Crop Recommendation and Yield Prediction for Agriculture using Data Mining Techniques

Aakunuri Manjula
Associate Professor, CSE, Jyothishmathi Institute of Technological Sciences, Karimnagar, Telangana State, India.

Dr. G.Narsimha
Professor, CSE, JNTUH, Telangana State, India. E-mail: narsimha06@gmail.com

ABSTRACT:

This structure is considered to predict the best harvest suitable for the agronomists' area. It also suggests farming strategies for crops such as diverse farming, spacing, irrigation, sow processing, etc. along with fertilizer and pesticide proposals. This is done on the basis of historical soil standards of the area and estimating crop and weather costs. Further, cost prediction is done based on Linear Regression to aid in ranking the crops recommended. India is defined as an agricultural country, where recommendations are given in traditional ways. In a Present-day, recommendations are based on farmers communicate between farmers, experts and various experts have a variety of recommendations. Recommendation can be provided to farmers who use past agricultural activities data. The application provides recommendations to farmers to determine the appropriate fertilizer and crop. This application can be used to increase crop yield and also recommend suitable crop.

Keywords: Data Mining, Crop Recommendation, Potassium, Nitrogen, Phosphorous, Crop Rotation

I. INTRODUCTION

India is second highest land area is more than 1.6 million square kilometers under agriculture. Most of the Indian population is involved in agriculture thus depends largely on the economy agriculture. India possesses the potential to be a great power in agriculture. Encourages agriculture raise poverty and rural development. Today in India agriculture is neglected leading to loss of hope of farmers in agriculture resulting in a rise in number of suicides among farmers. There is no such universal system to assist farmers in agricultural India. Hence, our rich collection previous agricultural data can be used for recommendation. Data extraction techniques and algorithms can be used to recommend crops as well as yield prediction. The recommendation will be based on nitrogen (N) Phosphorus (P), potassium values (K). This paper focuses to develop various algorithms that can help in building an effective recommendation system. This system recommends crops and fertilizers to farmers thus increase crop production.

Information extraction is the task of extracting information that is automatically structured from unorganized and / or semi-automated documents. A recommendation structure uses a number of different techniques. These structures can be grouped into two kinds: **content-based systems** and **collaborative filtering systems**. Content-based systems examine the characteristics of the recommended elements, while collaborative filtering systems recommend elements based on similarity measures between users and / or elements.

Content-based systems

Content based system examines the properties of elements for a recommendation, while collaborative filtering systems recommend elements based on similarity metrics between users and / or elements.

Collaborative filtering systems

The system should provide accurate recommendations to the super user as the number of users on the site increases. The filtering technology widely used in the recommended systems is "cooperative filtering". Compared to content-based filtering, a cooperative filtering system can normally filter out data that the system cannot consider and represent, and recommend recent information. Collaborative filtering approaches collect and evaluate a large amount of information about user behavior, creator preferences, and deduce what users will like based on their relationship to other users. The recommended cooperative filter system is scalability.

II. RELATED WORK

In current approaches, given the area and yield, the level of crop suitability appears in the different sub regions within the region [2]. Many geo-environmental factors such as soil, climate, slope, floods and erosion of hazards are considered [3]. But it is limited to very few crops. The results on other environmental factors were not good.

Mansi Shinde et al, developed both ANN and regression equations to predict cost to compare accuracy [1]. The end result will be the most appropriate sustainable crops in the province concerned as well as farming strategies to help grow the crop.

In this work, the following contributions were made:

- This system recommends crop by considering multiple features such as soil, weather and cost versus one feature.
- The cost factor is a very important factor considering the disasters that farmers now face. This is the first system that considered this feature to recommend crops.
- The farming strategy proposal algorithm is designed, which gives indicators for sowing seeds, spacing seeds, irrigation, harvesting, and nesting, which is one of its kind.
- An algorithm was designed to clean up strategic data using statistical methods to remove and separate agriculture strategies.

Soil quality is determined by NPK's value the soil. 'N' is nitrogen content in soil, 'P' is the content of phosphorus and "K" is the content of potassium the soil. Based on NPK contents, soil value can be predicted [4]. Nitrogen in the soil is responsible for the color the leaves of crop. If a low amount of nitrogen is found in the soil then the plants will have a slight yellowish leaves and quantity moderate or high will have green leaves. The content of phosphorus in the soil is responsible for Plant reproductive system. Its value will predict growth of fruits and flowers of plants. Potassium Soil content is responsible for its overall growth. Its value will predict the strength of plant roots will also determine the overall growth of the plant.

The data set for which we took the data the training group tried to apply the algorithms to it take the past data as a test group, and then view the output [5]. This output is compared with the actual output. Crop with maximum points can be recommended for Farmer. The market trend for crops is saved in database. While recommending more than one crop, the first factor that is selected will be the year factor that you will to be followed by the market factor and the ratio factor. To on recommend the crop to the user, we use Random Forest Algorithm.

