AUTOMATED CROP PREDICTION BASED ON EFFICIENT SOIL NUTRIENTS ESTIMATION USING NAIVE BAYES CLASSIFICATION ALGORITHM

Mrs. Soumya Dath G
Assistant Professor
Dept of Information Science and Engg
GSSSIETW, Mysuru

Bindu M
Information Science and Engg
4GW15IS062
GSSSIETW, Mysuru

Harshitha N
Information Science and Engg
4GW16IS016
GSSSIETW, Mysuru

N Sowndarya
Information Science and Engg
4GW16IS024
GSSSIETW, Mysuru

Sushmitha K S
Information Science and Engg
4GW16IS051
GSSSIETW,Mysuru

Abstract—Soil is an important ingredient of agriculture. There are several kinds of soil. Each type of soil can have different kinds of features and different kinds of crops grow on different types of soils. We need to know the features and characteristics of various soil types to understand which crops grow better in certain soil types. Machine learning techniques can be helpful in this case. In recent years, it is progressed a lot. Machine learning is still an emerging and challenging research field in agricultural data analysis. In this paper, we have proposed a model that can predict soil series with land type and according to prediction it can suggest suitable crops. Several machine learning algorithms such as weighted *k*-Nearest Neighbor (*k*-NN), Bagged Trees, and Gaussian kernel based Support Vector Machines (SVM) are used for soil classification. Experimental results show that the proposed SVM based method performs better than many existing methods.

Keywords—Soil series, Land type, Chemical feature, Geographical attribute, machine learning

I. INTRODUCTION

Agriculture is a main occupation in India and soil is an important ingredient of agriculture. There are several kinds of soil. Each type of soil can have different kinds of features based on soil nutrients like Nitrogen, Phosphorous and Potassium present in the soil and different kinds of crops grow on different types of soils. We need to know the features and characteristics of various soil types to understand which crops grow better in certain soil types. Machine learning techniques can be helpful in this case. This model can predict soil series with land type and according to prediction it can suggest suitable crops. By applying Naive Bayes Classification Algorithm this process becomes more efficient.

There are about 500 soil series in Bangladesh which is identified by Soil Resources Development Institute (SRDI). Soil series means group of soils which is formed from the same kind of parent materials and remains under the similar conditions of drainage, vegetation time and climate. It also has the same patterns of soil horizons with differentiating properties. Each of the soil has different names and it is named after its locality (e.g. Barisal series, Sara series, Isshwardi series, etc). In Bangladesh, it is a starting point of soil classification to create a platform for its correlation with international soil classification systems (FAO or USDA-United States Department of Agriculture) [4]. Soil series are given names after their locality for having a convenience so that a particular series can be differentiated with another. For instance, 'Barisal Series' does not mean that the category of that series only found in the region of Barisal city, Upazilla or district. It defines the properties of 'Barisal Series' soils and it is applied to all soils those have same kinds of properties. Here the series names are entirely a label [5].

Ramesh et al. [6] used dataset, collected from Soil Science & Agricultural department, Kanchipuram and National Informatics Centre, Tamil Nadu, India. They used random forest, Bayes Net, Naïve Bayes, J48. The dataset consists of 2045 samples of ten types of soil. Gholap et al. [7] used soil datasets from three regions (Khed, Bhor and Velhe) of Pune district, India. Dataset has total 1988 instances with 9 attributes. They focus on applying various algorithms such as Naïve Bayes, JRip, J48 (which is an open source java implementation of the C4.5 decision tree algorithm) for classification task. Devi et al. [8] used several classification algorithms like *k* -means, Random Tree and Apriori, for classifying the soil and predict the suitable crops. This paper collects datasets from the Agriculture University of Coimbatore district of India which has 250 samples of 32 different places. They used clustering technique to group data, and then they classified the data by the order of soil and places with Random Tree algorithm. Then they have applied apriori Mining process to generate an association rule for finding suitable crops for the specific soil. Ramesh, D. et al. [9] used the process of applying data mining techniques on the agricultural data of East Godavari district of Andhra Pradesh in India. This dataset has four input variables and they are Year, Rainfall, Area of Sowing and Production. The datasets have been collected from Indian Meteorological Department, Statistical Institution, and Agriculture department. They have used *k* -means clustering to form a table which has divided into 4 clusters ranging from 1 to 5 cm from the total number of raining years. Then they have taken minimum, mean and maximum values of those 4 clusters. Then they compared the results of the dataset of sowing and average production by applying *k*- means clustering and Multiple Linear Regression (MLR) techniques.

