**PRECISION AGRICULTURE - COIMBATORE BASED CROP SUGGESTION SYSTEM USING ML AND IoT**

**1. ABSTRACT**

Focusing on assisting the farmers in the Coimbatore district of Tamil Nadu with a goal to mitigate the traditional farming methods, a crop recommendation system is developed. A combination of machine learning models such as Random forest, KNN and Decision tree as an ensemble model is trained based on input parameters like soil and environmental characteristics. The trained model is further given real time inputs from the IoT system along with user inputs to give the output. The most appropriate crop to be grown is recommended to the farmer by considering all the pertinent soil based and environmental parameters.

**2. INTRODUCTION**

India is the world’s second largest producer of wheat and rice and world’s first producer of several other crops. So, it’s not surprising that agriculture rules more than half of India’s population economically.

By taking into account of the undulating nature of the environment and soil parameters that Indian agriculture relies on, farmers don’t find the most precise agriculture techniques for better yield. But they are unaware of such influencing parameters. They always prefer to trust their instincts and follow traditional farming methods. Such unsure practices may unfortunately cause the anticipated productivity and profitability become unsubstantial. Since soil is a non-renewable natural resource the soil health and land productivity may deteriorate overtime. This may lead to unpleasant implications on the farmer’s life as well as the status of food security in the specific region. In the thought of extenuating this situation, smart farming techniques can be imbibed.

A crop recommendation system developed for the Coimbatore region will be beneficial. In our system Machine learning models such as Random forest, KNN and Decision tree are combined to form an ensemble model. The ensemble model gives more prediction and efficiency than any of its models could achieve alone. This system is developed based on certain input parameters like soil pH, EC (Electrical Conductivity), type, moisture and environmental conditions like temperature and season will be beneficial. Sensor readings from IoT along with user inputs like …. is fed into the trained model to give the output. This system alleviates the conventional method by providing precise suggestion on the most appropriate crop to be grown in the farmer’s agricultural land in Coimbatore.

**3. MOTIVATION**

Coimbatore is a district in the state of Tamil Nadu. It is located on the banks of the Noyyal River and surrounded by the Western Ghats. Coimbatore is the second largest city in Tamil Nadu. The cultivable soil of the district is classified into 16 different soil series having different soil properties. All sorts of traditional farming practices do exist here. Farmers of Coimbatore are unaware of the heterogeneous nature of soil. There are differences in soil properties occurring even within short distances. Growing the same crop as the neighbouring farmer will not be as profitable as it is for them. Also, choosing a crop for cultivation based only on market trends will cause adverse loss financially and worsen the quality of soil.

**4. LITERATURE SURVEY**

**AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms**

**Zeel Doshi, Rashi Agrawal, Subhash Nadkarni, Prof. Neepa Shah**

Environmental parameters like temperature, rainfall, farm’s latitude, longitude, altitude and distance from the sea and soil characteristics such as pH value, soil type and thickness of aquifer and topsoil were used in this paper as attributes in the training set. Crops used were bajra, jowar, maize, rice and wheat, barley, cotton, groundnut, gram, jute, other pulses, potato, ragi, tur, rapeseed and mustard, sesame, soybean, sugarcane, sunflower and tobacco crops. They compared various machine learning models that have in built support for Multi-label classification (MLC): Decision Tree, K Nearest Neighbor (K-NN), Random Forest and Neural Network and obtained 90.2%, 88.78%, 90.43% and 91% efficiency respectively. Here we learnt the various models that are suitable for MLC and thus our proposed solution combines the power of all 3 models to get a better efficiency.

**Crop Recommendation System for Precision Agriculture**

**S.Pudumalar, E.Ramanujam, R.Harine Rajashree, C.Kavya, T.Kiruthika, J.Nisha**

The authors of this paper stated that a suggestion system based on site specific and soil parameters would be the most useful for farmers. The attributes considered in this paper: Depth, Texture, pH, Soil Color, Permeability, Drainage, Water holding and Erosion. The crops chosen by them include millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander. A recommendation system using an ensemble model with majority voting technique was proposed. The base learners used were Random tree, CHAID, K-Nearest Neighbor and Naive Bayes and it was observed that the prediction accuracy of the model was 88%. Ensemble model is one of the most preferred for this system according to these set of papers and therefore we decided to opt for it.

