WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

# Crop Recommendation through Machine Learning

*Mr. S. Vinod Kumar, V. Sai Surya Teja,*

GMR Institute of Technology, Rajam, India.
20341A05I2@gmrit.edu.in

ABSTRACT:

Agriculture is called the backbone of Indian economy. Agriculture sector serves as the source of raw material for non-Agricultural sectors. While agriculture's share in India's economy has progressively declined in the recent years to less than 15% due to the high growth rates of the industrial and services sectors. The population has also been increasing day to day and also the agriculture is not able to meet the demanded requirements for the increasing population. Previously, crop cultivation was undertaken based on farmer's hands-on expertise. However, the climate change has started to affect crop yields badly. Consequently, Farmers are unable to choose the right crop based on the soil and environmental factors. The research aims to solve the problem of crop prediction more effectively to ensure farmers' incomes and to increase the production effectively. Crop prediction is based on the soil, geographic and climatic attributes. Predicting suitable crop for cultivation is an essential part of agriculture, with machine learning algorithms playing a major role in such prediction in recent years.There are three common machine learning techniques: supervised, unsupervised, and reinforcement learning. This work uses supervised learning classification techniques for prediction. The principal contribution of this work is to find the best feature selection technique, with a classification method, to predict the most suitable crop for cultivation, based on factors such as soil and environment.

**Keywords:** Machine Learning, Agriculture, Crop Recommendation (or) prediction, K-Nearest Neighbor and Naive Bayes Random Tree, SVM, Bagging, Feature Selection.

## INTRODUCTION:

More than 60% of the land in the country is used for agriculture in order to meet the needs of 1.3 billion people Thus adopting new agriculture technologies is very important. This will lead the farmers of our country towards profit. Prior crop prediction and yield prediction was performed on the basis of farmers experience on a particular location. They will prefer the prior or neighborhood or more trend crop in the surrounding region only for their land and they don't have enough of knowledge about soil nutrients content such as nitrogen, phosphorus, potassium in the land. This study is to recommend the most suitable crop for particular land at an instant considering all the problem into account.

Crop prediction depends on geography of a region (e.g. hill area, river ground, depth region), weather condition (e.g. temperature, cloud, rainfall, humidity), soil type (e.g. sandy, silty, clay, peaty, saline soil), soil composition (e.g. PH value, nitrogen, phosphate, potassium, organic carbon, calcium, magnesium, sulfur , manganese, copper, iron) and harvesting methods. This study involves about different preprocessing techniques , and applying best suitable algorithm for crop prediction .The preprocessing techniques involved in this study are data cleaning(includes handling missing values),feature selection .Three feature selection methods – filter, wrapper, and embedded are used in the selection of attributes. Filter methods offer rapid execution, though wrapper methods have a better recognition rate. In this study, wrapper feature selection techniques are used to select the best attributes from the dataset, and classification to predict the most suitable crop for a particular piece of land using the selected attributes. Different classifier algorithms such as SVM(Support Vector Machine),K-NN(K-Nearest Neighbors),Decision Tree, Random Forest ,Gradient Boosted Decision Tree, Regularized Greedy Forest are studied in this paper.

## LITERATURE SURVEY:

***[1] Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1), 117-140.***

In this paper crop prediction is done through selecting best feature selecting methods among wrapper methods and selecting a best classifier algorithms in supervised learning based on accuracy results. Feature selection methods used in this paper are wrapper methods RFE,BORUTA,SFFS. The crop prediction is done through soil characteristics ,environmental characteristics. In the experimental analysis of this paper RFE with bagging classifier on KNN, Naïve bayes, decision Tree, SVM, Random Forest (accuracy after reduction 0.9272 ) outperforms the other combinations .

