INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1404-1411

e-ISSN : 2583-1062

Impact Factor: 5.725

GEOLOCATION BASED CROP PREDICTION SYSTEM

Prachi Lidbe[1], Maithili Gajbhiye[2], Srushti Meshram[3], Mrudul Deogade[4], Pranav Belorkar[5]

Prof Bhagyashree Dharaskar[6]

[1,2,3,4,5]Priyadarshini College of Engineering, India.

[6]Guide, Priyadarshini College of Engineering, India.

ABSTRACT

In India, where agriculture stands as the cornerstone of the economy and livelihoods for over 70% of the rural population, a critical challenge lies in the conventional farming practices.

Despite being the world's second most populous nation, farmers often engage in repetitive cultivation without exploring the potential benefits of diversifying their crops. Moreover, the arbitrary application of fertilizers, devoid of consideration for nutrient deficiencies, contributes to diminishing crop yields and exacerbates issues like soil acidification and topsoil degradation. Recognizing these challenges, our team has developed a sophisticated system employing machine learning algorithms to empower farmers with precise recommendations tailored to their specific land characteristics and prevailing weather conditions.

The core functionality of our system revolves around suggesting the most suitable crops for cultivation, factoring in crucial parameters such as soil content and weather patterns. Beyond crop recommendations, the system provides invaluable insights into the required content and optimal quantities of fertilizers, as well as the most suitable seeds for cultivation.

By offering this comprehensive guidance, our aim is to revolutionize traditional farming practices, allowing farmers to make informed decisions that not only enhance crop yields but also contribute to sustainable agricultural practices. This approach not only stands to increase farmers' profit margins but also addresses broader environmental concerns related to soil health and productivity.

Agricultural transformation in India hinges on the adoption of advanced technologies. Traditionally, farmers relied on personal experience and neighboring trends for crop predictions. However, our system breaks away from this reliance on anecdotal knowledge by harnessing the power of machine learning algorithms. This shift enables accurate and data-driven crop output projections, empowering farmers to make choices that optimize their agricultural endeavors. As we delve into various attribute selection techniques for crop prediction, this research marks a significant step toward the future of agricultural output prediction systems.

By embracing cutting- edge technology, we pave the way for a more sustainable and profitable future for Indian agriculture, benefiting both farmers and the broader economy.

Keywords: Indian Agriculture, Farming Practices, Crop Diversification, Machine Learning Algorithms, Sustainable Agriculture, SVM, Rainfall prediction Agricultural Technology, Technological Advancements

1. INTRODUCTION

Agriculture is a cornerstone of India's economy, utilizing over 60% of the country's land and providing sustenance for a population exceeding 1.3 billion. However, traditional farming practices, based on localized experiences, often lead to repetitive cultivation and inadequate nutrient management, resulting in diminished yields and soil pollution. Recognizing the need for a transformative approach, we have developed a system integrating machine learning algorithms to revolutionize agricultural practices.

Machine learning, a subset of artificial intelligence, is harnessed to recommend optimal crops based on specific weather parameters and soil content, drawing data from reliable sources. This comprehensive system, incorporating Support Vector Machine (SVM) and Decision Tree algorithms, not only suggests suitable crops but also provides invaluable information on required nutrients, approximate yield, seed quantity, and market prices, aiming to enhance profitability for farmers. In our active approach, highly optimized machine learning predictive algorithms leverage historical data for accurate yield estimations.

These supervised machine learning models, particularly classification algorithms, prove instrumental in achieving the system's objectives. Conversely, the passive approach, known as the blind approach, involves loading external crop datasets, preprocessing data, and training models using Decision Tree classifiers.

This method considers factors such as temperature, humidity, soil pH, and predicted rainfall for crop predictions. This dual-pronged approach operates at the farm scale, offering a sophisticated and data-driven solution to improve agricultural outcomes. Recognizing the importance of machine learning in enhancing yield predictions, we conducted a systematic literature review (SLR) to identify existing gaps in research and guide future studies in this domain.

Objectives

The project aims to determine crop yield by considering relevant features such as temperature, moisture, rainfall, and previous production data. Using regression models, the objectives include providing a user-friendly interface for farmers, identifying the best crops for specific soil and weather conditions, suggesting alternative crops to enhance productivity, and offering farmers information on fertilizers and pesticides. The overarching goal is to empower farmers with data-driven insights for more informed decision-making in agriculture.

