INTELLIGENT CROP RECOMMENDATION AND FERTILIZER PREDICTION

[1] RAGAVI R, [2] THIANESH G, [3] DHIVYAAN S, [4] GAYATHRI K

[1] *Student, Dept. of Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*
[2] *Student, Dept. of Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*
[3] *Student, Dept. of Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*
[4] *Professor, Dept. of Information Technology, Bannari Amman Institute of Technology, Tamil Nadu, India*

Abstract

The crop recommendation module employs machine learning algorithms to analyze soil type, climate conditions, and past crop performance, providing personalized suggestions for suitable crops in a given region. By considering local environmental characteristics, the system enhances the accuracy of its recommendations. This enables the system to forecast the ideal type and quantity of fertilizer needed for a specific crop and field. The user interface is designed for ease of use, ensuring accessibility for farmers with varying levels of technological proficiency. Farmers can input preferences, receive crop recommendations, and access detailed fertilizer prescriptions through a mobile or web-based application. Our Smart Agriculture solution aims to foster sustainable farming practices by promoting efficient resource use, reducing environmental impact, and enhancing overall farm productivity. This research proposes an intelligent system that uses a decision tree algorithm to anticipate fertilizer and recommend crops. To suggest appropriate crops for cultivation, the suggested approach examines several variables, including crop attributes, soil type, and climate. Furthermore, it forecasts the ideal kind and number of fertilizers needed for every suggested crop, supporting sustainable farming methods. The system's decision tree algorithm makes forecasts that are both accurate and efficient, which improves crop productivity and resource use. The field of crop recommendation and fertilizer prediction research aids in sustainable farming, global food security, rural development, and the agriculture sector's overall resilience to changing challenges by providing farmers with intelligent systems that incorporate data-driven insights.

Keywords: *AI, Decision tree algorithm, Crop Recommendation, Fertilizer prediction*

I. INTRODUCTION:

Modern agriculture faces the challenge of feeding a growing population while minimizing environmental impact. Integrating advanced technologies like machine learning offers promise in revolutionizing agricultural practices. This paper explores an intelligent agricultural system focusing on crop recommendation and fertilizer prediction using decision tree algorithms. Traditionally, agriculture relied on customary techniques. However, with increasing complexity in farming due to factors like soil erosion and climate change, precision agriculture solutions are imperative. Machine learning can evaluate vast amounts of data to identify hidden patterns, addressing these challenges effectively.

The proposed system targets crop selection and fertilizer application, traditionally based on anecdotal evidence and subjective judgment. In contrast, the machine learning-based approach utilizes data-driven decision-making for more informed practices. The decision tree algorithm, known for its interpretability and simplicity, forms the core of the system. By analyzing historical agricultural data encompassing various factors like crop characteristics and climate, decision trees optimize crop selection and fertilizer recommendations through iterative validation and refinement.

Beyond practical implications for farmers, this study contributes to sustainable agriculture discourse, showcasing technology-driven innovation. By optimizing agricultural decision-making, machine learning approaches offer tangible benefits for crop yield, resource efficiency, and environmental sustainability. Precision agriculture, driven by technology, marks a paradigm shift in farming. Machine learning algorithms, particularly effective in handling vast agricultural data, enhance decision-making across soil conservation, irrigation, crop management, and pest control.

This research analyzes an intelligent farming system focusing on crop suggestion and fertilizer prediction, crucial for agricultural management. By leveraging data-driven decision-making, the proposed approach aims to transform practices, providing tailored insights into farmers' unique circumstances. Crop selection and fertilizer application are pivotal in determining agricultural outcomes, affecting resource consumption, land use, and environmental impact. Precision nutrient management is crucial, as inadequate fertilization can lead to nutrient imbalances and environmental degradation..