***[2] Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. (pp. 1-6). IEEE.***

In this paper crop prediction done through two subsystems. One subsystem is fundamentally concerned with crop recommendation. In the first subsystem preprocessing and selection of different machine learning algorithms is done based on accuracies. Other Sub-system predicts the rainfall output of this subsystem is fed to sub-system – 1. In the experimental analysis decision tree classifier gives best accuracy among other classifiers. Random Forest has got highest accuracy (ie.90.43)

***[3] S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625-23641, 2022.***

This paper deals about the various feature selection methods in wrapper methods and classification algorithms to predict a crop. The dataset used in this paper is the Felin dataset. In this paper there is a systematic approach of process. Dataset Collection - > Pre-processing - > Feature Selection - > Classifiers. Random forest classifier(87.43) gave the best accuracy and performance metrics compared to other on felin dataset. Modified recursive feature elimination with random forest, performance metrics were at high level.

***[4] Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). 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) (pp. 114-119). IEEE.***

In this paper objective is to design a recommendation system for accurate crop selection based on the various soil, rainfall and surface temperature parameters. To improve crop productivity by providing predictions of high accuracy and efficiency through the ensemble technique. The collected data is initially subjected to preprocessing. Post dataset preprocessing, the dataset is divided into training set and test set samples. Each of the sample is trained and tested on the Random Forest, Naive Bayes and the Linear SVM algorithms. Voting Technique has been used as the combination method to provide the best accuracy. The average accuracy of crop classification into Kharif and Rabi crops is 99.91%. s.

***[5] Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.***

This paper proposes a system that uses the voting method to build an efficient and accurate model. Recommendation of crops is dependent on various parameters. Precision agriculture aims in identifying these parameters. Ensembling is one such technique that is included in such research works. The rules generated from the ensemble model is used to develop a RECOMMENDATION SYSTEM to test on the testing set. The tree to rules operator is used to induce rules directly from the CHAID and random tree. The prediction accuracy of model accounts to 88%.

***[6] PANDE, S. M., RAMESH, P. K., ANMOL, A., Aishwarya, B. R., ROHILLA, K., & SHAURYA, K. (2021, April). Crop recommender system using machine learning approach. (pp. 1066-1071),IEEE Xplore.***

In this paper Prediction of the crop for specific regions by executing various Machine Learning algorithms, with a comparison of error rate and accuracy. In this paper also discusses a GPS based location identifier to retrieve the rainfall estimation at the given area.

***[7] Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].***

The paper aims to solve the crucial problem of selecting the classifiers for the ensemble learning. A method to select a best classifier set from a pool of classifiers has been proposed. The proposal aims to achieve higher accuracy and performance. A method called SAD was proposed based on accuracy and classification performance. Using Q statistics, the dependency between most relevant and accurate classifiers is identified. The classifiers which were not chosen were combined to form the ensemble. This measure is supposed to ensure higher performance and diversity of the ensemble. Various methods such as SA (Selection by Accuracy), SAD (Selection by accuracy and Diversity) and NS (No selection) algorithm were identified.

## METHODOLOGY:

### *Data Collection:*

Data collection is the first step in creating a machine learning model. Collection of data involves pooling data by scraping, capturing and loading it from multiple sources.Data collection allows you to capture a record of past events so that we can use data analysis to find recurring patterns. From those patterns, you build predictive models using machine learning algorithms that look for trends and predict future Changes. Predictive models are only as good as the data from which they are built, so good data collection practices are crucial to developing high-performing models. The data needs to be error-free and contain relevant information for the task at hand.

### *Preprocessing:*

Preprocessing of data is important step in machine learning. The Preprocessing is used to convert the raw data into useful and efficient format. Preprocessing involves three techniques . i)Data Cleaning ii)Data Transformation iii)Data Reduction .

**i)Data Cleaning:**

Data Cleaning is process of handling the missing data, noisy data. Coming to missing data handling if there are multiple missing values within a tuple ignore the tuples else Fill the Missing values with appropriate value (by attribute mean or most probable value). Coming to the Noisy Data it cannot be interpreted by machines, It is generated due to faulty data collection, data entry errors etc. It can be handled through Binning method, Regression or Clustering.

**ii)Data Transformation:**

Data Transformation is a method used to normalize the range of independent variables or features of data. Scaling the features makes the flow of gradient descent smooth and helps algorithms quickly reach the minima of the cost function. There are two techniques standardization and min-max scaling.