2. LITERATURE REVIEW

Jetendra Shenoy et al.

this paper proposes a solution to minimize transportation costs by employing an IoT-based approach. The innovative strategy aims to streamline the communication chain between farmers and end-users by reducing the number of intermediary hops and agents. This approach, designed to benefit farmers, serves as a motivating factor for our project. The emphasis on cost reduction and efficiency through IoT applications underscores the potential for transformative solutions in agricultural logistics, influencing the direction of our own endeavors.

Rahul Katarya et al.

explores diverse machine learning strategies to enhance crop yield, delving into various artificial intelligence (AI) techniques, including machine learning algorithms and big data analysis tailored for precision agriculture. The paper elucidates the intricacies of a crop recommender system, employing methodologies such as KNN, Ensemble-based Models, and Neural networks. By comprehensively addressing different artificial intelligence techniques, the research provides a nuanced understanding of their applications in optimizing crop productivity. The focus on precision agriculture underscores the significance of leveraging advanced technologies to tailor recommendations and strategies for improved agricultural outcomes.

Zihao Hong et al.

Use Support Vector Machine (SVM) and Relevance Vector Machine (RVM), two supervised machine learning techniques, to accurately forecast soil quality. A smart wireless gadget that can sense soil moisture and meteorological data is incorporated into the investigation. Although the wireless gadget has an impressive accuracy rate of 95%, it is important to acknowledge an error rate of 15%. Notably, there hasn't been any testing done on the device's performance in real-time data settings, which calls for more research into its usefulness.

Ashwani Kumar Kushwaha et al.

explores innovative crop yield prediction methods, proposing suitable crops to enhance farmer profits and elevate agriculture sector quality. The research emphasizes the acquisition of substantial data, labeled as big data, encompassing soil and weather information. Utilizing the Hadoop platform and an agro algorithm, this study processes the extensive dataset to predict crops suitable for specific conditions, aiming not only to improve farmers' financial outcomes but also to enhance overall crop quality within the agricultural sector. The integration of advanced technologies signifies a transformative step towards sustainable and informed farming practices, bridging the gap between environmental factors and successful crop cultivation.

Li, Lecourt et al.

conducted a review study on fruit ripeness determination for optimal harvest timing and yield prediction. The paper addresses the challenge faced by Indian agriculturists in manually choosing fertilizer quantities, risking harm to plants and yield reduction. Emphasizing the importance of nutrient levels, the study provides an overview of diverse data mining frameworks applied to cultivate soil datasets for precise fertilizer recommendations, aiming to enhance agricultural practices and optimize crop yields.

Sabri Arik et al.

research focuses on assessing Soil Fertility and Plant Nutrient levels through the implementation of a backpropagation algorithm. The outcomes exhibit precision, contributing to enhanced soil properties. Despite its superior performance compared to conventional methods, the system is characterized by occasional inefficiency, slowing down, and instability. These limitations highlight areas for potential refinement, emphasizing the need for further optimization to achieve a more stable and efficient system in the evaluation of soil fertility and nutrient levels.

P. Vinciya et al.

this paper investigated the Multiple Linear Regression (MLR) is employed for crop analysis, complemented by the utilization of a Decision Tree algorithm and Classification to scrutinize over 362 datasets. The training dataset is meticulously categorized into organic, inorganic, and real estate classes, facilitating the accurate prediction of soil types. The outcomes generated by this system are not only precise but also deemed highly reliable, attesting to the effectiveness of the employed techniques in soil analysis and classification.

M.R. Bendre et al.

explores the prediction of crop yields involves a detailed analysis, leading to a categorization of crops based on data mining algorithms. The paper provides valuable insights into diverse classification rules such as Naive Bayes and K-Nearest Neighbour. Utilizing this information, our project conducts a thorough examination of these classification rules to identify their appropriateness for the specific dataset we are employing. Additionally, the exclusion criteria for our analysis excludes survey or traditional review papers, focusing specifically on related work discussed in this section.

