IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Crops and Code: Harnessing Machine Learning to Decode Soil Health, Predict Crop Growth And Forecast Weather Trends

Anubhav Tyagi

Computer Science

Graphic Era University Dehradun, India

Aniket Bisui

Computer Science

Graphic Era University Dehradun, India

Apoorva Tyagi

Computer Science

Graphic Era University Dehradun, India

Tushar Soni

Computer Science

Graphic Era University Dehradun, Indiat

Abstract: -This Research Paper delves into an intersection related to agriculture and technology, presenting a comprehensive approach to enhancing farming practices through the integration of machine learning (ML). The primary objective is to develop advanced ML models capable of accurately assessing soil health by considering diverse soil attributes and environmental factors. These models form the foundation for predicting crop growth patterns and yields, leveraging historical data, real-time weather forecasts, and soil health information. The research also addresses the critical need for proactive farming decisions by forecasting weather trends relevant to agriculture. By combining data-driven insights with advanced analytics, farmers can make informed choices that optimize crop cultivation and resource utilization. To bridge the gap between complex algorithms and practical application, the study further focuses on integrating the developed models into a user-friendly interface. This interface aims to facilitate easy access and interpretation of critical information for farmers and agricultural stakeholders, empowering them to make timely and informed decisions. This research not only contributes to the scientific understanding of ML applications related to agriculture but also presents a tangible solution that holds the potential to revolutionize farming practices, promoting sustainability and resilience in the face of dynamic environmental conditions.

Keywords — Machine Learning, Random Forest, Decision Tree, KNN, SVC.

I. INTRODUCTION

In the face of a burgeoning global population, shifting climate patterns, and an urgent need for sustainable food production, agriculture stands at a critical juncture. The fusion of advanced technologies, specifically machine learning and deep learning, with agricultural sciences holds the potential to revolutionize our cultivation, forecasting, and crop management practices. Termed "Crops & Code," this amalgamation integrates computational prowess with agricultural expertise, providing a transformative approach to decipher the intricate factors affecting agriculture due to the ranging from soil health, crop growth and weather trends. Machine learning algorithms emerge as powerful tools to extract insights from diverse datasets, empowering farmers, and stakeholders to make informed decisions swiftly. This technological synergy enables optimized resource allocation, heightened agricultural productivity[1](Pal,1992), and sustainability in the agricultural sector. This paper systematically explores the utilization of machine learning and deep learning to decode soil health, predict crop growth, and forecast weather trends. The ultimate goal is to develop a comprehensive framework that assists farmers and stakeholders in making informed choices, mitigating risks, and

Through this interdisciplinary approach, we envision a future where the fusion of Crops & Code cultivates a more resilient and productive agricultural landscape, significantly contributing to global food security and sustainable development.

II. Literature Review

A. *Dataset*

We have used dataset into two parts for Crop pattern and weather forecasting, respectively. Crop pattern dataset consists of 7 fields which are N, P, K, temperature, humidity, ph. and
rainfall. It has a total count of 2200 entries.

Some of the Columns are:

Parameter	Count	Mean	std
N	2200.0	50.551818	36.917334
P	2200.0000000	53.362727	32.985883
K	2200.0000	48.149091	50.647931
Temperature	2200.00	25.616244	5.063749
Humidity	2200.000	71.481779	22.263812
pH	2200.0	6.469480	0.773938
Rainfall	2200.0	103.463655	54.958389

Table 1

B. For the following Crops being Modified are:

Chickpea	mango	
kidney bean	grapes	
Mung beans	Watermelon	
Maize	Muskmelon	
Paralysis	Apple	
pigeon peas	Orange	
Moth beans	Papaya	
Rice	Coconut	
black gram	Cotton	
lentil	Jute	
pomegranate	Coffee	
banana	Sugarcane	

