International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 02 | February - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

CROP PREDICTION

Swathi. A, Assistant Professor, Department of Computer Science

St. Joseph's First Grade College, Jaylakshmipuram, Mysore-570012

ABSTRACT

Cereal crops such as rice, wheat, and different pulses account for the majority of India's food output. Predicting crop yields far ahead of harvest would assist policymakers and farmers in making informed decisions about agronomy, crop selection, and agricultural planning. Such forecasts will also assist related sectors in planning their logistical operations. The goal of the research is to create a machine learning model that can generate such predictions. The model is trained using a dataset that incorporates soil data from the previous decade, with features such as Ph value, temperature, and crop name, as well as labels such as crop yield. The model learns the link between the yield and variables such as soil type, location, and forecast using appropriate machine learning techniques.

INTRODUCTION

Agriculture has been defined as the science and practise of raising plants and livestock. Agriculture became one of the most important trends in the rise of sedentary human civilisation. The domestication of plants and animals is documented throughout agricultural history. Agriculture began on its own in many regions of the world. Agriculture's advancement allowed the human population to expand beyond what could be supported by hunting and gathering.

Soil fitness is defined as the capacity to fulfil essential environmental functions such as nutrient cycling, water filtering, and plant and animal habitat supply. Texture, depth, density, water penetration, preserving capacity, amount of natural remember, nutrient maintaining ability (CEC), and respiratory properties are all factors that influence soil fitness. Plants grow in a variety of environments. Air, water, nutrients, light, temperature, and growth space are all essential for plant existence, although the specifics vary greatly. Some flowers thrive in the wild, while others demand a tropical environment. Some plants can withstand harsh winter weather, while others may expire if the temperature goes below a particular level.

Soil is an important source of nutrients for plant development. Nitrogen (N), phosphorus (P), and potassium (K) are the three most important nutrients (K). NPK is the name of the group they form together. Calcium, magnesium, and sulphur are other important nutrients. Plants also require trace amounts of iron, manganese, zinc, copper, boron, and molybdenum, sometimes known as hint elements, because the plant only requires the simplest strains.

LITERATURE SURVEY

Each Software development requires the review interaction. The Survey interaction is expected to get the necessity for the product. The Survey likewise comprises of considering the current framework and furthermore reading the instruments required for the advancement of the product. A legitimate comprehension of the apparatuses is a lot of fundamental. Following is a concentrate of the data of the material gathered during the writing study. A writing review is a procedure of recognizing the issues in the current framework through research and proposing the advancement of the framework to take care of the issues of the current framework.

1. Crop Yield Prediction Using Machine Learning Algorithms

Authors: Aruvansh Nigam, Saksham Garg

Year: 2019

Findings

Agriculture is one of the major and the least paid occupation in India. Machine learning can bring a boom in the agriculture field by changing the income scenario through growing the optimum crop. This paper focuses on predicting the yield of the crop by applying various machine learning techniques. The outcome of these techniques is compared on the basis of mean absolute error. The prediction made by machine learning algorithms will help the farmers to decide which crop to grow to get the maximum yield by considering factors like temperature, rainfall, area, etc.

2. Use of Data Mining in Crop Yield Prediction

Authors: Shruti Mishra, Priyanka Paygude

Year: 2021

Findings

This paper focuses on implementing crop yield prediction system by using Data Mining techniques by doing analysis on agriculture dataset. Different classifiers are used namely J48, LWL, LAD Tree and IBK for prediction and then the performance of each is compared using WEKA tool. For evaluating performance Accuracy is used as one of the factors. The classifiers are further compared with the values of Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Relative Absolute Error (RAE). Lesser the value of error, more accurate the algorithm will work. The result is based on comparison among the classifiers.

3. Crop Yield Prediction Using Deep Neural Networks

Authors: Saeed Khaki, Lizhi Wang

Year: 2019

Findings

Our model was found to have a superior prediction accuracy, with a root-mean-square-error (RMSE) being 12% of the average yield and 50% of the standard deviation for the validation dataset using predicted weather data. With perfect weather data, the RMSE would be reduced to 11% of the average yield and 46% of the standard deviation. We also performed feature selection based on the trained DNN model, which successfully decreased the dimension of the input space without significant drop in the prediction accuracy. Our computational results suggested that this model significantly outperformed other popular methods such as Lasso, shallow neural networks (SNN), and regression tree (RT). The results also revealed that environmental factors had a greater effect on the crop yield than genotype

4. Crop Yield Prediction based on Indian Agriculture using Machine Learning

Authors: Potnuru Sai Nishant, Pinapa Sai Venkat

Year: 2020

Findings

In India, we all know that Agriculture is the backbone of the country. This paper predicts the yield of almost all kinds of crops that are planted in India. This script makes novel by the usage of simple parameters like State, district, season, area and the user can predict the yield of the crop in which year he or she wants to. The paper uses advanced regression techniques like Kernel Ridge, Lasso and ENet algorithms to predict

the yield and uses the concept of Stacking Regression for enhancing the algorithms to give a better prediction.

