

|  |  |
| --- | --- |
| PROJECT TITLE: | **SMART FARMING: MACHINE LEARNING FOR CROP SELECTION** |
| SUBMITTED BY: | **MUZAIB WANI**  BTech. CSC University of Kashmir |
|  | **&** |
|  | **TAWQEER NABI**  MSc (Agri.) Statistics SKUAST\_K |
|  |  |
| SUBMITTED TO: | **DR SYED NISSAR BUKHARI**  Scientist-D NIELIT Ministry of Electronics & Information |
|  |  |
| DATE OF SUBMISSION: | **19-02-2024** |

# Abstract

Agriculture contributes significantly to a country's economic development. The cultivation of most crops majorly depends upon the regional weather conditions and soil status of the region. So, the analysis of the agro-climatic conditions of a zone contributes significantly to deciding the right crop for the right land in the right season to obtain a better yield. Machine learning algorithms facilitate this process to a great extent for better results. Emerging technologies like Machine Learning (ML) can be used to improve the productivity of crops. ML techniques can help farmers improve crop yield by selecting the best crop for their land. In this work, a comparative study of various feature selection methods are carried out for crop prediction using classification. This project compares the performance of various ML algorithms in predicting suitable crop based on soil and weather parameters. Five supervised ML algorithms implemented and compared are Random Forest, Decision Tree, Support Vector Machine, Logistic Regression and K-Nearest Neighbour. The accuracy parameter is used to evaluate the performance of these algorithms in crop selection.

# Introduction

For a nation, one of the most important aspects of its growth revolves around its potential to produce food. For generations, the production of essential food crops has been correlated with agriculture. In reality, however, the rapid pace of population growth has, by far, been the single biggest preoccupation of our society. In doing so, the scope of agriculture has been greatly undermined, particularly in terms of land use and fertility. Given that the area of land under cultivation in this era of urbanization and globalization is unlikely to increase, the focus will have to be on making the most of what there is. In agriculture, crop cultivar prediction is a key factor. Although recent research has opened up statistical information on agriculture, few studies have investigated crop prediction based on historical data. However, owing to the unbridled use of fertilizers comprising nitrogen, potassium, and micronutrients, crop cultivation prediction is a challenge. In general, agro-climatic input parameters such as soil texture, rainfall, and temperature influence crop production. Input parameters for agriculture vary from region to region, and it is daunting to collect such information over large tracts of land. The vast datasets obtained can be used for crop prediction on a massive scale. Owing to the nature of the problems involved, there is a need to develop new machine learning methods for farming arable land and making the most of narrow land resources. Researchers in agriculture have been testing numerous forecasting methodologies to identify the most suitable crop for specific areas of land.

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. In this era of technology and data science, the agricultural sector stands to benefit greatly from properly implemented techniques. Feature selection and classification are critical machine learning techniques Feature selection has to do with selecting the most important attributes from a dataset. It involves picking a subset of appropriate attributes from a larger set of original attributes in terms of a predefined benchmark, such as classification performance or class separability, which plays a significant role in machine learning applications. There are three common machine learning techniques: supervised, unsupervised, and reinforcement learning. This work uses supervised learning classification techniques for prediction.

Due to the rapid population growth, the food requirement is increasing sharply which requires increased crop production. Cultivating suitable crops considering the soil, weather, and other parameters, can address this challenge, which comes under the category of precision agriculture. In the past few years, different data mining techniques have been used for making decisions in crop cultivation considering various environmental, soil, and other parameters. The most frequently used techniques can further be classified as statistical models and machine learning (ML) algorithms. In statistical methods, the organization of data is known in advance. On the other side, ML algorithms learn from existing data, known as 'dataset', and improve over time without explicit programming.

In ML, data is termed training and testing data where training data is utilized to train the model, and test data is used to make predictions.

In the past, various ML techniques have been used to address the various challenges in the domain of agriculture, e.g., weather prediction, plant disease classification, intelligent irrigation, yield prediction, crop selection, etc.

# Existing and Proposed System

Clearly, a farmer is the best decision maker in the selection and cultivation of crops. Today, however, cultivar prediction is done manually in laboratories, and farmers need the help of experts to determine the most suitable crop/s for a specific piece of land. The experts collect soil samples from a particular portion of land and test them in the laboratory, following which they offer suggestions on the ideal crop/s to be raised. Prediction takes time and selecting the most suitable crop/s is a complex task in agriculture. Manual prediction has largely failed, owing to climatic changes and environmental factors that affect crop cultivation. Accurate predictions of suitable crops for cultivation improve production levels. Crop prediction attributes are defined by multiple factors such as genotype, climate and the interactions between the two. Accurate crop prediction needs a fundamental understanding of the functional relationship between cultivation and interactive factors like the genotype and climate. Further, it requires both detailed datasets and efficient algorithms to examine these relationships. Justified by these facts, machine learning techniques are used in this study to predict the most suitable crop for a specific stretch of land, and this technique is ideal for considering factors like the soil and environmental conditions.

# Dataset Description

This work utilized an agricultural dataset that chiefly included soil characteristics and environmental factors, collected from the Kaggle website [here](https://www.kaggle.com/datasets/78c7595cbf5a98a8cc488e5b9f6911f216680e5a221e6704ba7be4d7ef42c753/download?datasetVersionNumber=1). The dataset contains 2200 rows, 7 attributes, where 4 attributes are soil characteristics and the remaining 3 environmental characteristics, respectively. The target class is the multiclass representation with 22 classes.

# References:

A. Suruliandi, G. Mariammal & S.P. Raja (2021) Crop prediction based on soil and environmental characteristics using feature selection techniques, Mathematical and Computer Modelling of Dynamical Systems, 27:1, 117-140, DOI: [10.1080/13873954.2021.1882505](https://doi.org/10.1080/13873954.2021.1882505)

Rani, S., Mishra, A.K., Kataria, A. *et al.* Machine learning-based optimal crop selection system in smart agriculture. *Sci Rep* **13**, 15997 (2023). <https://doi.org/10.1038/s41598-023-42356-y>

Patil P, Athavale P, Bothara M, Tambolkar S, More A. Crop Selection and Yield Prediction using Machine Learning Approach. Curr Agri Res 2023; 11(3). doi: <http://dx.doi.org/10.12944/CARJ.11.3.26>

Bhola, A., Kumar, P. (2023). Performance Evaluation of Different Machine Learning Models in Crop Selection. In: Muthusamy, H., Botzheim, J., Nayak, R. (eds) Robotics, Control and Computer Vision. Lecture Notes in Electrical Engineering, vol 1009. Springer, Singapore. https://doi.org/10.1007/978-981-99-0236-1\_16

[https://www.kaggle.com/atharvaingle/crop-recommendation-dataset. Accessed 19-02-2024](https://www.kaggle.com/atharvaingle/crop-recommendation-dataset.%20Accessed%2019-02-2024)

<https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/>. Accessed on 20-02-2024.