Shivnath Ghosh et al.

explores information is input into a Back Propagation Network to assess the test dataset. This network incorporates a hidden layer, enhancing its predictive efficacy for soil properties. The Back Propagation Network is utilized to construct a self-trained function for predicting soil properties with specified parameters, resulting in heightened accuracy compared to conventional methods. Despite its superior performance, occasional system slowdowns and inconsistencies in the output have been observed.

Vaneesbeer Singh et al.

investigates three methodologies— Decision Tree, Naive Bayes Classifier, and KNN Classifier—are employed to analyze soil conditions and forecast crop yield. However, to enhance accuracy, the study suggests the utilization of rule-based induction and SVM, highlighting their potential for more precise results. The comparison underscores the need for advanced techniques in soil analysis and yield prediction, guiding towards improved agricultural decision-making for optimal productivity.

Girish L et al.

investigates crop yield and rainfall prediction through a machine learning methodology, exploring various approaches within this domain. The study delves into distinct machine learning algorithms, including linear regression, SVM, KNN, and decision tree, assessing their efficiency in predicting both rainfall and crop yield. The comprehensive analysis leads to the conclusion that among these algorithms, SVM demonstrates the highest efficacy specifically in the context of rainfall prediction. This research contributes valuable insights into the comparative performance of diverse machine learning techniques, providing a nuanced understanding of their applicability in predicting essential agricultural factors such as rainfall and crop yield.

Chlingaryan et al.

investigates a comprehensive review study on the estimation of nitrogen status using machine learning. Their findings suggest that the rapid advancements in sensing technologies and machine learning techniques hold the potential for cost-effective solutions in the agricultural sector.

Elavarasan et al.

carried out a review of papers that focused on machine learning models for climatic parameter-based crop yield prediction. The study suggests expanding the search to include other factors affecting crop productivity and taking into account topics like crop management, animal management, water management, and soil management in their research.

3 DESIGN AND ANALYSIS

3.1 Requirement Analysis

The systematic review followed the guidelines by Kitchenham et al. (2007), involving three stages: planning, conducting, and reporting. Research questions were defined, and databases like Science Direct, Scopus, Web of Science, Springer Link, Wiley, and Google Scholar were used. Relevant studies were filtered and assessed using exclusion and quality criteria. Extracted data were synthesized to address research questions. In the planning stage, research questions, protocol, publication venues, search strings, and selection criteria were identified and validated. The search involved a broad automated query ("machine learning" AND "yield prediction") in six databases, yielding 567 studies. The search was refined using a complex string: (("machine learning" OR "artificial intelligence") AND "data mining" AND ("yield prediction" OR "yield forecasting" OR "yield estimation")). For crop prediction, external datasets are loaded, followed by data pre-processing. Decision tree classifier models are trained using factors like temperature, humidity, soil pH, and predicted rainfall. The system predicts crops based on input parameters, recommending the most suitable crop for cultivation. Details on required fertilizers, seeds, current market prices, and approximate yield are provided to assist farmers in choosing the most profitable crop.

