Crop Recommendation System

A Major Project Synopsis Submitted to

Rajiv Gandhi Proudyogiki Vishwavidyalaya

Towards Partial Fulfillment for the Award of

Bachelor of Engineering in Computer Science Engineering

Submitted by:
Abhishek Sharma(0827CS201012)
Alokit Sharma(0827CS201023)
Ankita Arya (0827CS201034)
Asit Joshi (0827CS201045)

Guided by:

Prof. Juhi Shrivastava
Computer Science and Engineering

Acropolis Institute of Technology & Research, Indore

Jan - Jun 202

1. Abstract

In India agriculture holds an incredibly predominant position in the expansion of our country's financial system. It is one of the fields which generates most of the employment opportunity in our country. Farmers, due to lack of their knowledg e about different soil contents and environment conditions, do not opt the exact crop for nurturing, which results into a major hinder in crop production. To eliminate this barrier, we have provided a system which offers a scientific approach to assist farmers in predicting the ample crops to be cultivated based on different parameters which affects the overall production. It also suggests them about several deficiencies of nutrients in the soil to produce a specific crop. It is in context of a website. We used the crop dataset which include parameters like temperature, rainfall, pH, and humidity for specific crops and applied different ML techniques to recommend crops with high accuracy and efficiency. Hence, it can be supportive for farmers to be furthermore extra versatile.

2. Introduction of the Project

Foreign nations have started using and implementing modern methods for the purpose of their profit. They already have gained such an upper hand in applying scientific and technological methods in the field of agriculture and farming to increase its quality of work that one can only imagine. Whereas, India still is holding on to the traditional approach towards farming and its technologies. As, we know, that singularly agriculture alone generates a huge percentage of revenue for our country. The income come is pretty handy, when we talk about the Gross Domestic Product value. Moving forward towards globalization, the need for food has grown exponentially. Farmer tries to increase the quantity of their production by adding different artificial fertilizers, which eventually results in a future environmental harm. But if the farmer knows exactly which crop to be sown according to different soil contents and environment conditions, then this will minimize the loss and results in efficient crop production . We have gathered a dataset, which consist of information about rainfall, climatic conditions and different soil nutrients. This will provide us better understanding of trends of crop production in consideration of geographical and environment factors. Our system also predicts the shortage of any particular components for growing a specific crop. Our predictive system can prove to be a great boon for agricultural industry. The trouble of nutrient insufficiency in areas, which occurred due to the reason of planting incorrect crop at wrong period of time, is abolished with the help of our predictive system. This results in scaling down the production efficiency of farmers. More and more scientific approach towards agriculture industry, will definitely take it to greater heights. We propose this system to provide farmers, knowledge about requirements of different minerals and climatic conditions, which are suitable for producing certain crops. Also, our project diverts our focus towards the lack of different minerals required to grow some crops, and proposes us the remedies to eliminate their shortage. Our system considers factors such as soil composition and climatic factors like temperature, rainfall and humidity.

3. Objective

Data set collection from various sources. Data parsing and cleansing technique is applied to make the raw data into processing data. The data collected is subject to machine learning system along with run time analysis makes an efficient crop value updation system. Usage of Ensemble of classifiers makes the model more robust and efficient. Ranking technique used in the project helps us to make efficient decisions. Creating a web application for user registrations and collection of data. The main objective is to obtain a better variety of crops

that can be grown over the season. The proposed system would help to minimize the difficulties faced by farmers in choosing a crop and maximize the yield. The model predicts the crop yield by studying factors such as rainfall, temperature, area, season, soil type etc.

4. Scope

This Project can be used to get the student performance with more accuracy than any other model published earlier and we can also make some mobile or web application based on the model.

## 5. Study of Existing System

There aren't many systems that are dedicated to Crop Detection . Some features of existing systems are :

The computational and data demands of structural price forecasting generally far exceed than what is routinely available in developing countries. Consequently, researchers often rely on parsimonious representations of price processes for their forecasting needs. Contemporary parsimonious form of price forecasting relies heavily on time series modelling. In time series modelling, past observations of the same variable are collected and analyzed to develop a model describing the underlying relationship. During the past few decades, much effort has been devoted to the development and improvement of time series forecasting models. Time series modelling requires less onerous data input for regular and up-to date price forecasting. Hence there is a need for better classification which would be an ensemble or hybrid classification model.

