Farmer Eco Friendly Recommendation System

Arfa Mahavish[1], A Rohith Reddy[2], K Vishnu Vardhan[3], D Sreevalli[4] *Assistant Professor, CSE, CMR Technical Campus, Hyderabad, India* [2]*Student, CSE, CMR Technical Campus, Hyderabad, India*
[3]*Student, CSE, CMR Technical Campus, Hyderabad, India*
[4]*Student, CSE, CMR Technical Campus, Hyderabad, India*

***Abstract*: In this paper, a real-time ML-based system was built for farmers' eco-friendly recommendation system, Agriculture is the largest source of livelihood in India. Climate plays a crucial role in agriculture production. It has a profound influence on crop growth, development, and yield. Choosing the best crop to the given climatic conditions can increase the yield of a farmer. Farmers need a guide to help them in terms of crop suggestions and recommendations based on the amount of rainfall, soil pH, humidity, and temperature. We want to aid the farmers in this aspect by suggesting to them the type of crop to cultivate and also recommend fertilizers for that crop. Our project aims to build a web application that uses a machine learning model to predict the suitable crop to the given weather conditions. We would also recommend the use of organic fertilizers to improve the fertility of the soil and produce ec0-friendly crops. Our model is trained using the MLP algorithm, this algorithm uses 100 hidden layers to give the result. We chose this algorithm because it provides greater accuracy than other algorithms. Our model is trained with 3000 records consisting of data of various crops. The accuracy of our model is 94.71%.**

***Keywords:* MLP Classifier, Crop Recommendation Dataset, Gradient descent, SoftMax**

1. Introduction

Agriculture, along with its related industries, is India's most important source of income. Agriculture is the primary source of income for 70% of rural households, with 82 percent of farmers being small and marginal. Understanding climatic conditions and cultivating appropriate crops are crucial for farmers to increase their income. Farmers will have a low yield due to a lack of understanding about which crops are best suited to the current weather conditions. To get a higher yield, this will increase the use of chemical fertilizers. Chemical fertilizers reduce soil fertility by depleting the organic carbon in the soil. Instead, organic fertilizers can be employed, which improves soil health. The cost of chemical fertilizers is much more than that of organic fertilizers.

Machine learning is an application of artificial intelligence (AI) that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. Machine learning focuses on the development of computer programs that can access data and use it to learn for themselves. The analysis starts with observations or data, such as examples, direct experience, or instruction, in order to look for patterns in data and make better decisions in the future based on the examples we offer. The main goal is for computers to learn on their own, without the need for human involvement, and to change their behavior accordingly. We can train our system to learn better crops based on temperature, humidity, rainfall, and soil pH by using Machine Learning.

2. Literature Survey

There is much research works in the area of crop prediction based on given weather conditions. Even though there are many research papers, there are no particular ground-level implementations available. Our project is apparently based on and similar to one such research paper namely “Crop Prediction System using Machine Learning” published as a Special Issue on Recent Trends in Data Engineering (Volume 4, Special Issue 5, Dec.-2017) [3]. This study proposes a strategy to lower transportation costs by reducing the number of middle hops and agents between farmers and end-users using an IOT-based method, which will benefit the farmer. Our project's motivation turns out to be this paper. We incorporate the processes described in the paper, as well as a prediction-based approach for recommending crops with the highest profit potential. A presentation titled “Agriculture yield prediction using predictive analytic techniques” was presented at the 2nd International Conference on Contemporary Computing and Informatics (ic3i) in 2016. This study addresses the building of several predictive models and theorises an exploratory data analysis. To identify and analyze the properties of each, a sample data set is gathered and several regression techniques are used. Linear, Multiple Linear, non-Linear, Logistic, Polynomial, and Ridge regression are some of the regression techniques described in this work. This paper provides a comparative analysis of various data analytics algorithms. This aids us in determining which algorithm is most suited to our proposed system. Finally, a review of many articles assisted us in selecting Multi-Layer Perceptron Classifier as the classifier in our suggested system. N.Heemageetha, “A survey on Application of Data Mining Techniques to Analyze the soil for agricultural purpose”,2016IEEE.
This paper discusses various data mining techniques like Market-based Analysis, Association Rule Mining, Decision Trees, Classification, and Clustering. It entirely covers the Data Mining concept. Various data mining algorithms such as Naive Bayes classifier, J48, K-Mean are explained in this paper. It also provides the classification of soil based on Naive Bayes, Genetic algorithm, Association Rule Mining. Eventually, it covers Clustering in soil databases. This paper helped us in understanding and analysis of different data mining algorithms and classification mechanisms. This will prove to be extremely beneficial while developing our project and will help in mining the dataset obtained

