Crop Recommendation System

Shivam Gupta[1]
School Of Computer Science and Engineering
Lovely Professional University, Phagwara, Punjab, India
shivam.12007464@lpu.in

Abstract - Automating agricultural processes involves the mechanization of tasks in farming, either with or without human intervention. In regions with limited domestic land availability, making informed crop choices based on the prevailing environmental conditions has become crucial. While INDIA possesses substantial agricultural knowledge and manual techniques, there is currently no system in place that detects environmental factors and provides crop recommendations to users. This paper presents a theoretical and conceptual framework for a recommendation system that integrates multiple models for collecting environmental data using Arduino microcontrollers, employs machine learning techniques like Naïve Bayes (Multinomial) and Support Vector Machine (SVM), utilizes an unsupervised machine learning algorithm such as K-Means Clustering, and incorporates Natural Language Processing, particularly sentiment analysis, in the domain of Artificial Intelligence. The system's goal is to offer highly accurate and efficient recommendations for crop selection, tailored to the specific conditions of the chosen land.
The need for such a system arises from the challenge of determining what crops to cultivate, not only on domestic but also on farming lands. This challenge stems from the uncertainty of environmental factors like temperature, water levels, and soil conditions, which are subject to change over time. To address these issues, the proposed crop recommendation system predicts the most suitable crop types for a given area by gathering and processing environmental data related to plant growth through trained sub-models within the overarching system.

Keywords—Agriculture, Machine Learning, Arduino, Recommender System, Farming

I. INTRODUCTION

Modern farms and agricultural operations differ significantly from those of previous decades, primarily due to the remarkable advancements in technology. These advancements include the integration of sensors, devices, machines, and information technology. Today's agriculture heavily relies on sophisticated technologies such as robots, temperature and moisture sensors, aerial imaging, and GPS technology, along with an array of complex IoT (Internet of Things) devices. These cutting-edge agricultural technologies empower both businesses and farmers to enhance their profitability, efficiency, safety, and environmental sustainability. The emergence of digital agriculture and its associated technologies has unlocked a wealth of new data opportunities. Remote sensors, cameras, and various interconnected devices have the capacity to continuously gather data across entire farms or land parcels, monitoring factors like plant health, soil condition, temperature, humidity, and more. The volume of data generated by these sensors is staggering. This wealth of information allows farmers to gain a deeper and more precise understanding of their agricultural circumstances through advanced technology. It informs them rapidly and accurately about the conditions on the ground. Environmental data collected by remote sensors is processed through algorithms and statistical analyses, which can be interpreted and leveraged by farmers for decision-making and farm management. The more data and statistical information collected, the more refined the algorithms become at predicting outcomes. The ultimate goal is for farmers to utilize these technologies to achieve improved crop yields by making more informed decisions in the field. By implementing a system that incorporates temperature, soil pH, and soil moisture detection, the captured data is processed using specific algorithms. The results are then fed into a centralized database linked to various research modules, enabling the main system to recommend the optimal crop types for the best possible yields in both small home gardens and large agricultural areas.

II. LITERATURE SURVEY

In this project's literature review, the research team diligently gathered and examined a wide array of sources, including patents, research papers, documents, newspapers, and magazine articles from diverse contexts. For instance, in a key research paper [1], the requirements and the drivers behind the shift toward precision agriculture [2] are explored, a shift motivated by the forces of globalization.
Precision agriculture involves site-specific farming and has demonstrated continuous improvements

over time. However, certain challenges persist. Notably, site-specific methodologies within such systems require vigilant oversight to achieve optimal results. Many outcomes are reliant on specific conditions, making the situation in farming indispensable. Any oversight or error may lead to significant damage to resources and crops.

The current research proposes a solution where major factors are simultaneously considered to simplify the system for users. Unlike other previous models, which often concentrate on individual parameters, this system accounts for all crucial factors essential for plant growth, processing them together using various algorithms. For example, tests are conducted to determine the rate of evaporation and its impact on plant growth when water is insufficient. A derived equation is presented as a result [3], expressed as ETo = K pan × E pan, where ETo represents reference crop evapotranspiration, Kpan denotes the pan coefficient, and E pan signifies pan evaporation. However, this equation is not without limitations [4]. It is primarily suited for smaller areas of land and may not be suitable for large-scale commercial cultivation, given its limited profit potential in smaller areas. Additionally, Sri Lanka experiences average rainfall that is generally favourable for various crop types. Hence, the water level alone is not a significant issue; other environmental factors play more pivotal roles in plant growth. These factors are interconnected, with temperature, for example, holding a critical role.

