

II. LITERATURE SURVEY

In [1] **Karan deep kauri,** have concluded that the paper will review that various applications of machine learning in the farming sector. And also provides an insight into the troubles faced by our Indian farmers and hoe these can be solved using these techniques. This method help in increasing the farming sector in the countries and apply the more machine learning applications. The algorithms used are artificial neural networks, Bayesian Belief Network, Decision Tree Algorithms, Clustering, Regression analysis. The disadvantage is less accuracy in term of performance.

In [2] **E. Manjula, S. Djodiltachoumy,** have concluded that the aim of paper is to propose and implement a rule based system and predict the yield production from the collection of previous data. The algorithms used are K- means algorithm, Clustering method. The disadvantage is it considered less data.

In [3] **Nishit Jain, Amit Kumar, At.al,** concluded that the paper helps in predicting crop sequences and maximizing yield rates and making benefits to the farmers. Also using machine learning applications with agriculture in predicting crop diseases, studying crop simulations, different irrigation patterns. The algorithms used are artificial neural networks, support vector machine. The disadvantage is exact accuracy is not specified.

In [4] **Raorane A.A, Dr. Kulkarni R. V,** have concluded that this method will help in estimating rain fall and investigate the reasons for getting lower yield. The algorithm used is regression analysis method. The disadvantage is that here the specific method is not specified.

In [5] **Ashwani Kumar Kushwaha, SwetaBhattachrya,** have concluded that this method will provide agro algorithm which helps in predicting suitable crop for the lands. And this helps in enhancing the quality of crop. The algorithm used is agro algorithm. The disadvantage is it results in less prediction of crops.

In [6] **T.R. Lekhaa**, "Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques" The paper hypothesizes analysis of Explorative Data and considers the design of different types of predictive models. A data set is taken as a sample data set, and different regression techniques are tried to recognise and examine each property. Specific regression methods discussed here are Multiple Linear, Linear, Non-Linear, Polynomial, Ridge regression and Logistic. Using this article, a comparative study of the different algorithms in data analytics. This helped in determining which algorithm is most appropriate to the proposed system.

III. SYSTEM ANALYSIS
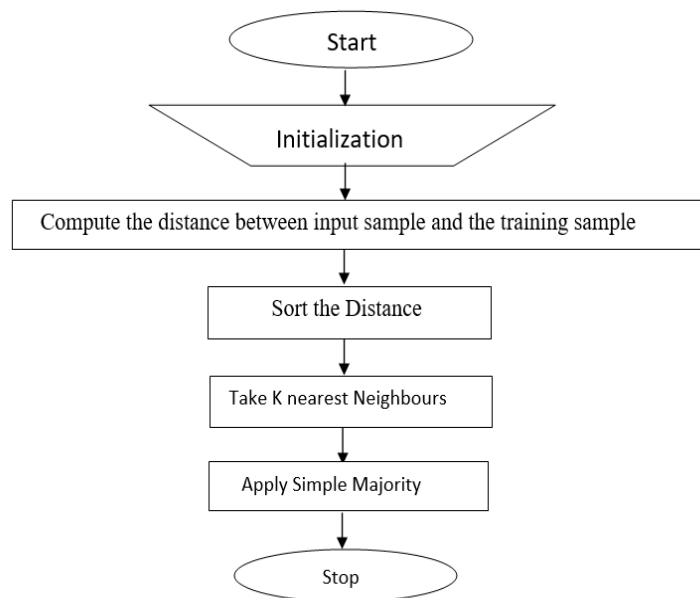
3.1 Existing System

Authors have designed a precise and effective model to estimate crop yield in term of climate change. Broader approaches using statistical data have been used in this model. However, simulated observation and regression analysis methodology have led to design the model. On the other hand, yield responses to different temperatures and precipitation are being simulated with CERES-Maize, which gives various results in the developing the model. Overall, although few factors that can affect the statistical model are not considered here, this model can help to predict yield crop production being responsive to climate change.
The existing system is completely implemented on the WEKA tool.

