**CROP RECOMMENDATION SYSTEM**

**Submitted for**

**CSET211 - Statistical Machine Learning**

Submitted by:

**E23CSEU0565 SRIMAY KUMAR**

**E23CSEU502 ESHA SHRIVASTAV**

**E23CSEU0562 ARUSHI SHAH**

Submitted to:

**PRASHANT KAPIL**

**July-Dec 2024**

**SCHOOL OF COMPUTER SCIENCE AND ENGINEERING**

A close-up of a logo

Description automatically generated

**INDEX**

|  |  |  |
| --- | --- | --- |
| Sr.No | Content | Page No |
| 1 | **Introduction** | 2 |
| 2 | **Related Survey** | 3 |
| 3  3.1 | **Dataset**  **Data Preprocessing** | 3  4 |
| 4  4.1  4.2 | **Methodology**  **Hardware and Software Requirements**  **Performance Metrics** | 4  5  5 |
| 5 | **Results and Analysis** | 6 |
| 6 | **Conclusions and Future Works** | 6 |

1. **INTRODUCTION**

Agriculture is vital for global sustenance and economic stability, especially in agrarian economies like India. Farmers often face challenges in selecting suitable crops, leading to low yields and financial losses due to the lack of reliable, data-driven tools.

The Crop Recommendation System addresses this issue using machine learning to predict the most suitable crop based on seven features: soil pH, nitrogen (N), phosphorus (P), potassium (K), temperature, humidity, and rainfall. An Artificial Neural Network (ANN) achieves a test accuracy of 98%, ensuring reliable recommendations.

Motivated by the need to improve productivity and sustainability, this system empowers farmers with informed decision-making. Its objectives include:

1. Boosting productivity through accurate recommendations.

2. Promoting sustainability by aligning crop choices with environmental conditions.

3. Reducing dependency on intuition-based methods.

This project highlights the potential of ANN in addressing agricultural challenges, advancing precision farming, and optimizing resource utilization.

**2. RELATED SURVEY**

This compilation explores nine research papers focusing on crop recommendation systems leveraging machine learning (ML) and IoT technologies to enhance agricultural productivity. Key features analyzed include soil properties (pH, nitrogen, phosphorus, potassium), environmental conditions (temperature, rainfall, humidity), and geographical factors. Prominent ML models used include Decision Tree, Random Forest, XGBoost, Naive Bayes, SVM, and ANN, with accuracies ranging from 92% to 99.31%.

Highlights include systems integrating IoT for real-time monitoring and feedback, hybrid models like WLSTM-IDCSO for precision and efficiency, and the introduction of XAI-CROP, which combines explainable AI with low MSE (0.9412) and high interpretability (R² = 94.15%). Many studies use ensemble techniques, like majority voting or optimization algorithms, to achieve high accuracy and scalability. These systems aim to maximize yields, improve sustainability, and provide location-specific recommendations, offering innovative solutions to address farming challenges and ensure data-driven agricultural advancements globally.

**3. DATASET**

* Dataset used here is [Crop Recommendation Dataset by Atharva Ingle](https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset) from Kaggle. It contains 7 features-
* N - ratio of Nitrogen content in soil , P - ratio of Phosphorous content in soil, K - ratio of Potassium content in soil, Temperature - temperature in degree Celsius, Humidity - relative humidity in %, Ph - ph value of the soil, Rainfall - rainfall in mm
* Based on these 7 features, the model predicts the most suitable crop (out of the 20 in dataset).

**3.1 DATA PREPROCESSING**

To prepare the dataset for training, several preprocessing steps were applied to ensure data consistency and improve model performance:

1. Checking for Null Values - No Null Values were found in any of the columns
2. Encoding string (object) values - Label encoding of the target variable ‘Crop’ was done since the model cannot process string values and needs numerical input.
3. Outlier Detection - Outlier detection using z-score is done to identify rows in the dataset where any feature deviates beyond +- 3.1 standard deviations. These rows (171 outliers) are filtered out, leaving a cleaned dataset for accurate analysis and modelling.
4. Train-test Split - Data was splitted into training set(80%) and testing(20%) as well as into features and target.
5. Scaling the Data - Data was scaled using Standard Scaler to speed up convergence and improve accuracy so that the features with larges ranges don’t dominate over others.

4. **METHODOLOGY**

* **Data Collection:** [**Crop Recommendation Dataset**](https://www.kaggle.com/datasets/atharvaingle/crop-recommendation-dataset) **from Kaggle was used.**
* **Data Preprocessing: Techniques such as Encoding, Outlier Detection, Scaling were used to preprocess data for model building.**
* **Model Architecture: The classic ANN /MLP model was chosen for the task and the architecture consists of 4 input layers to handle the input features. The activation function used in the input layers is ReLU to introduce non-linearity and improve learning. The model concludes with a single output layer, where the activation function is softmax.**
* **Training Process: The dataset was split into training (80%) and testing (20%) sets. Optimizer used was RMSprop since it provided better accuracy over adam and sgd, while sparse categorical crossentropy was used as the loss metric.**
* **Evaluation: The trained model was evaluated using metrics such as accuracy, confusion matrix, and classification reports.**

**4.1** **HARDWARE** **AND SOFTWARE REQUIREMENTS**

The project was developed and executed using the following hardware and software:

Hardware:  
• Laptop: MacBook M2 PRO

• Processor: M2 chip  
• RAM: 8 GB

Software:

* Programming Language: Python 3.11
* Libraries and Frameworks: TensorFlow, Keras, Sklearn, Pandas, NumPy, Matplotlib, Seaborn
* Development Environment: Jupyter Notebook
* Dataset Source: Kaggle

4.2 **Performance Metrics**

To evaluate the effectiveness of the classification model, the following performance metrics were used:

* + Accuracy: Measures the proportion of correctly predicted crops out of the total samples.
  + Confusion Matrix: Provides a detailed breakdown of true positives, true negatives, false positives, and false negatives.
  + Classification Report : Precision, Recall, and F1-Score:  
    – Precision: The proportion of true positive predictions among all positive predictions.

– Recall: The proportion of true positives identified out of all actual positives.

– F1-Score: The harmonic mean of precision and recall, balancing the trade-off between the two.

* + Loss Metrics: Training and validation loss were monitored to assess model convergence and detect overfitting.

These metrics provide a comprehensive assessment of the model’s performance.

**5. Results and Analysis**

The ANN model demonstrated high performance in classifying 20 crops.

Performance Summary:

* + - Accuracy: Training accuracy reached 99%, with a testing accuracy of 98%.
    - Confusion Matrix: Most predictions were correct, with only 1 misclassification observed.
    - Classification Report: The model achieved an average precision of 98.5%, recall of 98.3%, and F1-score of 98.3%.

6. **Conclusions and Future Works**

This project successfully demonstrated the application of advanced deep learning techniques for a Crop Recommendation System. By leveraging the ANN architecture and employing robust preprocessing strategies, the model achieved high accuracy (98%) across 20 crop classes. The results validate the potential of deep learning in addressing challenges related to crop selection and agricultural productivity.

Future Works

* Expanding the dataset to include more diverse crops to improve model robustness.
* Exploring advanced techniques such as attention mechanisms and ensemble methods to further enhance classification accuracy.
* Deploying the model for real-time crop prediction and integrating it into agricultural decision-making workflows.
* Addressing misclassification errors by incorporating domain-specific features and improving dataset quality.

These improvements can elevate the model’s practicality, ensuring its effectiveness in real-world applications for agricultural efforts.

GITHUB LINK : <https://github.com/AayushSharma1003/Crop-Recommendation-System>