Crop Detection Website

Submitted by

Group: GA18

Abhinav Choudhary - University Roll No: 2100290140002

Akshay Choudhary - University Roll No: 2100290140010

Akshit Shrawat - University Roll No: 2100290140019

Session: 2022-23(4th Semester)

Under the supervision of

Mrs. Vidushi Mishra

Associate Professor

KIET Group of Institutions, Delhi-NCR, Ghaziabad

DEPARTMENT OF COMPUTER APPLICATIONS

KIET GROUP OF INSTITUTIONS, DELHI-NCR, GHAZIABAD-201206

(April 2023)

ABSTRACT

Earlier, crop cultivation was undertaken on the basis of farmers' hands-on expertise. However, climate change has begun to affect crop yields badly. Consequently, farmers are unable to choose the right crop/s based on soil and environmental factors, and the process of manually predicting the choice of the right crop/s of land has, more often than not, resulted in failure. Accurate crop prediction results in increased crop production. This is where machine learning playing a crucial role in the area of crop prediction. Crop prediction depends on the soil, geographic and climatic attributes. Selecting appropriate attributes for the right crop/s is an intrinsic part of the prediction undertaken by feature selection techniques. In this work, a comparative study of various wrapper feature selection methods are carried out for crop prediction using classification techniques that suggest the suitable crop/s for land. The experimental results show the Recursive Feature Elimination technique with the Adaptive Bagging classifier outperforms the others.

TABLE OF CONTENTS

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 [Citation1–3] 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 [Citation4–7]. 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.

Literature Review

In a research carried out by Zaminur Rahman a comparative study of several machine learning techniques has been carried out. They have carried out the classification using the data of Bangladesh. Considered the six district soil data and used the geographical features for classification. They have used k Nearest Neighbour, Bagged tree and SVM finally compared the results of three algorithms and brought out a model for classifying the soil types and the suitable crop that can be cultivated in that particular soil type[3].Among the used three algorithms SVM has obtained the average accuracy.

In a research carried out by Leisa J.Armstrong a comparative study of data mining algorithms. They have used a large dataset extracted from the Australian Department of Agriculture and Food(AGRIC) to conduct the research[4].

In an approach carried out by Jay Gholap carried out a modal to classify the soil based on fertility. The dataset was collected from the soil testing laboratories of Pune District. They have used WEKA tool for developing an automated system[5].

Chiranjeevi M. N carried out a research for classifying the soil types so that it can be useful for the farmers for analyzing the type o soil and the crop that can be cultivated so that there will a good yield and profit. They have considered the data mining algorithms for classifying the soil.They have used algorithms such as J48 decision tree classifier and NaÃ¯ve bayes classifier among these two algorithms NaÃ¯ve bayes has obtained the maximum accuracy of 98%[6].

TECHNOLOGIES USED

The technologies used for developing this project are:

- **Django** - for developing the backend of the project

  Django is a high-level Python web framework that enables rapid development of secure and maintainable websites. Built by experienced developers, Django takes care of much of the hassle of web development, so you can focus on writing your app without needing to reinvent the wheel.

- **ReactJS** - for developing the frontend of the project

React (also known as React.js or ReactJS) is a free and open-source front-end JavaScript library for building user interfaces based on components. It is maintained by Meta (formerly Facebook) and a community of individual developers and companies.
React can be used as a base in the development of single-page, mobile, or server-rendered applications with frameworks like Next.js. However, React is only concerned with the user interface and rendering components to the DOM, so creating React applications usually requires the use of additional libraries for routing, as well as certain client-side functionality.

- **Scikit-learn** - for Preprocessing of dataset and classification.

Scikit-learn (Sklearn) is the most useful and robust library for machine learning in Python. It provides a selection of efficient tools for machine learning and statistical modeling including classification, regression, clustering and dimensionality reduction via a consistence interface in Python. This library, which is largely written in Python, is built upon NumPy, SciPy and Matplotlib.

HARDWARE & SOFTWARE REQUIREMENTS SPECIFICATIONS

Hardware Requirements

S. N.	Description
1	PC with 10 GB or more Hard disk.
2	PC with 2 GB RAM.
3	PC with core i3 or above processor.

Software Requirements

S. N.	Description	Type
1	Operating System	Windows 10 or 11 or Ubuntu 18.04 or above
2	Language	Python 3
3	Front End	React 17
4	IDE	Google Colab, VS Code,PyCharm
5	Browser	Chrome, Firefox, Edge

