

***Dissertation on***

**PerfectCrop**

**The right crop for your soil**

*Submitted in partial fulfilment of the requirements for the award of degree of*

**Bachelor of Technology in**

**Computer Science & Engineering UE18CS390A – Capstone Project Phase - 1**

***Submitted by:***

**Srish Srinivasan**

**Akash Kumar Rao**

**Vishruth P Reddy**

**Ishan Agarwal**

**PES1201800051**

**PES1201800089**

**PES1201800102**

**PES1201800291**

*Under the guidance of*

**Prof. Raghu B A**

Associate Professor

PES University

**January - May 2021**

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

FACULTY OF ENGINEERING

**PES UNIVERSITY**

(Established under Karnataka Act No. 16 of 2013) 100ft Ring Road, Bengaluru – 560 085, Karnataka, India



# PES UNIVERSITY

(Established under Karnataka Act No. 16 of 2013)

100ft Ring Road, Bengaluru – 560 085, Karnataka, India

**FACULTY OF ENGINEERING**

CERTIFICATE

*This is to certify that the dissertation entitled*

**‘PerfectCrop-The right crop for your soil’**

*is a bonafide work carried out by*

**Srish Srinivasan**

**Akash Kumar Rao**

**Vishruth P Reddy**

**Ishan Agarwal**

**PES1201800051**

**PES1201800089**

**PES1201800102**

**PES1201800291**

in partial fulfilment for the completion of seventh semester Capstone Project Phase - 1 (UE18CS390A) in the Program of Study - Bachelor of Technology in Computer Science and Engineering under rules and regulations of PES University, Bengaluru during the period Jan. 2021 – May. 2021. It is certified that all corrections / suggestions indicated for internal assessment have been incorporated in the report. The dissertation has been approved as it

satisfies the 6th semester academic requirements in respect of project work.

Signature

Prof. Raghu B A

Associate Professor

**Name of the Examiners**

Signature Dr. Shylaja S S

Chairperson

**External Viva**

Signature

Dr. B K Keshavan Dean of Faculty

**Signature with Date**

**1.** \_

**2. \_**

# DECLARATION

We hereby declare that the Capstone Project Phase - 1 entitled **“PerfectCrop-The right crop for your soil”** has been carried out by us under the guidance of Prof. Raghu B A, Associate Professor and submitted in partial fulfilment of the course requirements for the award of degree of **Bachelor of Technology** in **Computer Science and Engineering** of **PES University, Bengaluru** during the academic semester January – May 2021. The matter embodied in this report has not been submitted to any other university or institution for the award of any degree.

|  |  |  |
| --- | --- | --- |
| **PES1201800051** | **Srish Srinivasan** | **<Signature>** |
| **PES1201800089** | **Akash Kumar Rao** | **<Signature>** |
| **PES1201800102** | **Vishruth P Reddy** | **<Signature>** |
| **PES1201800291** | **Ishan Agarwal** | **<Signature>** |

# ACKNOWLEDGEMENT

I would like to express my gratitude to Prof. Raghu B A, Department of Computer Science and Engineering, PES University, for his continuous guidance, assistance, and encouragement throughout the development of this UE18CS390A - Capstone Project Phase – 1.

I am grateful to the project coordinators, Prof. Silviya Nancy, and Prof. Vineetha B for organizing, managing, and helping with the entire process.

I take this opportunity to thank Dr. Shylaja S S, Chairperson, Department of Computer Science and Engineering, PES University, for all the knowledge and support I have received from the department. I would like to thank Dr. B.K. Keshavan, Dean of Faculty, PES University for his help.

I am deeply grateful to Dr. M. R. Doreswamy, Chancellor, PES University, Prof. Jawahar Doreswamy, Pro Chancellor – PES University, Dr. Suryaprasad J, Vice-Chancellor, PES University for providing to me various opportunities and enlightenment every step of the way. Finally, this project could not have been completed without the continual support and encouragement I have received from my family and friends.

