International Journal of Integrated Science and Technology (IJIST)
Vol.1, No.4, 2023: 555-562

Modernizing Agriculture through ML and Deep Learning

Anuja Phapale[1*], Sakshi Shinde[2]
AISSMS IOIT Pune, India
Corresponding Author: Anuja Phapale; anuja.phapale@gmail.com

ARTICLEINFO

Keywords: Precision agriculture, Data analytics, Smart farming, Crop management

Received : 5, August
Revised : 18, September
Accepted: 20, October

ABSTRACT

Farming is an essential part of a country's economy, providing livelihoods for many people. Traditionally, farmers used less precise methods, which reduced productivity and took a lot of time. Precision farming, on the other hand, improves productivity by accurately planning the farming process. It involves predicting weather conditions, analysing soil, recommending crops, and determining the right amount of fertilizers and pesticides to use. Precision farming uses advanced technologies like IoT, data mining, data analytics, and machine learning to gather data, train systems, and make predictions. This technology reduces manual labour and increases productivity..

DOI: https://doi.org/10.59890/ijist.v1i4.683
https://journal.multitechpublisher.com/index.php/ijist

INTRODUCTION

Farming, or agriculture, is all about growing crops and raising animals. It's really important for a country's economy because it provides the materials for making things and the food we eat every day. Traditional farming methods have been used for a long time, but they aren't very precise, so they require a lot of hard work and take up a lot of time.

Now, there's something called Precision Agriculture, which uses digital technology like robots, sensors, and automation to make farming easier, more profitable, and better at decision-making. It helps farmers deal with the differences in their fields and the timing of planting and harvesting to make more money, get better yields, and improve the quality of what they grow.

But not all farms use this technology yet. Bigger, more valuable farms tend to use it more than smaller ones, and the adoption of Precision Agriculture varies from country to country and place to place. One big reason for this is that the technology can be expensive, and we need to find ways to make it more affordable for all farms.

Precision Agriculture relies on advanced technologies like the Internet of Things (IoT), which involves devices like sensors and smart gadgets that talk to each other and share information. These devices can help monitor the environment and soil, which can predict the health of crops and even figure out when to water them.

Machine learning is another big part of this. It makes farming more efficient and easier. It involves collecting data, creating models, and making predictions. Machine learning is particularly useful for dealing with complex farming issues.

THEORETICAL REVIEW

Crop growth is primarily influenced by the soil's macronutrient and trace mineral content of the soil. Soil being the broad representation of several environmental factors including rainfall, humidity, sunlight, temperature and soil ph. The use of a support vector machine and decision tree algorithm to distinguish the type of crop based on micronutrients and meteorological characteristics has been presented as an efficient means of predicting the crop. Three crops where selected such as rice, wheat and sugarcane. Based on certain observations details about micronutrients where been obtained.

The details were given to a classifier model, which predicted the crop based on the information provided. There are many different Machine Learning algorithms, so choosing just two is not enough. The SVM model had a higher accuracy score at 92% compared to the decision tree algorithm. In this research, the best out of these two models was chosen. However, there are other algorithms designed for classification tasks that should be explored, such as K Neighbours classifier, Logistic Regression, and Ensemble classifiers.

The research predicted a crop based on the data entered into the SVM model. Data is incredibly valuable, and it can provide more than just crop predictions. The proposed research goes beyond suggesting crops and uses the data to gather various information. This includes details like Growing Degree

Days, which measure heat units required for crop growth, and the specific amounts of nitrogen, phosphorous, and potassium needed per 200 lbs. of fertilizer for optimal crop growth.

Machine Learning algorithms like SVM and Decision Tree Classifier were used. However, in this research, a wider range of Machine Learning algorithms was applied, including Decision Tree, K Nearest Neighbor, Linear Regression, Neural Network, Naïve Bayes, and Support Vector Machine. This expanded the options compared to. Linear Regression was specifically used to predict crop production based on climate factors like rainfall, temperature, and humidity. It's worth noting that the accuracy scores for all these algorithms were below 90%.

