**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.

|  |  |  |  |
| --- | --- | --- | --- |
| **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. |

**System Architecture**

**Flow Chart**

Sensor data (IoT)

User Input (Location, Soil type)

Ensemble Model

Soil characteristics

**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.

**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

**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.