The component of crop rotation is adequate use of nitrogen, phosphorus and potassium through the use of green manure and fertilizers. Crop rotation also reduces the accumulation of pathogens and pests which occur frequently when one type is down produced [6]. Soil structure and fertility can also be improved alternating between different leguminous plants. Rotation is a component of poly culture. In the agricultural area, crop rotation is very necessary over the years.

III METHODOLOGY

1. Dataset Collection

The data sets include specific soil features that are collected for Warangal test area in the soil test laboratory, Warangal, Telangana State India. In addition, similar online general crop data sources were also used. Crops considered in our model include Pulses, Banana, Sugar Cane, Cotton, Vegetables, Rice, and White Corn. The number of cases of each crop available in the training set. Attributes of depth consideration, soil texture, soil color, permeability, drainage, water holding and corrosion. The following are the datasets used for training and testing the prediction model built. Crop yield data sets are obtained from fao.org

URL:
www.fao.org/soils-portal/soil-survey/soil-maps-and-databases/harmonized-worldsoil-database-v12/

2. System Architecture

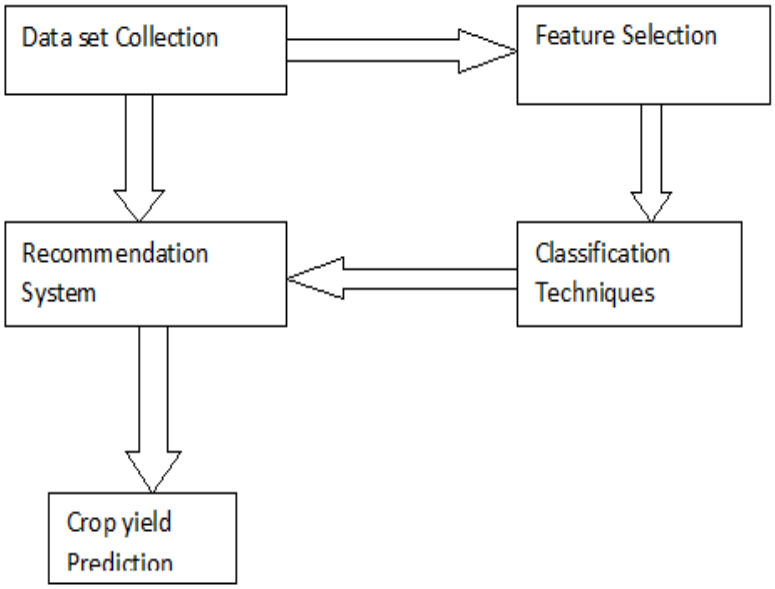

Figure 1: Proposed Architecture for Crop Recommendation

This application is very useful in terms of their applications as they provide each recommendation as well advantage to buy recommended products.

- ✓ It is useful for farmers to increase their crop production.
- ✓ It is user friendly.
- ✓ Requires less memory.

3. Crop Expectation using Ensembling methods

Ensembling is a data mining model used to integrate the power of two or more models to achieve better predictability and efficiency than any of its models can achieve on its own. In our system, any number of primary learners can be used. There must be at least two basic learners. Learners are chosen in a way that enables each other to be courteous as well. Top higher competition is a chance to improve forecasting. However, learners need to be free because when one or a few members make a mistake, the probability that the remaining members will correct this error will be high. Each learner builds himself in a model.

IV. Conclusion

India has a nation where agriculture plays a major role. In the flourishing of farmers, exploration of the nation. So our work will help farmers sow the right seeds based on soil requirements to increase productivity and gain profit from this technique. Thus, the farmer planting the right crop increases its yield as well increase the overall productivity of the nation. Our Future work is aimed at an improved data set of large number of attributes the yield prediction. This paper summarizes an effective recommendation system for fertilizers and crops based on NPK values and the region. If widely used, it will benefit farmers in terms of crop production. The application is very user friendly everyone can use it.

V. REFERENCES

[1] A Model for Prediction of Crop Yield, "**International Journal of Computational Intelligence and Informatics, Vol. 6: No. 4, March 2017**.

[2] Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India.

[3] Rakesh Kumar, M.P. Singh, Prabhat Kumar and J.P. Singh (2015), 'Crop Selection Method to Maximize Crop Yield Rate using Machine Learning Technique', International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM).

[4] (Soil Test Information | Recommended Fertilizers) Agri. Dept, "Soil Health Card ", February, 2015.

[5] Anshal Savla, Parul Dhawan, Himtanaya Bhadada, Nivedita Israni, Alisha Mandholia , Sanya Bhardwaj (2015), 'Survey of classification algorithms for formulating yield prediction accuracy in precision agriculture', Innovations in Information,Embedded and Communication systems (ICIIECS).

[6] Tapas Ranjan Baitharua, Subhendu Kumar Panib (2016), 'Analysis of Data Mining Techniques for Healthcare Decision Support

[7] Satish Babu (2013), 'A Software Model for Precision Agriculture for Small and Marginal Farmers', at the International Centre for Free and Open Source Software (ICFOSS) Trivandrum, India.

[8] http://www.geeksforgeeks.org/np-completenessset-1/

[9] http://www.tutorialspoint.com/

[10] https://data.gov.in/

[11] http://stackoverflow.com/