The work in question was primarily about implementing models using the dataset, and the next step is to create a web interface to make it user-friendly for common people. Currently, all the required data must be manually entered for the model to predict the crop. However, the proposed work simplifies this process by extracting temperature and humidity data through Web Scraping, eliminating the need for manual input. The proposed system offers an interactive web interface where the user only needs to specify average rainfall and soil pH value. Temperature and humidity details are automatically retrieved and used with the best model, which includes 10 different algorithms with hyperparameter tuning.

This approach aims to achieve an accuracy of 95.45%, thanks to the hyperparameter tuning of the algorithms, which was not part of. The results, along with additional information, are displayed on the web interface, making it easier for users to understand the predictions and data.

This study explores the use of existing deep learning techniques for weed detection. It highlights the importance of Machine Learning (ML) and Deep Learning algorithms for identifying weeds, with a particular focus on pre-trained models. Pre-trained models offer many advantages and can be utilized for image classification. The study also offers guidance on how to work with datasets to make them more efficient for building models. There are various public datasets available on different platforms for this purpose. The study suggests techniques like Image Resizing, data augmentation, and image segmentation to improve the accuracy of weed classification. Pre-trained models are particularly effective in increasing accuracy.

To perform deep learning techniques, the proposed model has adopted certain preprocessing steps, such as Image Resizing and data augmentation, before constructing the actual deep learning model for weed prediction.

Another algorithm used for identifying weeds in vegetable plantations is CenterNet. CenterNet follows a two-stage process: first, it collects and detects images of Bok choy plants, and in the second stage, it performs color-based segmentation to identify the weeds in the dataset. The images are collected from Nanjing, China, and have been augmented to expand the dataset size. CenterNet is used for both training and testing the images, but there's room for optimization to improve results.

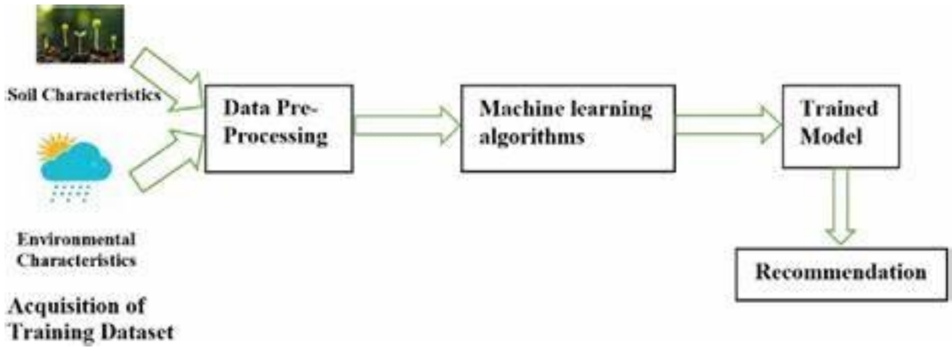


Figure 1. Conceptual Framework

METHODS

Module 1, which focuses on Crop Recommendation, utilizes several datasets obtained from Kaggle, including Crop recommender.csv, soil.csv, and scientific_names.csv.

The Crop recommendation dataset, in particular, is used for training the model due to its rich set of attributes essential for predicting the most suitable crops for specific conditions. After prediction, additional datasets like Soil names and Crop names are used to obtain information regarding the soil type and scientific name of the recommended crops.

The process of Crop Recommendation involves the following steps:

Step 1: Importing Libraries and Dataset

To harness the power of Machine Learning algorithms and preprocessing tools, specific libraries are imported. These libraries, including NumPy, pandas, pickle, matplotlib, seaborn, Label, and others, are crucial for efficient model building and prediction.

Step 2: Descriptive Analysis

Descriptive analysis is a crucial initial step to ensure the development of a robust predictive model. It provides an insight into the dataset's structure and helps in uncovering valuable information. After importing the dataset, the first task is to check for any missing values in each attribute. In the case of the crop recommendation dataset, it is found that the attributes are free of missing values, as indicated in Figure 3.1.4. This is a critical aspect, as missing data can significantly impact the accuracy of the predictive model. Subsequently, the data types of the attributes are identified and unique values in the dependent variable (Label attribute) are listed .