DISADVANTAGES OF EXISTING SYSTEM
● Efficiency is low.
● The existing system which recommends crop yield is either hardware-based being costly to maintain, or not easily accessible.
● Despite many solutions that have been recently proposed, there are still open challenges in creating a user-friendly application with respect to crop recommendation.
● More number of repeated work.

6. Project Description
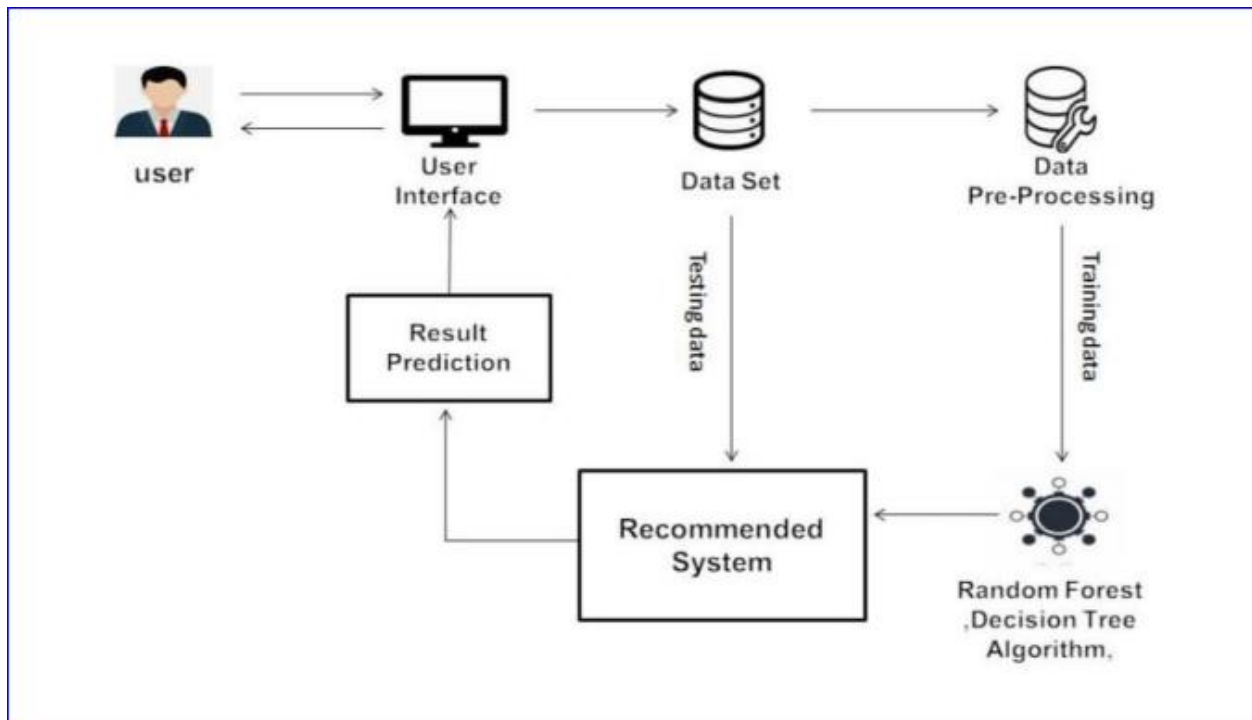
The data analysis technology is used to update the crop yield rate change. The concept of this paper is to implement the crop selection method so that this method helps in solving many agriculture and farmers problems. This improves our Indian economy by maximizing the yield rate of crop production. Different types of land condition. So the quality of the crops are identified using ranking process. By this process the rate of the low quality and high quality crop is also intimated. The usage of ensemble of classifiers paves a path way to make a better decision on predictions due to the usage of multiple classifiers. Further, a ranking process is applied for decision making in order to select the classifiers results. This system is used to predict the cost of the fertilizers for further. This project uses Ensemble of classifiers such as Decision tree and Random forest classifier. In addition, this project uses Ranking technique.

7. Expected Outcome

The project Handicrafts Online was developed by using Html, Sass, ReactJS, Stripe, Strapi and Javascript. The main aim of developing this project was to help the local artisans grow their businesses more and bring our traditional culture to light. Through our e-commerce website people can browse and view the local handicrafts made by our very own Indian Artisans and buy them at reasonable costs. We hope our project will provide a platform to all the local artists to display and sell their fine craftsmanship.