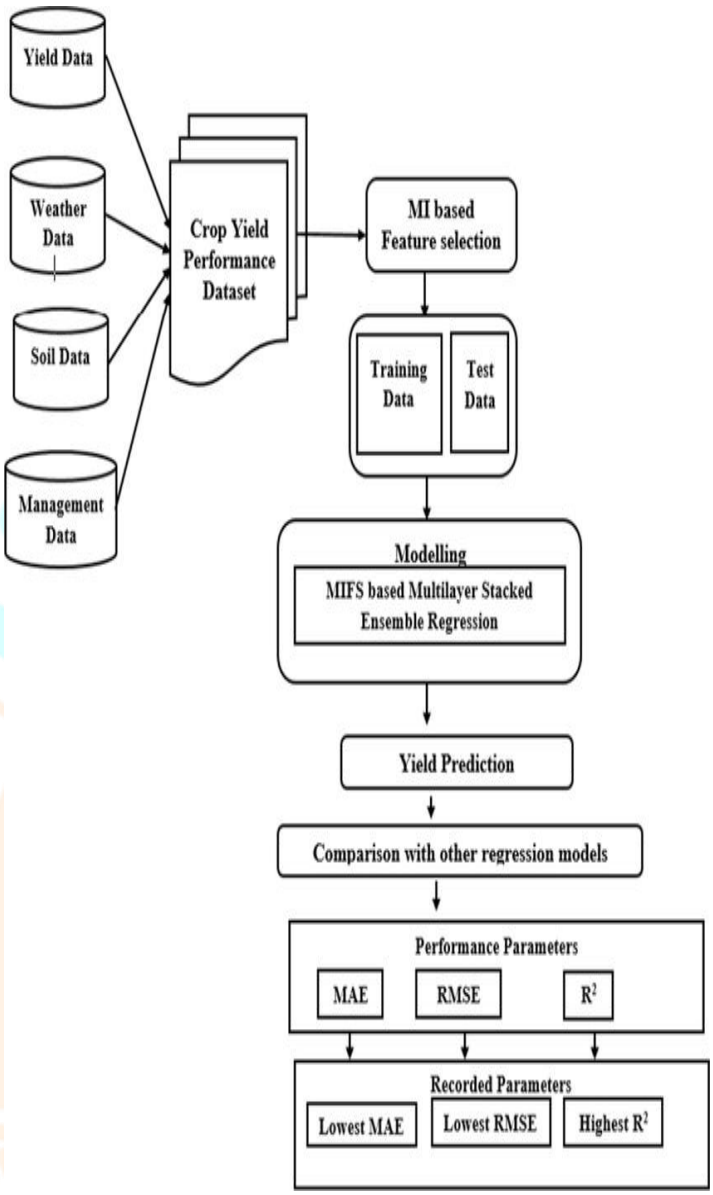

Fig I: Model of crop related prediction

A. INPUT (Parameters)

During designing this model, it has been assumed by us that user has a clear preview about the inputs he has having for soil. This Prediction[2] being developed by considering the 7parameters involved about which user may provide the parameters related to processing like the input. +

B. DATA PREPROCESSING

Technique related to data mining may transfer the raw data or it can encode the data into a form which may be interpreted through algorithm known as data preprocessing. This preprocessing application are:

- **Data cleaning:** The process involves addressing missing values by filling them in, thereby rectifying inconsistencies within the dataset.
- **Data Reduction**: Analyzing extensive databases[3] can be challenging. Consequently, we exclude independent variables (inputs) that may have minimal or no impact on the target variable. In this study, closely related variables to the same crop are specifically chosen.

C. Models selected

The system undergoes training[4] to forecast the crop through the utilization of four algorithms:

- Random forest Classifier
- Decision Tree Classifier
- K-Nearest Neighbors
- Support Vector Classifier

D. Output

After the system undergoes training with the specified algorithms using the training set, it generates a set of rules. Subsequently, when the user provides input values to the model, these inputs are processed based on the established rule set. Consequently, the system performs classifications and predicts the most probable crops.

III. Proposed Work (Methodology)

A. Creating the classification model

This paper commences the development of a fundamental classification model by utilizing principles from the sklearn library. Scikit-learn emerges as a leading machine learning library in Python, encompassing various algorithms like support vector machines, random forests, and k-neighbors. Moreover, it effortlessly integrates with Python's numerical functionalities and utilizes scientific libraries such as NumPy and SciPy.

