International Journal of Future Generation Communication and Networking
Vol. 13, No. 3, (2020), pp. 1896–1901

Crop Prediction Using Machine Learning

Kevin Tom Thomas[1], Varsha S[2], Merin Mary Saji[3], Lisha Varghese[4], Er. Jinu Thomas[5]

[1,2,3,4] UG students Department Computer Engineering, SAINTGITS College of Engineering, APJ ABDUL KALAM TECHNOLOGICAL University, Kerala, India
[5]Asst. Prof. Department Computer Engineering, SAINTGITS College of Engineering, APJ ABDUL KALAM TECHNOLOGICAL University, Kerala, India

Abstract

Agriculture is one of the most essential and widely practiced occupations in India and it has a vital role in the development of our country. Around 60 percent of the total land in the country is used for agriculture to meet the needs of 1.2 billion people, so improving crop production is therefore seen as a significant aspect of agriculture. Basically if we have a piece of land, we need to know what kind of crop can be grown in this area. Agriculture depends on the various soil properties. Production of crops is a difficult task since it involves various factors like soil type, temperature, humidity etc. If it is possible to find the crop before sowing it, it would be of great help to the farmers and the other people involved to make appropriate decisions on the storage and business side. The proposed project would solve agricultural problems by monitoring the agricultural area on the basis of soil properties and recommending the most appropriate crop to farmers, thereby helping them to significantly increase productivity and reduce loss. Our project is a recommendation system which makes use of different machine learning techniques such that it recommends the suitable crops based on the input soil parameters. This system thus reduces the financial losses faced by the farmers caused by planting the wrong crops and also it helps the farmers to find new types of crops that can be cultivated in their area.

Keywords: *K Nearest Neighbor, Decision Tree, K Nearest Neighbor with Cross Validation, Naive Bayes, Support Vector Machine.*

1. Introduction

Agriculture has a major role in the lives of every individual. From the olden times itself agriculture is considered to be one of the main practices practiced in India. In olden times, people used to cultivate crops in their own land in order to meet their requirements. Being treated as India's backbone, the agricultural sector has strengthened with the public's needs as the technology is improving. With the rapid population growth, these innovations are very much needed to meet the needs of every person.

Our country had undergone several fluctuations in the price of onions last year. The price of the onions increased from Rs. 26 per kg to Rs. 50 per kg in the month of August.[1] So most of the farmers decided to cultivate onion in their fields seeing this huge increase in price so that they could make large profits from their land. In some regions for example in Maharashtra this resulted in the abundant supply of onions while many other regions suffered a failed crop production and the farmers lost a large amount of money. This problem occurred due to many unfavorable conditions that prevented the growth of onions. A continuous shortage in the production of onions again in the next few months had a very bad effect on the lives of the common people. This happened because the middle-class people were not able to afford the huge price of onion which is a frequently used commodity in their houses.

The example above shows us that a decision of a farmer regarding which type of crop to grow in his land is generally depends on his intuition and many other factors such as making huge profits within a short period of time, lack of awareness about the demand in the market and when he overestimates a

ISSN: 2233-7857 IJFGCN

soil's potential to support the growth of a particular type of crop and many more. A wrong decision that is taken on the farmer's side could put a much bigger pressure on the financial condition of his family resulting in severe loss. For all this reason we can see a farmer's pressure regarding which crop should be grown in his land. So now the most important aspect is to design a recommendation system that predicts the type of crop that can be grown in a particular land and thereby helping the farmers. With this aim in mind we have decided to develop a system that takes in the soil parameters like N, P, K (Nitrogen, Phosphorus, Potassium) and the pH values and predicts the most suitable crop that can be grown in that region. The dataset already contains the NPK and pH values of the soil and the appropriate or the most suitable crop that can be grown in given soil.