METHODOLOGY

An algorithm is defined as a set or sequence of statements which are well defined or instructions to solve a particular problem. The instructions must be unambiguous in nature. The algorithm used to design this model is KNN. KNN is the most efficient algorithm for classification in neural network.

3.1 ALGORITHMS

3.1.1 KNN

K-nearest neighbours (KNN) algorithm is a type of supervised ML algorithm which can be used for both classification as well as regression predictive problems. However, it is mainly used for classification predictive problems in industry. The following two properties would define KNN well −

- **Lazy learning algorithm** − KNN is a lazy learning algorithm because it does not have a specialized training phase and uses all the data for training while classification.
- **Non-parametric learning algorithm** − KNN is also a non-parametric learning algorithm because it does not assume anything about the underlying data.

Working of KNN Algorithm

K-nearest neighbours (KNN) algorithm uses ‘feature similarity’ to predict the values of new data points which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. We can understand its working with the help of following steps −

Step 1 − For implementing any algorithm, we need dataset. So during the first step of KNN, we must load the training data from the exam taken data list and test data from the students currently being taking tests.

Test data set – Upload data based on user input dataset.

Train data set – collected data samples.

Step 2 − Next, we need to choose the value of K i.e., the nearest data points. K can be any integer.

Here k is taken from test data set, which acts like a centroid point.

Step 3 − For each point in the test data do the following −

- **3.1** − Calculate the distance between test data and each row of training data with the help of any of the method namely: Euclidean, Manhattan or Hamming distance. The most used method to calculate distance is Euclidean.
- **3.2** − Now, based on the distance value, sort the user exams preference data in ascending order.
- **3.3** − Next, it will choose the top K rows from the sorted array.
- **3.4** − Now, it will assign a class to the test point based on most frequent segments from the soil data of these rows.

Step 4 – End

In our project we compare k value with train data set using KNN methodology like Euclidean, Manhattan, where we cluster the values based on the nearest distance. Here we get nearest matches of the data preference to test data from train data set in turn we predict the content values from it.

PSEUDOCODE OF KNN

Nearest-neighbor algorithm

a) A pseudo code for the nearest neighbor algorithm is

ALGORITHM *Nearest-neighbor*$(D[1..n,1..n],s)$
//Input: A $n\times n$ distance matrix $D[1..n,1..n]$ and an index s of the starting city.
//Output: A list Path of the vertices containing the tour is obtained.
for $i \leftarrow 1$ to n do Visited $[i] \leftarrow$ false
Initialize the list Path with s
Visited $[s] \leftarrow$ true
Current $\leftarrow s$
for $i \leftarrow 2$ to n do
    Find the lowest element in row current and unmarked column j containing the element.
    Current $\leftarrow j$
    Visited $[j] \leftarrow$ true
    Add j to the end of list Path
Add s to the end of list Path
return Path

3.1.2 CNN

Artificial Neural Networks are quite effective when it comes to Machine Learning. Artificial Neural Networks are utilized for a variety of classification tasks, including picture, audio, and word categorization. For example, we utilize Recurrent Neural Networks, more exactly an LSTM, for predicting the sequence of words, and we use Convolution Neural Networks for picture categorization.

The neural network is a type of computer network. There are three sorts of layers in a standard Neural Network:

1. **Input Layers:** It's the layer where we feed data into our model. The entire number of characteristics in our data is equal to the number of neurons in this layer (number of pixels in case of an image).
2. **Hidden Layer:** The hidden layer receives the information from the input layer. Depending on our model and data size, there might be a lot of hidden layers. The number of neurons in each hidden layer might vary, although they are usually more than the number of characteristics. The output from each layer is generated by matrix multiplication of the preceding layer's output with that layer's learnable weights, then addition of learnable biases, and finally activation function, which makes the network nonlinear.
3. **Output Layer:** The hidden layer's output is then input into a logistic function like sigmoid or SoftMax, which turns each class's output into a probability score for each class.