Step 3: Data Visualization

Once the fundamental details about the dataset are gathered, the next step involves data visualization to provide a visual understanding of the data. A correlation matrix is used to showcase the relationships between attributes. It is essentially a table format that displays the correlation coefficients between the attributes. In the matrix, the attributes are represented in both the rows and columns, with the relationships between them clearly visualized. This visual representation can help identify patterns, dependencies, and correlations within the data, which can be instrumental in building an effective predictive model

Step 4: Outlier Detection and Outlier Treatment

Identifying and addressing outliers is an important step in ensuring the integrity of the dataset. Outliers are data points that significantly deviate from the norm and can potentially affect the accuracy of predictions. In this module, the proposed model employs two techniques for outlier detection: box plots and the Inter-Quantile Range Technique (IQR). Box plots are used to visualize outliers in specific attributes, with Figure 3.1.22 illustrating an example for the Soil pH attribute. The figure highlights values above 8.5 and below 4.5 as outliers. The IQR technique calculates the percentage of data outside the quantile range between 0.75 and 0.25. In some cases, outliers may be removed or retained, depending on their relevance to the business perspective. For the crop recommendation dataset, the outliers detected are considered useful and retained since the observations correspond to specific crop growth details obtained through experiments. As a result, all observations in the dataset are preserved to ensure the accuracy of predictions.

Step 5: Label Encoding

The "Label" attribute in the dataset, representing the dependent variable, is subjected to label encoding. This attribute contains non-numerical, categorical values representing crop names. As many classification models require numerical inputs, these non-numerical values are encoded into numerical values. This encoding ensures that the values can be processed by classification models for future predictions.

Step 6: Splitting the Data into Train and Test Sets

After addressing outliers and gaining a visual understanding of the dataset, the data is divided into training and testing sets. The training set is crucial for training machine learning algorithms, allowing them to learn to make accurate predictions. The testing set serves as a means to evaluate the model's accuracy once it is trained. The split ratio between training and testing data is 50:50.

Step 7: Model Building

The main objective of this module is to identify suitable crops using Machine Learning Classification algorithms. The proposed work employs ten different classification algorithms to determine the most effective model for future predictions. The steps involved in model building encompass importing the model from the library, defining the model, fitting the training and testing data to the model, testing the model on the testing dataset, and calculating confusion matrices and evaluation metrics. Figure 3.1.23 provides results for the Random Forest Classifier, showcasing predicted values for the testing set, training and testing set accuracy scores, and the overall accuracy score for the crop recommendation dataset.

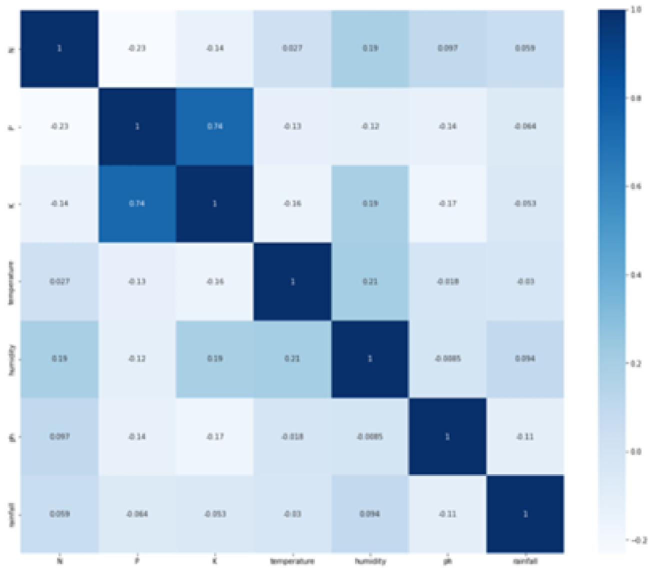
RESULT & DISCUSSION

Result

Table 1. Crop Recommendation dataset sample values

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice
...	...	...	...	...	...	...	...	...
2195	107	34	32	26.774637	66.413269	6.780064	177.774507	coffee
2196	99	15	27	27.417112	56.636362	6.086922	127.924610	coffee
2197	118	33	30	24.131797	67.225123	6.362608	173.322839	coffee
2198	117	32	34	26.272418	52.127394	6.758793	127.175293	coffee
2199	104	18	30	23.603016	60.396475	6.779833	140.937041	coffee

