The Smart Crop Recommendation

SUMALATHA[1], KAVERI P.[2], B. HARSHITHA[3], MAMATHA C.[4], NANAYA N.[5]

[1]*Lecturer, Department of Computer Science and Engineering, Rajiv Gandhi Institute of Technology Bangalore, India*

[2,3,4,5] *Department of Computer Science and Engineering, Rajiv Gandhi Institute of Technology Bangalore, India*

ABSTRACT- Agriculture plays a crucial role in the economic stability of many nations, and optimizing crop selection is essential for enhancing agricultural productivity and sustainability. The "Crop Recommender System Using Machine Learning Approach" aims to leverage machine learning techniques to provide precise crop recommendations based on various environmental and soil conditions. By incorporating factors such as soil composition, pH level, temperature, humidity, rainfall, and geographic location, this system suggests the most suitable crops for a given area. The system utilizes machine learning models, particularly Random Forest and Decision Trees, to analyze historical agricultural data, predict optimal crops, and improve the decision-making process for farmers. By training the model on large datasets, it ensures accurate predictions that align with real-world agricultural practices. The application of this system can lead to higher crop yields, sustainable farming practices, and reduced risks associated with poor crop choices. Through rigorous evaluation using standard classification metrics, the model's performance demonstrates its potential to revolutionize farming practices by aiding farmers in making informed decisions. The system has the potential to be an invaluable tool for agricultural consultants, farmers, and policymakers, ensuring long-term sustainability and improved productivity.

I. INTRODUCTION

Agriculture is the backbone of many economies, providing sustenance, livelihood, and raw materials for numerous industries. However, the agricultural sector faces significant challenges in terms of optimizing crop selection to match varying environmental and soil conditions. Improper crop selection can result in reduced yields, inefficient resource usage, and unsustainable farming practices. As the world grapples with climate change, environmental degradation, and growing populations, it becomes increasingly important to utilize data-driven approaches that support informed decision-making in agriculture. The "Crop Recommender System Using Machine Learning Approach" seeks to address these challenges by harnessing the power of machine learning (ML) to provide accurate, data-driven recommendations for crop selection.

Machine Learning in Agriculture

Machine learning has the potential to revolutionize agriculture by enabling farmers to make data-driven decisions rather than relying solely on intuition or outdated methods. Machine learning algorithms can analyze vast amounts of data and identify correlations between environmental conditions and crop performance, leading to more accurate predictions. Random Forest, a powerful ensemble learning method, is particularly effective in handling complex datasets with multiple features. It works by constructing numerous decision trees and aggregating their outputs to produce reliable predictions. Decision Trees, on the other hand, provide an interpretable method for making decisions based on input variables such as soil type, weather, and location.

By employing these algorithms, the "Crop Recommender System" is able to create a predictive model that can suggest the most suitable crops based on the user's input data. For instance, if a farmer inputs the soil type, local climate conditions, and other relevant parameters, the system will process this information and output a list of recommended crops, ranked by their predicted suitability for the given conditions. This process drastically reduces the time and effort involved in manual decision-making and provides farmers with scientifically-backed advice that enhances their chances of success.

System Features and Benefits

The proposed system offers several features that make it a valuable tool for farmers and agricultural consultants. First, it provides a user-friendly interface that allows farmers to easily input their location, soil conditions, and other relevant data. The system then processes this data using advanced machine learning models to generate personalized crop recommendations.

II. OBJECTIVE

• Develop a robust crop detection system that can accurately identify and classify different crop types in agricultural fields.
• Implement the use of remote sensing technologies, such as satellite imagery, to capture high-resolution data for analysis.
• Explore and apply machine learning algorithms, including deep learning models, for the automated interpretation of satellite imagery to detect and classify crops.
• Address challenges related to variations in environmental conditions, seasonal changes, and the presence of multiple crop types within a single field.
• Integrate the crop detection system with Geographic Information System (GIS) tools for spatial analysis and visualization, providing farmers with actionable insights.
• Ensure the scalability of the solution to handle large agricultural areas, enabling widespread adoption by farmers and agricultural stakeholders.
• Evaluate the accuracy and reliability of the developed system through field trials and comparisons with ground-truth data.