The main purpose of the proposed work is to create a suitable model for classifying various kinds of soil series data along with suitable crops suggestion for certain areas of certain Upazilla of Bangladesh. We have worked with soil series of six upazillas of Khulna district, Bangladesh.

series are recognized by machine learning methods using various chemical features and possible crops for that soil series are suggested using geographical attributes.

The rest of the paper is organized as follows. Section II describes the proposed methodology in details. In section III, experimental result are outlined. Conclusion is drawn in section IV.

II. PROPOSED METHODOLOGY

The System architecture of the proposed model is shown

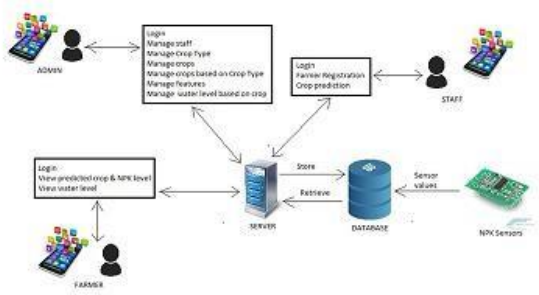

Fig. 1: Proposed System Architecture

The method involves two phases: training phase and testing phase. Two dataset are used: Soil dataset and crop dataset. Soil dataset contains class labeled chemical features of soil. Table I shows the details of the 12 chemical attributes of soil, used in our method.

VII CONCLUSION

A model is proposed for predicting soil series and providing suitable crop yield suggestion for that specific soil. The research has been done on soil datasets of six upazillas of Khulna region. The model has been tested by applying different kinds of machine learning algorithm. Bagged tree and K-NN shows good accuracy but among all the classifiers, SVM has given the highest accuracy in soil classification. The proposed model is justified by a properly made dataset and machine learning algorithms. The soil classification accuracy and also the recommendation of crops for specific soil are more appropriate than many existing methods. In future, providing fertilizer recommendation is our concern, also data of other districts will be added to make this model more reliable and accurate.

IX ACKNOWLEDGEMENT

We have calculated the mean and variance for twelve chemical attributes of various samples of soil database in order to find the characteristics of data. We have done that with respect of each class. We have found that for different kinds of soil, the number of independent attributes does not match with each other. Thus we can conclude that it cannot normally find any linear feature that can classify the soil series without using any machine learning algorithm.

Soil series and land type combine represents the soil class in the database. The machine learning methods are used to find the soil class (i.e. soil series and land type).

III. MODULES

1. Admin: Admin can login, manage staff, manage crops, manage crop type, manage crop type, manage features, manage water level based on crop.

2. Staff: Staff can login, farmer registration, crop prediction

3. Farmer: Farmer can login, view predicted crop and NPK level, view water level.

The data is stored in the database they can retrieve all these data and stored in the server.

IV METHODOLOGY

Naive Bayes Algorithm:

- It is a term in Bayesian statistics dealing with a simple probabilistic classifier based on applying Bayes theorem with strong(naive) independence assumptions.
- A Naive Bayes classifier assumes that the presence of articular feature of a class is unrelated to the presence of any feature.
- Depending on the precise nature of the probability model, Naive Bayes classifiers can be trained very efficiently in a supervised learning setting.

V ADVANTAGES

- Estimating the nutrient rate required to optimize crop productivity.
- Taking the differences in nutrient levels into account and acting on these levels to fix the particular crop to be planted in field.
- Diagnosis of nutritional deficiencies in soil directly in the fields at very low cost and it is flexible to all kinds of soil.

VI APPLICATIONS

- The project takes decades of field's data to analyze crop performance in various climates and new characteristics developed in the process. Based on this data they can build a probability model that would predict which genes will most likely contribute a beneficial trait to a plant.
- The application will help farmers to grow profitable crops.

We are grateful to Soil Resource Development Institute (SRDI), Government of the people's republic of Bangladesh, for providing valuable soil data of various upazillas of Khulna district, which helps us to conduct our experiment.

IV REFERENCES

[1] https://www.techopedia.com/definition/1181/data-mining. [Accessed date:19th August, 2018]

[2] http://www.library.arizona.edu/exhibits/swetc/azso/body.1_div.6.html . [Accessed date:19th August, 2018]

[3] Eswaran, H., Ahrens, R., Rice, T. J., & Stewart, B. A. (2002). *Soil classification: a global desk reference*. CRC Press..

[4] http://en.banglapedia.org/index.php?title=Bangladesh_Soil. Last Accessed date: 25th August, 2018.

[5] http://en.banglapedia.org/index.php?title=Soil_Series. Last Accessed date: 25th August, 2018.