Prevention of Soil Surface Humidity in Moderately Vegetated Fields Using a Machine Learning Based Regression Model

Dr.L.Sridhara rao[1], J.Akhil[2], Ch.Keerthi[3], N.Sagar[4]
[1,2,3,4] *Department of Information Technology, J.B Institute of Engineering and Technology Hyderabad, India*

***Abstract*— The soil surface humidity parameter over vegetated fields is of great importance for controlling water consumption; prevention of salinity caused by over-irrigation; efficient use of irrigation system and improving the yield and quality of the cultivated crop. However, determination of the soil surface humidity is very difficult on vegetated fields. In order to overcome this problem, polarimetric decomposition models and machine learning based regression model were implemented. The main purpose of this study is to predict soil surface humidity on moderately vegetated fields. The scenario mainly concentrates on weather forecasting, crop yield prediction and crop forecasting. These factors help the farmers to cultivate the best food crops and raise the right animals with accordance to environmental components. Also, the farmers can adapt to climate changes to some degree by shifting planting dates, choosing varieties with different growth duration, or changing crop rotations. For experimental analysis, the statistical numeric data related to agriculture is undertaken. Whereas, the clustering-based techniques and supervised algorithms are utilized for managing the collected statical data. Additionally, the suitable classification methods like Support Vector Machine(SVM),neural networks are employed for better classification outcome.**

Index Terms: **Salinity, SVM, Logistic regression, Bayesian Network, Random Forest.**

I.INTRODUCTION

Agriculture is the backbone of Indian economy. The yield obtained primarily depends on weather conditions as rainfall patterns largely influence cultivation methodologies. With this context, farmers and agriculturalists require spontaneous advice proposition in predicting future reaping instances to maximize crop yield. Due to insufficient involvement of technology, the throughput of agriculture is yet to reach its full glory. Every farmer is interested in knowing the yield he/she could expect at the harvest period and hence, yield prediction is an important aspect for them. Over the years, farmers have an idea about the pattern in yield as per innate human intuition. However, rainfall as a major driver for crop raising can extensively rattle intuitive yield prediction by controlling some of the soil and environmental parameters related to the crop growth. Moreover, the right kind of soil to be employed for a crop is only known to the farmer only by on-paper advice and makes it difficult for him/her to trial and test on crop investment. The proposed architecture provides a computational dimension to enhance knowledge about the yield before the crop sowing period. It is made possible through a data driven hybrid model. Since the model performs a joint prediction of both rainfall and soil features on the yield, it is termed as a hybrid model. This model reads from data of past period and learns through computational processes like ARIMA, Exponential Sequencing and Recurrent Neural Networks. The model accepts soil features, fertilizer concentration and sowing period from the end user as test data for the trained model. The output is displayed as the expected yield in tons per hectare, for the given parameters. The soil features include the concentration of Nitrogen, Phosphorous and Potassium in the soil and fertilizer, and the mean pH value of the soil. The rainfall analysis model forecasts rainfall quantity in future time by learning and analyzing huge sets of past data, using ARIMA and Exponential Sequencing. On the other hand, a Recurrent Neural Network (RNN) is constructed for analyzing and learning crop yield patterns on the soil types. The RNN also develops predictions on the test

data using the results of rainfall predictions from the previous model flow. This approach is highly beneficial to farmers, agriculturalists, local self-governments, and Tahsildars to observe and allocate capital for agriculture and crop raising. The proposed approach also tries to solve the misery of high suicide rate of farmers in the country.

II.EASE OF USE

A. Selecting a Template (Heading 2)

First, confirm that you have the correct template for your paper size. This template has been tailored for output on the A4 paper size. If you are using US letter-sized paper, please close this file and download the Microsoft Word, Letter file.