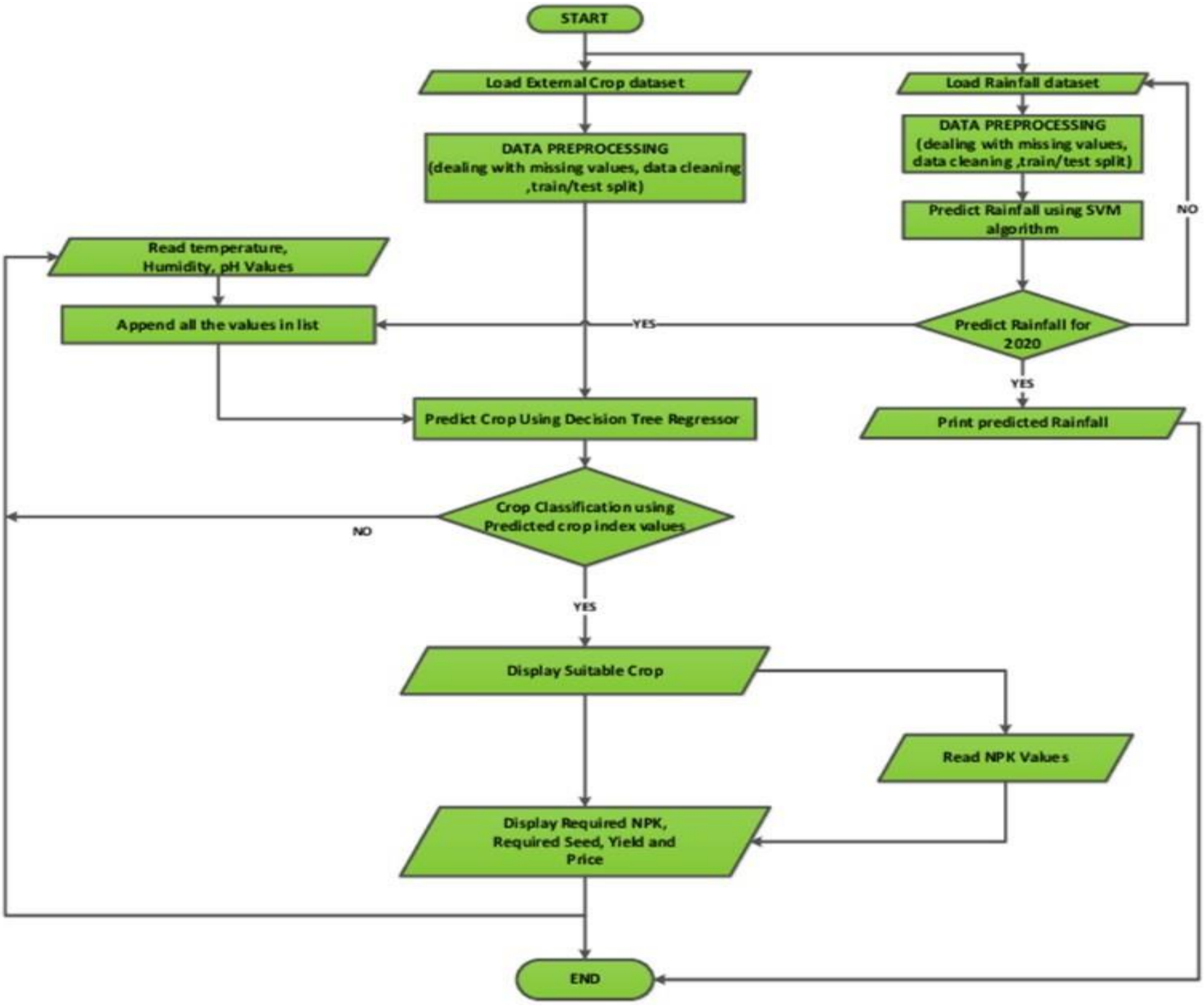


Fig. 1 Crop Prediction Flow Diagram

3.2 Components Used

1. Input Image

The "Input Image" for a Geolocation-Based Crop Prediction System comprises crucial components such as geographic information, climate data (temperature, humidity), soil characteristics (pH, fertility), topography details, historical crop performance, rainfall predictions, sensor data, user-provided information, crop rotation history, and remote sensing imagery. These diverse elements collectively form the basis for accurate predictions tailored to the specific conditions of a given location. The system utilizes this comprehensive input to train machine learning models, offering farmers personalized and data-driven recommendations for optimal crop selection and cultivation practices, ultimately enhancing yield and profitability.

2. N, P, K value

In the Geolocation-Based Crop Prediction System, the N (Nitrogen), P (Phosphorus), and K (Potassium) values are critical components influencing crop recommendations. These values are indicative of the soil's nutrient content, crucial for understanding its fertility. The system considers these nutrient levels in the soil, using them as parameters in its predictive models. By assessing N, P, and K values, the system determines the suitability of the soil for specific crops, offering insights into the required fertilizers for optimal cultivation. This information is invaluable to farmers, guiding them on the precise amounts of nitrogen, phosphorus, and potassium needed for their chosen crops, thereby enhancing agricultural productivity and sustainability. The integration of N, P, and K values underscores the system's ability to provide comprehensive and tailored guidance for efficient crop management.