a) Four different algorithms are used namely J48, LAD Tree, IBK, LWL.
b) The dataset is collected from the various websites like indiaagristat.com and india.gov.in/data-portal-india.
c) All the analysis of the data set was done using "WEKA (Waikato Environment for Knowledge Analysis)" [8]. It is open source software which is written in JAVA programming language and developed at the University of Waikato, New Zealand. It is used for solving data mining and machine learning problems

3.1.1 Disadvantages of existing system

*   There is no particular prediction algorithm used to predict the crop yield.
*   Accuracy is very low.
*   The system is for predicting crop yield but there is a lack of predicting algorithms.
*   Presence of Variance and bias due to that predictions may become wrong.

```
           ┌─────────────┐
          (    Start     )
           └─────────────┘
                 │
                 ▼
         ╱─────────────────╲
         ╲  Initialization  ╱
          ╲───────────────╱
                 │
                 ▼
    ┌──────────────────────────────────────────────────────┐
    │ Compute the distance between input sample and the      │
    │              training sample                           │
    └──────────────────────────────────────────────────────┘
                 │
                 ▼
         ┌───────────────────┐
         │  Sort the Distance │
         └───────────────────┘
                 │
                 ▼
         ┌─────────────────────────┐
         │ Take K nearest Neighbours│
         └─────────────────────────┘
                 │
                 ▼
         ┌─────────────────────┐
         │ Apply Simple Majority│
         └─────────────────────┘
                 │
                 ▼
           ┌─────────────┐
          (    Stop      )
           └─────────────┘
```

Figure 3.3.1 KNN Classifier Algorithm for Crop Prediction

Algorithm Description for Crop Prediction:

Step 1: Start

Step 2: Initialize K value

Points to remember while selecting the value of K

A very low value for k such as K=1 or K=2, can noisy and lead to the effects of outliers in the model.

Step 3:

Compute the distance between Input sample and Output sample i.e. Calculate the Distance of K number of neighbors by using the KNN metric namely Euclidean Distance.

$$d(x,\ y)\ =\ \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

Step 4: Sort the Distance

Step 5: Take the Nearest neighbors

(Take the nearest neighbors as per the calculated Euclidean Metric, among these K neighbors count the number of data points in each category, Assign the new data points to that category for which the number of the neighbor is maximum.)

Step 6: Apply Simple Majority.

Step 7: End

IV. GENERAL ARCHITECTURE

4.1 General Architecture for Crop Prediction
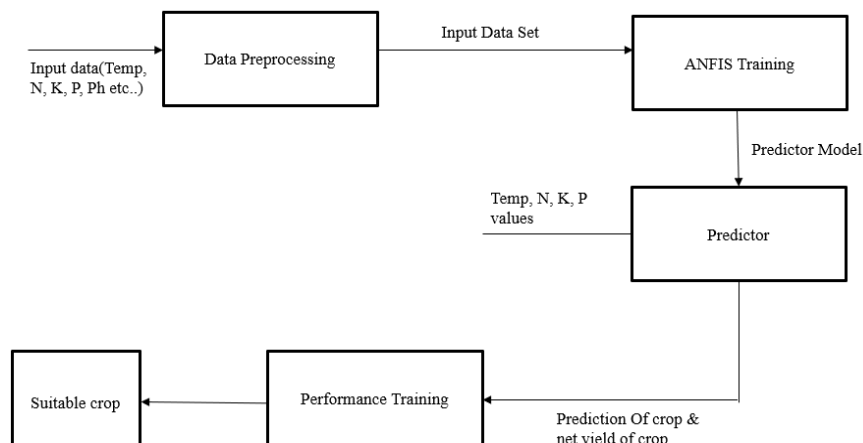
Figure 4.1. Architecture for Crop Prediction

Description: Architecture diagram will provide the outline of the paper. This Model collect the data of temperature, N, P, K, PH etc..., the data will be pre-processed and the input data is went to ANFIS training and the data will be predicted, through thus the paper will get the prediction of the Crop.