# ABSTRACT

Our project comes under the domain of Precision Agriculture. It helps farmers make informed decisions with regards to the kind of crop they must invest in to get good returns. The aim of this project is to build a predictive model to recommend the most suitable crop to grow based on the various parameters that influence the fertility of the soil with the help of machine learning algorithms. A farmer or a horticulturist toils all day and keeps himself really busy throughout the year by looking after his crops and providing them with the right amount of water and nutrients expecting only one thing in return. And this is good crop yield. But if the farmer or the horticulturist makes a mistake in the very first step by choosing the wrong crop to cultivate, then the all the remaining steps in the process is absolutely useless. So, our main goal in this project is to give the farmer or the horticulturist a sensible start to their cropping season by helping them chose the right crop to grow in their lands so that they are able to obtain a very good yield at the end of the harvest season. The focus will be restricted to a very small part of India and we will be making use of the historic crop yield data for that particular region and try to incorporate machine learning algorithms that can ingest this data in order to build accurate models that could recommend the most ideal crop to be grown when fed with the properties associated with the soil and atmosphere of that region.

## TABLE OF CONTENT

##### Chapter No. Title Page No.

1. **INTRODUCTION**
2. **PROBLEM STATEMENT**
3. **LITERATURE REVIEW** 
   1. Keerthan Kumar, T.G., Shubha, C. and Sushma, S.A., **Random**

**Forest Algorithm for Soil Fertility Prediction and Grading**

**Using Machine Learning**. *International Journal of Innovative*

*Technology and Exploring Engineering (IJITEE), 2019*.

* 1. Kumar, R., Singh, M.P., Kumar, P. and Singh, J.P., 2015, May. **Crop**

**Selection Method to maximize crop yield rate using machine**

**learning technique**. In 2015 international conference on smart

technologies and management for computing, communication,

controls, energy and materials (ICSTM) (pp. 138-145). IEEE.

* 1. Palanivel, K. and Surianarayanan, C., 2019. **An approach for prediction  
     of crop yield using machine learning and big data techniques**.   
     International Journal of Computer Engineering and Technology,  
     10(3), pp.110-118.
  2. Patil, A., Kokate, S., Patil, P., Panpatil, V. and Sapkal, R., 2020. **Crop   
     Prediction using Machine Learning Algorithms**. International   
     Journal of Advancements in Engineering & Technology, 1(1), pp.1-8.

1. **PROJECT REQUIREMENTS SPECIFICATION**
2. **HIGH LEVEL DESIGN DOCUMENT**
3. **SYSTEM DESIGN**
4. **IMPLEMENTATION AND PSEUDOCODE**
5. **CONCLUSION OF CAPSTONE PROJECT PHASE-1**
6. **PLAN OF WORK FOR CAPSTONE PROJECT PHASE-2**

**REFERENCE/BIBLIOGRAPHY**

**APPENDIX A DEFINITIONS, ACRONYMS AND ABBREVIATIONS**

## CHAPTER 1

**INTRODUCTION**

Our project comes under the domain of Precision Agriculture. Therefore, it is very important to understand what Precision Agriculture means at its core. Precision Agriculture is a farming management system that is based on the use of modern technologies at every stage of the agriculture process. It is based on observation, measurement and response to the variability in crops. The aim of Precision Agriculture is to develop a Decision Support System for the management of a farm with the goal being maximization of returns with the efficient and judicious use of resources. In simpler terms, the goal is to increase the harvest through the efficient usage of seeds, fertilizers and pesticides. With the involvement of advanced technology, decisions are made purely based on data thereby reducing the risk of failure to a large extent as decisions are no longer made based on intuition and luck. And by making use of resources in an efficient way, we are saving the environment from the rapid depletion of its natural resources.

