Bachelor of Science in Computer Science & Engineering


# A user-friendly plant recommendation android application for cultivation based on machine learning

by

Kaniz Fatima Tonni

ID: 1604053

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

August, 2021

**Chittagong University of Engineering & Technology (CUET)**
**Department of Computer Science & Engineering**
**Chattogram-4349, Bangladesh.**

---

# Thesis Proposal

Application for the Approval of B.Sc. Engineering Thesis/Project

**Student Name** : Kaniz Fatima Tonni Session : 2019-2020
**ID** : 1604053

**Supervisor Name** : Dr. Mahfuzulhoq Chowdhury
**Designation** : Assistant Professor
Department of Computer Science & Engineering

**Department** : Computer Science & Engineering
**Program** : B.Sc. Engineering

**Tentative Title : A user-friendly plant recommendation android application for cultivation based on machine learning**

# Table of Contents

# List of Figures

# List of Abbreviations

**CHAID** Chi-square automatic interaction detection. 2

**EC** Electrical conductivity. 2

**KNN** K-Nearest Neighbor. 4, 6

**NPK** Nitrogen (N), phosphorus (P) and potassium (K).. 2

**OC** Organic carbon. 2

# 1 Introduction

Due to its rich delta soil, ample water, and many agricultural seasons, Bangladesh has long been known for its prodigious development of the region's cash crops of jute, rice, etc. With numerous problematic circumstances, such as a dense population, Bangladesh's status as a market leader has evolved through time, and the country is always seeking methods to increase total agricultural production through a range of measures.

Technology is one such instrument that may help you achieve your goals – and it's even more effective when utilized in a local setting. Also, it has reached every corner of the world. Nowadays, rural people also use digital devices and smartphones. One of the most significant benefits of smart technology and the digital era is the ability to check factors and market prices/demand from afar while also receiving useful technical agricultural advice. Bangladeshi farmers have prospered for decades utilizing low-tech methods. Therefore, any new technology provided to help them better their livelihoods must be easily accessible. Aside from accessibility, many of these farmers just like their "old ways," and it's worth exploring why some of them may still be feasible technological progress alternatives. Farmers in hard-to-reach areas who may profit tremendously from low-cost, high-quality farming technology frequently lack the means, expertise, and training to apply it. Another issue that must be considered is the cost. Because the cost of cultivating short-term and long-term plants depends on the size of the region.

As a result, several research articles suggest a method for selecting crops based on soil characteristics, climate, and worldwide locations [1] Several machine learning-based frameworks have been created to assist farmers in determining which crops to yield depending on soil characteristics. Not only were basic machine learning frameworks built, but the work also extended across the Android system [2].

It would be a huge benefit to individuals working in the agricultural sector if there was a system that could help you figure out which crop to grow based on criteria like soil type, parameters, season, production time, area size, and yielding cost. This type of system can give a complete idea for cultivating any plants, such as

approximate profit for a specific budget, cultivation method, suitable plants for a specific type of soil. So, depending on soil parameters, production time, and cost, we offer a recommendation system using machine learning and develop an android application based on it.

# 2 Background and Present State

Many works have been established on crop recommendation. To accomplish these tasks, machine learning was a key study.

In paper [1], the authors used ensemble model with majority voting techniques utilizing random tree, CHAID, K-Nearest Neighbor, and Naive Bayes as the learner to select the appropriate crop based on soil parameters. Also in [3], the authors used an ensemble model utilizing Support vector machines, Naive Bayes, Multilayer Perceptron (Artificial Neural Network), Random forest as learners. In both papers, Depth, Texture, PH, Soil Color, Permeability, Drainage, Water Holding, and Erosion are the attributes evaluated in the parameters. The limitation of these papers is that the authors only utilized soil parameters to suggest a crop. Moreover, organic matters or nourishment substances such as pH, Fe, Mn, OC, EC, P, N, K, Zn, B, etc were not considered in the parameters.

The authors of the paper [4] used fuzzy rules using the LEM2 method. For both discretized and fuzzified datasets, experimental results give greater prediction accuracy. The authors only focused on the soil parameters with characteristics such as pH, Fe, Mn, EC, OC, Zn, B, P, K, N, etc. There is no mention of other factors like season, short-term or long-term crops, approximate cost, etc.