**Crop Prediction Using Machine Learning**

**Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese, Er. Jinu Thomas**

In this paper, the input dataset included environmental factors like temperature, humidity etc and soil parameters such as N, P, K and pH of the soil. The algorithms: kNN, kNN with cross validation, Decision Tree, Naïve Bayes and SVM were considered for evaluation and their respective efficiencies were 85%, 88%, 81%, 82% and 78%. It was inferred that kNN with cross validation was the ideal model and thus we decided to adopt kNN as one of the base learners in our ensemble model. The idea of adding IoT sensors to measure soil properties in their scope for improvement section inspired us to include the same in our project.

**Improving Crop Productivity Through Crop Recommendation**

**Harshitha L, Rashmi M N, Vaishnavi I Dodamani, Yamini S**, **Nandini M S**

In this paper, nutritional features in soil like pH values, organic Carbon, iron, zinc, nitrogen, phosphorus, sulphur have been used as it is assumed that chemical analysis of the soil helps in improving crop production. Other attributes include rainfall and temperature. Algorithm chosen by them is Naïve Bayes Algorithm.

**Implementation of Machine Learning Algorithms for Crop Recommendation Using Precision Agriculture**

**K. R. Akshatha, K. S. Shreedhara**

In this paper, the considered dataset includes soil based attributes like depth, texture, pH, soil color, permeability, drainage, water holding capacity and erosion. The crop recommendation system is built using an ensemble model – majority voting technique, using a combination of models like Random tree model, K-Nearest Neighbor model and Naïve Bayes model. The crops recommended by mapping with appropriate soil parameters are millet, groundnut, pulses, cotton, vegetables, banana, paddy, sorghum, sugarcane, coriander.

**Machine learning approach: Recommendation of suitable crop for land using meteorological factors**

**Srilakshmi A, Madhumitha K, Geetha K**

This project aims to get the best solution for crop prediction and recommendation system by comparing machine learning algorithms like random forest, decision tree and SVM. Accuracy obtained for random forest is 95.09%, higher than the rest of the models. The datasets are trained using SVM – grid search and random forest algorithms. The soil parameters taken into account for predicting crop based on environmental factors are temperature, humidity, pH and rainfall. The dataset consists of 3100 rows x 5 columns obtained from online that required no preprocessing. It was also found that combination of SVM and decision tree gave an accuracy of 91.8%, randomized search used on random forest gave 94.7% accuracy. Hence, it is concluded that random forest algorithm performs the best accuracy when compared to other models. Future work comprises of including more soil parameters to perform crop prediction.

**Agricultural Crop Yield Prediction Using Artificial Neural Network Approach**

**Miss Snehal S.Dahikar, Dr.Sandeep V.Rode**

This paper is aimed at making the farmers attain maximum crop yield at minimum cost using ANN – feed forward back propagation technique. The yield for a crop is predicted based on the soil specific parameters like pH, depth, nutrient contents like nitrogen, phosphate, potassium, organic carbon, magnesium, sulphur, manganese, copper, iron and environment-based parameters like temperature, and rainfall. It is concluded that ANN can be a beneficial tool for prediction.

**Soil classification and crop recommendation using IoT and Machine Learning**

**Varsha A, Midhuna V M, Ritly**

The paper proposes to predict a specific crop grown in a soil. The dataset collected IoT system are soil moisture, soil pH and gas content of soil. Firstly, data mining techniques are applied on the sensor values then machine learning steps are applied. A sophisticated front end web interface is also developed for the farmers to view the recommended crop for their soil. The idea of adding IoT sensors to measure soil properties inspired us.

**Recommendation System for Crop Identification and Pest Control Technique in Agriculture**

**Avinash Kumar, Sobhangi and Chittaranjan Pradhan**

It proposes a recommendation system using soil type, average rainfall, average temperature, pH and its corresponding crop type as attributes for 15 different crops. They have used 3 different algorithms individually to figure out which one provides the best accuracy (SVM – 89.66%, Decision Tree – 86.80%, Logistic Regression – 86.04%). It also provides details of Pests and its management for the recommended crop using Information Retrieval.

