Electrical and Automation Engineering

Vol: 2(1), 2023

REST Publisher; ISBN: 978-81-956353-5-1

Website: http://restpublisher.com/book-series/eae/

DOI: https://doi.org/10.46632/eae/2/1/5

A Comprehensive Study on Crop Recommendation System for Precision Agriculture Using Machine Learning Algorithms

*** Keerthi Reddy Vutukuru, Varshini Vanmalli, N. Gireesh, M. Venkata Naresh**

Sree Vidyanikethan Engineering College, Tirupati, Andhra Pradesh, India.

*Corresponding Author Email: vkeerthireddy2002@gmail.com

Abstract. *Crop recommendation is among the most crucial components of the field known as precision agriculture. The crop recommendations were formulated after considering a wide range of distinct considerations. The term "precision agriculture" refers to a method of contemporary farming that makes use of information on soil features, soil kinds, and other factors, such as crop yields, and weather conditions to provide farmers with recommendations regarding the types of crops that would be most beneficial to grow on their farms to achieve the highest possible levels of both yield and profit. The crop recommendation system is currently being built with the use of machine learning methods including Random Forest, Gradient Boosting, XG Boost, Light GBM, SVM, and decision tree. The process of identifying the best crop to produce is aided by the information that consists of qualities such as potassium (K), phosphorous (P), and nitrogen (N) as well as temperature, humidity, pH, and rainfall. This application will be of assistance to farmers in enhancing agricultural output, minimizing the deterioration of soil on the cultivated ground, reducing the number of chemicals used in agricultural production, and increasing the productivity of cultivated land with available water resources. Also, this application recommends the top 5 prioritized crops to farmers. XG Boost achieved the best accuracy of 99.500 out of all these classifiers that were utilized in this paper. It also achieved the highest recall score of 99.545455, a precision score of 99.564935, and an F1 score of 99.545669.*

Keywords: *Precision agriculture, Dataset, Machine learning algorithms.*

1. INTRODUCTION

In India, farming is not seen as a form of commercial enterprise, but it does have a significant influence on the sociallives of the people who are engaged with it. Agriculture is important to a substantial number of people, either through direct or indirect means. Intuition as well as other extraneous considerations, such as generating fast profits, a lack ofinformation regarding market needs, exaggerating the ability of a soil to support a certain crop, and so forth, tend to muddle the decision that a farmer makes regarding which crop to cultivate. The farmer can put a huge burden on his family's financial situation if he makes a move that is extremely foolish and goes against his better judgement. Perhaps this is one of the numerous factors that contribute to the many incidents of farmer suicide that are reported inthe media daily. If so, then this might be one of the causes. In a nation like India, where farming and the industries that are closely tied to it contribute about 20.4% of the total value added to the product (GVA), An inaccurate assessment of the circumstances could lead to a detrimental impact not just on the farmer's family, but additionally on the whole economy of the region. Because of this, we consider the predicament that a farmer has when deciding whatkind of crop to cultivate during a given season to be a highly significant one. The development of a system that may offer Indian farmers predictive insights and assist them in making an educated choice regarding which crop to cultivateis an urgent task that has to be completed as soon as possible. Bearing this in view, we are inclined to suggest a system,an intelligent system, that would take into account environmental characteristics such as rainfall, temperature, and humidity as well as soil characteristics (pH value and nutrient concentration) before advising the user on the crop thatwould be best suited to their needs. Our project's overriding goal is to alleviate the challenges that are currently beingexperienced by farmers as well as make it possible for those farmers to cultivate the appropriate crop while avoidingundesirable outcomes through the application of efficient solutions derived from machine learning methods. Many different strategies for crop recommendation are discussed. The following is a list of the key contributions made by the approach that is being presented:

- Provide suggestions to the farmer about the types of crops that would work best in the region so that he does not suffer any financial losses.

- Based on the data from the prior year(s), provide a study of the profitability of the various crops.
- To reduce the possibility of people making incorrect assumptions about the crops.
- To eliminate the risk factors associated with producing crops that are not suitable for production and to automate the process of crop recommendation.

The following is the organizational structure of the research work that has been done: The work is broken up into five sections: Section 2, which discusses previous studies in the same field; Section 3, which gives a comprehensive description of the techniques provided; The explanation of the outcomes may be found in Section 4, while the concluding and potential future applications are presented in Section 5.