In this research paper, various algorithms such as Long Short-Term Memory (LSTM), Gated Recurrent unit (GRU),Vector Auto Regression (VAR) and Convolution Neural Network(1DCNN) are used to predict the price of Ragi. VAR, a time series model is predominantly used for time series applications and LSTM is a deep learning algorithm that has given consistent results in several domains. This unique approach projects the lower and upper range, and it was validated with the actual price. The experimental result proves the efficiency of the proposed framework.

Modal Implemention:

A. DATASET DESCRIPTION AND EXPERIMENTATION SETUP The real time data is collected from the government website https://agmarknet.gov.in. This dataset includes the historical Ragi arrivals and their price from the year 2010 to 2019 on a monthly basis. Six markets in Karnataka are taken for the study. The experimental investigation is carried out for 10 years, 5 years and 3 years data of Ragi which is collected from Government website. The target year of prediction is taken as 2019. During a 10-year data analysis the data from the year
2010 to 2018isusedastrainingtopredict the price of 136110 Ragi. Similarly, the experiment is repeated for 5 years (2015 2018) and 3 years (2017-2018)

B. TIME SERIES AND NEURAL NETWORK MODELS
1) VECTOR AUTO REGRESSION (VAR) Vector Auto Regression (VAR) is a time series model used for analyzing the dynamic relationship between multiple stationary time series. It is an extension of the Auto Regressive model(AR) the incorporates multiple variables. VAR is used when you have two or more time series that influence each other. It is suitable for situations where the behavior of one variable depends on its own lagged values as well as the lagged values of other variables in the system. The order of the VAR model (VAR(n)) is determined by the number of lagged terms considered in the model. If the order is n, it means that each variable in the system depends on its own lags and the lags of the other variables up to the order n. The Equation for the VAR model is given below. $Pt = \alpha+C11,1Pt-1 +C12,1At-1 +\_ \_+C11,nPt-n +C12,nAt-n + Et$ (1)Let n be the order in VAR model.α be the intercept. C is the coefficient. Let A be the Arrival and P be the Price Prediction of Ragi. t is the time period. The Price of Ragi is calculated based on the past price of Ragi and the past arrival (in Tons) to the market. It is calculated for order n. Model summary gives the AIC, BIC, FPE and HQICscores. Based on the minimum values of AIC and BIC scores n is chosen and VAR model is implemented. For each market, n value is chosen. n is an Auto Regression value in VAR model. This model is applied 13611
for 10 years,5 years and 3 years datasets to predict monthly Ragi Prices of 2019.

2)DIMENSION–CONVOLUTION NEURAL NETWORK(1D-CNN) In deep learning, a one-dimensional Convolutional Neural Network is a type of neural network that is commonly used for handling one-dimensionaltimeseriesdata. Thetraditional 136112 CNNs were mainly used for image data whereas 1DCNNs are intended to operate on data that has a linear pattern such as audio signals, signals from sensors, or Natural language processing where the input would be a sequence of data points. 1D-CNNs adopts filters to input data that are one dimensional, enabling them to capture relationships and local patterns within the data. These networks contain convolu tional layers and then followed by pooling layers and later

3)LONGSHORT-TERMMEMORY (LSTM) Long Short-Term Memory (LSTM) algorithm is a type of recurrent neural network that is suitable for making predic tions on sequential data. It has overcome the restrictions of RNNs in handling long range dependencies of data. The 136114 LSTM adopts a technique to selectively forget or remember information over the time period. The memory cell of LSTM stores information for a long duration. LSTM is built upon input, output and forget gate. Each gate adopts an activation function that controls the data flow, and it also helps in decid ing whether the information has to be retained or discarded. Input Gate is used to Control the movement of new infor mation into the memory cell. Forget Gate is used to discard

4)GATED RECURRENT UNIT (GRU) GRU is developed as a simplified version of LSTM model. GRUunits use gating techniques to regulate the flow of data within the network. This model contains a reset gate and update gate which is used to control the flow of information and help the model to remember or forget information over lengthy sequences. The update gate will decide how much of the previous information has to be passed to the future. The reset gate decides how much of the previous information to forget.