2. LITERATURE SURVEY

[1] Dr. V. Geetha et al. (2020) employed the Random Forest algorithm to analyze crop growth in various climatic conditions. Their study utilized a crop growth dataset to train and test the algorithm, demonstrating Random Forest's superior accuracy compared to other methods. **[2] Zeel Doshi et al. (2018)** developed AgroConsultant, an intelligent system aiding Indian farmers in crop decision-making based on factors like sowing season, soil characteristics, and environmental variables. It incorporates two subsystems: crop recommendation and rainfall prediction, enhancing crop sustainability. **[3] Vaishnavi et al. (2019)** proposed a system providing real-time crop predictions to agriculturists, leveraging data mining and big data techniques. Their approach offers personalized recommendations, leading to increased crop yields. **[4] Ms. Kavita et al. (2020)** predicted crop yields for India using data from 1950 to 2018. Their study compared various algorithms and found Decision Trees to provide the highest accuracy, benefiting small farmers by enabling precise crop yield estimation. **[5] Mrs. R. Usha Devi et al. (2022)** developed a recommendation system using Logistic Regression, Naive Bayes, and Random Forest classifiers to suggest crops based on soil qualities. Their goal is to enhance agricultural production by guiding farmers in crop selection using machine learning algorithms. **[6] Priyadharshini A et al. (2022**) proposed a system aiding farmers in choosing the right crop by leveraging real-time weather data and profit analysis. Their sequential model, employing Linear Regression and Neural Networks, achieved high accuracy in crop recommendation. **[7] Dr. Y. Jeevan Nagendra Kumar et al. (2021)** utilized a Supervised Machine Learning approach, specifically Random Forest, for crop yield prediction. Their system, trained on historic data, accurately predicts crop yields based on weather and pesticide factors, benefiting Indian farmers with increased profits. **[8] Meeradevi et al. (2017)** designed an application predicting crop production based on rainfall and temperature data. Their model assists farmers in selecting crops optimized for their region's conditions, improving overall crop yield. **[9] Sonal Agarwal et al. (2021)** developed a model combining Support Vector Machine and Deep Learning techniques to predict cost-effective crops with high accuracy, aiding farmers in maximizing profits. **[10] N. Manjunathan et al. (2019)** built a Support Vector Machine-based model for rice crop yield prediction, accompanied by a web application for user interaction. Their system facilitates accurate crop yield prediction, empowering farmers with informed decision-making. **[11] Kusum Lata et al. (2019)** applied four classification algorithms to improve crop yield prediction, with Random Tree outperforming others. Their work aims to modernize agriculture data accessibility, enhancing crop yield prediction in Maharashtra. **[12] D. Jayanarayana Reddy et al. (2016)** proposed a crop yield prediction system utilizing various Machine Learning algorithms. Their model, after data pre-processing and feature extraction, achieved reduced relative error, enhancing crop yield estimation. **[13] Saeed Khaki et al. (2018)** trained Deep Neural Networks for yield prediction, utilizing genotype and environment effects. Their model, optimizing prediction accuracy while minimizing overfitting risks, achieved accurate results using specific network configurations.

3. OBJECTIVES AND METHODOLOGY:

The project aims to develop an intelligent agricultural system utilizing decision tree algorithms for crop recommendation and fertilizer prediction, with a focus on optimizing resource utilization, enhancing crop productivity, and promoting sustainability. Specific objectives include designing a robust data collection framework to gather diverse agricultural data encompassing soil properties, climate conditions, crop characteristics, agronomic practices, and historical yield outcomes.

To streamline later analysis and modeling, the project aims to preprocess and curate the gathered agricultural data to ensure consistency, completeness, and quality. Using historical agricultural data, the project seeks to identify patterns, relationships, and decision rules governing optimal crop choices and fertilizer applications based on contextual factors such as soil type, climate, and agronomic preferences. Decision tree-based models will be developed for crop recommendation and fertilizer prediction. The project plans to integrate these models into an intelligent agriculture system, providing farmers and stakeholders with personalized insights and recommendations through an intuitive interface. Evaluation of the system's efficacy and performance will involve benchmarking against current methods, utilizing real-world agricultural datasets, and assessing critical performance metrics like accuracy, precision, recall, and F1-score.

The chosen methodology involves multiple crucial phases, starting with gathering agricultural data from various sources including government agencies, open data repositories, and agricultural research organizations. The collected data will undergo thorough preprocessing to address issues like outliers, missing values, and inconsistent data, ensuring its readiness for modeling activities. Decision tree algorithms, including variations like CART (Classification and Regression Trees) and C4.5, will be employed to create models for crop recommendation and fertilizer prediction. These models will be trained using supervised learning techniques on preprocessed agricultural data. Once developed, the models will be integrated into the intelligent agriculture system to provide stakeholders with actionable insights. Ultimately, the project aims to advance the field of intelligent agriculture by offering practical solutions for resource optimization and sustainable farming management. Real-world agricultural datasets will be utilized for validation through holdout validation, cross-validation techniques, and rigorous assessment of performance metrics.

4. DATA COLLECTION:

A crucial first step in machine learning and data-driven decision-making is data collection. This step involves compiling data from numerous sources to create a dataset that can be used to develop, test, and validate machine learning models. Here, we downloaded the dataset named “crops.csv” from an online website named Kaggle which contains several datasets that can be downloaded and used for free. There are other websites too to download datasets like Socrata, and GitHub, etc., There are certain steps that we followed before downloading the dataset. Before planning, we established goals and the necessary data kinds (categorical data, numerical data, text data, image data, or video data). Machine learning models can be trained using the data collected. Models create patterns and relationships from data, which they then use for classification or prediction. The dataset needs to be cleaned to get rid of any errors, outliers, and missing values. It has been carried out to ensure that the dataset generates more accurate findings. The dataset should also not contain any biased data as this could prevent the entire dataset from accurately representing the dataset. In this phase, we evaluated the dataset's completeness to ensure that it had all the necessary data. Also, the data has to be relevant because irrelevant data can instigate noise into the model.