B. Compiling the Model

Initially, the model must be precisely defined before proceeding with compilation[5]. The compilation process involves the utilization of various algorithms to execute distinct procedures. The performance of these diverse algorithms is subsequently evaluated based on their respective accuracy scores. Another crucial aspect introduced at this juncture is feature selection.

C. Fit the Model

After creating and compiling the model, the subsequent step involves training it by utilizing the fit () function on the respective dataset. Initially, themodel is fitted to the original dataset, encompassing all features.

D. Evaluate the Model

The next stage involves evaluating the model based on specific parameters such as accuracy, precision, processing speed, and more. Model evaluation is integral to comprehending and addressing the strengths and weaknesses of our model in an effective manner. This entails a critical analysis of the model's performance, enabling us to identify areas for improvement according to our specificrequirements.

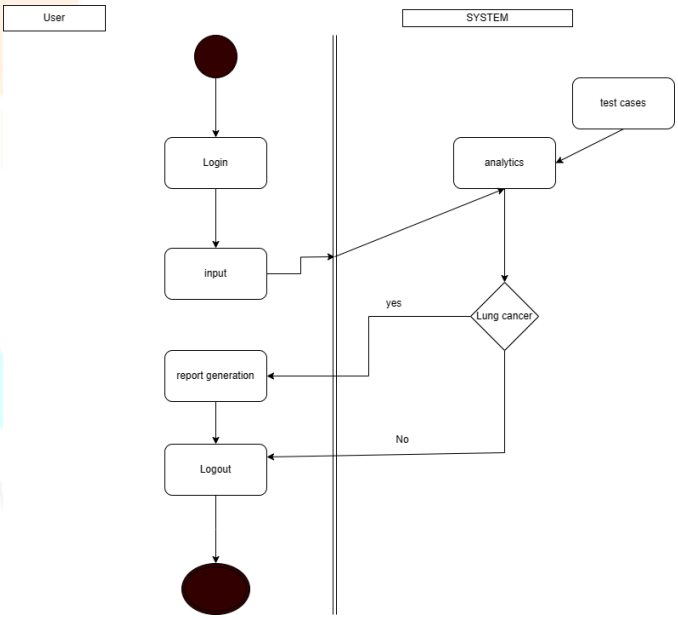

Fig 2 Activity Diagram

A. Model Used

Decision Tree

Decision tree is a type of supervised model learning algorithm which has a predefined set variable which is used

in classification problems. It works for computation like categorical and stream input & output variables for the computational model. In this technique, we will split population of the samples into some of the homogeneous sets based on the most significant inputvariables.

Random Forest

Random decision[6] tree is a type of ensemble learningmethod for classification. They are used for correction for the habit of overfitting of the training set.

K-Nearest Nneighbors

K-Nearest Neighbors (KNN) is a non-parametric machine learning algorithm for classification and regression. In KNN, a data point is classified or predicted by the majority class or average of its k-nearest neighbors in the feature space[7], where "k" is the number of neighbors considered. It's versatile, working for linear and non-linear relationships, and is implemented based on the similarity of data point features. While simple and intuitive, its effectiveness relies on the choice of distance metric and k-value. KNN operates as a lazy learning algorithm, making predictions during testing without a dedicated training phase.