In the paper [5], the authors proposed a framework for obtaining the dirt and yield boundaries, as well as guiding those in determining what harvests are appropriate. To recommend crop soil parameters with attributes of nitrogen, phosphorous, and potassium, and pH, as well as yield boundaries, were taken into account. The limitation of this paper is that the authors only utilized soil parameters and yield boundaries to suggest a crop.

In the paper [6], the authors used Naive Bayes, Random forests, Linear Support Vector Machine as base learners in an ensemble framework. The authors gave primary significance to soil type, pH value, NPK content and porosity of the soil,

average rainfall, surface temperature, and sowing season to determine the crop. The limitation of this paper is that the authors did not mention approximate cost or profit for crop cultivation.

In the paper [2], the authors implemented an application system. Farmers may use the proposed model's extensive recommendation set to improve their crop choices based on individual characteristics including location, farm size, temperature, rainfall, and a dataset on a variety of crops. The authors considered a large number of factors to determine the crop for cultivation. But the application has no feature for cost prediction, profit prediction, long-term or short-term, and no guidelines for cultivation.

Most of the related work previously done focused on soil type, parameters, season, average rainfall, etc. In contrast with them, we would like to take other elements into account: cost, approximate profit and long-term or short-term. It will be able to give an idea of the user's spending budget and calculation of subsequent profit. So, we propose to develop a plant recommendation system using machine learning. The proposed system is expected to recommend plants based on soil parameters, farm size, sowing season as well as cost budget, approximate profit and long-term or short-term information. The expected output will be a list of the plant(s) with guidelines for cultivation and approximate profit.

# 3 Specific Objectives and Possible Outcomes

The primary objective of this work is to create a plant recommendation system using machine learning. The key objectives and possible outcomes of the work are mentioned below:

- To develop a system using machine learning that can predict plants based on multiple factors such as soil parameters, farm size, sowing season, cost budget, approximate profit and long-term or short-term information.
- To provide some guidelines to cultivate regarding the plant that is recommended.

- To develop a user-friendly android application for the proposed system.

# 4 Outline of Methodology

Our suggested model's main goal is to select a plant-based on a set of characteristics such as soil parameters, production time, season, and cost. The workflow diagram for the system we suggested is shown in Figure 4.1.

## 4.1 Dataset collection

Soil characteristics, farm size, planting season, location, cost budget, and long-term or short-term information are all included in the dataset. The dataset must be large enough to train the ensemble model, as well as the system, which will be trained by the training dataset and evaluated with the input. [7].

## 4.2 Processing data

The Naive Bayes technique may be used to evaluate data based on user input of short-term or long-term plants, as well as season. The challenge was chosen because the Naive Bayes method evaluates if a data item fits in a certain category or not. It may be used in text analysis to determine whether or not words or phrases belong to a preset category. This method's output should be a reduced dataset version of the preceding main dataset depending on length and season.

## 4.3 Recommended system

The recommended system is divided into two parts. The machine learning approach is used in both parts. Our systrem uses machine learning techniques to determine which plants are appropriate.

The first section of the system works on an ensemble model. The model, which will be trained utilizing the dataset and tested with the input. [7]. This model consists of three base learners respectively, Random forest, KNN, and Decision tree.

The second section of the system determines the plants that are suitable for yielding based on the cost factor. A machine Learning approach called Linear