3. Model Training for prediction

In the Geolocation-Based Crop Prediction System, the Model Training for prediction is a pivotal stage that leverages machine learning algorithms to enhance the accuracy of crop predictions. During this phase, the system utilizes historical data, including geographic information, climate variables, and soil characteristics, to train predictive models

INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1404-1411

e-ISSN : 2583-1062

Impact Factor: 5.725

such as the Decision Tree Classifier. The models learn intricate patterns and relationships within the data, enabling them to make informed predictions about suitable crops for a specific location. The training process involves iteratively adjusting the model parameters until it achieves optimal performance in predicting crop outcomes. By using advanced algorithms, the system enhances its ability to recognize complex dependencies and variations in environmental factors, contributing to more precise and context-specific crop recommendations. The effectiveness of the Geolocation-Based Crop Prediction System relies significantly on the thoroughness and accuracy of this model training process, ensuring that farmers receive reliable insights for successful and sustainable crop cultivation.

4. Land for Scaling

"Land for Scaling" component is crucial for implementing and scaling the predictive models developed. This involves considering the geographical extent and characteristics of the land where the crop predictions will be applied. The system must account for variations in soil types, climate conditions, and topography across different regions. Additionally, the scalability of the predictive models is vital to ensure their applicability to diverse agricultural landscapes. By addressing the "Land for Scaling" component, the system can be tailored to suit the specific needs and nuances of various farming environments, allowing farmers across different regions to benefit from accurate and localized crop recommendations. This emphasis on scalability ensures the practicality and effectiveness of the Geolocation-Based Crop Prediction System on a broader scale, contributing to its utility in diverse agricultural settings.

5. Technology and Software Details Hardware requirements:

An intel core 3 processor with 8th generation, a minimum of 4 GB of RAM

Software requirements:

Python [PyCharm community edition/Jupyter notebook], GUI library PyQt5, TensorFlow Keras API, Pandas, etc.

4 PROJECT DESCRIPTION

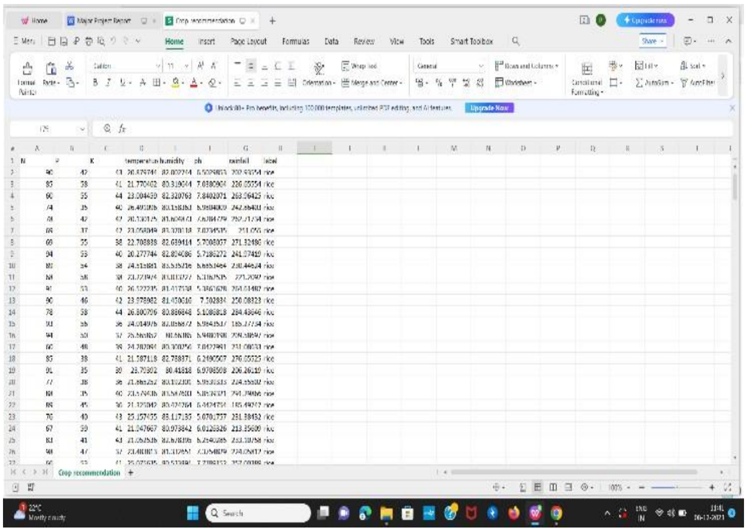
4.1 Design of propose of work

The K-Nearest Neighbour (KNN) algorithm is used by the proposed Geolocation-Based Crop Prediction System to determine which crop would be best for a given plot of land based on the composition of the soil and meteorological factors like rainfall, temperature, humidity, and soil pH. By comparing new instances to preexisting data, KNN is a straightforward machine learning method that quickly classifies new instances. It is non-parametric and does not require identifying distribution parameters. However, it is a lazy learner, storing training data and classifying during the testing phase, making it time-consuming and memory-intensive.

The system incorporates supervised machine learning algorithms, focusing on classification for tasks like rainfall prediction (using the SVM algorithm) and crop prediction (using the Decision Tree algorithm). Data collection involves efficiently gathering information from various sources, including government websites and agricultural databases, to form a dataset with attributes such as soil pH, temperature, humidity, rainfall, crop data, and NPK values for accurate predictions. The annual rainfall prediction specifically involves loading external datasets, preprocessing the data, and training the model using an SVM classifier with an RBF kernel.

4.2 Working Modules

Module 1 (Collection of Dataset): The dataset is made up of different types of crops which is suitable for yeilding.



INTERNATIONAL JOURNAL OF PROGRESSIVE RESEARCH IN ENGINEERING MANAGEMENT AND SCIENCE (IJPREMS)

www.ijprems.com
editor@ijprems.com

Vol. 04, Issue 05, May 2024, pp: 1404-1411

e-ISSN : 2583-1062

Impact Factor: 5.725

Fig. 2 Dataset for Crop Prediction System

Module 2 (Testing Dataset): Figure displays an example of test result obtained on testing Dataset over different models.