2. LITERATURE REVIEW

The crop suggestion technique for the formers is based on support vector machines (SVM) was developed via collaborative efforts by V. Yalavigi, A. Aravatagimath, and D. Modi (2021). In the study that they did [1], they concluded that it is important to analyse the profit that can be made from a certain crop, as this may help reduce the amount of money lost by farmers and boost overall output. The support vector machine, or SVM, approach is utilized here for classification purposes. This allows for the categorization of the many properties of the soil and the prediction of the crop that would be most suited. Anaconda Navigator is used to doing a simulation of the proposed algorithm tostudy the properties of the soil and make appropriate crop recommendations. To classify data, the SVM method is being examined. To determine whether or not the suggested method is effective, an accuracy and confusion matrix will be produced. P.K. Ramesh, S.M.Pande, B.R. Aiswarya, A. Anmol, K. Shaurya and K. Rohilla are the individuals responsible for this work (2021), The authors suggest that farmers may benefit from a yield forecast system that is both practical andeasy to use. A smart phone application would be used to link farmers as part of the system that is being suggested. The position of the user may be determined through the use of GPS. The user supplies the area and the kind of soil as inputs, and the most lucrative crop list is selected by the algorithms using machine learning, or the production of a crop is predicted based on the crop that the user chooses. In this case [2], Predictions of agricultural yields are made using a number of different machine learning techniques, such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Multivariate Linear Regression (MLR), K-Nearest Neighbor (KNN) and Random Forest (RF). G. Chauhan and A. Chaudhary (2021) suggested the crops for increasing agricultural production depending on the soil, weather, humidity, and rainfall, as well as other factors. It's good for the economy generally and farmers specifically, and it helps keep food prices down. Using machine learning techniques like Decision Tree and Random Forest, this article [3] demonstrates how to determine which crop would thrive in a given soil type based on the available data sets. Shubha C, Keerthan Kumar T G and Sushma S (2019) offered a method for the purpose of evaluating the soil's gradeand determining which crops would thrive there based on its basic qualities. Specific soil quality as determined by itsnutrient content is the goal, whereas feature variables include the rich soil nutrients (e.g., electrical conductivity, pH,hydrogen power, organic carbon, etc.). EC stands for electrical conductivity, pH stands for the power of hydrogen, and OC stands for organic carbon. In this [4] study, the dataset is pre-processed, to forecast the soil quality, a regression technique is utilized, and the RMSE (Root Mean Square Error) is calculated. Additionally, we utilized several classification algorithms to determine the best crop to recommend, and we discovered that in regards to accuracy, Random Forest is the best option. Many studies that are recommended for agriculture development have been conducted by M. R. Kumar and D. J. Reddy (2021), and the objective was to develop a reliable method for classifying crops. Weather-based agricultural production estimation, disease-based crop classification, and growth-stage-based crop categorization are just a few examples. The authors of this paper [5] used several ML approaches were used in crop production estimation, and a comprehensive study of their accuracy was provided.

3. PROBLEM STATEMENT

According to the findings of a recent study, crop yield predictions in the past were determined by taking into accountthe specific crops and levels of cultivation experience of individual farmers. But currently, as a consequence of unforeseen situations such as unseasonal rainfall, the employment of different pesticides, and climatic conditions, the projections made by the farmers are not correct, which results in decreased crop yields. This is recognized as a significant challenge facing the agriculture industry at present and has to be addressed. Because of this, a crop recommendation system for precision agriculture that uses a machine learning model was developed and implementedas a result. When it comes to contemporary farming, precision agriculture is all about using data regarding the characteristics of the soil, the types of soil, crop yield data, and weather conditions to provide farmers with recommendations regarding the types of crops that would be most beneficial to grow on their farms to achieve the highest possible levels of both yield and profit. This method can help lower the number of crop failures, which in turnwill assist farmers in making more educated decisions about their agricultural plan.

4. PROPOSED MODEL

In this paper, excellent crop recommendations were developed through the application of ML (machine learning) classifiers. Random Forest, Decision Tree, Gradient Boost, XG Boost, Support Vector Machine, and Light GBM arethe machine learning algorithms that are currently available. A comparative performance study of several machine learning models is examined, and the crop is advised. Figure 1 depicts the overall plan for the technique that has been described here. The following procedures were carried out during the course of putting the project into action to accomplish the desired outcomes.