**Recommendation System Using Neural Networks**

**Tanmay Banavlikar, Aqsa Mahir, Mayuresh Budukh, Soham Dhodapkar**

The designed system deploys soil moisture sensor, temperature sensor, humidity sensor embedded on Raspberry Pi to detect the physical parameters of the soil. Artificial Neural Networks concept is used to train the datasets to suggest the appropriate crop. The attributes for training the datasets include soil moisture content, humidity, temperature and the apt crop.

**Prediction of crop yield and fertilizer recommendation using machine learning algorithms**

**Devdatta A. Bondre, Santosh Mahagaonkar**

This paper describes how to predict crop yield based on soil nutrients and recommends fertilizers. The datasets are collected for attributes such as fertilizer, location and crop yield. Crops considered are rice, wheat, soybean, sunflower, cotton, sugarcane, tobacco and dry chilli. Random Forest and Support Vector Machine were used for Crop Yield Prediction and fertilizer recommendation. They arrived at an accuracy of about 86.35% for support vector machine and 99.47% for random forest which inspired us to use random forest as one of the model for our project.

**Crop Recommendation System through Soil Analysis Using Classification in Machine Learning**

**Dr.A.K.Mariappan, Ms C. Madhumitha, Ms P. Nishitha, Ms S. Nivedhitha**

This paper predicts crop suitability for particular location and recommends the most suitable fertilizer. They collected data sets for soil and crops. The algorithm used here is Random forest and KNN algorithm.

**Agro based crop and fertilizer recommendation system using machine learning**

**Preethi G, Rathi Priya V, Sanjula S M, Lalitha S D, Vijaya Bindhu B**

This paper describes about the method that takes the soil and pH sample as inputs to recommend the suitable crop and fertilizer as the solution. So, the soil information is collected through sensors and the data is transmitted from the Arduino through Zigbee and WSN (Wireless Sensor Network) to MATLAB. Analysis of the soil data and processing is done with help of ANN (Artificial Neural Network) and crop recommendations is done using SVM (Support Vector Machine). We gathered some IoT ideas from this paper.

**Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques.**

**Soumya Sri Attaluri, Nowshath K Batcha, Raheem Mafas**

This paper provides a crop recommendation system with a prime motive of creating economic welfare of farmers. Factors such as cost of planting, cost of harvesting, rainfall, crop demand, cost of seed, cost of fertilizer and yield of crop are considered to generate a more accurate prediction of whether harvesting a crop will yield profits or not. Dataset for this research is sourced from Indian government websites published by different ministries related to agriculture.