**iii)Data Reduction:**

### *Feature Selection:*

Datasets contain redundant information that harms the classification task. Feature selection is a major task in data analytic research, where datasets have a large number of attributes .Feature selection is selection of required attributes from the total attributes. It allows the machine learning algorithm to train faster, it decreases the complexity of the model, it makes interpretation easier. It also maximizes the model's accuracy when choosing the right subset, and prevents overfitting. Three types of feature selection methods are used in the selection of attributes: filter, wrapper, and embedded.

**1)Wrapper Feature selection :**

Wrappers require some method to search the space of all possible subsets of features, assessing their quality by learning and evaluating a classifier with that feature subset. The feature selection process is based on a specific machine learning algorithm that we are trying to fit on a given dataset. It follows a greedy search approach by evaluating all the possible combinations of features against the evaluation criterion. The wrapper methods usually result in better predictive accuracy than filter methods. So in this work wrapper methods are used.
Wrapper Feature Selection methods
i)SFFS(Sequential Forward Feature Selection)
ii) RFE (Recursive Feature Elimination)

**i)Sequential Forward Feature Selection:**

This is an iterative method wherein we start with the best performing variable against the target. Next, we select another variable that gives the best performance in combination with the first selected variable. This process continues until the preset criterion is achieved.

**ii)Recursive Feature Elimination:**

Given an external estimator that assigns weights to features (e.g., the coefficients of a linear model), the goal of recursive feature elimination (RFE) is to select features by recursively considering smaller and smaller sets of features. First, the estimator is trained on the initial set of features and the importance of each feature is obtained either through a coeff attribute or through a feature importance attribute. Then, the least important features are pruned from the current set of features. That procedure is recursively repeated on the pruned set until the desired number of features to select is eventually reached.

## CLASSIFICATION TECHNIQUES

### *Support Vector Machine:*

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane.

### *Naive Bayes:*

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.

### *Decision Tree:*

A Decision tree is a flowchart-like tree structure, where each internal node denotes a test on an attribute, each branch represents an outcome of the test (different decisions), and each leaf node (terminal node) holds a class label.
While implementing a Decision tree, the main issue arises that how to select the best attribute for the root node and for sub-nodes. So, to solve such problems there is a technique which is called as Attribute selection measure or ASM. By this measurement, we can easily select the best attribute for the nodes of the tree. There are two popular techniques for ASM, which are Gini Index, Information gain.

### *Bagging:*

A Bagging classifier is an ensemble meta-estimator that fits base classifiers each on random subsets of the original dataset and then aggregate their individual predictions (either by voting or by averaging) to form a final prediction. Further, Bagging takes votes for each sample to improve the performance of prediction.

### *SubSystem-2 (for Rainfall prediction)*

Each and every crop has its own rainfall requirement ,If this requirement is not met ,the crop yield will suffer .On the other hand ,if surplus rainfall is available the yield may again undergo negative consequences .Hence Rainfall is a very important factor for the growth of any crop. For this reason, we decided to implement this sub-system, which predicts the rainfall (in mm) (or else) place some GPS sensors which monitor the rainfall of the specific region. However through prediction we need to collect rainfall data over past few days historical data.
For the subsystem we use meteorological data as training dataset .

### *Data Preprocessing:*

Similar to the data Preprocessing step done for sub-system –1 here the missing values are eliminated first by replacing with large negative values (-9999).

***Linear Regression:***

Linear Regression is a supervised learning approach that is used to predict a quantitative response (y) from a predictor variable (x) by making use of statistical measures . Once the trained dataset is fitted to the linear regression algorithm, we get rainfall predictor model .

## RESULTS AND DISCUSSIONS:

***Dataset Description:***

This work utilized an agricultural dataset that chiefly included soil characteristics and environmental factors .The dataset contains 1000 instances and 16 attributes .The target class is the multiclass representation with 9 classes.

***Results-1:***

RFE with bagging classifier on KNN, Naïve bayes, decision Tree, SVM, Random Forest (accuracy –after reduction 0.9272 ) outperforms the other combinations .