The overarching goal of previous research has been to predict the best crop type. However, once the predicted crop type is cultivated, the system's role traditionally concludes. In contrast, the proposed system in this paper features a feedback mechanism. Even after suggesting the optimal crop type, the system continues to monitor plant growth and provides feedback on any nutrient deficiencies in the farm. This proactive approach enables users to take necessary precautions promptly.

III. METHODOLOGY

1. Dataset Collection

To ensure the healthy growth of plants and achieve a bountiful harvest in India, specific environmental conditions like temperature, humidity, soil pH, sunlight, and soil moisture need to be met. These conditions vary depending on the plant varieties being cultivated. Our initial dataset has been sourced from multiple Indian agricultural resources, including the Department of Agriculture in India [5], agricultural literature, Indian agricultural websites [6], and various reports and research papers. This initial dataset has played a crucial role in training our crop recommendation model, ultimately enhancing its accuracy for the Indian agricultural context.

Table.1. Sample Dataset

	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

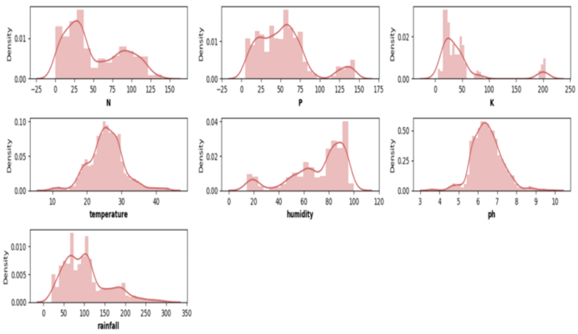

Fig.1. Data Plot of features

2. Collecting Environment Factors To perform a comparative analysis and make predictions based on the initial dataset, we needed to gather crucial environmental factors. We employed Arduino microcontrollers [7] for collecting these environmental parameters. Four sensors were utilized, as both temperature and humidity readings were obtained from a single microcontroller. These sensors included a sunlight intensity sensor, soil moisture sensor, soil pH sensor, and a humidity and temperature sensor. They were connected to an Arduino Wi-Fi module, and the collected data was transmitted to a database. Subsequently, the acquired data underwent a cleansing process and was subjected to clustering and other algorithms to prepare the values for the subsequent stages of the crop recommendation system. The processed data was also stored in the database for further analysis and use.

3. Prediction of the crop

Given the variation in environmental conditions from one region to another, a machine learning model is employed to determine the most suitable crop type for a specific plot of land. To train this crop recommendation model, we utilize data collected from the Arduino sensors, and we rely on machine learning algorithms [8] to identify the crop with the highest probability of successful growth. The selection of the best crop type is carried out through the application of Naïve Bayes and Support Vector Machine algorithms. These algorithms analyse various factors, including humidity, temperature, soil moisture, pH levels, and sunlight, to make recommendations. The system primarily suggests four crop types by assessing the factors mentioned above using these two machine learning algorithms. In addition to Naïve Bayes and Support Vector Machine, we also make use of Random Forest, XGBoost, and Logistic Regression in our modelling process.

4. Monitoring and Feedbacking the result

The main objective of this proposed system is to determine the most suitable crop types based on the environmental conditions of a chosen land plot. However, achieving a probability of more than 90% for these recommended crops can sometimes be influenced by soil conditions or other alterations in the selected land. To mitigate the impact of these variables on crop predictions, we have integrated a farmer's feedback system into the system.

Once a crop type is recommended, the farmer is prompted to provide regular details and feedback through a mobile application. This ongoing interaction helps guide the farmer in taking necessary precautions and adapting to changing conditions. The feedback system within the mobile application, linked to the chosen crop type, plays a crucial role in enhancing the overall accuracy and reliability of the product over time.

IV. RESULTS AND DISCUSSIONS

The proposed system was successfully put into practice on a selected land, and data were collected from the sensors at hourly intervals on a typical sunny day, as shown in Table I. Through various tests and the data gathered, the system can recommend the best crop for maximum yield. The system's accuracy is refined through the incorporation of farmer feedback, where invalid data is disregarded. For instance, if farmers consistently provide negative feedback for growing strawberries in Galle, the system adapts and enhances the accuracy of its recommendations, no longer suggesting strawberry cultivation in that region. Farmers can input their feedback in their native language, with custom libraries implemented to identify the language, primarily Sinhala and English, using natural language processing algorithms. Based on the current environmental conditions of the chosen land, the system suggests the top four suitable crop types to the farmer. Figure 2 displays a sample result provided to the farmer once the environmental factors are input. The PLU code is an attribute used for unique crop identification. The overall accuracy of the proposed system exceeds 92%. With continued use, as the system accumulates more data, the accuracy continues to improve. Over time, farmers can expect to achieve over 95% accuracy from the system.