PSUEDOCODE OF CNN

```
PSEUDOCODE OF CONVOLUTIONAL LAYER
1 for (l = 0;l<L ;l ++){
2       for (m = 0;m<M ;m ++){
3         for (n = 0;n<N ;n ++){
4               sum = bias[l];
5               for (k = 0;k<K ;k ++){
6                       for (s1 = 0;s1 <S1;s1 ++){
7                               for (s2 = 0;s2 <S2;s2 ++){
8                   sum+=weight[k][l][s1][s2] × input [k][m + s1][n + s2];
9               }}}
10        output [l][m][n] = activation_func(sum);
11 }}}
```

PROPOSED FRAMEWORK

Initially, system will be provided with the soil data like Ph level, location of the user will be automatically retrieved by the system. Once the required data has been feed to the system it will give prediction of crops which can be grow in the soil. System will also provide the information’s like seed required for the acres, investment, and profit on the crops.

Advantages:

- Easy to give prediction of crop
- Better accuracy than the existing system
- Time consumption is reduced

HARDWARE AND SOFTWARE REQUIREMENTS

Hardware Requirements

- Processor : Intel i5 2.53GHz
- Hard Disk : 30GB
- RAM : 4 GB or above

Software Requirements

- Operating system : Windows 8 and above
- Coding Language : .Net
- IDE : MS Visual Studio

Detailed Design

- Definite plan begins after the framework configuration stage is finished and the framework configuration has been guaranteed through the audit. The objective of this stage is to foster the interior rationale of every one of the modules recognized during framework plan.
- In the framework plan, the attention is on recognizing the modules, though during nitty gritty plan the emphasis is on planning the rationale for the modules. At the end of the day, in framework plan consideration is on which segments are required, while in definite plan how the segments can be executed in the product is the issue.
- The design process is frequently split into two phases: system design and detailed design. Top-level design is another term for system design. At the most basic level, the focus is on determining which modules are required for the system, their specs, and how they should be linked. This is referred to as top-level design or system design. The internal design of the modules, or how the module's standards may be met, is decided at the second level. This level of design is also known as detailed design or logic design

Data Flow Diagram

DFD graphically addressing the capacities, or cycles, which catch, control, store, and circulate information between a framework and its current circumstance and between segments of a framework. The visual portrayal makes it a decent specialized apparatus among User and System architect. Design of DFD permits beginning from a wide outline and extend it to a pecking order of nitty gritty charts. DFD has frequently been utilized because of the accompanying reasons:

- Logical information flow of the system

- Determination of physical system construction requirements
- Simplicity of notation
- Establishment of manual and automated systems requirements

Data Flow Diagram

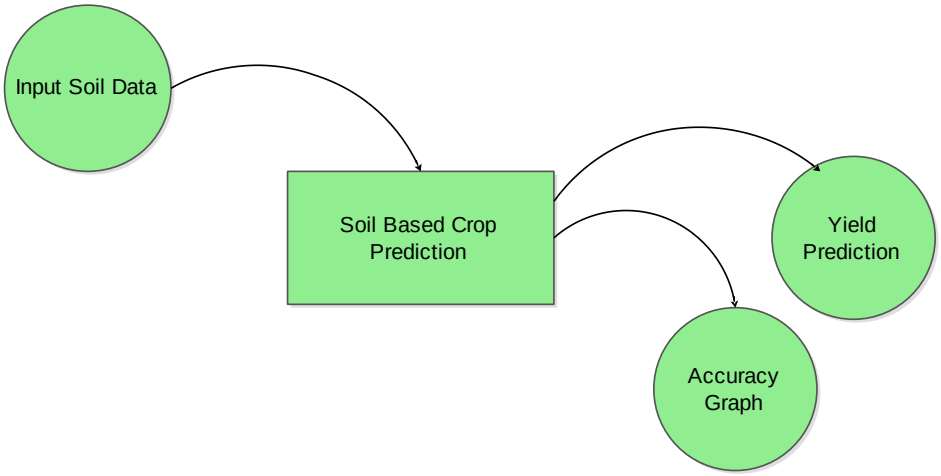
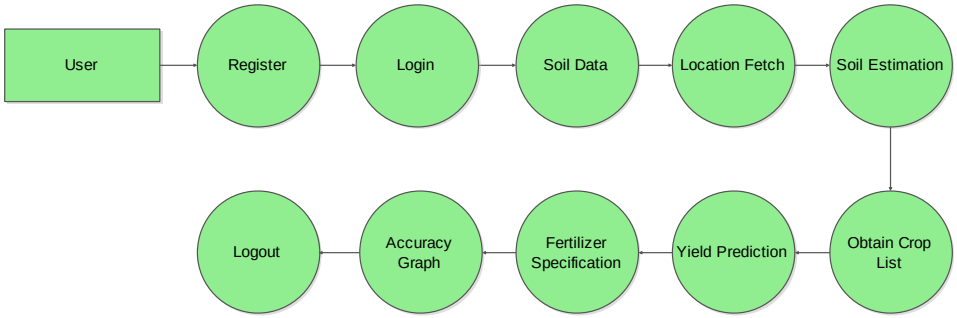
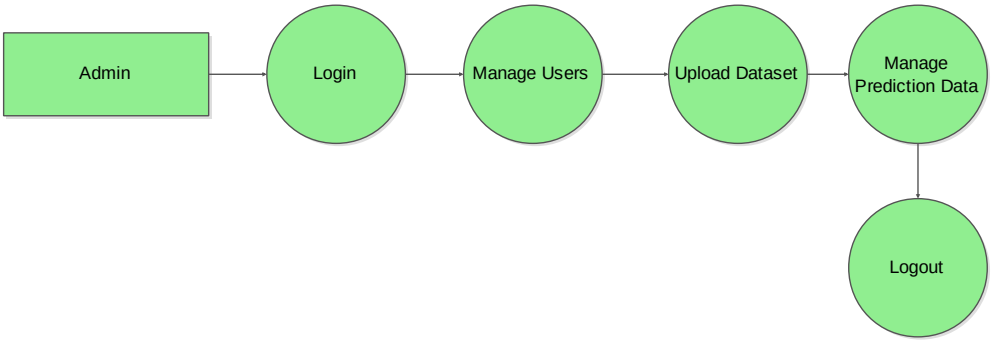