All combinations of feature selection methods with classifiers are applied. Feature selection methods applied are RFE,BORUTA,SFFS and classifiers used are K-NN,SVM,NAIVE BAYES,DECISION TREE and also BAGGING with above classifiers.

***Results-2:***

In this along with classification we perform subsystem-2 Rainfall predictor model. For classification we got 90% for decision tree and 89% for K-NN and 90.43% for Random forest classifier.

## CONCLUSION:

In this paper , successfully showed a procedure for crop recommendation system .The main work involves around chooses which can be easily used by farmers all over India .This system would assist farmers in making an informed decision about which crop to grow depending on a variety of environmental and geographical factors. We have also proposed method for rainfall prediction either through the GPS Sensors or predicting the current rainfall through regression. This secondary system which is rainfall predictor predicts the rainfall also can be useful individually for the farmers .

The model proposed in this paper can be further extended in future to give decisions about crop rotations .This would ensure maximized yield as the decision about which crop to grow would now also depend upon which crop was harvested in previous cycle . Furthermore crop demand and supply as well as other economic indicators like farm harvest prices and retail prices can also be considered as parameters to the Model .This would ensure prediction based on environmental ,geographical also economic aspects.

**REFERENCES:**


1. Suruliandi, A., Mariammal, G., & Raja, S. P. (2021). Crop prediction based on soil and environmental characteristics using feature selection techniques. Mathematical and Computer Modelling of Dynamical Systems, 27(1), 117-140.
2. Doshi, Z., Nadkarni, S., Agrawal, R., & Shah, N. (2018, August). AgroConsultant: intelligent crop recommendation system using machine learning algorithms. In 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA) (pp. 1-6). IEEE.
3. S. P. Raja, B. Sawicka, Z. Stamenkovic and G. Mariammal, "Crop Prediction Based on Characteristics of the Agricultural Environment Using Various Feature Selection Techniques and Classifiers," in IEEE Access, vol. 10, pp. 23625-23641, 2022
4. Kulkarni, N. H., Srinivasan, G. N., Sagar, B. M., & Cauvery, N. K. (2018, December). 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) (pp. 114-119). IEEE.
5. Pudumalar, S., Ramanujam, E., Rajashree, R. H., Kavya, C., Kiruthika, T., & Nisha, J. (2017, January). Crop recommendation system for precision agriculture. In 2016 Eighth International Conference on Advanced Computing (ICoAC) (pp. 32-36). IEEE.
6. PANDE, S. M., RAMESH, P. K., ANMOL, A., Aishwarya, B. R., ROHILLA, K., & SHAURYA, K. (2021, April). Crop recommender system using machine learning approach. (pp. 1066-1071),IEEE Xplore.
7. Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].
8. Eswari, K. E., & Vinitha, L. (2018). Crop yield prediction in Tamil Nadu using Baysian network. International Journal of Intellectual Advancements and Research in Engineering Computations, 6(2), 1571-1576.
9. Sriram Rakshith.K, Dr.Deepak.G, Rajesh M, Sudharshan K S, Vasanth S, Harish Kumar N, "A Survey on Crop Prediction using Machine Learning Approach", In International Journal for Research in Applied Science & Engineering Technology (IJRASET), April 2019, pp( 3231- 3234)
10. Bandara, P., Weerasooriya, T., Ruchirawya, T., Nanayakkara, W., Dimantha, M., & Pabasara, M. (2020). Crop recommendation system. International Journal of Computer Applications, 975, 8887.
11. Liying Yang (2011), 'Classifiers selection for ensemble learning based on accuracy and diversity' Published by Elsevier Ltd. Selection and/or peer-review under responsibility of [CEIS].
12. Paja, K. Pancerz, and P. Grochowalski, ''Generational feature elimination and some other ranking feature selection methods,'' in Advances in Feature Selection for Data and Pattern Recognition, vol. 138. Cham, Switzerland: Springer, 2018, pp. 97–112.
13. Reddy, D. A., Dadore, B., & Watekar, A. (2019). Crop recommendation system to maximize crop yield in ramtek region using machine learning. International Journal of Scientific Research in Science and Technology, 6(1), 485-489.