from sensors employed remotely. Await Kumar, Shiv Kumar "Prediction of production of crops using K-Means and Fuzzy Logic", IJCSMC, 2015.

This paper presents a technique for predicting current-year crop production. It uses the K-Means data mining technology to calculate crop productivity. In addition, this system employs a fuzzy logic-based prediction method. Fuzzy logic is a rule-based prediction logic in which a set of rules is applied to the land for agricultural purposes, such as rainfall and crop production. This study provides a thorough understanding of how K-Means can be applied to data analysis. We will use the set of rules to anticipate which crop would return the most profit based on previous years' crop costs and current soil and weather data, similar to how they have applied the set of rules in the form of fuzzy logic. The "Food and Agriculture Organization of the United Nations" published "Soil pollution: a Hidden Reality." In addition to crop prediction, our study advocates for the use of organic fertilizers over synthetic fertilizers. After a thorough review of numerous agricultural research publications, it became clear that artificial fertilizer use has caused soil fertility to decline rapidly over time.

3. Methodology

A.Dataset Creation

A Dataset was created by collecting a set of 3000 records that were taken using the various crop yielding results. Each record was labeled with crop name, temperature, humidity, rainfall, pH, Nitrogen, Phosphorus, and potassium. These records are saved in a CSV file. This dataset was built by augmenting datasets of rainfall, climate, and fertilizer data available for India. We are going to show you a dataset that will allow you to develop a predictive model that will recommend the best crops to produce on a specific farm based on numerous criteria.

The data fields in our dataset are:

N - the ratio of Nitrogen content in the soil
P - the ratio of Phosphorous content in the soil
K - the ratio of Potassium content in the soil
Temperature - the temperature in degree Celsius
Humidity - relative humidity in %
pH - ph value of the soil
Rainfall - Rainfall in mm

B. System Architecture

With the help of Labels in the dataset, a real-time crop recommendation ML model was developed using the MLP algorithm. The data taken from the various farmers are kept under data preprocessing. Because the deep learning algorithm cannot handle noisy, incompatible, or missing data, the crop suggestion model's dataset must be preprocessed. Outliers and errors are present because the data is noisy: In the sense that it represents the difference in data, incompatible or inconsistent data.
features: Incomplete data refers to a lack of features or attribute values in a dataset. The basic steps involved in any data preprocessing phase include data cleaning, data Integration: data transformation and data reduction are all things that may be done with data.
After the data preprocessing the data is split into test and train data the amount of training data we took is 80%, and the remaining is test data. The trained data is sent into the classification model and test data is sent into the test model. Now both the classification model and test model are sent into the API model.
On the other hand, the web app takes input from the farmer or any user, those inputs are sent to the server, the data is sent to the API model and the predicted results are sent back to the server. From the server, the results are displayed to the farmer.

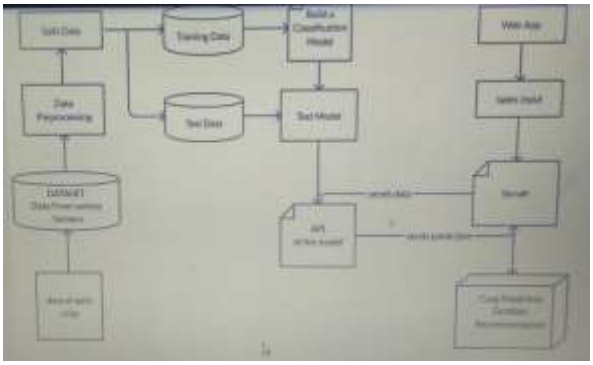

Figure 1: System Architecture

Figure 2: Plotting data: Considering X: Rainfall, Y: Humidity.

