Determine the Soil Nutrients to Find the Crop Yields Using Data Mining Algorithms

P.Sakthi Murugan
Assistant Professor in Dept of CS
Kaamadhenu Arts and Science College
Sathyamangalam,Erode, Tamilnadu.
sakthigasc@gmail.com
Dr.P.Prabhusundhar
Assistant Professor in Dept of CS
Gobi Arts and Science College Gobichettipalayam,Erode
drprabhusundhar@gmail.com
B.Kiruthika
Research Scholar
Gobi Arts and Science college
Gobichettipalayam,Erode

Abstract— Crop cultivation used to be done by farmers who had hands-on experience. However, crop yields have begun to suffer as a result of climate change. As a result, farmers are unable to select the appropriate crop/s based on soil and environmental parameters, and the process of manually predicting the appropriate crop/s of land has frequently failed. Crop prediction accuracy leads to higher crop production. The agriculture is fully based on the soil wealth, climatic conditions, irrigation, quality of seeds, harvesting etc. It is really important all around the world. Only a seasoned farmer can recognise the type of soil and select the appropriate crop for it. Predicting the soil type and its surrounding environment for a specific field is crucial for future crop yields. This study focuses on using a range of data mining techniques to estimate future soil conditions and boost crop production. Clustering, OneR, and J48 are among the data mining techniques used in the study. These data mining techniques are used to systematically anticipate and analyse soil bearing in order to improve crop output. This technique would be more useful to farmers in identifying the type of the soil and its riches, which, in turn, would help them choose the crop that is best suited to their soil and produces the highest yield.

Index Terms – Soil Nutrients, Spatial mining, J48, OneR, Crop yields.

1. INTRODUCTION

One of the most significant parts of a country's development is its ability to generate food. Agriculture has been linked to the production of key food crops for generations. In actuality, though, the tremendous rate of population expansion has been our society's single most pressing concern. Agriculture's reach has been severely limited as a result, notably in terms of land utilisation and fertility.Given that the amount of land under agriculture is unlikely to expand in this period of urbanisation and globalisation, the focus will have to be on making the most of what is available. Crop cultivar prediction is crucial in agriculture[1-3].Early prediction is possible through collection of previous experience of the farmers, weather conditions and other influencing factors and; store it in a large database. The common input parameters are rainfall, temperature, humidity, potassium,

phosphorous, nitrogen .To find the suitable crop best fit for the soil. Generate the algorithm called Best fit crop prediction to find the suitable crop to get profitable yields.

The first report on the use of computers in agriculture was published in 1983 [4]. To overcome the existing difficulties in agriculture, various approaches have been proposed, ranging from databases [5] to decision support systems [6]. Systems that use AI have been determined to be the most excellent performers in terms of accuracy and robustness among these solutions. Agriculture is a dynamic domain in which it is impossible to generalise conditions in order to propose a common answer.

The purpose of this study is to increase agricultural yield while also learning more about the crops by taking into account their climatic circumstances. The solution is to use machine learning algorithms to create prediction models. In crop yield and fertiliser prediction, machine learning systems estimate the production more accurately. In the past, predictions were made based on previous experience.

Formers are unable to accurately anticipate yield due to severe changes in meteorological conditions. As a result, by incorporating technology such as machine learning algorithms, the model can accurately predict the outcome. The trained model, which is based on the gathered data set and predicts the yield with a more accurate outcome, can be clearly understood from the data set.

The model is developed to increase agricultural economic growth by analysing aspects such as season, temperature, humidity, type of soil, and its composition. The best option for predicting crop production prediction is to use data mining techniques to create the model. Data mining [7] is a process for analysing acquired data from many perspectives and extracting usable information. Crop yield forecast is one of the most difficult difficulties that every farmer in our country faces.

The key challenge is to compare machine learning algorithms such as Random Forest and Backpropagation to determine which is the more accurate of the two.

Another difficulty is to develop an efficient model with a more precise algorithm for predicting crop output and fertiliser application amout ratio

2. LITERATURE SURVEY

Machine learning techniques have been used in agriculture by a number of researchers in recent years. The following is an overview of the use of several machine learning algorithms in the field of agricultural yield prediction from soil analysis over the last few years.

Niketa et al, 2016 [8] proposed a data science technique for recommending crops in order to improve productivity.The data from prior years is utilised to forecast the data for the following year. Machine learning techniques are used to analyse the results with high accuracy.

Pallavi V. Jirapure, Prof. Prarthana A. Deshkar, 2016 [9] discussed qualitative data analysis and how it aids in the development of Artificial Neural Networks (ANN) models. This agricultural data is used to create an information system that benefits both previous and current customers. The author of this paper talked about data mining techniques including regression and classification.

The backpropagation algorithm, according to Shivnath Ghosh and Santanu Koley (2014) [10], has a three-layered architecture.The input layer, hidden layer, and hidden layer are the three layers.

I In the agriculture sector, yield prediction is a major issue. If a farmer wants to know how much he can expect, he is about to find out. Analyze the many relevant attributes dataset, such as pH, location, and the value from which the soil's alkalinity is calculate and determined. In addition, the

percentage of nutrients or datasets such as N, P, and K Location are employed, as well as the usage of third-party data.

APIs for weather and temperature, kind of soil, nutrient value of soil in that region, and amount of rainfall in that region are examples of such applications.[11] The terms "region" and "soil" are used interchangeably.

Data mining techniques are used to extract relevant information from raw data, according to Ch Anwar ul Hassan et al, 2018 [12]. Six distinct classifiers were compared by the author.Logistics Regression (LR), Decision Tree (DT), Nave Bayes (NB), K-Nearest Neighbours (KNN), Support Vector Machine (SVM), and Random Forest are the algorithms used (RF). According to the findings, the random forest algorithm is a superior regression technique for prediction.

3. **SOIL NUTRIENT ANALYSIS METHODOLOGY**

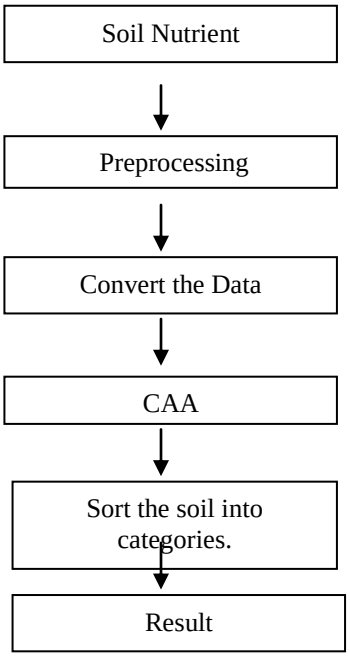
In this system , analyze the soil nutrition parameters are mentioned in the Table 1.

ATTRIBUTES	DESCRIPSION
N	Nitrogen
K	Potassium
P	Phosphorus
Hum	Humidity of the soil
RF	RainFall
TemP	Temperation
pH	PH value of the Soil

Table 1: Soil Nutrition data set attributes.

Proposed System Work Flow

The soil nutrients are input parameters, and the dataset has been preprocessed using the Crop Analysis algorithm [CAA] to transfer the data to the field and predict the crop that is suited for the soil.



Calculate the amount of available phosphorus.

$$Available\ Phosphorus\ (mg/kg) = \frac{CxW}{ODW}$$

Where:

C = Phosphorus concentration (g/2.5 mL) from chart/equation

ODW stands for Oven-Dried Sample Weight (g)

Dilution factor = 14

Controlling soil pH is essential for optimal soil use and productivity. The most effective and practical technique to raise the pH of acid soils is to add lime. To lower soil pH, utilise elemental sulphur, iron sulphate, or aluminium sulphate. We'll concentrate on improving soil pH because acidity is a problem.

Table 2. Rating the pH

pH	**Ratings**
<4.5	Strongly acidic
4.6–5.5	Highly acidic
5.6–6.5	Moderately acidic
6.6–6.9	Slightly acidic
7.0	Neutral
7.1–8.0	Slightly alkaline
8.1–9.0	Moderately alkaline
9.1–10.0	Strongly alkaline
10.1–11.0	Very strongly alkaline

3.1Measurement of Soil Moisture

The gravimetric method was used to measure soil moisture in each field and weather station. Uhland cores were used to obtain soil samples from fields. Using sampling cores with dimensions of 6 cm 8 cm at a depth of 0 to 6 cm, composite soil samples were obtained from three distinct places in individual fields, and GPS coordinates were recorded. Weighed wet soil samples from the fields were oven-dried for 48 hours at 105 degrees Celsius.

The amount of water lost during drying and the dry weight of the soil sample were used to calculate the gravimetric water content. By multiplying gravimetric water content (m^3 m^{-3}) by bulk density (g cm^3), volumetric water content (VWC) of soil was obtained [13]

4. PROCEDURE FOR FINDING A CROP

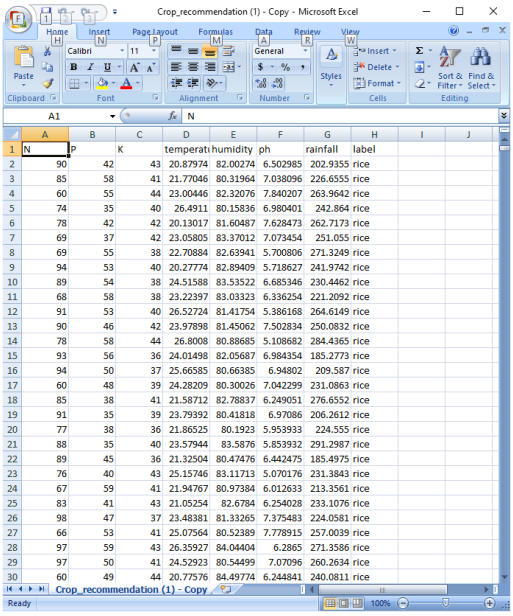
The basic crop-finding process is outlined below. Soil nutrients are an input parameter that should be analysed to determine which crop is best suited for the soil and should be planted to maximise production..

i. Import the Soil Dataset
ii. Data Pre-processing
iii. Model Structure Learning with Cross-Validation
iv. Trained Model.
v. Performance Assessment.

vi. Calculate *Accuracy*, *Recall*, *Precision*, *Specificity* and *F-Score* for each model Conclude the crops as a result.
vii. Comparative Analysis of final results for each trained model

5. RESULT AND DISCUSSION

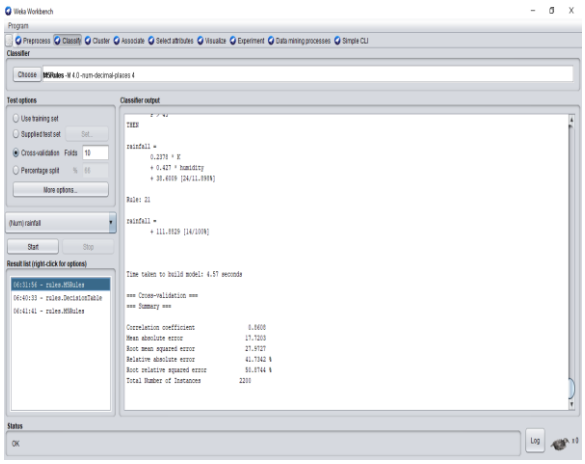
The dataset is used to evaluate the experimental results. Using the Crop Analysis Algorithm method, this dataset has 2200 samples that can be used to predict crops that are best suited for the soil.



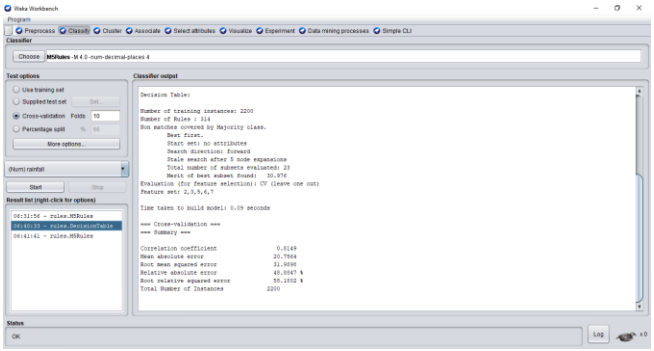
N	P	K	temperature	humidity	ph	rainfall	label
90	42	43	20.87974	82.00274	6.502985	202.9355	rice
85	58	41	21.77046	80.31964	7.038096	226.6555	rice
60	55	44	23.00446	82.32076	7.840207	263.9642	rice
74	35	40	26.4911	80.15836	6.980401	242.864	rice
78	42	42	20.13017	81.60487	7.628473	262.7173	rice
69	37	42	23.05805	83.37012	7.073454	251.055	rice
69	55	38	22.70884	82.63941	5.700806	271.3249	rice
94	53	40	20.27774	82.89409	5.718627	241.9742	rice
89	54	38	24.51588	83.53522	6.685346	230.4462	rice
68	58	38	23.22397	83.03323	6.336254	221.2092	rice
91	53	40	26.52724	81.41754	5.386168	264.6149	rice
90	46	42	23.97898	81.45062	7.502834	250.0832	rice
78	58	44	26.8008	80.88685	5.108682	284.4365	rice
93	56	36	24.01498	82.05687	6.984354	185.2773	rice
94	50	37	25.66585	80.66385	6.94802	209.587	rice
60	48	39	24.28209	80.30026	7.042299	231.0863	rice
85	38	41	21.58712	82.78837	6.249051	276.6552	rice
91	35	39	23.79392	80.41818	6.97086	206.2612	rice
77	38	36	21.86525	80.1923	5.953933	224.555	rice
88	35	40	23.57944	83.5876	5.853932	291.2987	rice
89	45	36	21.32504	80.47476	6.442475	185.4975	rice
76	40	43	25.15746	83.11713	5.070176	231.3843	rice
67	59	41	21.94767	80.97384	6.012633	213.3561	rice
83	41	43	21.05254	82.6784	6.254028	233.1076	rice
98	47	37	23.48381	81.33265	7.375483	224.0581	rice
66	53	41	25.07564	80.52389	7.778915	257.0039	rice
97	59	43	26.35927	84.04404	6.2865	271.3586	rice
97	50	41	24.52923	80.54499	7.07096	260.2634	rice
60	49	44	20.77576	84.49774	6.244841	240.0811	rice

Soil Nutrient Dataset

M5Rules algorithm use this dataset to build the model in 4.57 seconds

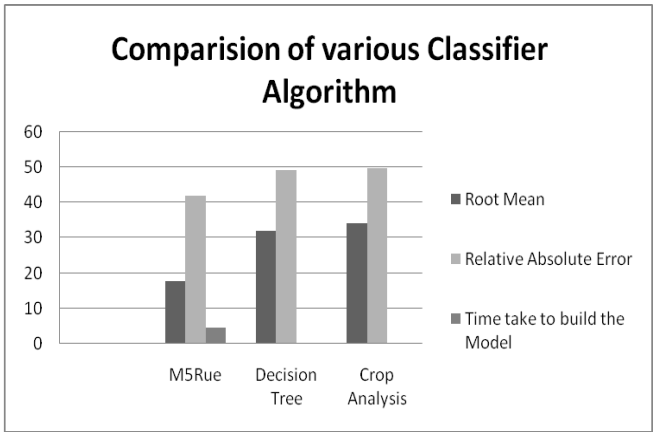


Decision Tree Classifier use the dataset to build the Model in 0.09 Seconds



Compared using Various Classifier Algorithms

Algorithm	Root Mean Square	Relative Absolute Error	Time take to build the Model
M5Rue	17.7203	41.7342	4.57
Decision Tree	31.9898	48.8847	0.09
Crop Analysis	33.9999	49.56	0.08



Result for Classifier Algorithm

Compared to the M5Rule and Decision Tree algorithm, the Crop Analysis algorithm is very efficient for the soil nutrient to predict the right crop for the soil.

6.CONCLUSION

We introduced an algorithm for crop prediction based on soil nutrient characteristics in this research. CA algorithm is superior to M5 Rule and Decision Tree. With an accuracy of 0.93 percent, the CA is better at predicting soil nutrients and locating crops.

In the future, construct a crop prediction system to determine the best crop for the soil. This method will suggest the best fertilizer based on the soil sample dataset and cropping pattern.

7. REFERENCE

[1] S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In International conference on machine learning 1 (2001), pp. 74–81.

[2] H. Liu and Y. Lei, Toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005), pp. 491–502. doi:10.1109/ TKDE.2005.66.

[3] A. Maruf, M. Abdullah, and S. Shatabda, iRSpot-SF: Prediction of recombination hotspots by incorporating sequence based features into Chou's Pseudo components, Genomics 111 (2019), pp. 966–972. doi:10.1016/j.ygeno.2018.06.003.

[4] D.N. Baker, J.R. Lambert, J.M. McKinion, ―GOSSYM: A simulator of cotton crop growth and yield,‖ Technical bulletin, Agricultural Experiment Station, South Carolina, USA, 1983.

[5] P. Martiniello, "Development of a database computer management system for retrieval on varietal field evaluation and plant breeding information in agriculture," Computers and electronics in agriculture, vol. 2 no. 3, pp. 183-192, 1988.

[6] K. W. Thorpe, R. L. Ridgway, R. E. Webb, "A computerized data management and decision support system for gypsy moth management in suburban parks," Computers and electronics in agriculture, vol. 6 no. 4, pp. 333-345, 1992.

[7] Snehal S.Dahikar, Dr.Sandeep V.Rode ,"Agricultural Crop Yield Prediction Using Artificial Neural Network Approach ",International journal of innovative and research in electrical,instrumentation and control engineering, volume 2,Issue 2,2014.

[8] Niketa Gandhi et al ," Rice Crop Yield Forecasting of Tropical Wet and Dry Climatic Zone of India Using Data Mining Techniques" ,IEEE International Conference on Advances in Computer Applications (ICACA) , 2016.

[9] Pallavi V, Jirapure, Prof. Prarthana A. Deshkar, "Quantitative data analysis using Regression method for Agricultural data", Futuristic Trends in Research and Innovation for social Welfare, World Conference, 2016.

[10] Shivnath Ghosh, Santanu Koley , "Machine Learning for Soil Fertility and Plant Nutrient Management using Back Propagation Neural Networks ", IJRITCC, 2014

[11]Crop Yield Prediction and Efficient use of Fertilizers, S. Bhanumathi ; M. Vineeth ; N. Rohit, 2019 International Conference on Communication and Signal Processing (ICCSP)

[12] Ch Anwar ul Hassan, Muhammad Sufyan Khan, MunamAli Shah, "Comparison of Machine Learning Algorithms in Data classification", International Conference on Automation and Computing (ICAC), 2018.

[13]Reynolds, S.G. The gravimetric method of soil moisture determination Part I: A study of equipment, and methodological problems. J. Hydrol. **1970**, 11, 258–273.