**COMPARING WITH EXISTING SYSTEM**

|  |  |  |  |
| --- | --- | --- | --- |
| **Paper Name** | **Methodology used** | **Advantages** | **Disadvantages** |
| AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms  Zeel Doshi, Rashi Agrawal, Subhash Nadkarni, Prof. Neepa Shah | Decision Tree  (90.2%)  K Nearest Neighbor (K-NN)  (88.78%)  Random Forest  (90.43%)  Neural Network  (91%) | Compared various algorithms and chose the best one.  Accuracy of the model is very high. | Absence of sensors for input making it essential for users to type all the inputs.  Multiple models are not used together. |
| Crop Recommendation System for Precision Agriculture  S.Pudumalar, E.Ramanujam, R.Harine Rajashree, C.Kavya, T.Kiruthika, J.Nisha | Ensemble Model (Voting Based classifier)  Random tree, CHAID, K-Nearest Neighbor and Naive Bayes  (88%) | Combined the power of multiple models.  Use of varying soil parameters. | Environmental parameters are not used in this model. |
| Crop Prediction Using Machine Learning  Kevin Tom Thomas, Varsha S, Merin Mary Saji, Lisha Varghese, Er. Jinu Thomas | KNN  (85%)  KNN with cross validation  (88%)  Decision Tree  (81%)  Naïve Bayes  (82%)  Support Vector Machine (SVM)  (78%) | Compared various algorithms and chose the best one.  Tried two different methods for KNN. | Absence of sensors for input making it essential for users to type all the inputs (mentioned in their scope for improvement) |
| Improving Crop Productivity Through Crop Recommendation  Harshitha L, Rashmi M N, Vaishnavi I Dodamani, Yamini S, Nandini M S | Naïve Bayes | Nutrient content of the soil is given importance in this system. | Use of only one model in the entire solution. |
| Implementation of Machine Learning Algorithms for Crop Recommendation Using Precision Agricuture  K. R. Akshatha, K. S. Shreedhara | Ensemble model with majority voting technique, using a combination of models like Random tree model, K-Nearest Neighbor model and Naïve Bayes model. | -Sufficient soil specific attributes are considered.  - Ensemble model gives more prediction and efficiency than any of its models could achieve alone. | - The combination of these models may not be complementing each other so hence might not produce high accuracy in recommendation. |
| Machine learning approach: Recommendation of suitable crop for land using meteorological factors  Srilakshmi A, Madhumitha K, Geetha K | Random forest, decision tree and SVM | -Ample number of datasets are collected for training and testing for better accuracy.  - Aimed at deciding which model has highest accuracy: Random forest -95.09%. | -Less soil parameters are considered.  -Only static data is provided to the trained model. |
| Agricultural Crop Yield Prediction Using Artificial Neural Network Approach  Miss Snehal S.Dahikar, Dr.Sandeep V.Rode | Artificial Neural Networks | -Good number of soil-based attributes like different soil nutrients used.  -Use of ANN gives best results for pattern recognition. | -Using ANN alone does not give higher accuracy when compared to ensemble models for prediction systems. |
| Soil classification and crop recommendation using IoT and Machine Learning  Varsha A, Midhuna V M, Ritly | NA | -Use of IoT for getting real time datasets  -Well developed front end for farmer’s usability. | - Preprocessing of inputs from IoT system (sensors) every time are time consuming and makes the system complex. |
| Recommendation System for Crop Identification and Pest Control Technique in Agriculture  Avinash Kumar, Sobhangi and Chittaranjan Pradhan | SVM Classification (89.66%)  Decision Tree (86.80%)  Logistic Regression (86.04%) | All major parameters are included for recommendation system.  Used different algorithms to choose the one with highest accuracy.  Apart from identification, Pest control techniques is also suggested. | Since the user manually input the values, he may be unaware of the exact values which might mislead. |
| Recommendation System Using Neural Networks  Tanmay Banavlikar, Aqsa Mahir, Mayuresh Budukh, Soham Dhodapkar | Artificial Neural Networks | The input data are collected using sensors to get the accurate value which gives the correct recommendation. | Few more parameters could be taken into account for more effective recommendation. |
| Prediction of crop yield and fertilizer recommendation using machine learning algorithms  Devdatta A. Bondre, Santosh Mahagaonkar | Support vector machine  (86.35%)  Random Forest  (99.47%) | Compared various algorithms and chose the best one |  |
| Crop Recommendation System through Soil Analysis Using Classification in Machine Learning  Dr.A.K.Mariappan, Ms C. Madhumitha, Ms P. Nishitha, Ms S. Nivedhitha | Random Forest  and  KNN algorithm | Compared various algorithms and chose the best one | Absence of sensors for input making it essential for users to type all the inputs. |
| Agro based crop and fertilizer recommendation system using machine learning  Preethi G, Rathi Priya V, Sanjula S M, Lalitha S D, Vijaya Bindhu B | Artificial Neural  Networks and Support Vector  Machine | Gathered ideas  For using soil sensors | Accurate results are not mentioned by using artificial neural networks and environmental parameters are not used properly. |
| Crop Plantation Recommendation using Feature Extraction and Machine Learning Techniques.  Soumya Sri Attaluri, Nowshath K Batcha, Raheem Mafas | KNN, ANN (regression model) | Simulations are performed using to improve the model. | Market price of the crop and yield of the crop are major impacting factor |

**5. DEFINITION**

**5.1 SYSTEM ARCHITECTURE**

Arduino Uno R3 and Sensors

ESP8266 Wi-Fi Module

Sensor data (IoT)

Sensor data (IoT)

User Input (Location, Soil type)

Ensemble Model

Soil characteristics

Server

Input to the trained system

**5.2 ARCHITECTURE DEFINITION**

* Firstly, data is collected from Soil Survey department and the Tamil Nadu Agriculture University.
* The data collected is then pre-processed to remove missing values and outliers to transform raw data into a structured format to improve accuracy of the system.
* After pre-processing, data is normalized to give equal priority to all the attributes.
* It is then provided as input to the ensemble model for training.
* User input and input from the IoT sensors is supplied to the resulting trained model and the output is collected.
* The recommended crop is then presented as the output.