The idea of precision farming originated in the 1980s in the United States of America. Precision agriculture was the most important development of the third wave of modern agricultural revolutions. In the first agricultural revolution that took place between the 1900 to 1930, people saw a large increase in mechanized agriculture. This resulted in each farmer producing food that was more than sufficient to satisfy the hunger of 26 people. Later in the 1960s, people witnessed the second agricultural revolution in the form of Green Revolution which gave birth to genetic modification. At this stage, farmers were able to produce food that was more than sufficient to satisfy the hunger of 156 people. And finally, with rapid technological advancements and its integration in the field of agriculture gave birth to the third agricultural revolution which was widely known as Precision Agriculture. The others synonyms for Precision Agriculture include “Precision Farming”, “Satellite Farming” and “Site Specific Crop Management”. Input recommendation map for fertilizers was the first outcome in this domain and this was based on the grid soil sampling that was done. However, it was in its very early stages and was not practiced much. But with the advent of smartphones, high speed networks and enormous amount of satellite data, precision agriculture has become very popular and has seen a steep growth in the last 5 years.

As we all know, a field has heterogenous zones and with the aid of technology, we can identify these zones and manage their variability. Therefore, it is important to have some knowledge about the technologies being used. The GPS has been one of the most important enabling factors for the practice of precision farming. The farmer’s ability to precisely locate his/her position in the field led to the development of spatial variability maps for variables such as crop yield, topography, organic matter content, moisture levels, nitrogen levels, pH, EC, Mg, K and many more. Data similar to the ones mentioned in the previous sentence are extracted using sensor arrays mounted on GPS-equipped combine harvesters.

These sensors work in real time thereby collecting a wide spectrum of information ranging from chlorophyll levels to plan water status, along with multispectral imagery. This data is used in combination with satellite images through variable rate technology which include seeders, sprayers, etc. in order to achieve optimal distribution of resources. In order to avoid the possibility of an overlap or an underlap during the process of sowing seeds or the application of fertilizers and pesticides, onboard computers and GPS-navigators are used in vehicles. Digital maps and variable rate applications are being used for fields based on variable characteristics and the calculation of fertilizer dosage for every individual zone respectively. With the help of recent technological advancements, real time sensors placed in soil can wirelessly transmit data without the need of human intervention post sensor installment in the soil. For the remote monitoring of fields, drones and satellites have been put to use. Unmanned aerial vehicles are not very expensive and can be controlled by pilots without the need of any prior experience. These drones are equipped with multispectral cameras that can capture multiple images of the field. In addition to the traditional red, green and blue values, there are other values such as near infrared and red edge spectrum values in order to process and analyze vegetation indices such as NDVI maps. These drones also provide geographic information such as elevation which can be further used to build topography maps. These topographic maps are used to identify the correlation between crop health and the topography, the findings of which can be used to optimize crop inputs such as herbicides, growth regulators, fertilizers, water, etc. through variable rate applications.

A lot of sensors have been, predominantly wireless ones have been used in order to numerically capture the influence of field indicators such as moisture, pressure, rainfall, temperature, etc. And finally, with the help of applications that are hosted either as a mobile application or a web application, all the data that has been captured will be analyzed in order to obtain some meaningful insights that can be used to manage farms in an efficient manner.

Coming to the complexity involved in practicing precision farming, it is a little complex because most of the technologies that are being used are new and therefore requires a skilled workforce in order to make good use of this technology. For instance, a person who lacks the required skills will find it difficult to analyze a satellite image or repair an onboard computer. But at the same time, technological solutions that are simple do exist and can be accessed by every farmer. Some of these simple solutions include weather sensors, wireless modems, etc.

The next very important thing to look into is the cost involved in the incorporation precision farming in your farms or agricultural fields. At present, some of the sophisticated equipment and software are quite expensive and therefore the precision farming technologies are mostly seen only in large farm owned by affluent farmers.

But it is a well-known fact that with increasing developments in technology, it only becomes more affordable and easier to use. With this natural dynamic, the focus is on enabling easy access to these technologies with very little cost involved.

Stepping into the future, precision agriculture is something that will be unavoidable and it makes absolutely no sense in not making use of it because it is extremely profitable. In one of the blogs by the name onesoil.ai, we learnt that the American farmers are able to save on a large amount of money that they spend on agriculture through the incorporation of precision farming in their fields. To be precise, they save between 11 and 39 thousand dollars on an average per year. This blog also states that the farmers in Tanzania make use of their mobile phones to collect loans, arrange for contracts and to process other payments. The decisions may be based on decision-support systems that are based on big data. But ultimately it is the farmer’s call where is decision is influenced by business value and the impact created on the environment. But this too is being taken over by AI systems based on machine learning and ANNs.