2. Literature Review

[2] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture"

The crops that were considered in the model for prediction include coriander, pulses, cotton, paddy, sorghum, groundnut, sugarcane, banana and vegetables. Different attributes of the soil were considered in order to predict the crop, which included pH, depth, erosion, permeability, texture, drainage, dater holding and soil color. The technique used was ensembling, which combined the power of using two or more different models for better prediction. The ensembling technique used was called the Majority Voting Technique.

[3] R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique"

The crops were inspected and graded depending on an examination to estimate crop yielding. This categorisation is found from different data mining algorithms. This paper provides a perception into various grouping rules, such as K-Nearest Neighbour and Naive Bayes. By making use of this document, we evaluated the classification rules and established which all will match the set of data we will be using in our project.

[4] T.R. Lekhaa, "Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques"

The paper hypothesizes analysis of Explorative Data and considers the design of different types of predictive models. A data set is taken as a sample data set, and different regression techniques are tried to recognise and examine each property. Specific regression methods discussed here are Multiple Linear, Linear, Non-Linear, Polynomial, Ridge regression and Logistic. Using this article, we obtain a comparative study of the different algorithms in data analytics. This helped in determining which algorithm is most appropriate to the proposed system.

[5] Viviliya, B. and Vaidhehi, V., "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms"

The attributes in the dataset included the soil type, groundwater level, rainfall, water availability, temperature of one dataset and the other dataset included the potassium, phosphorus, and nitrogen values, fertilizers, soil pH and organic carbon value. The dataset was preprocessed using basic preprocessing tasks. Naive Bayes and J48 classifiers were used for the crop recommendation. The final recommendation was done using association rules based on the results obtained from the classifiers. The model was trained using 10-cross validation. The testing was done based on different metrics like the Accuracy, ROC Area, Recall, Precision, F-Measure etc.

3. Proposed System

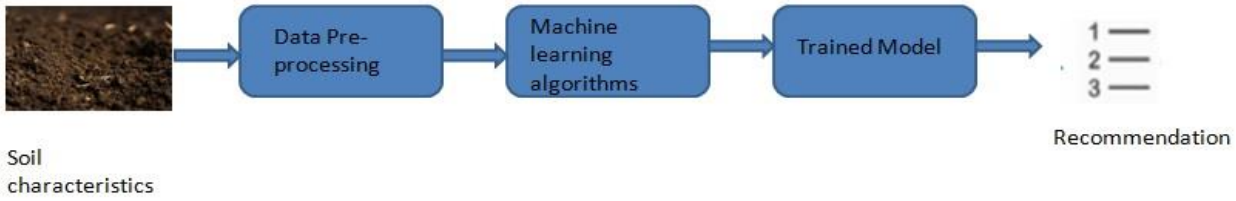

FIGURE I. Model of the system

The project proposes a model which can predict the crop based on the soil nutrient values (NPK values) and pH given as the input.

A. Acquisition of training dataset:

The accuracy of a machine learning algorithm may depend on the number of parameters used and to the extent of correctness of the dataset [6]. Our dataset contains the N, P, K, and pH values of different kinds of soils as attributes and it also contains the corresponding crops that can be grown in that soil as label. Thus, by using an appropriate machine learning algorithm we can train the dataset to predict the most suitable crop that can be grown under the given input parameters.

B. Data preprocessing:

Data preprocessing is the second step and it contains two steps. Original dataset can contain lots of missing values so initially all these should be removed. Missing values are denoted by a dot in the dataset and their presence can deteriorate the value of entire data and it can reduce the performance. So, to solve this problem we replace these values with large negative values which will be treated as outliers by the model. Generating the class labels is the second step. Since we are using a supervised learning method, for each entry in the dataset there should be a class label which is created during the preprocessing step.