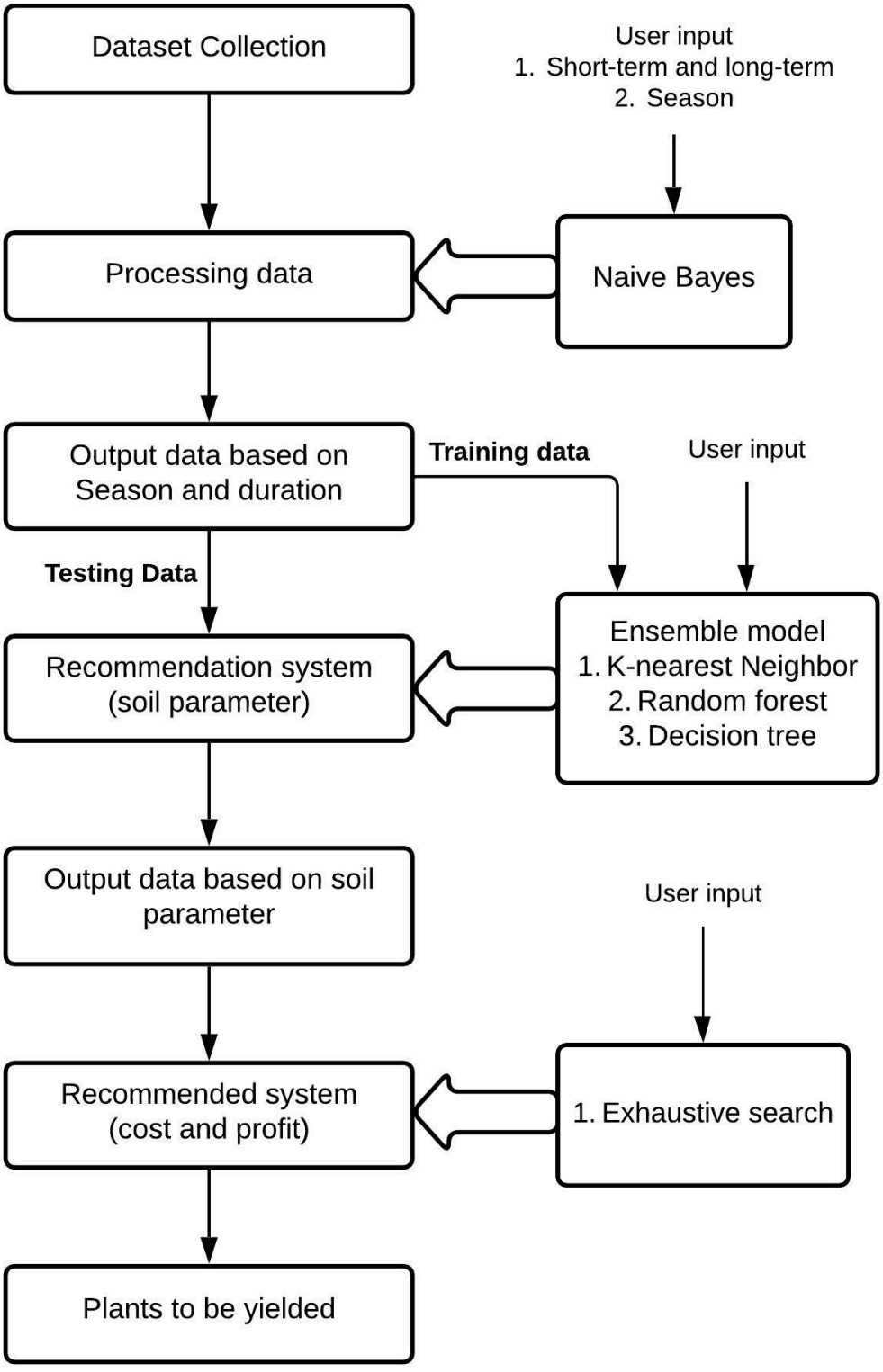


Figure 4.1: Workflow diagram of recommended system

Regression, which uses an algorithm, may readily handle the issue. The Gradient Descent Method is a basic algorithm that finds the best answer from a set of data

and aids prediction [8]. We propose using a machine learning technique called, Exhaustive search to anticipate suitable plants based on the cost factor.

## 4.4 Ensemble model

An Ensemble is a data mining model that combines the capabilities of multiple models to achieve greater prediction and efficiency than most of the other models alone could. In our system, we employ the Majority Voting approach, which is one of the most well-known Ensemble techniques. In the voting procedure, any number of base learners can be used. It is necessary to have at least two base learners. The pupils are chosen such that they are both capable and complimentary to one another. The more competition there is, the more likely it is that a right prediction will be made. Learners, on the other hand, must be complimentary, since if any of the learners make a mistake, the odds of the remaining members rectifying the mistake are great. Each student makes a self-portrait. The model is trained using the supplied training data set [3]. The learner for the system model is decided to be K-Nearest Neighbor, Random Forest, and Decision Tree.