Therefore, in conclusion, precision agriculture was witnessed for the very first time in the form of satellite images, weather prediction, variable rate fertilizer applications and crop health indicators. Its main goal is to optimize the field-level management with regard to:

1. **Crop Science**: establishing a connect between the farming practices followed and the crop needs
2. **Environmental Protection**: bringing down environmental risks and footprint of farming
3. **Economics**: Reducing overall expenditure through efficient usage of resources such as fertilizers.

With precision farming up and running, farmers will be able to

1. **Make improved decisions**
2. **Improve the inherent quality of the farm products**
3. **Enhance marketing of farm products**
4. **Improve relationships with landlords and local money lenders**

So, our focus in this project is to improve the decision making involved in the process of crop selection with the help of machine learning algorithms by taking into account the soil properties and the surrounding atmospheric conditions.

**CHAPTER 2**

**PROBLEM STATEMENT**

Our aim in this project is to build a recommendation system that recommends the most suitable crop to be grown given the properties of the soil and the surrounding atmospheric conditions of a specific location. We will be narrowing down our focus onto a specific part of India. The properties of the soil that are to be taken into consideration could include its nutrient contents and their quantities. The nutrients could include both macronutrients as well as micronutrients. Macronutrients are those nutrients that are produced by the soil in relatively larger quantities. Some examples of macronutrients include N, P, K, Ca, S, Mg, C, O, and H. On the other hand, micronutrients are those nutrients that are produced by the soil in relatively smaller quantities. Some examples of micronutrients include Fe, B, Cl, Mn, Zn, Cu, Mo, and Ni. These elements are present in the form of salts beneath the soil and are absorbed by the plants in the form of ions. Other soil properties include pH, EC, moisture, temperature, etc. The atmospheric conditions to be taken into consideration could include rainfall, temperature, humidity, etc. On collecting the required data, we plan to use machine learning algorithms in order to build robust models that can make forecasts with a very good accuracy level. Our project comes under the domain of precision farming. Precision farming practices can reduce the amount nutrients and other crop inputs used by a large extent while boosting the yield by a large margin. Farmers can thus obtain great returns on their investments and can also save big on fertilizer, pesticide, water and other resources. The next big advantage of practicing precision farming is the prevention of malefic impact on our environment by using the right quantity of chemicals and thereby conserving the quality of soil and ground water.

**CHAPTER 3**

**LITERATURE SURVEY**

**Paper1**: Keerthan Kumar, T.G., Shubha, C. and Sushma, S.A., **Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning**. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019*.

The authors of [1] have proposed a machine learning based solution for the analysis of imperative soil parameters and their influence on the kind of crops that could be suitably grown in a given soil. The various soil nutrients are treated as the independent variables and the grade of the soil is the target variable.  The regression algorithm along with Root Mean Square Error were employed to predict the rank of a soil and on applying a few classification algorithms for the purpose of crop recommendation, they found that Random Forest was the most accurate model.

A large portion of the Indian Economy is heavily dependent on the agriculture sector and therefore necessitates the need for accuracy in predicting the yield of crops much before the seeds are sown. In order to have a good yield, it is important that the soil is rich in the required nutrients. So the main goal in this project was to rank a soil sample by examining its nutrient contents (Macronutrients and Micronutrients) and then recommend the most suitable crop that could be grown in this soil.

**Module-1: Grading of Soil**

The contents of various soil nutrients such as EC, K, pH, Mn, Zn, S, P, and B are considered as the independent variables and the grade of the soil is considered as the target variable. So, a Multi-Variate Linear Regression model was built to predict the fertility of soil on a scale of 5.

A linear combination of the independent variables was chosen as the hypothesis function. The cost function chosen was:



Xi = vector of independent variables

hθ = hypothesis function

Yi = True value of the response variables

m = normalizing parameter

The Gradient Descent algorithms was adopted to minimize the cost function. Then hypothesis testing was carried out on the test dataset in order to check for the model’s correctness and efficiency and the Root Mean Square Error was used to determine accuracy of the model.