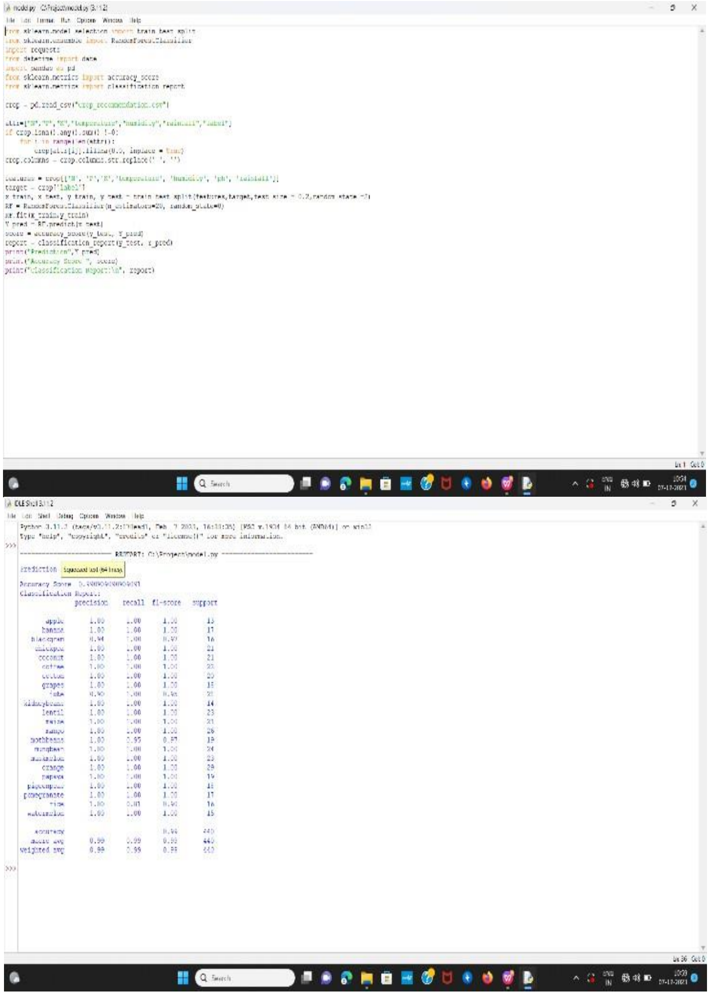


Fig. 3 Classification Report For Testing Data

5 PROJECT IMPLEMENTATION

5.1 Module Implementation

Data Collection – Acquiring data stands as a paramount method for efficiently gathering and quantifying information from a range of sources, including government websites, VC Form Mandya, and APMC websites, to compile an approximate yet comprehensive dataset for the system. This dataset is meticulously curated to incorporate key attributes essential for the prediction process, encompassing parameters such as soil pH, temperature, humidity, rainfall, crop data, and NPK values. In the context of forecasting annual rainfall, the specific emphasis is on retrieving and incorporating historical rainfall data from preceding years. This thorough data collection ensures a robust foundation for the subsequent stages of analysis and prediction within the system.