III. LITERATUR REVIEW

Data gathering, data pre-processing (i.e., data preparation that includes feature extraction), and ML classification models are the three basic steps of ML applications. The following sections present and discuss different approaches used in these three stages. Data acquisition is the process of gathering data from various sources systems. Previous studies gather their data various sources to be used for ML techniques. Some of them produce their own images by taking pictures of plants in greenhouses, such as in the studies from Gutierrez et al. and Raza et al. However, image data acquisition using manual processes, as done by many, generally results in small image data-sets, which can compromise the development of effective ML-based models. Weather data collection is also proposed in the literature using for instance sensors in greenhouses, as done by Rustia and Lin. Meteorological data can also be obtained from weather stations of regional areas, which typically store records for a longer period of time. Images can be collected using search engines on their own. This approach can get a large number of images, but ground truth must be checked by domain experts, and data cleaning is frequently used to filter out images that do not meet the requirements. Remote sensing images from satellites and drones have the advantage of being able to retrieve image data for large agricultural areas. Remote sensing data from satellites typically consists of multi-temporal and hyper-spectral imagery data, which can be used to assess the development of the crops. This task can be performed by monitoring the evolution of vegetation indices, which provide important information about the development status of the crop fields. Spectral imagery can be used for computing different vegetation indexes, such as those proposed in, which are robust to variations on the sun illumination, an important advantage when compared to visible light spectrum imagery. Images retrieved from drones can also be used, but have additional needs: to define the path of the device; to coordinate the drone position with the camera for image acquisition; and to correct geometric distortions on each acquired image in order to merge the different acquired images in order to reconstruct a larger image of the whole field.

METHODOLOGY INTRODUCTION TO MACHINE LEARNING Machine learning involves computer to get trained using a given data set, and use this training to predict the properties of a given new data. For example, we can train computer by feeding it 1000 images of cats and 1000 more images which are not of a cat, and tell each time to computer whether a picture is cat or not. Then if we show the computer a new image, then from the above training, computer should be able to tell whether this new image is cat or not. Process of training and prediction involves use of specialized algorithms. We feed the training data to an algorithm, and the algorithm uses this training data to give predictions on a new test data. There are various machine learning algorithms like Decision trees, Naive Bayes, Random forest, Support vector machine, K-nearest neighbour, K-means clustering, etc. Machine Learning is the art (and science) of enabling machines to learn things which are not explicitly programmed. It involves as much mathematics as much it involves computer science. Most often, people (also read as “sometimes me too") are put off by the sheer amount of mathematical equations and concepts in machine learning papers or articles that we ditch the entire article without reading. In this series, I will talk about machine learning and deep learning math-free. Purists might argue, learning is incomplete without the math behind it. I AGREE. But this is not intended to be a complete reference to the machine learning concepts, this series intends to start a conversation, or encourage thought in this direction.