C. Machine Learning Algorithm:

Different machine learning algorithms are being used in order to make comparisons. The different algorithms used are as follows:

a. k-Nearest Neighbor:

In this algorithm, the input provided will be the k nearest training examples of the dataset and the output will depend on whether it is a classification or regression problem. Basically, it works based on the minimum distance from the given input value which is soil values to the trained values to find the nearest k neighbors and afterwards those with majority is taken to be as the output prediction to predict the crop label[7]. To find which is most similar to the given instance distance measure is used. Mostly, by default Euclidean distance is used as a distance measure. It is calculated by the given formula,

$$d(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

where d is the Euclidean distance, x is the new point, y is the existing point, i is the input attribute and n are the number of data points.

b. Decision Tree:

A decision tree is a non-parametric method of supervised learning technique. Throughout the process a tree like structure is formed. In this, the dataset is broken down to build upon the tree subsequently. Finally, the resulting output is a tree like structure with both decision nodes and leaf nodes. Decision nodes can either have two or more branches while the leaf nodes indicate the final nodes representing classification or regression result. The topmost node is the root node and the one with higher gain (or gini index) value is taken to be the root. Decision trees have the ability to classify both categorical and numerical data [8].

c. kNN with Cross Validation:

Cross validation is referred to as a technique that is used for finding out the accuracy regarding a particular type of classifier and is performed either by making use of a number of testing or training partitions of the given data. Cross validation technique is carried by partitioning the dataset randomly into 'k' groups. Amongst these groups one out of the k groups will be used for testing and the rest k-1 groups are used for training. This process goes on repeating until the whole k groups are used as the test set.

d. Naive Bayes:

This classifier has features that are statically independent to one one another. Most of the other classifiers predict some amount of correlation between the features but Naive Bayes models its different features as independent features given its class. This implements a restriction on the given data, but in practice naive bayes have more sophisticated techniques to use and enjoy some theoretical support for improving its efficiency [9]. Naive Bayes classifiers can take different high dimensional features with very less number of training data and they are also very highly scalable classifiers.

e. Support Vector Machine (SVM):

An SVM that is Support Vector Machine is an example of a supervised machine learning model which has many learning algorithms that analyzes the data that is used for solving both classification and regression problems. We are given some training samples where each sample is marked such that it belongs to one or other of the two initially given categories. Support Vector Model algorithm creates a model where it allots new samples to any of the given categories. An SVM represents many examples that are taken as dots in space such that the samples belonging to different groups are partitioned with a gap between them.

D. Trained model:

Trained models are obtained after applying the dataset to the machine learning algorithms. Our paper suggests a crop prediction system which is based on the KNN or K-Nearest Neighbour algorithm. Soil properties such Nitrogen, Phosphorus, Potassium, pH value, etc. are given as input to the model. The algorithm will look for a crop which will have the value closest to the inputted values. The output will be all the crops which are suitable for the inputted values. The project is proposed to be implemented as a mobile application. A mobile application is considered as the number of people using smartphones are rising as by the passage of time.

Taking into consideration kNN, this algorithm essentially looks over the whole dataset for similarities. The result is calculated based on the most comparable or closest values. Because of its higher convergence speed and simplicity this algorithm is preferred over other algorithms. The input for the algorithm is the soil properties such as Nitrogen, Phosphorus, Potassium, pH value, etc. The entire dataset

in this algorithm is divided under a number of classes or possible outcomes depending on the number of records in the dataset. The algorithm predicts that the given input falls into which class one falls. We determine the value of k which is to be considered for the number of nearest neighbours. The value of k is determined according to the number of records present in the dataset. The result's accuracy is dependent on assessing the correct k value which should not be too high or too low. With 1850 record entries in our dataset the value for k is taken as 10. The k nearest neighbours, that is the k minimum distances, are used to predict the crop for the perfect soil type. Likewise, other models can be trained using the given dataset.

4. Results

All processing is carried out on a Windows 10 system having hardware configuration of Intel core i5 processor with an internal RAM of 8GB and a CPU speed of 2GHz. The algorithms that were used for testing were kNN, kNN with Cross Validation, Decision Tree, Naive Bayes and SVM, the accuracies obtained were 85%, 88%, 81%, 82% and 78% respectively.

kNN with cross validation has the highest accuracy and thus can be used for implementation in the final system.