Fig.level 0 dfd



User level 1 dfd



Admin level 1 dfd

CONCLUSION

The project uses a machine learning approach in which a historic soil dataset is supplied as a training dataset that is used to forecast the crop that may be grown in that soil based on its Ph value. Additionally, the system will automatically collect the user's location. It will be preprocessed after gathering associated details and values. The data that has been tested will be compared to data that has been trained. A new dataset will be compared to the training dataset using the KNN algorithm, with the closest data being utilized to forecast crop features. The result will be shown, which will include a crop name, the number of seeds necessary for the specified acres, and a profit estimate. It is suggested to describe soil using ML approaches based on the following applications and methodologies.

FUTURE ENHANCEMENT

The system can be enhanced further to add following functionality:

1. Accuracy of the result can be increased by adding more datasets of soil.
2. By putting this device on a drone with extra software assistance, the average soil nitrogen content over a broad region may be calculated.
3. Detection of nutrient deficit in leaves, as well as other issues that may be detected using image processing, can be done in the same system with the right software and no additional hardware.

REFERENCES

[1] Yao Zhang, MinZan Li, LiHua Zheng, Yi Zhao, Xiao Shuai Pei “Soil nitrogen content forecasting based on real-time NIR spectroscopy” Key Laboratory of Modern Precision Agriculture System Integration Research, Ministry of Education, China Agriculture University, QinghuaDonglu, Haidian,Beijing 100083, People’s Republic of China, 2016.

[2] Rafael F. Muñoz-Huerta, Ramon G. Guevara-Gonzalez, Luis M. Contreras-Medina, Irineo Torres-Pacheco, Juan Prado-Olivarez andRosalia V. Ocampo-Velazquez, “A Review of Methods for Sensing the Nitrogen Status in Plants: Advantages, Disadvantages and Recent Advances**”**; Sensors, published:16 august 2013.

[3] Sk Al Zaminur Rahman, kaushik Chandra Mitra and S.M. Mohidul Islam “Soil Classification Using Machine Learning Methods and Crop Suggestion Based on Soil Series” International Conference of Computer and Information Technology (ICCIT), 21-23 December, 2018.

[4] Amrutha A, Lekha R and Sreedevi, **"**Automatic soil nutrient detection and fertilizer dispensary system",In International Conference on Robotics: Current Trends and Future Challenges (RCTFC) 2016**,** Department of Electrical and Electronics Engineering, R V College of Engineering, Bangalore, India.

[5] Reshma U N, Prithvi P Bangera, Chethana H C, Kavya Nadig N C, Keerthi D S, **"**Raspberry Pi based Soil Parameters Monitoring Device using Sensors**"** Department of Electronics and Communication, Malnad College of Engineering, Hassan, India. International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 6 Issue V, May 2018.

[6] Rushika Ghadge, Juilee Kulkarni, Pooja More, Sachee Nene, Priya R L, "Prediction of crop yield using machine learning", International Research Journal of Engineering and Technology (IRJET), Volume: 05 Issue: 02, Feb-2018.