Figure 3: Plotting data: Considering X: Humidity, Y: Temperature.

C. Feature Extraction

Feature extraction is a broad phrase that refers to strategies for creating combinations of variables to get around these issues while still accurately representing the data. Many practitioners of machine learning feel that well optimized feature extraction is the key to building good models.

D. Training and Testing

The data collected is divided into two sets i.e. training set and testing set. The data to which the model is fitted is known as training data. The process of training a model requires the help of an algorithm that learns from the training data. We have different algorithms in Machine Learning, Deep Learning, and Neural Networks. In our project, we are using the machine learning algorithm MLP to train the model and then compare the accuracies they provide when test data is provided.

To test the accuracy of the model we use a Testing mechanism. The data collected is divided into training and testing data. Data is tested by using testing data set and new data sets and comparing the result in form of accuracy, precision, and recall. Testing the model in our project involves giving the test data, extracting the features of test data.

E. Classification

The Class MLP classifier uses Back Propagation to train a multi-layer perceptron (MLP) method. MLP uses two arrays to train: array X (3000 samples, 4 features), which contains the training samples as floating-point feature vectors, and array y (samples,) which contains the target values (class labels) for the training samples. By using SoftMax as the output function, the MLP classifier offers multi-class classification. The multi-layer perceptron is a supervised learning algorithm that learns from data. It can learn a non-linear function approximator for classification or regression given a collection of features X=x1,x2,...,xm and a target y. The input layer, on the left, is made up of a group of neurons that represent the input features. Temperature, humidity, soil pH, and rainfall are all inputs for our model. Each neuron in the previous hidden layer transforms the values from the previous layer with a weighted linear summation w1x1 + w2x2 + w3x3 + w4x4 + a non-linear activation function (x1=temperature, x2=humidity, x3=soil pH, x4=rainfall), followed by a non-linear activation function (x1=temperature, x2= The public attributes are contained in the multi-layer perceptron in scikit learn. The terms _coefs_ and _intercepts_ are used interchangeably. The weight matrix at index I represents the weights between layer I and layer i+1, and coefs_ is a collection of weight matrices. intercepts_ is a list of bias vectors, where the vector at index i represents the bias values added to layer i+1. MLPClassifier uses 100 hidden layers by default. We have 30 classes of output in our system.

F. SoftMax

SoftMax is a function that takes a vector z of K real numbers as an input and normalizes it into a probability distribution made up of K probabilities proportional to the input numbers' exponentials.
The formula defines the SoftMax function.

G. Regularization

The MLP Classifier employs the parameter alpha for regularization terms, which helps to minimize overfitting by penalizing excessive weights. The graphic below shows various decision functions as a function of alpha.
In our MLP Classifier, we chose alpha = 0.0001.

H. Gradient Descent

MLP uses Stochastic Gradient Descent, Adam, or L-BFGS to train. Stochastic Gradient Descent (SGD) is a method for updating parameters based on the gradient of the loss function for a parameter that requires adaptation, i.e.

$$w \leftarrow w - \eta(\alpha \partial R(w)/\partial w + \partial Loss/\partial w)$$

where η is the learning rate that governs the parameter space search step size. The network's loss function is called Loss.

I. Activation Formula

The hyperbolic tan is the default activation function. It's written as

$$\sigma(\mathbf{z})_i = \frac{e^{z_i}}{\sum_{j=1}^{K} e^{z_j}} \text{ for } i = 1, \ldots, K \text{ and } \mathbf{z} = (z_1, \ldots, z_K) \in \mathbb{R}^K$$

4. Feasibility Study

A feasibility study involves deciding whether the project we are taking up can be done or not. The decision is taken based on two criteria i.e. cost required and the value of the project. Before starting on with the project, a detailed study of the background of the project is required. The positive and negative outcomes of the project should be analyzed beforehand. If the commencement of the project is done without a proper feasibility study, it might lead to wastage of money invested in the project or might lead to the failure of the project.

A. Technical Feasibility

Technical feasibility involves the evaluation of the hardware and the software requirements of the proposed system. In this project, the technologies involved are Machine learning algorithms for building the model. These are necessary for fulfilling the requirements of the proposed project. The language used is Python. PyCharm is the IDE that will be used in the project. A basic understanding of the inbuilt libraries such as Librosa, NumPy, Scikit-Learn, SciPy, Pandas, Statistics, etc. is required.