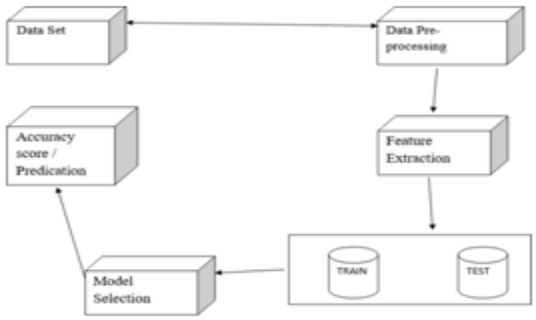
B. Maintaining the Integrity of the Specifications

The template is used to format your paper and style the text. All margins, column widths, line spaces, and text fonts are prescribed; please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

III.FUNCTIONAL REQUIREMENTS

- Data collection
- Data pre-processing
- Feature extraction
- Evaluation model

IV. SYSTEM ARCHITECTURE



V.ALGORITHM

A. Algorithm-random forest

Random forest is a type of supervised machine learning algorithm based on ensemble learning. Ensemble learning is a type of learning where you join different types of algorithms or same algorithm multiple times to form a more powerful prediction model. The random forest algorithm combines multiple algorithm of the same type i.e. multiple decision trees, resulting in a forest of trees, hence the name "Random Forest". The random forest algorithm can be used for both regression and classification tasks.

B. Support Vector Machine:

“Support Vector Machine” (SVM) is a supervised machine learning algorithm which can be used for both classification or regression challenges. However, it is mostly used in classification problems. In the SVM algorithm, we plot each data item as a point in n-dimensional space (where n is number of features you have) with the value of each feature being the value of a particular coordinate. Then, we perform classification by finding the hyper-plane that differentiates the two classes very well (look at the below snapshot). Support Vectors are simply the co-ordinates of individual observation.

C. Logistic Regression

Logistic regression is a statistical technique used to predict probability of binary response based on one or more independent variables. It means that, given a certain factors, logistic regression is used to predict an outcome which has two values such as 0 or 1, pass or fail, yes or no etc.

D. Bayesian network

A Bayesian network is a probabilistic graphical model which can be used for statistical analysis of the attributes for a given dataset.

E. Artificial Neural Networks

Artificial Neural Network is one of the most used technique for prediction models, ANN is based on the structure and features of Neural Networks, the imitation of human brain.

In this the primary computational units are called as neurons, these neurons are connected together in

layers, where the data is passed in as the input the network is trained throughout with special equations called as the activation functions. The applications of neural network is widely used for agricultural practices.

VI.MODULES

A. DATA COLLECTION:

Data used in this paper from Kaggle website we have data of weather report and soil condition based on that we are predicting . This step is concerned with selecting the subset of all available data that you will be working with. ML problems start with data preferably, lots of data (examples or observations) for which you already know the target answer. Data for which you already know the target answer is called labelled data.

B. DATA PREPROCESSING:

Our Organize your selected data by formatting, cleaning and sampling from it.

- Cleaning: Removing Unwanted Columns
- Formatting: Filling missing fields
- Sampling: Balancing the unbalanced data

C. FEATURE EXTRATION:

Next thing is to do Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes.

D. EVALUATION MODEL:

Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future.

VII. FUTURE ENHANCEMENT

This paper describes crop yielding prediction ability of the algorithm. In future we can determine the efficient web enhancement or application based on their accuracy metrics that will helps to choose an efficient algorithm for crop yield prediction.

VIII.CONCLUSION

The Results shows that we can attain an accurate crop yield prediction using the Random Forest algorithm. Random Forest algorithm achieves a largest number of crop yield models with a lowest models. It is suitable for massive crop yield prediction in agricultural planning. This makes the farmers to take the right decision for right crop such that the agricultural sector will be developed by innovative ideas.

REFERENCE

[1] Japneet Kaur, “Impact of Climate Change on Agricultural Productivity and Food Security Resulting in Poverty in India”, Università Ca' Foscari Venezia, 2017, pp. 16-18, 23.

[2] Anupama Mahato, “Climate Change and its Impact on Agriculture”, International Journal of Scientific and Research Publications, Volume 4, Issue 4, ISSN 2250-3153, April 2014, pp. 4-5.

[3] Pratap S. Birthal, Md. Tajuddin Khan, Digvijay S. Negi and Shaily Agarwal, “Impact of Climate Change on Yields of Major Food Crops in India:Implications for Food Security”, Agricultural Economics Research Review, Volume 27 (No. 2), pp. 145-155, July-December 2014.

[4] Rasul G, Q. Z. Chaudhry, A. Mahmood and K. W. Hyder, “Effect of. 2840Temperature Rise on Crop Growth & Productivity”, Pakistan Journal of Meteorology, Volume 8, Issue 15, 2011, pp. 7-8