MODULES IN PROJECT

Homepage

- Description

Prediction page

- Take input from the user
- Predict the crop

Registration page

Login page

User prediction history

- Details of users history predictions

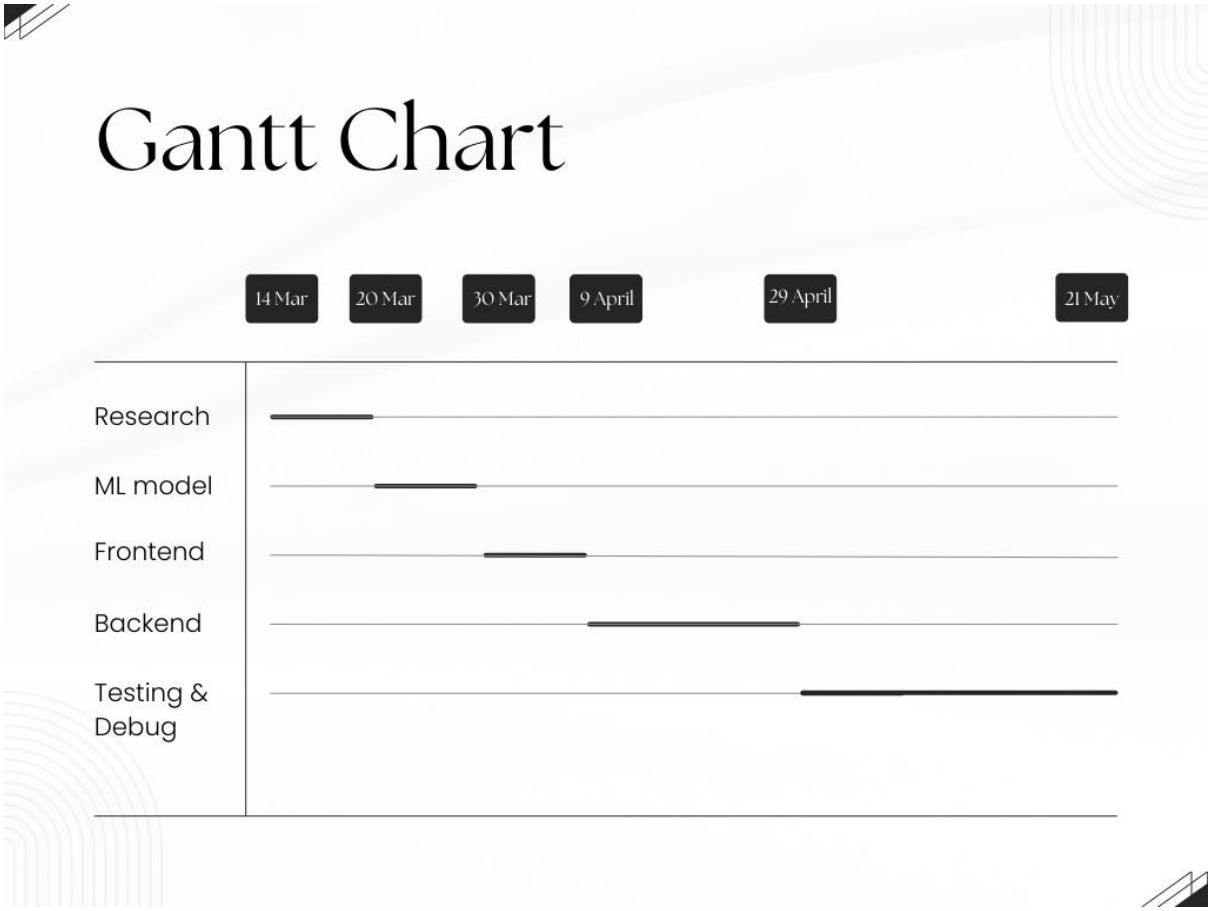
REPORTS

This project will have the following reports:

- Pre-processing of dataset
    - EDA
    - Feature Engineering
    - Encoding Features
    - Imputations
    - Standard Scaling

- Comparison of accuracy of various ML models used for intent classification
    - Random Forest Classifier
    - Gaussian Naive Bayes Model
    - K-Nearest Neighbours (KNN) Model

GANTT CHART


Gantt Chart
14 Mar
20 Mar
30 Mar
9 April
29 April
21 May
Research
ML model
Frontend
Backend
Testing &
Debug

References

1. S. Das. *Filters, wrappers and a boosting-based hybrid for feature selection*. In International conference on machine learning 1 (2001), pp. 74–81. [Google Scholar]
2. H. Liu and Y. Lei, *Toward integrating feature selection algorithms for classification and clustering*, IEEE Trans. Knowl. Data Eng. 17 (2005), pp. 491–502. doi:https://doi.org/10.1109/TKDE.2005.66. [Crossref], [Web of Science ®], [Google Scholar]
3. A. Maruf, M. Abdullah, and S. Shatabda, *iRSpot-SF: Prediction of recombination hotspots by incorporating sequence based features into Chou's Pseudo components*, Genomics 111 (2019), pp. 966–972. doi:https://doi.org/10.1016/j.ygeno.2018.06.003. [Crossref], [PubMed], [Web of Science ®], [Google Scholar]
4. I. Kononenko, *Estimating attributes: Analysis and extensions of RELIEF*. European conference on machine learning 784 (1994), pp. 171–182. [Crossref], [Google Scholar]
5. J. Pavel Pudil, N.N. Choakjarernwanit, and J. Kittler, *Feature selection based on the approximation of class densities by finite mixtures of special type*, Pattern Recognit 28 (1995), pp. 1389–1398. doi:https://doi.org/10.1016/0031-3203(94)00009-B. [Crossref], [Web of Science ®], [Google Scholar]
6. J. Novovicová, P. Pudil, and J. Kittler, *Divergence based feature selection for multimodal class densities*, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996), pp. 218–223. doi:https://doi.org/10.1109/34.481557. [Crossref], [Web of Science ®], [Google Scholar]
7. H. Liu and H. Motoda, *Feature Selection for Knowledge Discovery and Data Mining*, 1st ed., Springer, US, 1998. [Crossref], [Google Scholar]
8. M. Dash, K. Choi, P. Scheuermann, and H. Liu. F*eature selection for clustering-a filter solution*. IEEE International Conference on Data Mining, 2002, Proceedings., Maebashi City, Japan, pp. 115–122. [Crossref], [Google Scholar]