### 4.4.1 K-Nearest Neighbor

Both classification and regression may be done with the K-Nearest Neighbor method. KNN is a basic method that stores all previous instances and uses a similarity score to classify new ones. This algorithm was selected because it may outperform more sophisticated classifiers despite its simplicity, and it is utilized in a range of applications.

### 4.4.2 Random Forest

The random forest algorithm is a decision tree variant in which you create a large number of decision trees using training data and then fit your new data into one of the trees as a random forest. It simply connects your data to the nearest tree on the data scale by averaging it. Random forest models are useful because they solve the problem of decision trees "pushing" data points into a category needlessly.

### 4.4.3 Decision Tree

A decision tree is a supervised learning method that is ideal for classification issues because it can precisely rank classes. It operates like a flow chart, dividing data points into two comparable categories at a time, starting with the "tree stem" and on to "branches" and "leaves," where the categories become increasingly finitely similar.

## 4.5 Exhaustive search

A machine learning technique of seeking for the most ideal hyper parameters by evaluating whether each candidate is a good match is known as exhaustive search, or brute-force search. The reason for choosing this technique is,

- All alternatives are weighed.
- The one that makes the most sense.

It's best used when the database is small and accuracy is more critical than cost or comp speed.

## 4.6 Android application

We propose creating an android application based on the recommended system that we decided to establish using machine learning. Our application will be user-friendly in the sense that it will be simple, convenient to use, and reliable. The flowchart in Figure 4.2 depicts the application's workflow.

- The app will provide a log-in or sign-up system so that users can save their data.
- Our app will have two activities: Profile and Homepage. Profile, where users can manage history and account details. Homepage, where users can system to get a suggestion for choosing plants. The recommendation system will work in three steps.
- In the first step, the user will be asked to provide if he/she wants to choose