**6. REQUIREMENTS**

**6.1 HARDWARE REQUIREMENTS:**

**IoT:**

* + Arduino Uno R3
  + Temperature and Humidity Sensor
  + Soil Moisture Sensor
  + Soil pH Sensor
  + ESP8266 Wi-Fi Module

**Operating Systems:** Windows 8 or above

**Memory:** Minimum 4 GB RAM

**6.2 SOFTWARE REQUIREMENTS:**

* + Arduino IDE 1.8.0
  + Python
  + Jupyter Notebook

**7. PROPOSED METHODOLOGY**

**7.1 DATA PRE-PROCESSING**

Data pre-processing is a technique used in data mining, machine learning which is used to transform the raw data into useful datasets that can be used to perform various tasks. It is generally the first step of any learning model.

Data collected from different sources generally tend to have some missing values or outliers. So to remove all this, data pre-processing is performed.

The rows containing missing values can be removed in case of large datasets. Otherwise, mean for numeric data and mode for categorical data can be used as a suitable replacement.

An outlier is an object that deviates significantly from the other objects in the dataset. This can affect the accuracy of the model and hence must be removed.

**7.2 FEATURE SCALING**

Feature scaling is a method used to normalize data in each attribute. This is mainly used to keep all attributes in the same value range to avoid giving higher priority to attributes with higher values.

The most appropriate method to do this is:

**Standardization (Z-score Normalization)**

X’ =

Where – mean of the sample

– standard deviation of the sample

**7.3 ENSEMBLE LEARNING TECHNIQUE**

Ensemble learning technique in machine learning integrates multiple models to achieve increased performance. It combines the power of multiple ML algorithms so that one model can correct the errors of another and predict results with much higher accuracy. Two or more models can be used as the base learners in this technique. The base learners should be chosen such that each of them perform well individually so that the combined model gives better results.

In our project, we are planning to use k-Nearest Neighbours, Decision Tree and Random Forest as our base learners for the Majority Voting technique. Majority Voting is an ensemble technique that is one of the best models to solve classification problems. In this, the training set is supplied to all the models and each of them gets trained individually. The inputs for prediction is then fed to all the models separately. Each prediction result is counted as a vote and the output with the maximum number of votes is given as the final result.

**7.4 LEARNERS USED IN THE MODEL**

**7.4.1 RANDOM FOREST:**

Random Forest is a popular algorithm that belonging to the supervised learning technique. It basically involves the process of combining multiple classifiers to solve a complex problem and to improve the performance of the model, which is generally called as ensemble learning. Random Forest combines the use of many decision trees. It takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output.

**GENERAL WORKING:**

1. K random data points are selected from the training set.
2. The decision trees for the selected data points is constructed.
3. Choose the number N for decision trees that you want to build.
4. Repeating of step 1 & 2.
5. The category of decision trees that wins the majority votes in the predictions will be assigned for the new data points.

**7.4.2 DECISION TREE**

Decision Tree is a non-parametric supervised machine learning technique used for solving both classification and regression problems. It often mimics a human’s thinking ability while making a decision. It is a tree-like structure where the internal nodes represent the features of a dataset; the branches act as the decision rules, and leaf nodes are the final result of the decisions. This uses different attribute selection measure to decide on where to split. They are namely Gini Index, Information Gain, Reduction in Variance and Chi Square. The tree algorithm depends on the selected attribute measure.

**GENERAL WORKING:**

1. Start the tree with root node S (complete dataset)
2. Find the best attribute using Attribute Selection Measure.
3. Divide S into subsets
4. Generate decision tree node (which contains the best attributes)
5. Recursively continue from step 3

**7.4.3 K-NEAREST NEIGHBOUR:**

K-NN is a non-parametric algorithm, that stores the dataset and at the time of classification, it performs an action on the dataset. For this reason, it is also called a lazy learner algorithm. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

**GENERAL WORKING:**

1. Number K of the neighbours is selected or calculated.
2. Calculating the Euclidean distance of K number of neighbours
3. Considering the calculated Euclidean distance, take the K nearest neighbours.
4. Among these k neighbours, counting the number of the data points in each category.
5. The category that has the maximum number of neighbour, will be assigned to the new data point.