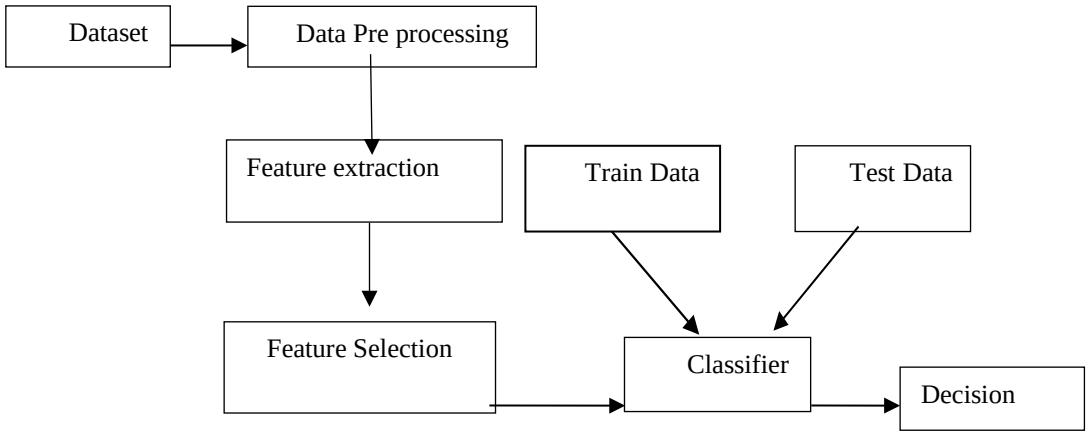

FIGURE 1. Representing Block diagram for the implementation

Dataset Selection

The selection of a dataset is the first stage through this method of putting the style into action. Dataset that we utilizedfor this project is a statistical dataset containing structured data that was acquired from Kaggle. This dataset was used in this project. The full extent of the dataset includes (2200, 7). The dataset [9] contains a variety of data variables, including those for nitrogen (N), phosphorus (P), potassium (K), pH, humidity, temperature, rainfall, and label.

Data Pre-Processing

Pre-processing data is essential as it prepares the raw data for use in ML (machine learning) classifiers and helps eliminate noise. This makes it an essential step. Verifying the reliability of the dataset we are working with is an essential step that must be taken first. The dataset must not be lacking any elements, and if it is, any missing values must be remedied by having the right values inserted in their place as soon as possible. In addition to this, the data has to be analyzed to determine whether or not its characteristics follow a normal distribution. It is necessary to get rid of the outliers. Checking the skew value of the features is necessary, and if the features are found to have skewness, the features in question need to be normalized through the use of transformations. We created bar graphs, scatter plots, box plots, and other similar visualizations to display the data and determine whether or not there are any patterns or trends that we can identify that will be useful when putting our idea into action.

5. FEATURE EXTRACTION AND SELECTION

We must narrow our choices down to those characteristics that are required to choose the kind of plant to cultivate. To do this, we came up with the idea of developing a correlation matrix, which displays the linear connection that each attribute has with the other features. Since the features aren't closely linked with one another, we can see in the matrix below that none of the features should be dropped. If features are strongly correlated with one another, then that feature must be lowered. Here is the correlation matrix for the characteristics in our dataset, which can be found here

Classifiers

The following step is to provide the classifier with the dataset that has previously been pre-processed, feature extracted, and selected. The classifier will divide the data into training and test sets. The data from the split sets is then tested and trained using the appropriate procedures associated with each classifier. The following is an explanation of the classifiers that are utilized in this undertaking. The classifiers that are used in this project are explained as follows:

FIGURE 2. Correlation matrix

Decision Tree

This is a form of supervised learning in which a tree structure is used to represent both attributes and corresponding labels. Decision tree training data is used to create a training model that may be used to forecast the category or valueof target variables based on learned decision rules derived from historical data. The decision tree may be viewed as consisting of two distinct parts: the decision nodes and the leaves. The leaves are the results or final product. Each node in the tree represents a test case for some attribute, and each edge from a node represents a possible answer to that test. This process repeats itself several times for each offspring tree whose origin is a newly created leaf node.