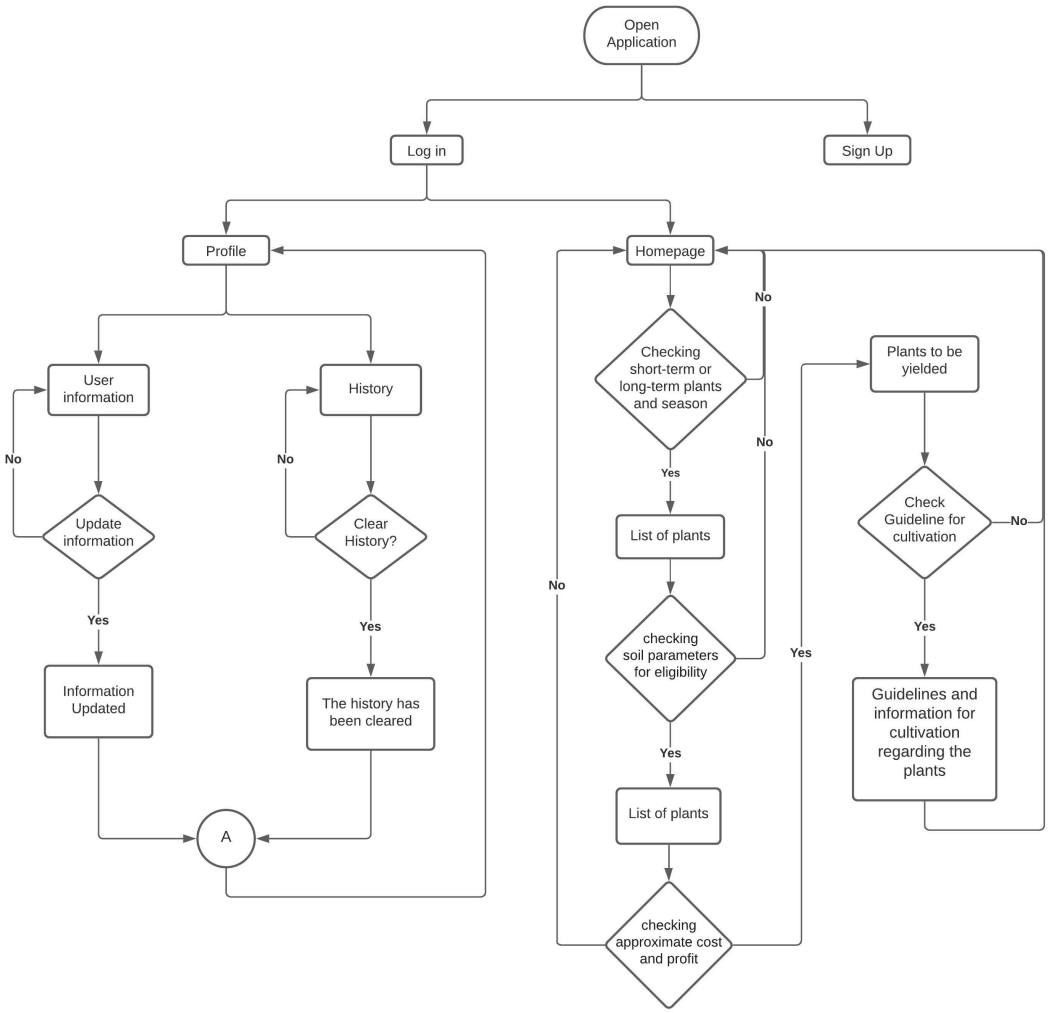


Figure 4.2: Workflow diagram of android application

long-term plants or short-term plants and the growing season. The expected output is a list of plants.

- In the second step, the user will be asked to provide parameters of the soil.
- In the third step, the user will be asked to provide the budget for the plantation. The output will show a list of the suitable plant(s).
- A guideline on cultivating the recommended plant will be provided to the user.

# 5 Required Resources

## 5.1 Hardware resources

- Personal computer
- Android device

## 5.2 Software resources

- Python
- React native framework
- Django

# 6 Cost Estimation

| | |
|---|---|
| • Hardware devices | Tk 90000 |
| • Softwares | Tk 5000 |
| Total cost | Tk. 95000 |

**CSE Undergraduate Studies (CUGS) Committee**

**Reference :**

**Meeting No :** **Resolution No :** **Date :**

_____

**Signature of the Student**

_____

**Signature of the Supervisor**

_____

**Signature of the Head of the Department**

# References


[1] D. A. Reddy, B. Dadore and A. Watekar, 'Crop recommendation system to maximize crop yield in ramtek region using machine learning,' *International Journal of Scientific Research in Science and Technology*, vol. 6, no. 1, pp. 485–489, 2019 (cit. on pp. 1, 2).

[2] H. Salpekar *et al.*, 'Design and implementation of mobile application for crop yield prediction using machine learning,' in *2019 Global Conference for Advancement in Technology (GCAT)*, IEEE, 2019, pp. 1–6 (cit. on pp. 1, 3).

[3] R. K. Rajak, A. Pawar, M. Pendke, P. Shinde, S. Rathod and A. Devare, 'Crop recommendation system to maximize crop yield using machine learning technique,' *International Research Journal of Engineering and Technology*, vol. 4, no. 12, pp. 950–953, 2017 (cit. on pp. 2, 6).

[4] A. Rajeswari, A. S. Anushiya, K. S. A. Fathima, S. S. Priya and N. Mathumithaa, 'Fuzzy decision support system for recommendation of crop cultivation based on soil type,' in *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, IEEE, 2020, pp. 768–773 (cit. on p. 2).

[5] G. Suresh, A. S. Kumar, S. Lekashri and R. Manikandan, 'Efficient crop yield recommendation system using machine learning for digital farming,' *International Journal of Modern Agriculture*, vol. 10, no. 1, pp. 906–914, 2021 (cit. on p. 2).

[6] N. H. Kulkarni, G. Srinivasan, B. Sagar and N. Cauvery, 'Improving crop productivity through a crop recommendation system using ensembling technique,' in *2018 3rd International Conference on Computational Systems and Information Technology for Sustainable Solutions (CSITSS)*, IEEE, 2018, pp. 114–119 (cit. on p. 2).

[7] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika and J. Nisha, 'Crop recommendation system for precision agriculture,' in *2016 Eighth International Conference on Advanced Computing (ICoAC)*, IEEE, 2017, pp. 32–36 (cit. on p. 4).

[8] A. Chandrashekhar, U. Desai and P. Abhilash, 'Cost prediction using gradient descent algorithm,' in *Journal of Physics: Conference Series*, IOP Publishing, vol. 1706, 2020, p. 012 038 (cit. on p. 6).