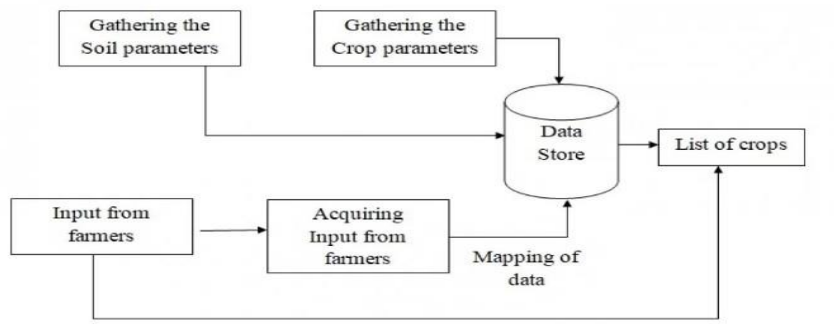


Fig. 4 Collection of Data

Data Preprocessing: - After gathering datasets from various sources, it is imperative to undergo dataset preprocessing before initiating model training. The culmination of this preprocessing phase is data cleaning, commencing with the reading of the collected dataset. During this stage, redundant attributes are identified and eliminated from the datasets, with these attributes deemed irrelevant for crop prediction. To enhance accuracy, the focus is on discarding undesirable attributes and handling datasets containing missing values by either removal or filling with nonessential "nan" values. Subsequently, the emphasis shifts to printing the anticipated precipitation. The process initiates with loading an external dataset that includes historical rainfall data. Following this, the loaded dataset undergoes preprocessing, as detailed in the data preprocessing section. Once the data preprocessing is completed, an SVM classifier with an RBF kernel is employed to train the model, and subsequently, the classifier is fitted to the training set. The mathematical expression for the radial basis function is then elucidated.

We conducted system tests using diverse datasets obtained directly from individual farmers, providing insights into the conditions of their respective lands. The terrain characteristics considered encompass variations in pH, humidity, and NPK levels. The annual rainfall predictions exhibit consistency. Based on the algorithm's analysis of weather conditions, soil composition, and expected rainfall, it suggests the most suitable crop for cultivation. Additionally, the system provides details on the quantity of seed required for cultivating the recommended crop in kilograms per acre and offers information about essential fertilizers, including nitrogen (N), phosphorus (P), and potassium (K), in grams per hectare.

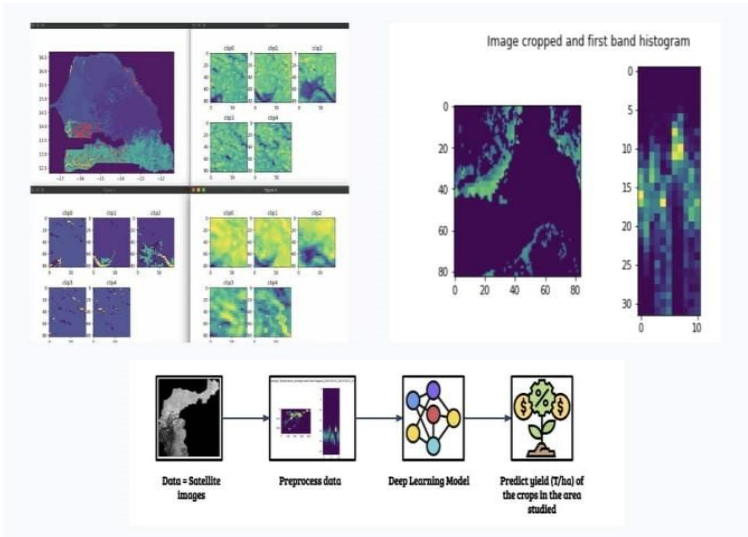


Fig. 5 Images of Crop Prediction System

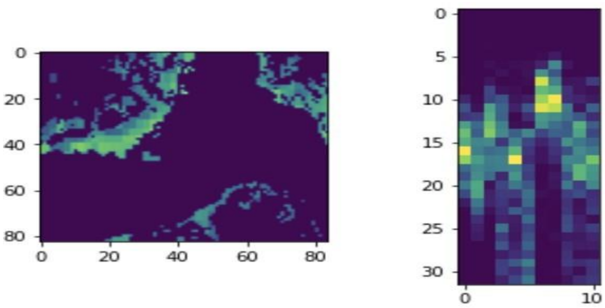


Fig. 6 Data distribution in dataset and crop Image

6 CONCLUSION

Due to the current limited adoption of technology and proper analysis by our farmers, there exists a risk that they might select inappropriate crops for cultivation, leading to a reduction in their income. To address this, we've developed a user-friendly system featuring a graphical user interface (GUI) that can identify the most suitable crop for a given plot of land. This system also provides valuable information, including nutrient requirements, seed quantities, anticipated yield, and market prices. By empowering farmers with this user-friendly tool, we aim to guide them in making informed decisions and thereby advance the agricultural sector.

To enhance crop cultivation efficiency, we conducted a comparative analysis of three distinct supervised machine learning models—KNN, Decision Tree, and Random Forest—to predict the optimal crop for a specific plot of land.