4.1 DATA PREPROCESSING

4.1.1 Lowercasing

We have converted the uppercase letters to lowercase letters to make sure that the model should not treat them differently as both are the same.

4.1.2 Removal of special characters and punctuations

As punctions and special characters like ‘$’,’&’ etc., don’t carry any necessary information, we have removed them to make the dataset compact.

4.1.3 Handling newlines and blank spaces(whitespaces)

In addition to the above process, we have also removed extra newlines to support text formatting.

4.1.4 Tokenization

We split the text document into a list of words that are known to be tokens. Each token in the text serves as a unique unit of meaning. We consider this step as very important because it enables you to treat each word as a separate object, enabling fine-grained analysis and processing of the text. The machine learning model uses each token as a feature to make predictions. For example, if we have the sentence “we have scheduled a meeting this evening”, after tokenization the output willbe as [“we”,” have”,” scheduled”,” a”,” meeting”, ”this”, ” evening”]. Here, ‘we’, ‘a’, and all other words are known as tokens.

4.1.5 Removal of Stop Words

In this step, we have removed the words that occur very frequently in text and also don’t carry any necessary information in it. Performing this step helps in dimensionality reduction which improves the model performance further. In English, the words that frequently occur are ‘the’, ’and’, ’or’, ’in’, ’is’ etc., for example, if we have a sentence “today is a good day”, after stopping words removal the output will be as [‘today’, ‘good’, ‘day’]. ‘is’ and ‘a’ has been removed.

4.1.6 Stemming and Lemmatization

Lemmatization and stemming are methods for breaking down words to their root or basic forms. The main objective is to combine words with comparable meanings, which can aid in feature engineering and feature reduction. When you wish to minimize dimensionality but don't necessarily need proper words, stemming is a good approach. Stemming, while useful for grouping word variants, can result in nonsensical forms like "improv." In contrast, lemmatization transforms words into meaningful dictionary forms, preserving grammatical correctness and interpretability. For instance, "improves" and "improving" become "improve." Compared to stemming, lemmatization consistently generates more meaningful results. Both techniques are integral in standardizing terms and reducing feature space, crucial for training machine learning models. This optimization significantly enhances the overall effectiveness of spam detection systems.

5. FEATURE EXTRACTION:

We have created numerical features by converting the preprocessed text data. These numerical features are considered to be the inputs for our machine-learning model. This technique includes the term TF-IDF which we have implemented. It's a numerical statistic that is frequently used in information retrieval and natural language processing to assess the significance of a word inside a document. For text-based machine learning applications which include text categorization and information retrieval, TF-IDF is especially helpful. The frequency of a particular word in a document is quantified by the TF. It measures a term's significance within the confines of a single document. To avoid bias against longer documents, TF values are often standardized. TF can be calculated as

TF (t, d) = Number of times t comes in d / Total num of t in d

Where t stands for terms and d stands for document IDF quantifies a term's prominence within a corpus of documents. It aids in locating terms that are comparatively uncommon or unique in the corpus. IDF can be calculated as

IDF (t, d) =log (Number of documents in D/Number of documents in term t+1) +1

Where t stands for terms and D stands for corpus TF-IDF can be calculated as

TF-IDF (t, d, D) =TF (t, d) *IDF (t, D)

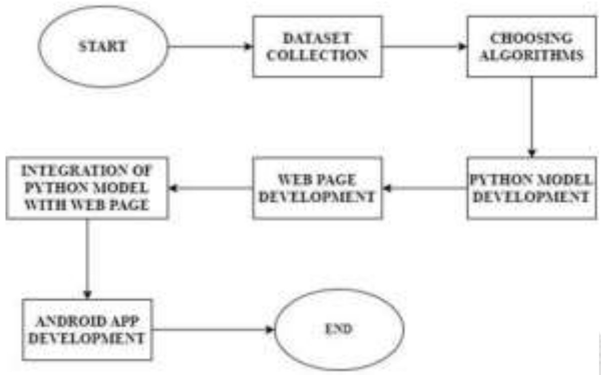
The terms that contain higher TF-IDF values are noted as important to a document.