We think in today's world and in this economy it's very important to support the local craftsmen and admire their fine art and so this e-commerce website will not only help them but also help people buy their fine art sitting at home and ordering at their ease instead of going to shop for them specially.

8. Algorithm to be used

RANDOM FOREST ALGORITHM:

Random Forest is a popular machine learning algorithm that belongs to the supervised learning technique. It can be used for both Classification and Regression problems in ML. It is based on the concept of ensemble learning, which is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model. As the name suggests, "Random Forest is a classifier that contains a number of decision trees on various subsets of the given dataset and takes the average to improve the predictive accuracy of that dataset." Instead of relying on one decision tree, the random forest takes the prediction from each tree and based on the majority votes of predictions, and it predicts the final output. The greater number of trees in the forest leads to higher accuracy and prevents the problem of over fitting. Random Forest works in two-phase first is to create the random forest by combining N decision tree, and second is to make predictions for each tree created in the first phase.

The Working process can be explained in the below steps:

Step-1: Select random K data points from the training set.

Step-2: Build the decision trees associated with the selected data points.

Step-3: Choose the number N for decision trees that you want to build.

Step-4: Repeat Step 1 & 2.

Step-5: For new data points, find the predictions of each decision tree, and assign the new data points to the category that wins the majority votes.
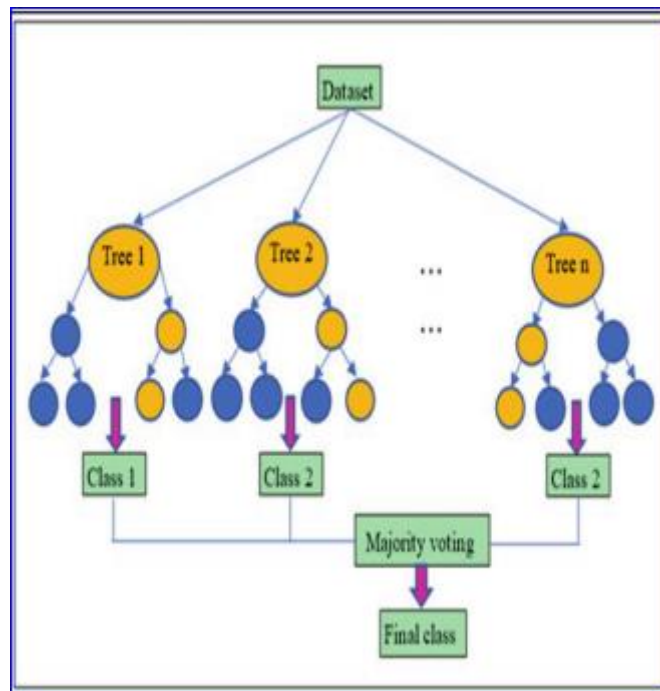
DECISION TREE:

Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems. It is a tree-structured classifier, where internal nodes 18 represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome. In a Decision tree, there are two nodes, which are the Decision Node and Leaf Node. Decision nodes are used to make any decision and have multiple branches, whereas Leaf nodes are the output of those decisions and do not contain any further branches. The decisions or the test are performed on the basis of features of the given dataset. It is a graphical representation for getting all the possible solutions to a problem/decision based on given conditions. It is called a decision tree because, similar to a tree, it starts with the root node, which expands on further branches and constructs a tree-like structure. In order to build a tree, we use the CART algorithm, which stands for Classification and Regression Tree algorithm. A decision tree simply asks a question, and based on the answer (Yes/No), it further split the tree into subtrees.

The complete process can be better understood using the below algorithm:
Step-1: Begin the tree with the root node, says S, which contains the complete dataset.
Step-2: Find the best attribute in the dataset using Attribute Selection Measure (ASM).
**Step-3:** Divide the S into subsets that contains possible values for the best attributes.
Step-4: Generate the decision tree node, which contains the best attribute.
Step-5: Recursively make new decision trees using the subsets of the dataset created in step - 3. Continue this process until a stage is reached where you cannot further classify the nodes and called the final node as a leaf node.