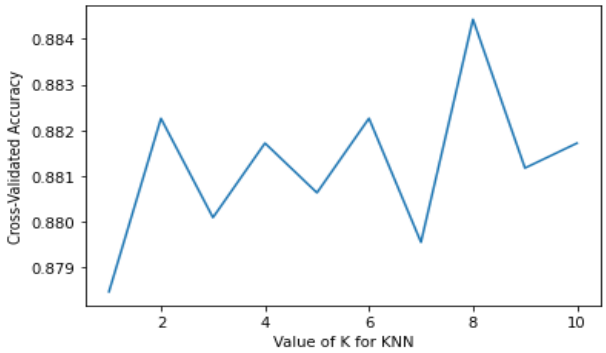

FIGURE II. Accuracy graph for KNN with cross-validation

There are many advancements for our project compared to the previous papers. In our project we are taking a large dataset therefore we can get the details regarding a greater number of crops. So more number of crops that can be grown in different soil conditions can be predicted. We have used different machine learning models in our project. Different models show different accuracies so we can select the best among them in order to do the accurate predictions. In this manner we get the results in a speedy way. We have built our project in such a way that it is easily accessible to all the farmers and with the advancement in technology we can incorporate more features into it. Since we are using the machine learning model of higher predicting accuracy, our project gives best results. The occurrence of natural disasters like flood and soil erosion can change the overall composition of the soil and our recommendation system provides a better way to predict the suitable crops in the changed soil conditions. The UI of the project is designed such that it is easily understandable by the common people.

5. Conclusion and Future Work

The proposed system takes the soil N, P, K, and pH values into consideration and determines which are the best productive crops that can be grown in that suitable soil conditions. Since the system lists all potential crops it helps the farmer determine which crop to be grown in their area. This system thus helps the farmer to decide on the maximum profitable crop and also helps in finding new crops that can be cultivated which have not been cultivated till that time by the farmer. In the future, this system can be implemented further using IOT to get the real time values of the soil. In the farm, the sensors can be

ISSN: 2233-7857 IJFGCN

installed to collect information about the current soil conditions, and the systems can therefore increase the accuracy of correctness of the results. Hence, farming can be done in a smart way.

References

[1] T. Edwin, "Onion, tomato price spike: season not the only reason", Nov. 13, 2017. [Online]. Available: https://www.thehindubusinessline.com/economy/agri-business/onion-tomato-price-spike-season-not-the-only-reason/article9957255.ece#. [Accessed Feb. 22, 2020].
[2] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, "Crop recommendation system for precision agriculture," 2016 Eighth International Conference on Advanced Computing (ICoAC), Chennai, 2017, pp. 32-36. doi: 10.1109/ICoAC.2017.7951740.
[3] R. Kumar, M. P. Singh, P. Kumar and J. P. Singh, "Crop Selection Method to maximize crop yield rate using machine learning technique," 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), Chennai, 2015, pp. 138-145. doi: 10.1109/ICSTM.2015.7225403
[4] T.R. Lekhaa, "Efficient Crop Yield and Pesticide Prediction for Improving Agricultural Economy using Data Mining Techniques", International Journal of Modern Trends in Engineering and Science (IJMTES), 2016, Volume 03, Issue 10.
[5] Viviliya, B. and Vaidhehi, V., "The Design of Hybrid Crop Recommendation System using Machine Learning Algorithms". International Journal of Innovative Technology and Exploring Engineering, 2019, 9(2), pp.4305-4311.
[6] Z. Doshi, S. Nadkarni, R. Agrawal and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6. doi: 10.1109/ICCUBEA.2018.8697349
[7] "k-nearest neighbors algorithm- Wikipedia", available at https://en.wikipedia.org/wiki/K-nearest_ neighbors_algorithm, visited in February 2018.
[8] "How Decision Tree Algorithms work" available at dataaspirant.com/2017/01/30/how-decision- tree-algorithm-works
[9] Naive Bayes classifier available at https://en.wikipedia.org/wiki/Naive_Bayes_classifier