**Module-2: Crop Recommendation**

Machine learning algorithms such as Support Vector Machines, Random Forest Classification and Decision Tree were applied and based on the Root Mean Square Error value the best model was chosen. This model will then be subjected to evaluation where the accuracy of the model would be determined by passing real-time data.

A Random Forest Classifier is a supervised machine learning algorithm that builds multiple decision trees and combines them together in order to have improved accuracy and stability in the predictions made.  The most important feature is used to split a node and then recursively the next most important features is looked for from the subset of the remaining features thereby generating a highly accurate classifier with wide diversity.

In order to split a node, only a selective group of features are selected among all the features. An element of randomness is introduced through the use of random thresholds for the feature set. A Random Forest Classifier applies a technique known as bootstrap aggregation or bagging to the tree learners. From the training set, random sampling with replacement was performed and to each of these samples, trees were fit.

Then a voting is performed among all the predictions output by all the trees in order to arrive at the final result. To ensure that the variance is low and at the same time the bias is also kept low, the bootstrapping procedure was applied.  If the trees are not related to each other, then the average of the outputs produced by these trees are more robust to noise but in the case of a single tree, the prediction made can be very easily influenced by the noise.

Therefore, the idea behind using different samples from the training sets was to develop trees that are highly uncorrelated. The number of trees used in a random forest classifier is usually in the range of a few hundreds to several thousands and this number is heavily dependent on the size and nature of the training set.

**Conclusions from [1] are**

* In soil grading, Linear Regression was found to be the most efficient algorithm with a very less Root Mean Square Error value.
* In the case of crop recommendation, the Random Forest Classifier proved to perform better compared to Support Vector Machine and Gaussian Naive Bayes.

**Learnings from [1] are**

* The Random Forest Algorithm is based on ensemble learning and proved to be a very effective algorithm for classification.
* The basic idea is to build multiple decision trees from randomly selected subsets of the data. And then when a new data instance comes in, it is put through all these decision trees and a majority vote is taken in order to give the instance its final classification.
* Each tree as individual entities might not be ideal, but as a group they can perform really well.
* Since there are a large number of trees, the existence of any errors or uncertainties associated with any of the trees are taken care off by this algorithm.

**CHAPTER 4**

**PROJECT REQUIREMENTS SPECIFICATION**

# Introduction

Our project comes under the domain of Precision Agriculture. It helps farmers make informed decisions with regards to the kind of crop they must invest in to get good returns.

# Project Scope

The aim of this project is to build a predictive model to recommend the most suitable crop to grow based on the various parameters that influence the fertility of the soil.

This project enables the farmers to grow the most suitable crop by factoring in various soil characteristics like N, P, K contents and pH and atmospheric conditions like temperature, humidity and rainfall. This results in greater yield of crop and therefore, stabilizing their financial status.

In this project, the focus is on analysing the existing data and employing suitable models in order to give the best recommendations possible to the farmers. On the other hand, we will not be diving too deep into the implementation of how the data will be extracted but we will be researching about the methods used to collect the same. One of our data sources is only limited to 22 crops but we will make an effort to find more data in order to make this product more robust with regard to its recommendation power.

# Product Perspective

Usually, farmers and horticulturists don’t have a firm idea as to what is the best crop to be grown due to the limited knowledge of the soil parameters and the surrounding conditions. This often results in poor yield of crops which impacts the farmers financially thereby instigating the farmers to take extreme measures like committing suicide.

Therefore, our project is a return of favour to all those hard-working men and women who slog all day at the fields so that we are able to consume nutritious food.

# Product Features

Our product ingests parameters that describe the soil and its surrounding atmospheric conditions such as N, P, K, temperature, humidity, pH and rainfall as input and outputs the name of the most suitable crop that could be grown in order to achieve maximum yield and have a successful harvest season.

# Operating Environment

Our plan is to build a web application that provides a simple user interface for the farmers to interact with in order to make informed decisions with regards to crop selection. We will also be developing a mobile version of this web application in order to increase the usability of this application with all the hardware limitations of each person accounted for. This mobile adaptation will be for android devices only.