V. IMPLEMENTATION

5.1 Steps for Execution

Step 1: Open Anaconda Prompt and run the code
Step 2: After installing the NumPY, Sklearn, Flask, SQL Client e.t.c and running the code, it displays the IP Address Link.
Step 3: Copy the link and browse in the selected browser.
Step 4: Next it displays page as follows

Figure 5.1 Input Screen

Fig 5.1 describes the Home page which consists of Soil Properties such as Nitrogen, Phosphorous, Potassium, Ph Value and Temperature. Based on various analyses the parameters above are taken as input and prediction have been undertaken.
Step 5: Taking input from user.
Step 6: By clicking on Analyse the proposed model predict the specific Crop with respect to the given Soil Properties and Temperature.

5.2 Module Description

- Collection of Data Sets Module
- Applying KNN Algorithm Module
- Prediction of Crop Module

5.2.1 Collection of Data Sets Module:

Collected the Data set from Indian Agricultural Statistics and load a dataset using Pandas library, and apply the following algorithms and then find the best one for the specific dataset by accuracy evaluation methods.

Figure 5.2.1: Collection of Data Sets Module

Figure 5.2.1. describes the Collection of dataset from Indian Agricultural Statistics which consists of Features & Labels where feature namely N, P, K, pH level of soil, Temperature and from these parameters name of the crop & net yield of the crop can be predicted.

5.2.2 Applying KNN Algorithm Module:

The data sets have been collected and refined based on commonality uses such as Location, Crop, Area, Soil type, Temperature, Humidity etc. From these parameters name of the crop and net yield rate of the crop can be predicted. Based on various analyses the parameters Temperature, soil contains, PH are taken as input and prediction have been undertaken. The attribute soil type specifies the type of soil in a particular region such as Coastal alluvia's. By using KNN algorithm, the particular crop has been analyzed and predicted by taking various parameters into an account such as PH, Temperature, P, K, and N. By analyzing and predicting the crop name and price of particular crop can be found out. This helps the farmers to take correct decision to sow the crops such that yield rate can be increased. The inputs are classified into features as x and the output crop type will be trained to labels as y. By the reference of previous data the KNN algorithm will find the nearest points to the given input (the features such as temperature, PH) and it will be further classified and predicted as output crop.

5.2.3 Prediction of Crop Module:

This paper represents a review of KNN technique for the early Crop prediction. KNN analysis is used for predicting the unknown parameter from the known parameters. In this work the proposed paper are considering Temperature, PH, K, N, P as input parameters which are the main parameters to be considered for a good crop yield, although there are many other factors that can be considered. The unknown value of the nearest neighbours by calculation of Euclidean distance between them. Then the proposed paper would be able to predict crop as follows for given Temperature, PH, K, N, P parameters.

Figure 5.2.3. Prediction of Crop

Description:

There are three varieties of peas that will suit your garden and cooking needs:
Pisum savitum, which includes both types of garden peas: sweat peas ((inedible pods) and snow peas (edible flat pods with small peas inside).
Pisum macroncarpon, snap peas (edible pods with full-size peas).

Pea plants are easy to grow, but have a very limited growing season. Furthermore, peas do not stay fresh long after harvest, so enjoy them while you can!

VI. RESULTS AND DISCUSSIONS

The output design is one which meets the client (end user) requirements and presents the information clearly. The output is the direct information to the user. The output of this proposed paper is the crop name that is suitable for particular type of soil.

Result

Figure 6.1: Output for Crop Prediction

Description:

After giving soil properties and temperature values just by clicking on Analyse button the Proposed system analyses and predict the suitable Crop along with net Yield with description of the specific crop as shown in Fig 6.1 with respect to the parameters given.

Figure 6.1 Background running of Anaconda prompt

Description: The proposed work used the Anaconda tool for execution. After running the code it displays one IP Address Link paste it in browser. Figure 8.1.2 is the Background running of Anaconda prompt it displays the user given parameters along with crop name as label as shown.