Random Forest

The Random Forest classifier takes the average of many decision trees that have been used to analyse various subsetsof the information that has been provided. This contributes to improving the dataset's ability to make accurate predictions. The random forest model does not rely on just one decision tree; rather, it takes into account the predictionmade by each tree in the forest and calculates the final output based on which tree's prediction won the majority of votes. This ensures that the model is as accurate as possible. When there are more trees in a forest, the level of precisionthat can be reached and the likelihood that the forest will not be over-treed are both increased.

Support Vector Machine (SVM)

The Support Vector Machine, more frequently referred to as SVM, is an algorithm or model for supervised machine learning that may be used to tackle classification issues as well as regression problems. Its name comes from the fact that it is commonly referred to as the "Support Vector Machine." But, the majority of the time we put it to use in classification competitions. SVM is typically portrayed as a collection of training data points scattered in an area thathas been partitioned into groups using an understandable gap that is as large as feasible. The value of each feature is determined by the value of a specific coordinate in the support vector machine (SVM) method. This means that each data item is represented as a point in an n-dimensional space. The next phase in the process of classification is to locate the hyperplane that most effectively divides the two categories from one another.

Gradient Boosting

One of the most widely used ensemble learning modeling approaches is boosting. It is used to construct robust classifiers by combining several less reliable classifiers. It starts by developing a primary model by using the many different training data sets that are accessible, and then it proceeds to discover the errors that are present in the primary model. When the error was found, a second model was built, and after it was complete, a third model was added to the procedure at that point. In this way, this process of introducing more models will continue until we acquire a complete training data set from which a model can accurately predict. This process will be repeated until we have a model that can predict correctly.

Light GBM

The Light GBM model is a gradient-boosting method that uses decision trees as its foundation. Its goal is to achieve more efficiency in the operation of the model while also reducing the amount of space that is required to run it. It does this by employing two cutting-edge methods: When compared to previous methods of boosting, LightGBM divides the tree not along its level splits but rather along its leaf levels. It decides to develop on the leaf with the greatest potential for delta loss. The level-wise technique has a higher loss, but the leaf-wise algorithm has a lesser loss since the leaf is fixed. If a tree is grown from the inside out, it may make the model more sophisticated, but it may also cause overfitting

in datasets with few observations.

The XG Boost classifier came out on top when compared to the other classifiers utilized in this study, achieving the best overall results. Below is an explanation of the XG Boost model, as well as the testing and training processes.

XG Boost

XG Boost is a package that facilitates optimally distributed gradient boosting that was created for the training of machine learning models in a manner that is both efficient and scalable. It belongs to the category of practices known as "ensemble learning" that combines the predictions of a number of less accurate models in order to get a more accurate prediction. It can handle enormous datasets and obtain reliable results in a variety of tasks related to machine learning, including classification and regression, which can be linked to its meteoric growth in popularity. In addition, it can handle large datasets with ease. XG Boost's ability to handle observations from the real world that contains the absence of data in an effective manner is one of its most notable qualities. This enables the algorithm to do so without requiring a substantial amount of pre-processing on the data.

Decision

The final element of the block diagram is titled "Decision," and it consists of nothing more than a recommendation regarding the crop that should be grown based on the results that were generated by the classifier.

6. RESULTS

The information presented herein is the experimental results of the crop recommendation system that implemented in this project. Results from all the methods, i.e., Random Forest, Decision Tree, SVM, Gradient Boost, XG Boost and Light GBM are compared, and a better model is detected based on the accuracy is given by those algorithms The performance here is nothing but the accuracy in predictions, and thus the accuracy for the various algorithms are checked. The accuracy of these classifiers is considered based on the Confusion

Matrix and Classification Report gave by that classifier after testing.

Model comparison

The comparison of the models can be accomplished by contrasting the various algorithms' precision, recall, F1 score, and accuracy values. The model comparison table for all of the classifiers that were involved in this study can be found in the table below.