# General Constraints, Assumptions and Dependencies

* Good quality network connection to use the web application
* Good quality network connection to install the mobile application
* After the installation is complete, network connection is not required to make use of the application.
* A minimum of 4 GB of RAM is a must for the smooth functioning of the application.
* The mobile must have Android as its operating system.

# Risks

The data which we are utilizing must be from a reliable source as farmers will be investing their time, efforts and resources in growing the crop recommended by our model with the aim of maximizing their profits. We must also ensure that the mobile application is lightweight so that it can function efficiently even if there’s a fluctuation in the network connectivity.

# Functional Requirements

* After finishing with all the installation and setup, the user needs to input the soil and atmospheric parameters requested by the application.
* The application validates the parameters input by the user and raises an exception in case of an erroneous input. It then prompts the user to change the value and this continues until all the parameters are correctly input by the user.
* The application passes these values to the machine learning models and returns to the user the name and details of the crop that is most suitable to be grown based on the results obtained from the analytics.

# Hardware Requirements

A good quality network connection is necessary for using the web application and also for downloading the mobile application. However, on successfully installing the mobile application, network connection is no longer required to use the same. The mobile running the application is required to have a minimum of 4 GB of RAM and must have android as its operating system.

# Performance Requirement

1. Smartphone

* Android Operating System
* 4 GB of RAM (Minimum)
* 5.5 inch display (Minimum)
* Good quality network connection (Wi-Fi or Cellular Data)

1. PC

* Windows 10
* 4 GB of RAM (Minimum)
* Good quality network connection (Wi-Fi or Ethernet)
* Web Browser like Chrome, Firefox

# Security Requirements

The client will have to create an account in order to make use of our application. The client’s login credentials will be stored in an encrypted format and it will be made sure that no other user is able to compromise any other fellow user’s account.

**CHAPTER 5**

**HIGH LEVEL DESIGN DOCUMENT**

**CHAPTER 6**

**SYSTEM DESIGN**

**CHAPTER 7**

**IMPLEMENTATION AND PSEUDOCODE**

**CHAPTER 8**

**CONCLUSION OF CAPSTONE PROJECT PHASE-1**

**CHAPTER 9**

**PLAN OF WORK FOR CAPSTONE PROJECT PHASE-2**

**REFERENCE / BIBLIOGRAPHY**

**[1]** Keerthan Kumar, T.G., Shubha, C. and Sushma, S.A., **Random Forest Algorithm for Soil Fertility Prediction and Grading Using Machine Learning**. *International Journal of Innovative Technology and Exploring Engineering (IJITEE), 2019*.

**[2]** Kumar, R., Singh, M.P., Kumar, P. and Singh, J.P., 2015, May. **Crop Selection Method to maximize crop yield rate using machine learning technique**. In 2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM) (pp. 138-145). IEEE.

**[3]** Palanivel, K. and Surianarayanan, C., 2019. **An approach for prediction of crop yield using machine learning and big data techniques**. International Journal of Computer Engineering and Technology, 10(3), pp.110-118.

**[4]** Patil, A., Kokate, S., Patil, P., Panpatil, V. and Sapkal, R., 2020. **Crop Prediction using Machine Learning Algorithms**. International Journal of Advancements in Engineering & Technology, 1(1), pp.1-8.

**APPENDIX A DEFINITIONS, ACRONYMS AND ABBREVIATIONS**

GPS: Global Positioning System Mo: Molybdenum

pH: power of Hydrogen Ni: Nickel

EC: Electrical Conductivity

Mg: Magnesium

K: Potassium

NDVI: Normalized Difference Vegetation Index

AI: Artificial Intelligence

ANN: Artificial Neural Networks

N: Nitrogen

P: Phosphorus

K: Potassium

Ca: Calcium

S: Sulphur

Mg: Magnesium

C: Carbon

O: Oxygen

H: Hydrogen

Fe: Iron

B: Boron

Cl: Chlorine

Mn: Manganese

Zn: Zinc

Cu: Copper