6. DATA SPLITTING

We have split the preprocessed dataset into training and testing data in a ratio of 80:20. The machine learning model is trained using the data from the subset ‘training set’. This part of the data is used by the model to learn the patterns and relationships. The performance of the final model is assessed using the testing set. It is a chunk of the data that has been withheld and has not been viewed by the model during training or validation. Apart from the training and testing set, we have a validation set. This set helps us in selecting a good model among others. It evaluates the performance of the model at the time of training.

Also, it prevents overfitting. We have been very conscious when we split the dataset because it is good to randomize the rank of the samples to make sure that the sets reflect a vast number of examples. Also, to maintain the class distribution in each split, we have utilized stratified sampling. This made sure that each set accurately represented the proportions of the entire class. Data splitting can also be a component of data preprocessing in machine learning, where a dataset is divided into training and testing subsets to evaluate the performance of a model. Overall, data splitting is a common technique in various fields to handle and process data more effectively.

7. SYNTHETIC PROCEDURE OF THE PROPOSED WORK:



8. PREDICTION:

In the intelligent agricultural system project, the developed predictive model offers personalized recommendations for crop selection and fertilizer application based on a comprehensive analysis of agricultural data. The model precisely forecasts the best crops and fertilizer kinds and amounts suited to certain soil characteristics, climatic circumstances, and agronomic methods by utilizing decision tree algorithms and sophisticated data preprocessing techniques. The model shows its usefulness in increasing crop productivity, resource utilization, and sustainability in agriculture through thorough examination and validation, including performance measures, cross-validation, and comparison analysis. The system offers farmers and other agricultural stakeholders actionable insights that enable them to make well-informed decisions that optimize yields, reduce input costs, and limit environmental impacts.

9. CONCLUSION

In conclusion, the intelligent agricultural system project has effectively created a predictive model that, using extensive data analysis, provides individualized crop suggestions and fertilizer projections. Through the use of sophisticated preprocessing methods and decision tree algorithms, the system enables farmers to maximize crop management strategies, increase yield, and advance agricultural sustainability. The method exhibits its efficacy in furnishing practical insights for well-informed decision-making via meticulous assessment and ongoing enhancement. All things considered, the project is a major step forward in using technology to tackle the intricate problems of contemporary farming, opening the door to a more resilient, effective, and ecologically sustainable agricultural industry.

10. REFERENCES

[1] Elavarasan, D., & Vincent, P. D. (2020). Crop yield prediction using deep reinforcement learning model for sustainable agrarian applications. IEEE access. 53-54

[2] Hemageetha, N., &Nasira, G. M. (2019). Analysis of soil condition based on pH value using classification techniques. IOSR J ComputEng (IOSR-JCE), 50-54.

[3] Bhargavi, P., &Jyothi, S. (2022). Soil classification using data mining techniques: a comparative study. International Journal of Engineering Trends and Technology, 2(1), 55-5W9.

[4] Rajeswari, V., &Arunesh, K. (2019). Analysing soil data using data mining classification techniques. Indian journal of science and Technology, 9(19), 1-4.

[5] Noble, I., Bolin, B., Ravindranath, N., Verardo, D., &Dokken, D. (2020). Land use, land use change, and forestry. Environmental Conservation, 28(3), 284-293.

[6] Li, S., Peng, S., Chen, W., & Lu, X. (2017). INCOME:
Practical land monitoring in precision agriculture with sensor networks. Computer Communications, 36(4), 459-467.

[7] KiryushinV I, The management of soil fertility and productivity of agrocenoses in adaptive-landscape farming systems, Eurasian Soil Sci, 52 (2019) 1137–
1145

[8] Nordjo R E & Adjasi C K, Integrated soil fertility management (ISFM) and productivity of smallholder farmers in the northern region of Ghana (2019).
[9] Suchithra M S & Pai M L, Improving the prediction accuracy of soil nutrient classification by optimizing extreme learning machine parameters, Inf Process Agric 7(1) (2020) 72–82.
[10] Toriyama K, Development of precision agriculture and ICT application thereof to manage spatial variability of crop growth, J Soil Sci Plant Nutr, 66(6) (2020) 811– 819.
[11] Dharmaraj V & Vijayanand C, Artificial intelligence (AI) in agriculture, Int J Curr Microbiol Appl Sci, 7(12) (2018) 2122–2128.
[12] Blackman N J & Koval J J, Interval estimation for Cohen's Kappa as a measure of agreement, Stat Med, 19(5) (2000) 723–741
[13] Apat S K, Mishra J, Raju K S & Padhy N, A study on smart agriculture using various sensors and agrobot: A case study, Smart Intel Comput Appl, 1 (2022) [14] Toseef M & Khan M J, An intelligent mobile application for diagnosis of crop diseases in Pakistan using fuzzy inference system, Comput Electron Agric, 153 (2018) 1–11.