TABLE 1. Model Comparison

Sl. No	Classifier	Accuracy	Precision	Recall	F1score
1	XG Boost	99.500	99.564935	99.545455	99.545669
2.	Random Forest	99.273	99.386364	99.318182	99.309901
3.	Light GBM	99.091	99.107143	99.090909	99.088460
4.	Gradient Boosting	99.318	98.700758	98.636364	98.622704
5.	Decision Tree	98.864	98.231372	98.181818	98.170455

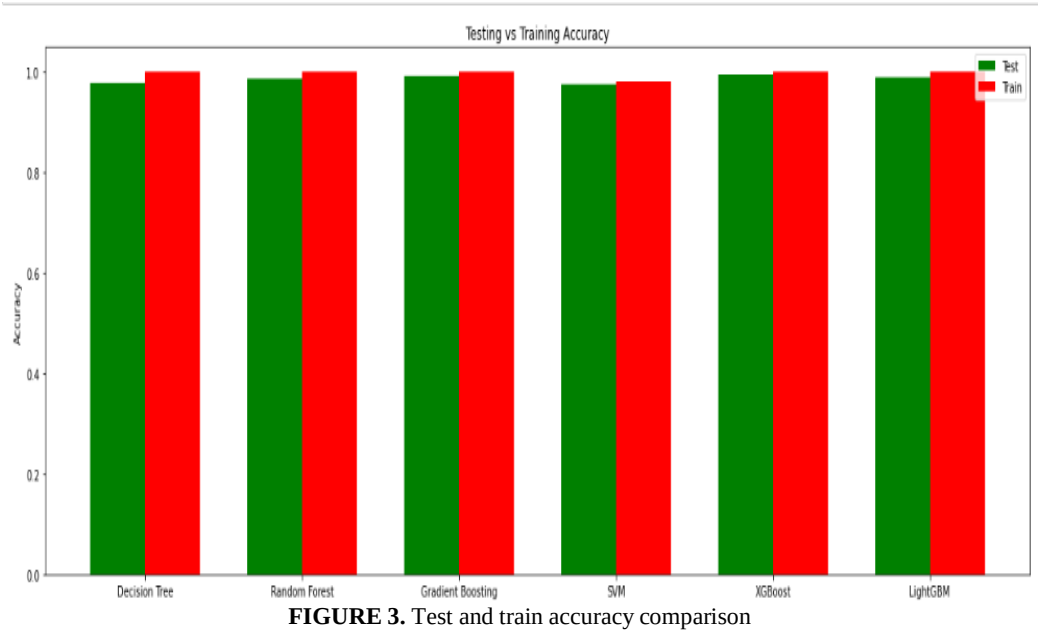

FIGURE 3. Test and train accuracy comparison

It can be seen from the table and the figure that the XG Boost algorithm provides the highest accuracy, precision, and recall scores compared to the other algorithms. The F1 score it generates is also the highest. Because of this, the XG Boost algorithm is used for the improved crop recommendation model.
The below are the results of XG Boost classifier

```
new_data_scaled = scaler.transform(a)
prediction = xgb_model.predict(new_data_scaled)
prediction=le.inverse_transform(prediction)

print("The Suggested Crop for Given Climatic Condition is:", prediction)
```

```
The Suggested Crop for Given Climatic Condition is: ['coconut']
```

FIGURE 4. Recommended Crop by XG Boost

Prioritized Recommendation of Crops

Instead of proposing a single crop, the algorithm that we built advises the top 5 priority crops so that farmers have more freedom in their decision making.