After rigorous evaluation, we found that the crop forecast dataset achieved the highest accuracy using the Random Forest Classifier, reaching an accuracy of 99.32% for both Entropy and Gini Criterion. In contrast, the K- Nearest Neighbor model exhibited the lowest accuracy at 97.04%, while the Decision Tree Classifier's accuracy fell between KNN and Random Forest Classifier. Notably, the Decision Tree Entropy Criterion demonstrated a slightly lower accuracy of 98.86% compared to the Decision Tree Gini Criterion in terms of accuracy.

7 FUTURE SCOPE

To acquire essential information, we can obtain GPS coordinates of the land and utilize the government's rain forecasting system, enabling us to predict crops by leveraging GPS positions. Furthermore, the implementation of a model can effectively address concerns related to food shortages and surpluses. Gathering requisite details through GPS coordinates and accessing the government's rain forecasting system allows us to streamline crop anticipation. Additionally, the development of a model serves as a proactive measure to manage and mitigate issues of food scarcity and excess production.

8 REFERENCES

[1] Dahikar S and Rode S V 2014 Agricultural crop yield prediction using artificial neural network approach International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering vol 2 Issue 1 pp 683-6.

[2] Suresh A, Ganesh P and Ramalatha M 2018 Prediction of major crop yields of Tamilnadu using K-means and Modified KNN 2018 3rd International Conference on Communication and Electronics Systems (ICCES) pp 88-93 doi: 10.1109/CESYS.2018.8723956.

[3] Medar R, Rajpurohit V S and Shweta S 2019 Crop yield prediction using machine learning techniques IEEE 5th International Conference for Convergence in Technology (I2CT) pp 1-5 doi: 10.1109/I2CT45611.2019.9033611.

[4] Nishant P S, Venkat P S, Avinash B L and Jabber B 2020 Crop yield prediction based on Indian agriculture using machine learning 2020 International Conference for Emerging Technology (INCET) pp 1-4 doi: 10.1109/INCET49848.2020.9154036.

[5] Kalimuthu M, Vaishnavi P and Kishore M 2020 Crop prediction using machine learning 2020 Third International Conference on Smart Systems and Inventive Technology (ICSSIT) pp 926-32 doi: 10.1109/ICSSIT48917.2020.9214190.

[6] Geetha V, Punitha A, Abarna M, Akshaya M, Illakiya S and Janani A P 2020 An effective crop prediction using random forest algorithm 2020 International Conference on System, Computation, Automation and Networking (ICSCAN) pp 1-5 doi: 10.1109/ICSCAN49426.2020.92623 11.

[7] Pande S M, Ramesh P K, Anmol A, Aishwaraya B R, Rohilla K and Shaurya K 2021 Crop recommender system using machine learning approach 2021 5th International Conference on Computing Methodologies and Communication (ICCMC) pp 1066-71 doi: 10.1109/ICCMC51019.2021.9418351.

[8] Sellam V, and Poovammal E 2016 Prediction of crop yield using regression analysis Indian Journal of Science and Technology vol 9(38) pp 1-5.

[9] Bharath S, Yeshwanth S, Yashas B L and Vidyaranya R Javalagi 2020 Comparative Analysis of Machine Learning Algorithms in The Study of Crop and Crop yield Prediction International Journal of Engineering Research & Technology (IJERT) NCETESFT – 2020 vol 8 Issue 14.

[10] Mahendra N, Vishwakarma D, Nischitha K, Ashwini and Manjuraju M. R 2020 Crop prediction using machine learning approaches, International Journal of Engineering Research & Technology (IJERT) vol 9 Issue 8 (August 2020).

[11] Gulati P and Jha S K 2020 Efficient crop yield prediction in India using machine learning techniques International Journal of Engineering Research & Technology (IJERT) ENCADEMS – 2020 vol 8 Issue 10.

[12] Gupta A, Nagda D, Nikhare P, Sandbhor A, 2021, Smart crop prediction using IoT and machine learning International Journal of Engineering Research & Technology (IJERT) NTASU – 2020 vol 9 Issue 3.

[13] R Sujatha,Dr P.Issaki,"A Study of Crop Yield Prediction Using Data Mining Techniques" IEEE 2016.

[14] Jeetendra Shenoy, Yogesh Pingle, "IOT in agriculture",IEEE 2016 .