B. Economic Feasibility

For the reckoning ability of a new project economic feasibility is the most familiar method. It is also called cost analysis. This assessment commonly comprises a cost/ benefits analysis of the project. The most important factors for the study of the project are Cost and Time.

C. Operational Feasibility

Operational feasibility is used to analyze if we can solve a real-life problem with the proposed system. It assesses the practical nature of the project. It provides us an opportunity for grabbing the advantages of the opportunities and helps to meet all the requirements

$$g(z) = \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$$

which are provided by the customer for the development of the project.

5.Testing

A Testing Plan

A test plan is the first step in the testing process. This plan defines all of the testing-related actions that must be completed, as well as the timelines, resources, and testing requirements. The stated test cases are executed during unit testing, and the actual result is compared to the expected output. The test report and the error report are the final outputs of the testing phase.

B Test Data

We have used 20 percent of the collected data for testing. We have shuffled the data before splitting it into test data and train data, then used the

train_test_split (x, y, test size=0.2, shuffle=True) method from the sklearn learn module in python.

C Test Report

The module is working appropriately given the farmer entered valid input into the web application. Every component of the system is working fine as per the flow mentioned above. We have used the accuracy score (y_test, y_predict) method from the sklearn. metrics module in python as a metric to measure the accuracy of the model.

The accuracy obtained for the model is 91.12 percent.

D Error Report

If the client does not provide information in the deciding request, the client will receive error messages. Error reduction is used to deal with both routine and unexpected errors.

The model's error was 8.88 percent, which suggests that just 9 out of 100 predictions were incorrect.

E Back Propagation

For training Multi-layer Perceptron's, backpropagation is a supervised learning approach. Using a concept known as the delta rule or gradient descent, the backpropagation algorithm looks for the least value of the error function in weight space. The weights that minimize the error function are therefore regarded as a learning problem solution.

We initialize weights with some random numbers or any variable for that fact when creating a Neural Network. Now, obviously, we are not superhuman. As a result, it isn't necessary that the weight values we choose be correct or that they are the best fit for our model.

We chose some weight values at the start, but our model output differs significantly from our real output, resulting in a large error value. What we need to do is figure out how to get the model to modify the parameters (weights) so that the mistake is as small as possible.

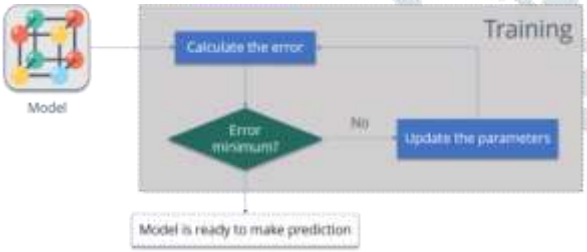

Figure 4: A model for error calculation

6.Results and Discussions

The information gathered is divided into two groups: training and testing. The MLP model's training dataset is used to create the crop suggestion, prediction model. During mode development, the activation function and hidden layers are carefully chosen to deliver the best results. The model is provided test data once it has been developed to calculate the error and accuracy. Gradient descent backpropagation alters the error value by allocating weights to the neurons based on the error value. The model predicts and suggests the crops to be sown with an accuracy of roughly 94.71 percent when the inputs are submitted into the developed model via the user interface.

Figure 5: Home Page

Figure 6: Entering the input data

Figure 7: Predicted crop and quantity of chemical fertilizers is shown.

Figure 8: Quantity of organic fertilizers is shows

Figure 9: Recommendation of fertilizers

7. Conclusion

Prioritizing the agricultural sector will be crucial in the future. Despite the fact that Deep ANN has been utilized in a number of projects, its performance is still improving. The Deep Neural Network has a good processing capability in general,

and its performance with the small dataset cannot be improved. As a result, when Deep Learning is applied to an ANN, the performance of the ANN is improved by using an additional dataset. Current agricultural methods utilize an excellent forecasting system in an attempt to resolve the issues. The precision crop recommendation model is constructed in such a way that it can address the issues that farmers face. As a result, the crop prediction model efficiently selects the optimal crop for cultivation-independent of seasonal variances. The recommendations provided by the GUI greatly aid farmers in selecting which crop is most suited to their field. The system could be enhanced in the future by incorporating hybrid methods for proposing fertilizers to be applied on a timely basis in order to maximize profit and yield.