VII. Comparisons

7.1 Comparison of Existing and Proposed System

While comparing Proposed System and Existing system, proposed system is more accurate than the existing system. The existing system is statistical method which does not provide accurate results and cannot predict easily. The existing system is based on the climate factor where as the proposed method considers various factors like climate, temperature, humidity and other chemical component data sets it is more efficient and economical.
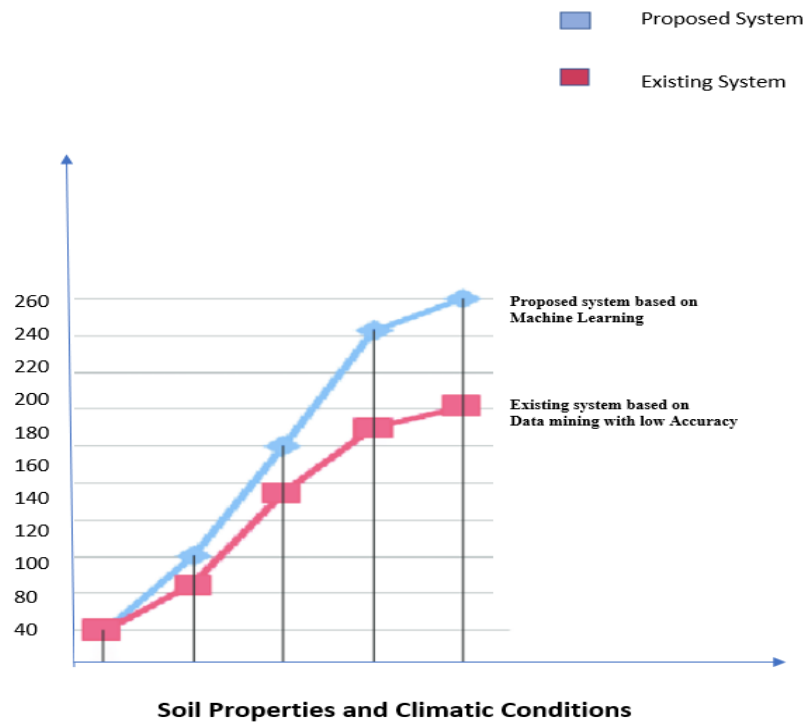

Figure 7.1 Comparison of Existing and Proposed system for crop prediction

Description:

The existing system is statistical method which does not provide accurate results and cannot predict easily. The existing system is based on the climate factor where as the proposed method considers various factors like Soil properties, climate, temperature, humidity and other chemical component data sets it is more efficient and economical. This Proposed system can further recommend which crop has to be yielded based on the soil properties, climatic conditions and PH of the soil.

VIII. Conclusion

This paper presents a Machine learning framework for the task of crop 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. The Proposed Model uses modern representation learning ideas for crop yield prediction, and successfully learn much more effective features from raw data compared with the hand-crafted features that are typically used. The existing system is statistical method which does not provide accurate results and cannot predict easily. The existing system is based on the climate factor where as the proposed method considers various factors like Soil properties, climate, temperature, humidity and other chemical component data sets it is more efficient and economical. This paper propose a dimensionality reduction approach based on KNN classifier algorithm for model provides us with the state-of-the-art prediction accuracy and will have great impact in sustainable agriculture and food security.

IX. Future Enhancements

The Proposed Paper Model believe that the system will be able to help framers to take the right decision of cultivating the right crop. A farmer can plant different crop in different districts based on the system recommendation. So, every farmer will get the chance of maximizing their yield and profit by using the system. The main goal is to produce more with less as even being a developing country; and almost using all the 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.

- Proposed system uses "KNN algorithm" to find the specific crop & net yield using Machine Learning.
- In data science the proposed paper have many algorithms for classification such as Naive Bayes, SVM, Decision Tree, ID3 etc…
- In future proposed model can add more algorithms to find outputs and algorithms can be compared to find the efficient algorithm.

REFERENCES

[1] Karan deep Kauri,2019, "Machine Learning: Applications 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", Interna tional 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 Bhattachrya, 2016, "Crop yield prediction using Agro Algorithm in Hadoop", 14(3),pp. .,106-114.

[6] T.R. Lekhaa, "Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques".

[7] Viviliya, B. and Vaidhehi, V., "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms".