Extemporize Agriculture Yield with Predictions Based on Water and Soil Properties using Multivariate Analytics and Machine Learning Algorithm

V. Sudha, S. Mohan

Abstract – Big data have rapidly developed in agriculture that increase the extensive farmers to get extraordinary ideas for future crop development based on soil and water is an essential key factor of agriculture to predict data. In future farmers apply the extracts PCA (Principal component analysis) in the agriculture for crops to find best yield. The level of data is analyzed using PCA and PLS (Projection to latent structure) datasets like Crop data, Soil properties and Water properties such as Linear Regression, Multi Linear Regression, Discriminate Analysis, Partial Least Square, Least squares algorithm is used instead of Multivariate Curve Resolution. Predictive analytics can be used to make a modular decision in farming by observation of actual time data on crops, soil and water data received from different agriculture sources would contain multi-dimensions, the entire content is needed for performing analysis. Multivariate data analysis based on Partial Least Square Programming model that identifies the cropping pattern to get maximum yield correlated. The motive of the work is to compare various techniques which give the maximum accuracy of crop.

Keywords: PCA; PSL; Support Vector Machine; Multivariate; Correlate; Data Analytics.

I. INTRODUCTION

The prediction about the yield is very important to the farmers to get highest Productivity. The prediction of yield can be extracted from the PCA (Principal Component Analysis) and PLS (Partial Least Squares) through our researchers the sharp observation of data analytics that has to be increased the yield so that we have a complicate problem to the big data for better observation. The major plan of principal component analysis (PCA) is to reduce the originality of a dataset having many variables correlated with each other, either heavily or slightly, from the remaining variation provide in the dataset, up to the maximum extent. Projection to Latent Structures regression observatory and result may be useful to first review this information in Elementary concepts. The varieties of samples analyzed in this title are also characterized in common linear sample, Generalized Linear Models, and General systematic Regression. In Cuddalore district agriculture maintains to be the top most group in the economic growth of Crop Paddy, Cumbu, Ragri, Cholam, Redgram, Gingerly, Groundnut, Greenogram, Cashew nuts, Sugarcane, Coconut.

II. RELATED WORK

Sririsha Adamala presented and implemented in this study papers with big data applications and water Resources engineering by using machine learning, research [1] the water managers, is showing big promise in many water related application such as planning optimum water systems, detecting ecosystem changes through big remote sensing and geographical information system, forecasting, predicting, detecting natural and man-made calamities, scheduling irrigations, mitigating environmental pollution, analyzing atmosphere alterations of impacts etc.

S. P. P. Kolhe, [2] Information technology is mostly used in modern agricultural irrigation systems to improve the food production. In this paper, we present the material for various purposes about modern irrigation system with the use of information technology in irrigation system at irrigation sector of agriculture, such as weather parameter, water availability, and water requirements of crops to create water supply schedules. Management of water availability in irrigation scheduling, crop management automated irrigation.

S. Kumawat, Sensor Based Automatic Irrigation System and Soil pH Detection using an Image Processing Development of agriculture is necessary. [3]. Lack of rain and water scarcity is the major problems that are faced by the farmers. The main objective of this paper is providing an automatic irrigation system, thereby saving the time, money & power of the farmer. The traditional farmland irrigation techniques require manual intervention can be minimized. Nutrient availability and plant growth are affected by the acidity and basicity of the soil, which is determined by soil pH property.

Haryana, Evaluation of ground water quality for irrigation [4] the quality of groundwater, water samples was collected from all the villages. Water sample were collected and analyzed for pH, EC, soluble cations (CA, mg, Noah and k) and anions. The values of sodium adsorption ratio (SAR) and residual sodium carbonate (RSC). Spatial maps of EC, pH, RSC, SAR, Water quality of groundwater, Following details are used for irrigation. The analysis of various parameters indicated that quality of groundwater is not safe for irrigation purpose.

The techniques of data mining are [5] Though data mining techniques in agriculture is extremely popular, its application in agricultural soil knowledge sets field is very young to wear data processing system product and domain Specific
data processing application software are Obtainable however; data processing in agricultural soil Knowledge set may be a comparative a young analysis field. The information from the dataset is extracted and transformed to an understandable structure, which is the ultimate goal of data-mining. This paper offers internet based answer for soil testing laboratories yet as free messages to farmers that contains soil testing code important chemicals required for the crop.

The standard statistical analysis [6] they are time consuming and expensive. Agriculture research has been profited by technical advance such as automated data mining. Data mining can be used for solving complex data, which in turn can be used for improving accuracy and effectiveness in classification of large soil. Data mining process involves extracting information from dataset and transforming into an understandable structure that can be used for further studies. The soil test is to analyze the nutrient content, composition and other characteristics. Soil classification deals with the categorization of soil into different soil classes as very Low, low, medium, high and very high of nutrient found in the soil and the basis of these classes fertilizers are recommended for a soil sample.

Soil samples have been collected from [7] the samples of the locations spread uniformly over the study area in order to examine the infiltration rate of soils and its impacts on the overall crop production process. Double ring infiltration was used to carry out the experimental study. Infiltration rates were taken at 0 to 70 minutes of 10 minute intervals. The assessment of infiltration rate was processed by laboratory analysis of soil samples for the particle size distribution the recommended values for crop production. This will assists to develop the structure and maintain the soil potentials.

The paper by the author Abdul–Rahman, S., & Mohamed, A. (2010) Explains “Support Vector Machine” [8] based on the classification of the soil types. Soil Classification includes steps like image acquisition, image preprocessing, and feature extraction and classification. The texture features soil image is extracted using the low pass filter, Gabor filter and using the color quantization technique. Mean amplitude, Standard deviations is taken as the statistical parameter. Classification of soil is the dissolution to soil sets to particular group having a like characteristics and similar manners. Higher agricultural products are very much depends on the soil characteristics, identification and classification is very much important. Soil type helps to avoid agricultural product quantity loss.

J.Gholap presented and implemented information system Performance tuning of Algorithm for prediction of soil fertility. [9].Soil Test is the analysis of a soil sample to determine nutrient content, composition and other characteristics. Soil dataset containing soil test result has been used to apply various classification techniques in data mining.

H. U. Leena, B. G. Premasudha, and P. K. Basavaraja,[10] “Sensible approach for soil fertility management using GIS cloud, “presented an overview of the current state of the art of web-based tools and technologies for processing large dataset. These latest technologies allows to observe and survey a wider collection of agricultural related data such as different crops, soil type, variety of seeds, climatic conditions, and a mixture of socio economic variables concurrently and accurately. Integration of GIS technology into cloud together emerges as a crucial model for capturing uploading and maintaining the new promising web applications in agricultural sector.

K. Arunesh and V. Rajeswari, [11] has presented an efficient strategy for Lime soil status analysis using techniques. Agriculture is one of the most important research fields because variety of data is available in this field for researches. Techniques of data analysis including natural trees, statistical machine learning and other analysis methods.

M. N. Chiranjeevi and R.B.Nadagoudar developed a framework [12] Farmer must find which crop is suitable to cultivate in a particular soil sample. The farmers can get it from data mining Techniques.V. A. Bharadi “Analysis and Prediction in Agricultural Data Using Data Mining Techniques, [13] suggested the agricultural yield primarily depends on environmental factors such as rainfall, temperature and geographical topology of the particular region. These factors along with some other influence the crop cultivation.

N. Sneha and J. Majumdar, “Translated copy of Tank of [14] presented an efficient strategy for crops cultivation of crop yield prediction. It is the most important factor where the farmers need some prior information about the crop yield before sowing seeds in their fields with available requirements. Agriculture will face many challenges due environmental changes and natural calamities. To maximize the particular crop provided the prior information about the crop for the farmers can be done by analyzing historical crop data.

M. P. K. Devi, “Enhanced Crop Yield Prediction and Soil Data Analysis Using Data Mining [15] suggested a crop yield prediction and soil data analysis using data. Productivity of crop is one of the main issues in agriculture. The farmer is interested to know how much yield is about to expect. In earliest days, the yield prediction was carried out only by the experience of the farmers. As technology improved many ways that lead to know the productivity of crop. Crop models and decision tools are increasingly used in agricultural field to improved production competence. Crop yield predict analysis requires a model of how crops respond to soil factors.

K. Samundeeswari and K. Srinivasan, “Data Mining Techniques [16] presented an efficient techniques can be developed and tailored for solving complex soil data sets using data mining to improve the effectiveness and accuracy of the classification of large soil datasets.

III. PROPOSED WORK

A. Variation of Crop Yield with Soil
The soil of the Cuddalore district is classified as the black, red, ferruginous and arenaceous. They are again subdivided into clays, loam and sands. Black soils are observed. The soil is a major factor of agriculture Production. Dissimilarity of soil cans affected the crop production.

Table 3.1 General Statement for Soil Status

<table>
<thead>
<tr>
<th>Statement for Soil Status</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
</tbody>
</table>

Published By: Blue Eyes Intelligence Engineering & Sciences Publication

Retrieved from: F8137088619/201908BEIESP
DOI: 10.35940/ijeat.F8137.088619
1505
Table 3.2 Soil Attributes

<table>
<thead>
<tr>
<th>SYMBOL</th>
<th>DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>Sand</td>
</tr>
<tr>
<td>Sl</td>
<td>Silty Clay</td>
</tr>
<tr>
<td>C</td>
<td>Clay</td>
</tr>
<tr>
<td>Sl</td>
<td>Sandy loam</td>
</tr>
<tr>
<td>Cl</td>
<td>Clay loam</td>
</tr>
<tr>
<td>S</td>
<td>Silty Loam</td>
</tr>
<tr>
<td>L</td>
<td>Loam</td>
</tr>
<tr>
<td>Ls</td>
<td>Loamy sand</td>
</tr>
<tr>
<td>Sr</td>
<td>Sand clay</td>
</tr>
<tr>
<td>Sr</td>
<td>Sand clay</td>
</tr>
</tbody>
</table>

Figure 3.1 Soil Type

The figure 3.1 shows the chemical compositions of soil and its ratio of mixing with soil. The chemicals above disturb the growth of the crops depending upon its mixing ratio with soil.

Table 3.3 Soil PH Range

<table>
<thead>
<tr>
<th>Attribute</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ultra Acidic</td>
<td>pH&lt;3</td>
</tr>
<tr>
<td>Extremely Acidic</td>
<td>3.5-4.4</td>
</tr>
<tr>
<td>Very Strongly Acidic</td>
<td>4.5-5.0</td>
</tr>
<tr>
<td>Strongly Acidic</td>
<td>5.1-5.5</td>
</tr>
<tr>
<td>Moderately Acidic</td>
<td>5.6-6.0</td>
</tr>
<tr>
<td>Slightly Acidic</td>
<td>6.1-6.5</td>
</tr>
<tr>
<td>Neutral</td>
<td>6.6-7.3</td>
</tr>
<tr>
<td>Slightly Alkaline</td>
<td>7.4-7.8</td>
</tr>
<tr>
<td>Moderately Alkaline</td>
<td>7.9-8.4</td>
</tr>
<tr>
<td>Strongly Alkaline</td>
<td>8.5-9.0</td>
</tr>
<tr>
<td>Very Strongly Alkaline</td>
<td>&gt;9.0</td>
</tr>
</tbody>
</table>

Table 3.4 PH Level

<table>
<thead>
<tr>
<th>ACIDIC</th>
<th>NEUTRAL</th>
<th>ALKALINE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LESS 6</td>
<td>7.5-8</td>
<td>8.6-9.0</td>
</tr>
</tbody>
</table>

Table 3.5 Soil Key Parameters

B. Variation of Crop Yield with Water

The Geological processes role and significant role in cultivation. The humidity is one of the most considerable factors of geological processes.

Table 3.6 Water Range

<table>
<thead>
<tr>
<th>Class</th>
<th>C1</th>
<th>C2</th>
<th>C3</th>
<th>C4</th>
<th>C5</th>
</tr>
</thead>
<tbody>
<tr>
<td>EC</td>
<td>&lt;1.5</td>
<td>1.5-3</td>
<td>3.5</td>
<td>5-10</td>
<td>&gt;10</td>
</tr>
<tr>
<td>Water</td>
<td>Normal</td>
<td>Low</td>
<td>Medium</td>
<td>Saline</td>
<td>High</td>
</tr>
<tr>
<td>Quality</td>
<td>Waters</td>
<td>Salinity</td>
<td>Salinity</td>
<td>Water</td>
<td>Salinity</td>
</tr>
<tr>
<td></td>
<td>Waters</td>
<td>Water</td>
<td>Water</td>
<td>Water</td>
<td>Water</td>
</tr>
</tbody>
</table>

C. Principal Component Analysis

PCA is a familiar strategy who explore important concept of variation in a model of random curves. This major source of variation are represented by functional principal component analysis. PCA can be used within huge reports as a bilinear interpretable overview of the important message had a multidimensional table, projection method because instructions carried by the original variables and projects them onto a smaller number of latent variables. Matrix is representation a model with a given number of components equations.

\[
X = TP^T + E
\]

T -> Score Matrix
P -> Load Matrix
E -> Error Matrix

D. Partial Least Squares

Projection of Latent Structure both the X-and Y-matrixes at the same time to find the least variables in X that will best to predict the least variables in Y. The PLS components are parallel to principal components they are referred to as Factors. PLS magnifies the equivalent in centre of X and Y.

E. Support Vector Machines

SVM is a machine learning (Supervised Learning) technique. Which means nonlinear attributes and it is used for classification as well as regression problem. Linear classification is anyway easy to implement as analyse to nonlinear classification.

F. PLS Based Field Crop Yield

PLS is a statistical method actual choice of relation to Principal component regression, alternatively finding hyper plane of maximal variance between the reaction and impartial variables, it initiate a linear regression model by predicted variables. This method of work is used to realise classification regression problems. When the dataset is complex and large then we frequently go for linear version.
IV. PERFORMANCE EVALUATION

Descriptive statistics provides some simple and valuable plotting to medium size datasets. We receive and get a data from various farmers of villages and create the sequential datasets. Actually data of cultivation about field Crops, Soil and Water. Generated From – The reports of farmers to the Government.

![Figure 4.1 Various Crops Yield](image1)

![Figure 4.2 Various Soil Samples and Crop](image2)

![Figure 4.3 Histogram of Pulses](image3)

A. PLS based Soil

Obviously the PLS method was better than various models for the plan of maximum yield from soil status it would be a valuable tool to conform with sample is need future analysis. The character of PLS factor was identified using cross validation during the time of soil or water.

![Figure 4.4 Graphical view of Water](image4)

V. RESULT AND DISCUSSION

The above data sets are finalized in this paper is achieved from the agriculture soil Testing Laboratory, Cuddalore District, Tamilnadu, India. Field sampling gives us the primary data for soil survey.

![Figure 5.1 Graphical view of PCS Range](image5)

Dataset has 13 attributes such village name as Ec, pH, N, P, K, Zn, Cu, Fe, Mn, B, Ca, Mg, and S. As before the primary data for water and Field Crop also acquired, observed and collected directly from the farmers. The following dataset has 15 attributes of water. In this work we have proposed and analysis of the soil, water and field crop data using classification techniques and prediction techniques to predict the status for maximized yield. We have reported comprehensive study of various classification Algorithms with the Principal component Analysis, Partial Least Squares Regression, Descriptive
statistics must perform efficiently. In future we decided to purpose a system to predict the suitable Field Crop for a particular soil and water.

![Figure 5.2 Overall Statistical Statements of Water and Field Crop](image)

**VI. CONCLUSION**

The intention of this paper is to intend the multivariate analytics and Machine Learning Algorithms; we forecast a statement to improve the agriculture yield. For that prediction we gather reports from farmers of each village based on water and soil properties. This will also offer the well-organized information regarding the soil and water level like, pH, Ec, N, P, K, Zn, Cu, Fe, Mn, B, Ca, Mg, and S along with the proper implication. Finally, it is accomplished that, with this projected system one can save manpower and water to extemporize the production.

**REFERENCES**