This project aims to predict the type of crop to be sown for given climatic conditions using the Back-Propagation classification algorithm.

Dataset collected is preprocessed like label-encoded, feature extraction is done, and missing values were handled for better accuracy.

It also recommends the type of fertilizers to be used for the crop predicted. We promote the usage of organic fertilizers by comparing the cost of fertilizers per acre, for artificial fertilizers and organic fertilizers.

8. Reference

[1] 1.17. neural network MODELS (SUPERVISED). (n.d.). Retrieved May 27, 2021, from https://scikit-learn.org/stable/modules/neural_networks_supervised.html

[2] Abhishek, L., & Bharath, R. (1970, January 01). [Pdf] automation in agriculture using IOT and machine learning: Semantic scholar. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/Automation-in-Agriculture-Using-IOT-and-Machine-Abhishek-Barath/c26faae9c2ded577ad516ed23ccc6846a93f0d74

[3] Agriculture yield prediction using predictive analytic techniques. (n.d.). Retrieved May 27, 2021, from https://ieeexplore.ieee.org/document/7918789/references#references

[4] Chlingaryan, A., Sukkarieh, S., & Whelan, B. (1970, January 01). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review: Semantic scholar. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/Machine-learning-approaches-for-crop-yield-and-in-A-Chlingaryan-Sukkarieh/78f232ab1d6001acc6c332b75777c0d531d0df0b

[5] Crane-Droesch, A. (1970, January 01). Machine learning methods for crop yield prediction and climate change impact assessment in agriculture: Semantic scholar. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/Machine-learning-methods-for-crop-yield-prediction-Crane-Droesch/039c1148106113a9f723e3bab22a07d7fad00a33

[6] Khan, M. (2019, September 12). Deploy your machine learning model using flask - made easy now. Retrieved May 27, 2021, from https://medium.datadriveninvestor.com/deploy-your-machine-learning-model-using-flask-made-easy-now-635d2f12c50c

[7] Meeradevi, & Salpekar, H. (1970, January 01). Design and implementation of a mobile application for crop yield prediction using machine learning: Semantic scholar. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/Design-and-Implementation-of-Mobile-Application-for-Meeradevi-Salpekar/604793487324adda5a8aaa6c746a0f5375b51d49

[8] Omkar, B. (2014). Machine learning-based crop prediction system using multi-linear regression. Retrieved May 27, 2021, from https://www.scribd.com/document/391407747/Machine-Learning-based-Crop-Prediction-System-Using-Multi-Linear-Regression

[9] Panchamurthi, S. (1970, January 01). Soil analysis and prediction of Suitable crop for agriculture using Machine Learning: Semantic Scholar. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/Soil-Analysis-and-Prediction-of-Suitable-Crop-for-Panchamurthi/8db22afb356d255491282ac1d748d35f72b62728

[10] Priya, P., Muthaiah, U., & Balamurugan, M. (1970, January 01). Predicting yield of the crop using a machine learning algorithm. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/PREDICTING-YIELD-OF-THE-CROP-USING-MACHINE-LEARNING-Priya-Muthaiah/377fd4e165a74e6aa2a36424f614cf8fc56981e5

[11] Reshma, S., & Pillai, A. (1970, January 01). Impact of machine learning and internet of things in agriculture: State of the art: Semantic scholar. Retrieved May 27, 2021, from https://www.semanticscholar.org/paper/Impact-of-Machine-Learning-and-Internet-of-Things-Reshma-Pillai/e7893b131d13d997777330a091c324d9108c43f8

[12] Sklearn.neural_network.mlpclassifier¶. (n.d.). Retrieved May 27, 2021, from https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

[13] S. Nagini, T. V. R. Kanth and B. V. Kiranmayee, "Agriculture yield prediction using predictive analytic techniques," 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), 2016, pp. 783-788, DOI: 10.1109/IC3I.2016.7918789.