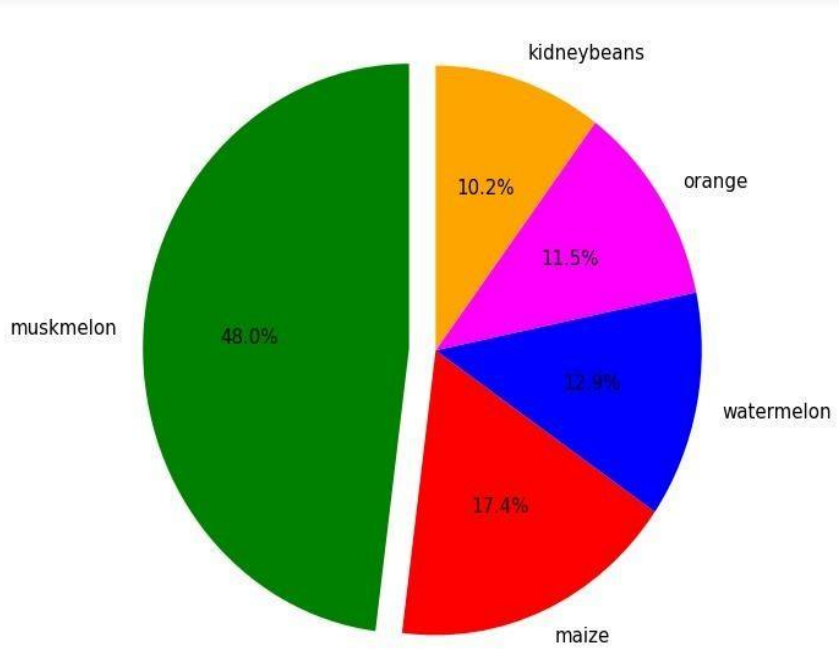

FIGURE 5. Pie chart representing Top 5 prioritized Crops

7. CONCLUSION AND FUTURE SCOPE

This paper and our implementation support the farmer in picking the suitable crop by delivering insights that are not generally kept track of by normal farmers; as a consequence, the likelihood of crop failure is lowered and total productivity is raised. Also, it avoids them from incurring losses, and additionally, this model advises the top 5 crops in order of priority, so that the farmers may make a flexible option without being dependent on a single crop that is recommended. By the use of the XGBoost, our final numbers for accuracy, precision, recall, and F1 score all came in at 99.5, 99.5475, 99.54935, and 99.545669, respectively. The initiative provides a lot of potential for conducting forward-looking tests. In the future, the project may be carried out via the website. Due to its high degree of adaptability in terms of growth, the project will be able to undergo revisions in the not-too-distant future whenever there is a need for doing so. By employing image processing and providing photographs as input, it is possible to identify the illnesses that affect plants. a quicker and more accurate suggestion of crops for the farmers, immediately. Using more advanced forms of

machine learning algorithms to produce more accurate results from machine learning algorithms than is currently possible. To make a more informed crop recommendation to farmers, additional factors like texture, permeability, drainage, water holding capacity, soil colour, and erosion should also be taken into account.

REFERENCES

[1]. D. Modi., A. V. Sutagundar., V. Yalavigi and A. Aravatagimath, "Crop Recommendation Using Machine Learning Algorithm" in 2021 5th International Conference on Information Systems and Computer Networks (ISCON),2021, pp.15,doi:10.1109/ISCON52037.2021.9702392.

[2]. S. M. PANDE., P. K. RAMESH., A. ANMOL., B. R. AISHWARYA., K. ROHILLA and K. SHAURYA, "Crop Recommender System Using Machine Learning Approach" in 2021 5th International Conference on Computing Methodologies and Communication (ICCMC),2021, pp.10661071, doi:10.1109/ICCMC51019.2021.941835.

[3]. G.Chauhan and A. Chaudhary, "Crop Recommendation System using Machine Learning Algorithms" in 2021 10th International Conference on System Modeling Advancement in Research Trends (SMART), 2021, pp.109112,doi:10.1109/SMART52563.2021.9676210.

[4]. Keerthan Kumar T G., Shubha C., Sushma S A, "Random Forest Algorithm for Soil Fertility Prediction and Grading using Machine Learning" International Journal of Innovative Technology and Exploring Engineering (IJITEE)ISSN: 2278-3075, Volume-9 Issue-1, November 2019.

[5]. D. J. Reddy and M. R. Kumar, "Crop Yield Prediction using Machine Learning Algorithm" in 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS),2021, pp.14661470, doi:10.1109/ICICCS51141.2021.9432236.

[6]. Z. Doshi., S. Nadkarni., R. Agrawal and N. Shah, "AgroConsultant: Intelligent Crop Recommendation System Using Machine Learning Algorithms" in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697349 Pune, India.

[7]. S.Pudumalar., E.Ramanujam., R.Harine Rajashree., C.Kavya.,T.Kiruthika and J.Nisha, "Crop Recommendation System for Precision Agriculture" in 2016 Eighth International Conference on Advanced Computing (ICoAC), January 2017, DOI:10.1109/ICoAC.2017.7951740

[8]. Tom M. Mitchell, Machine Learning, India Edition 2013, McGraw-Hill Education.

[9]. https://www.kaggle.com/datasets/atharvaingle/croprecommendation-dataset