Table 2. Correlation Matrix



Discussion

The discussion on this topic revolves around the utilization of Machine Learning (ML) and Deep Learning techniques for weed detection, particularly focusing on the effectiveness of pre-trained models in achieving accurate classifications. The study underscores the advantages of pre-trained models, which have been previously trained on extensive datasets for tasks like image recognition, thus significantly reducing the need for training from scratch and serving as a solid foundation for weed classification. Data preparation is identified as a critical step to enhance model accuracy, with techniques like Image Resizing, data augmentation, and image segmentation recommended for optimizing the dataset used for deep learning model training. Furthermore, the study suggests the use of publicly available datasets, simplifying the training and evaluation process.

The study introduces the CenterNet algorithm for weed identification, which comprises two stages involving image collection and detection, followed by color-based segmentation. While the algorithm is straightforward, there is a recognition of the potential for optimization to improve the precision of weed detection. Additionally, the Resnet152V2 algorithm is proposed in the study due to its capacity to deliver high accuracy. It features special layers designed to generate highly accurate predictions, a crucial aspect in weed detection for agriculture where misclassifications could lead to the misuse of herbicides or resources. Beyond accuracy, the discussion touches upon the practical implications of precise weed detection in agriculture, such as optimized herbicide usage, cost reduction, and increased crop yields.

CONCLUSIONS AND RECOMMENDATIONS

The Crop Recommendation module encompasses a series of vital steps to build an effective predictive model for suggesting optimal crops based on environmental and soil factors. It commences with data importation and descriptive analysis to ensure data quality and understanding. Data visualization, particularly the use of correlation matrices, aids in identifying data patterns and relationships. The detection and treatment of outliers play a pivotal role in preserving relevant data points that contribute to accurate predictions. Label encoding is essential to transform non-numerical crop names into numerical values, facilitating model processing. The data is then divided into training and testing sets to enable model training and evaluation. The heart of the module lies in model building, where ten classification algorithms are explored, and the best-performing model is determined. These steps are underpinned by good data practices, feature engineering, and ongoing model improvement.

FURTHER STUDY

For further research, it is recommended to explore methods for adaptability to changing environmental conditions, perhaps integrating real-time data sources and climate forecasting. Implementing advanced outlier treatment techniques, such as anomaly detection and robust modelling, can improve data quality and model accuracy. Additionally, considering the impact of socio-economic and market factors on crop recommendations could provide a more comprehensive approach to agricultural decision support. Collaboration with agronomists and domain experts can help in fine-tuning the model to real-world agricultural scenarios. In conclusion, ongoing research should focus on making crop recommendation systems more dynamic, responsive, and comprehensive to address the evolving challenges in modern agriculture.

ACKNOWLEDGMENT

I would like to express my heartfelt gratitude to my colleagues and peers who generously shared their insights and suggestions during the course of this research. Their valuable input significantly contributed to the quality and depth of this work. I would also like to extend my thanks to the institutions and organizations that provided financial support, enabling the successful completion of this study. Their assistance was instrumental in bringing this research to fruition.

REFERENCES

Margit Paustian, Ludwig Theuvsen, Adoption of Precision Agriculture Technology gies by German Crop Farmers, Springer, 2016.

Tanja Groher, Katja Heitkämper, Achim Walter, Frank Liebisch, Christina Um stätter, Status Quo of Adoption of Precision Agriculture Enabling Technologies in Swiss Plant Production, Springer, 2020.

Mohammed Riyadh Abdmezie, D. Tandjaoui, Imed Romdhani, Architecting the Internet of Things:State of the Art, Springer, 2016.

Ravesa Akhter, Shabir AhmadSofi, Precision Agriculture using IoT Data Analytics and Machine Learning, Elsevier, 2021.

Abhinav Sharma, Arpit Jain, Prateek Gupta, Vinay Chowdary, Machine Learning Applications for Precision Agriculture: A Comprehensive Review, IEEE, 2020.

Andreas Kamilaris, F. Prenafeta-Boldu, Deep Learning in Agriculture: A Survey, Elsevier, 2018.

Mohamed Torky, Abdoul Ella Hassanein, Integrating Blockchain and Internet