Fig.2. Sample that is outputted to user

Table.2. System Metric Result

Name of the Module	Accuracy
Temperature and Humidity detection model	92%
2Sunlight intensity, soil pH and soil humidity detection model	95%
Crop recommending system	90%
Monitoring and feedback system.	96%

V. CONCLUSION

In the context of a modern Indian agricultural landscape characterized by limited space and varying levels of agricultural expertise, all factors are taken into consideration from the perspective of both the farmer and the crops. The farmer is provided with comprehensive guidance from the initial stages of cultivation to the point of harvesting. It is imperative to possess a deep understanding of the factors impacting cultivation and how to effectively manage them before selecting the appropriate crops to grow. This system automates the processing of these critical factors to recommend the most suitable crop type for cultivation. After the crops are planted, the farmer is encouraged to provide regular feedback at monthly intervals. This feedback contributes to the system's self-improvement over time, enhancing its accuracy as more data is accumulated. The system eliminates the need for specialized agricultural expertise, resulting in reduced maintenance. Therefore, implementing this system has no significant financial impact on users. Given the limited land space available to individuals in India, data collected by sensors indicates that this proposed system exhibits an accuracy rate of over 95% and is suitable for both rural and urban areas. Over time, with the accumulation of a significant amount of data and self-training of the system over approximately one year, the accuracy is expected to increase significantly. Experiments conducted over 4-6 months have confirmed the reliability and accuracy of the system. In a world where there is a growing demand for innovative agricultural consultancy solutions, this system is poised to lead the way in India..

VI. FUTURE WORKS

The system can be enhanced with a range of additional features. Currently, it accepts key environmental factors as inputs and recommends suitable crops for cultivation. In the next phase, an automation component can be incorporated to respond to user feedback. This extension could allow for the control of humidity, water levels, and other variables based on the farmer's requirements. While the system presently takes all environmental factors as inputs, an algorithm could be introduced as an extra capability to predict one factor using two others. For instance, it could predict soil pH levels based on soil moisture and sunlight, potentially reducing the initial sensor setup costs and making maintenance more accessible.

References

[1] Lakshmi.N, Priya.M, Sahana Shetty, and Manjunath C.R, Crop Recommendation System for Precision Agriculture, vol. 6 Reading, IND: International Journal for Research in Applied Science & Engineering Technology, 2018. [Online] Available: www.ijraset.com.Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park.

[2] Remi Schmaltz, "What is Precision Agriculture", April 2017. [Online]. Available: https://agfundernews.com/what-is-precision-agriculture.html [Accessed Feb.23, 2020].

[3] C. Brouwer and M. Heibloem, Irrigation Water Management: Irrigation Water Needs, manual 6 Reading, ITALY: Food and Agriculture Organization of the United Nations, 1987. [Online]Available:http://www.fao.org/3/s2022e/s2022e07.htm#TopOfPage.

[4] A. de Carbon, "PRECISION AGRICULTURE: ITS BENEFITS AND LIMITATIONS", carrhure, 2019 [Online]. Available: https://www.carrhure.com/precision-agriculture-benefits-limitations/ [Accessed: Feb.25, 2020].

[5] Department of Agriculture Sri Lanka (2015), 'Crop Suitability Recommendation for Grama Niladhari Divisions in Sri Lanka', Natural Resources Management Center, Peradeniya.

[6] Marie ionnotti (2020), 'Outdoors & Gardning', The Spruce Web Site

[7] WolkWriter (2019), 'Periodically measure environmental conditions and send them to WolkAbout IoT Platform to monitor the environment remotely.', ARDUINO PROJECT HUB

[8] Simon Tavasoli (202), '10 Machine Learning Algorithms You Need to Know', Simplilearn Solutions

[9] Sunil Ray (2015), '6 Easy Steps to Learn Naive Bayes Algorithm with codes in Python and R', Analytics Vidhya

[10] Rohith Gandhi (2018), 'Support Vector Machine — Introduction to Machine Learning Algorithms', Toward Data Science Web Site