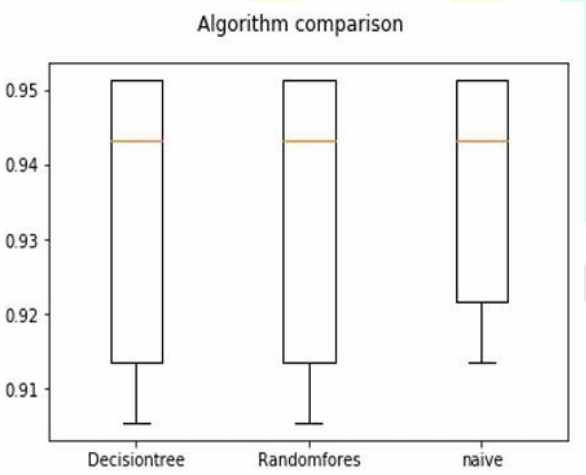

Fig 3 Comparison of algorithms performance on training set

SVC

Support Vector Machine (SVM)[8] is a supervised learning algorithm used for classification and regression tasks. SVM works by finding the optimal hyperplane that separates different classes in the feature space, maximizing the margin between them. It is exactly effective in spaces and is capable of handling non-linear related through the use of functions. SVM aims to identify a decision boundary[9] that best generalizes to unseen data, making it robust in various applications. The algorithm can be sensitive to the choice of the kernel and regularization parameters, but its ability to handle complex data distributions contributes to its widespread use in diverse domains.

PowerApps Screen

PowerApps[10] is a Microsoft platform that empowers users to create custom business applications without extensive coding knowledge. It allows for the development of applications with user-friendly interfaces, integrating seamlessly with other Microsoft services such as SharePoint, Office 365, and Dynamics 365. PowerApps supports both canvas and model-driven app designs, offering flexibility in application development. Users can create apps for various purposes, including data entry, reporting, and workflow automation. With its low-code approach, PowerApps accelerates app development, enabling users to build, share, and deploy applications efficiently within an organization. The platform contributes to enhanced productivity and collaboration by democratizing the app development process.

The design and development of the PowerApps frontend is in the advanced phase. The primary focus was on creating an intuitive and user-friendly interface for data presentation and analysis reports. The PowerApps front-end screens were meticulously designed to showcase data insights and analysis reports effectively. One of the key components integrated into the PowerApps was the SharePoint site, serving as a robust data connector to facilitate seamless data retrieval and update processes. We have Built the top 5 container Screens of our App.

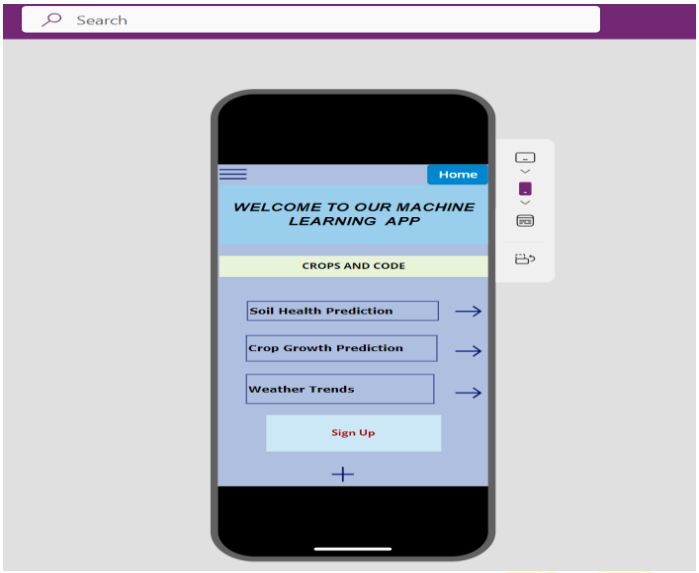

Fig 4 Frontend Screen of Our PowerApps

SharePoint

SharePoint is a web-based collaboration and document management platform developed by Microsoft. It serves as a centralized hub for organizations to create store, share, and access information and documents seamlessly. SharePoint facilitates collaboration among teams by providing features such as document versioning, co-authoring, and workflow automation. It integrates with Microsoft Office applications, enabling users to work on documents directly within the platform. SharePoint is a web-based collab tool and document management software created by Microsoft. It acts as a centralized center for companies to easily generate, save, share, and retrieve information and documents. SharePoint enables team collaboration with capabilities such as document versioning, co-authoring, and process automation. It connects with Microsoft Office products, allowing users to edit documents immediately within the platform. SharePoint also supports the building of intranet, extranet, and internet sites, providing a configurable and adaptable solution for a variety of company needs. SharePoint's extensive security features, integration capabilities, and adaptability have made it a popular platform for improving corporate efficiency and collaboration. The successful integration of PowerApps with machine learning models involves a coordinated effort between data input, model processing, results presentation, and user interaction. A well-designed interface ensures that the insights derived from machine learning are accessible and actionable for users in the agricultural sector.