Support Vector Machine

SVM is a method of ML that generates an optimal hyperplane or decision boundary that can segregate dimensional spaces into classes, for the purpose of putting the new data in correct

category in future. With the help of support vectors, we create hyperplane. The hyperplane thus generated has two support vectors each on either side of the hyperplane. The support vectors are nothing but lines that is drawn passing through two data points on either side, which is closest to hyperplane. The accuracy of this model is about 96.08 %. Thus, this model turns out to be more accurate than Random tree algorithm. The "Fig.7" is depiction of SVM classifier.
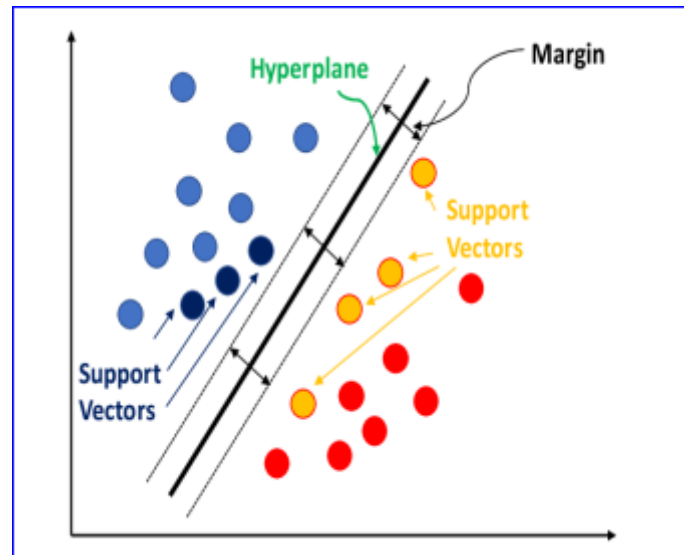
Logical Regression

Logical regression is the one of the ML algorithm whose approach is to model bond between dependent along with independent variables. It is used to solve categorical data It is straightforward to implement and very fundamental and efficient model. The accuracy of logical regression model is 95.22 %. Its accuracy is better than random tree algorithm, but is lower than SVM. Thus, this model is discarded.

Python:

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

● `Python is Interpreted` – `Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.`

● `Python is Interactive` – `You can actually sit at a Python prompt and interact with the interpreter directly to write your programs.`

● `Python is Object-Oriented` – `Python supports Object-Oriented style or technique of programming that encapsulates code within objects.`

● `Python is a Beginner's Language` – `Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.`

Python Features: `Python's features include –`
● `Easy-to-learn` – `Python has few keywords, simple structure, and a`

clearly defined syntax. This allows the student to pick up the language quickly.
- Easy-to-read – Python code is more clearly defined and visible to the eyes.
- Easy-to-maintain – Python's source code is fairly easy-to-maintain.
- A broad standard library – Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
- Interactive Mode – Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
- Portable – Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
- Extendable – You can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
- Databases – Python provides interfaces to all major commercial databases.
- GUI Programming – Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
- Scalable – Python provides a better structure and support for large programs than shell scripting.

Flask:

Flask is a micro web framework written in Python. It is classified as a microframework because it does not require particular tools or librariesIt has no database abstraction layer, form validation, or any other components where pre-existing third-party libraries provide common functions. However, Flask supports extensions that can add application features as if they were implemented in Flask itself. Extensions exist for object-relational mappers, form validation, upload handling, various open authentication technologies and several common framework related tools.

9. **Conclusion**

Agriculture is key to our country's economy, so even the smallest investment done in this sector has a tremendous effect on our country altogether. Therefore we need to be more serious about it. Due to the lack of scientific knowledge about different factors affecting crops, the farmers of our country tend to face a lot of challenges in selecting right crops to grow. Hence, face a loss in their profit, due to less productivity. But our system will provide them with a ray of hope to grow crops which will earn them at most profit. The quality as well as quantity of their production will increase exponentially. Also it will also help them in maintaining nutrients content in the soil. Both the quantity and quality will be increased

10.   References

1. https://www.irjet.net/archives/V9/i11/IRJET-V9I1110.pdf

2. https://www.tcetmumbai.in/IT/HME-HSD/PBL/PBL%20REPORT%2020-21.pdf

3. https://www.theseus.fi/bitstream/handle/10024/159502/Thesis.pdf?sequence=2&isAllowed=y

4. Manpreet Kaur, Heena Gulati, Harish Kundra, "Data Mining in Agriculture on Crop Price Prediction: Techniques and Applications", International Journal of Computer Applications, Volume 99– No.12, August 2014.

5. J. Meng, "Research on the cost of agricultural products circulation and its control under the new normal economic development," Commercial Times, no. 23, pp. 145147, 2016.