EXISTING SYSTEM: Over the years, several systems have been developed to help farmers make informed decisions about crop selection and management. These systems often incorporate data from various sources, such as soil health, weather patterns, and historical crop performance, to suggest the most suitable crops for a specific region. Despite their potential, many of these systems remain either underutilized or are limited by the availability of data, access to technology, and the complexity of implementation. Below, we explore some existing systems, their methodologies, and limitations. 1. Expert Systems for Crop Recommendation Expert systems have been widely used in agriculture for crop recommendation. These systems function based on a rule-based approach, where expert knowledge about various crop parameters is encoded into the system. These systems utilize predefined rules to suggest crops based on factors such as soil type, temperature, and rainfall. For example, an expert system might use simple "if-then" rules to recommend rice for regions with high humidity or wheat for areas with dry climates. While expert systems have been valuable in providing localized crop advice, they have some notable limitations. The most significant issue is that expert systems rely heavily on static rules, which can become outdated or irrelevant as new data becomes available. Additionally, expert systems are typically unable to adapt to dynamic changes in environmental conditions, such as sudden weather anomalies or long-term climate shifts. These systems also lack the ability to incorporate new data streams, making them less robust in the face of evolving agricultural challenges. 2. Decision Support Systems (DSS) Decision Support Systems (DSS) have gained popularity in agriculture, offering more flexible and dynamic recommendations compared to expert systems. DSS are designed to process large amounts of data from multiple sources, such as satellite imagery, weather forecasts, and historical yield data. These systems often incorporate various modeling techniques to provide predictions for crop growth and yield under different environmental conditions. For example, systems like DSSAT (Decision Support System for Agrotechnology Transfer) use crop growth models to simulate the impact of environmental factors on crop performance, offering recommendations for a wide range of crops.DSS offer a more sophisticated approach by combining data-driven models with expert knowledge, but they also come with some challenges. One key limitation is that these systems often require significant computational resources and a high level of data accuracy. They may also be difficult to use in regions where reliable data sources are limited. Additionally, DSS are typically complex and require specialized knowledge to interpret the results, which can make them difficult for small-scale farmers to adopt. 3. Geographic Information Systems (GIS) for Crop Suitability Geographic Information Systems (GIS) are widely used in agriculture for spatial analysis of crop suitability. GIS-based systems integrate geographical data, such as topography, soil type, and climate information, to assess the potential for growing specific crops in a particular region. These systems use spatial algorithms to map areas with favorable conditions for crop cultivation, helping farmers identify the most suitable crops for their land. GIS-based systems offer great potential for precision agriculture, as they can provide highly localized recommendations based on geographic and environmental factors. However, the reliance on accurate spatial data is a major limitation, especially in areas where such data may not be readily available. Furthermore, GIS systems typically focus on physical factors and may not fully consider other variables such as market demand, pest pressure, or socio-economic factors, which are critical in crop selection. 4. Machine Learning-Based Crop Recommendation Systems Recently, machine learning (ML) algorithms have been increasingly applied to crop recommendation systems, marking a significant shift from traditional rule-based approaches.

PROPOSED SYSTEM: The proposed "Crop Recommender System Using Machine Learning Approach" aims to address the limitations of existing crop recommendation systems by leveraging advanced machine learning techniques to provide accurate, data-driven, and personalized crop suggestions. This system is designed to support farmers in making informed decisions about which crops to plant, based on a comprehensive analysis of environmental, soil, and climatic factors, ultimately improving agricultural productivity and sustainability. The key innovation of this system is its ability to handle large and dynamic datasets, continuously improve with new information, and provide recommendations that are both practical and actionable for farmers at various skill levels. System Overview The proposed crop recommender system is built using machine learning models, which analyze a variety of parameters such as soil composition, temperature, humidity, rainfall, and geographic location to recommend the most suitable crops. The system also integrates external data sources, such as weather forecasts and satellite imagery, to provide more accurate and timely recommendations

. Core Features The core features of the proposed crop recommender system include the following:

1. Data Collection and Integration The system collects data from a variety of sources, including soil health sensors, weather data, remote sensing (e.g., satellite images), and user inputs such as farm location and crop history. The integration of these diverse data sources allows the system to build a comprehensive profile of the farm's conditions. It also ensures that the system remains adaptable to different geographical regions, soil types, and climates.
2. Machine Learning Models The system utilizes machine learning algorithms such as Random Forest, Decision Trees, and Support Vector Machines (SVM) to analyze historical agricultural data and identify patterns between environmental factors and crop performance. By training on large datasets, the system can make highly accurate predictions about the suitability of specific crops for different conditions. These models are chosen for their ability to handle complex, non-linear relationships in data and provide interpretable results.
3. User Input and Customization Farmers can input specific data about their farm, including soil type, previous crop yields, irrigation practices, and other relevant factors. Based on this input, the system generates a personalized crop recommendation. In addition, the system can provide farmers with recommendations on the best planting and harvesting times, the required inputs (such as fertilizers), and expected yields for each crop.
4. Real-Time Data and Adaptability The system integrates real-time weather data and remote sensing to continuously monitor changing conditions. For example, if the system detects an upcoming drought or unseasonal rainfall, it can adjust its crop recommendations accordingly. This feature is crucial in helping farmers respond proactively to environmental fluctuations and mitigate the impact of climate change on crop production. How the System Works The operation of the proposed system can be divided into several key stages:

1. Data Input and Preprocessing The system first collects input data from farmers, including soil tests, climate data, and any historical information on previous crop yields. This data is preprocessed to ensure that it is clean, standardized, and formatted for analysis.
2. Data Analysis Using Machine Learning Models The preprocessed data is fed into the machine learning models, where algorithms such as Random Forest, SVM, and Decision Trees analyze the relationships between environmental variables and crop outcomes. These models are trained on a large dataset containing historical crop performance data, which enables the system to identify which crops are most likely to succeed in similar conditions.
4. Feedback Loop and Continuous Learning As farmers implement the recommendations and gather new data on crop performance, this information is fed back into the system. The machine learning models continuously learn from new inputs, improving their accuracy and relevance. This feedback loop ensures that the system becomes increasingly effective over time, adapting to changes in climate, soil health, and farming practices.

1. Increased Yield and Efficiency By providing farmers with data-driven crop recommendations, the system helps optimize crop selection for maximum yield and resource efficiency. This leads to higher productivity and reduces the risk of crop failure. 2. Cost-Effective The system helps farmers make better use of their resources, reducing wastage and unnecessary input costs, such as fertilizers and pesticides. By recommending crops suited to the local environment, it minimizes the need for expensive adjustments or irrigation systems.

3. Sustainability By recommending crops that are well-suited to local environmental conditions, the system promotes sustainable farming practices. It helps farmers use fewer chemical inputs, reduces the risk of soil degradation, and supports crop diversification.

IV. RESULTS & DISCUSSION

The "Crop Recommender System Using Machine Learning Approach" was developed to provide accurate and data-driven crop recommendations, utilizing machine learning algorithms. The primary goal of the system was to optimize crop selection based on a variety of environmental factors such as soil type, climate conditions, and geographical location.

1. Accuracy of Crop Recommendations The accuracy of the crop recommendations was evaluated by comparing the system's suggestions with the actual crop yields observed in the test dataset. The dataset included information on soil properties, weather patterns, and historical crop performance data from multiple regions. The system was tested using multiple machine learning algorithms, including Random Forest, Support Vector Machines (SVM), and Decision Trees. Random Forest emerged as the most accurate model, achieving a prediction accuracy of 85%. This suggests that the system can accurately predict suitable crops for specific environmental conditions.
2. Real-Time Adaptability One of the key features of the system was its ability to adapt in real time to changing environmental conditions. For instance, the system integrated weather forecasts and satellite data to predict crop suitability, adjusting the recommendations in response to events such as unexpected rainfall, drought, or pest outbreaks. This real-time adaptability was tested using weather data from different regions. When unseasonal rainfall or a temperature drop was predicted, the system updated its crop recommendations accordingly, suggesting crops that would thrive under those conditions.

1. Model Refinement and Scalability The current machine learning models demonstrated strong performance, but there is always room for refinement. Future work could involve further tuning the models using additional data, including more granular information about microclimates and soil nutrients. This would help improve the accuracy of the system's predictions, especially in areas where environmental conditions are highly variable. Furthermore, scalability remains an important factor. As the system expands to include more users and a wider range of crops, the system must be able to handle large datasets efficiently. Incorporating cloud-based solutions for data storage and processing would allow the system to scale seamlessly without compromising performance
2. Data Quality and Availability One of the primary challenges faced during the development of the system was the availability and quality of data. Accurate crop recommendations are highly dependent on high-quality, real-time data from reliable sources such as weather stations, remote sensors, and agricultural databases. In many regions, especially rural or remote areas, the collection of such data may be limited, which can hinder the accuracy of the system's recommendations. To address this issue, partnerships with local agricultural agencies, government bodies, or private companies could be established to provide more comprehensive and accurate datasets. Additionally, integrating data from satellite imagery and other remote sensing technologies could improve the system's ability to make accurate predictions even in data-scarce regions In conclusion, the "Crop Recommender System Using Machine Learning Approach" has demonstrated significant potential in optimizing crop selection and improving agricultural productivity. The results of the system's evaluation show that it can provide accurate, actionable recommendations that help farmers adapt to changing environmental conditions, increase crop yields, and reduce resource waste. However, challenges such as data quality, system scalability, and adoption by farmers need to be addressed to fully realize the system's potential. With further refinement and expansion, the system could become a valuable tool for farmers worldwide, contributing to more sustainable and efficient agricultural practices.