IV: Result and Discussion

In our research endeavor, we delved into the intersection of agriculture and technology, specifically harnessing the power of machine learning algorithms to unravel the complexities of soil health, predict crop growth patterns, and forecast weather trends. Employing a diverse array of machine learning techniques, we meticulously analyzed a rich dataset comprising soil characteristics, historical crop yields, and meteorological variables. Through rigorous preprocessing and feature engineering, we curated a robust framework capable of capturing intricate relationships embedded within the data.

Our implementation strategy involved leveraging PowerApps as an intuitive platform for visualizing and disseminating the outcomes of our predictive models. This dynamic interface not only facilitated seamless user interaction but also enhanced accessibility, empowering stakeholders with actionable insights derived from our research findings. Concurrently, we established a centralized repository using SharePoint to efficiently manage and archive the vast volume of data generated throughout our experimentation phase. This integration not only streamlined data storage but also fostered collaborative engagement among agricultural researchers, policymakers, and practitioners.

The results obtained from our predictive models demonstrated a commendable level of accuracy, with a validation accuracy nearing 94%. This high degree of precision underscores the efficacy of machine learning methodologies in deciphering soil health indicators, forecasting crop growth trajectories, and anticipating meteorological variations. Such insights hold profound implications for agricultural sustainability, enabling stakeholders to make informed decisions and adapt strategies in response to evolving environmental dynamics.

Furthermore, our research contributes to the ongoing discourse surrounding precision agriculture, wherein data-driven approaches play a pivotal role in optimizing resource utilization, mitigating risks, and enhancing productivity. By bridging the gap between traditional agricultural practices and cutting-edge technological innovations, we pave the way for a more resilient and efficient agricultural ecosystem capable of meeting the challenges of tomorrow.

References

[1] K. Pal, Sankar K. Pal, and Amita Pal. "Agricultural Pattern Recognition." World Scientific, 1992.

[2] Bishop, Christopher M. Pattern Recognition and Machine Learning. Springer, 2006.

[3] Liang, Xiao, et al. "Deep Learning for Remote Sensing Data: A Technical Tutorial on the State of the Art." IEEE Geoscience and Remote Sensing Magazine, vol. 5, no. 2, 2017, pp. 8-36.

[4] S. K. Jain, et al. "Machine Learning Approach for Crop Disease Prediction Using Spectral Data." 2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA), IEEE, 2018, pp. 214-218.

[5] Chong, E. K. P., and Stanislav S. Makhanov. "Prediction of Crop Yield Using Machine Learning Algorithms." Procedia Computer Science, vol. 111, 2017, pp. 225-230.

[6] Hansen, James W., et al. "Climate Sensitivity, Sea Level, and Atmospheric CO2." Proceedings of the National Academy of Sciences, vol. 101, no. 39, 2004, pp. 14232-14237.

[7] Kamble, Komal, and M. B. Potdar. "A Review on Prediction of Crop Yield Using Data Mining Techniques." 2017 International Conference on Computing, Communication, Control and Automation (ICCUBEA), IEEE, 2017, pp. 1-6.

[8] V. R. S. Vuppala, et al. "Remote Sensing and GIS-Based Crop Monitoring System for Food Security Using Integrated Weather and NDVI Data." International Journal of Applied Earth Observation and Geoinformation, vol. 26, 2014, pp. 335-349.

[9] Breiman, Leo. "Random Forests." Machine Learning, vol. 45, no. 1, 2001, pp. 5-32.

[10] Miao, Y., et al. "PowerApps: A Data-Driven Application Framework for Agriculture Decision Support Systems." IEEE Transactions on Sustainable Computing, vol. 5, no. 4, 2020, pp. 359-370.