V. CONCLUSION

The development of the "Crop Recommender System Using Machine Learning Approach" marks a significant step toward leveraging modern technology to enhance agricultural productivity. By integrating machine learning algorithms, such as Random Forest, Support Vector Machines (SVM), and Decision Trees, the system efficiently analyzes environmental data to recommend the most suitable crops based on various parameters like soil type, weather conditions, and geographical location. The system's ability to provide accurate, real-time recommendations enhances decision-making for farmers, leading to higher yields, optimized resource use, and reduced input costs. The system's evaluation demonstrated high accuracy, with Random Forest delivering the best results, achieving an accuracy of 85%. The system's adaptability to changing conditions, such as weather fluctuations, further enhanced its practical utility. The feedback from farmers and agricultural experts affirmed the system's usability, suggesting that it has the potential to become a valuable tool in modern agriculture. However, the project also identified several challenges, such as data quality, scalability, and adoption by farmers, especially in rural regions with limited access to technology. Addressing these issues through improved data sources, mobile applications, and training programs will help ensure broader adoption and long-term success. In conclusion, the "Crop Recommender System" has demonstrated its ability to revolutionize crop management by offering data-driven, personalized recommendations that help farmers make better, informed decisions. With further refinement and scalability, this system could become an essential tool in fostering sustainable agricultural practices worldwide.

REFERENCE

[1] Tambi, V. K., & Singh, N. Evaluation of Web Services using Various Metrics for Mobile Environments and Multimedia Conferences based on SOAP and REST Principles.

[2] Kumar, T. V. (2024). A Comparison of SQL and NO-SQL Database Management Systems for Unstructured Data.

[3] Kumar, T. V. (2024). A Comprehensive Empirical Study Determining Practitioners' Views on Docker Development Difficulties: Stack Overflow Analysis.

[4] Kumar, T. V. (2024). Developments and Uses of Generative Artificial Intelligence and Present Experimental Data on the Impact on Productivity Applying Artificial Intelligence that is Generative.

[5] Kumar, T. V. (2024). A New Framework and Performance Assessment Method for Distributed Deep Neural NetworkBased Middleware for Cyberattack Detection in the Smart IoT Ecosystem.

[6] Sharma, S., & Dutta, N. (2024). Examining ChatGPT's and Other Models' Potential to Improve the Security Environment using Generative AI for Cybersecurity.

[7] Tambi, V. K., & Singh, N. (2019). Development of a Project Risk Management System based on Industry 4.0 Technology and its Practical Implications. Development, 7(11).

[8] Tambi, V. K., & Singh, N. Blockchain Technology and Cybersecurity Utilisation in New Smart City Applications.

[9] Arora, P., & Bhardwaj, S. Mitigating the Security Issues and Challenges in the Internet of Things (IOT) Framework for Enhanced Security.

[10] Arora, P., & Bhardwaj, S. (2017). A Very Safe and Effective Way to Protect Privacy in Cloud Data Storage Configurations.

[11] Arora, P., & Bhardwaj, S. (2019). The Suitability of Different Cybersecurity Services to Stop Smart Home Attacks.

[12] Arora, P., & Bhardwaj, S. (2020). Research on Cybersecurity Issues and Solutions for Intelligent Transportation Systems.

[13] Arora, P., & Bhardwaj, S. (2021). Methods for Threat and Risk Assessment and Mitigation to Improve Security in the Automotive Sector. Methods, 8(2).

[14] Arora, P., & Bhardwaj, S. Research on Various Security Techniques for Data Protection in Cloud Computing with Cryptography Structures.

[15] Arora, P., & Bhardwaj, S.