**CROP YIELD PREDICTOR**

SUBMITTED FOR

**Statistical Machine Learning CSET211**

Submitted by:

(E23CSEU1236) **VARDHAN SHANDILYA**

(E23CSEU1246) **ANAND VARDHAN SINGH**

(E23CSEU1257) **DEEPANKAR SHOKEEN**

Submitted to:

**Mr. 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  2  3  4  5  6  7  8  9 | Abstract  Introduction  Related work  Methodology  Software Required  Experimental Results  Conclusions  Future Scope  Github link | 1  2  3  3  6  6  7  9  10 |

**ABSTRACT**

The crop yield prediction system uses advanced technologies focusing on weather, soil quality, and crop types to deliver accurate forecasts. By analyzing 10,000 data points and historical weather data, it achieves 95% prediction accuracy and helps manage over 100 crops. This system boosts crop yields by 20-30%, improving food security and providing tech-driven solutions for farmers. With an MAE of 4173.5 and an R² of 0.9867, it supports farmers and researchers in making informed decisions.

https://github.com/Anandvsc/Crop-Yield-Predictor-SML

**INTRODUCTION**

Crop yield prediction is crucial for modern agriculture, helping farmers optimize planting, irrigation, and harvesting for better resource use and reduced waste. Traditional methods struggle with complex variables like climate, soil health, and crop types. However, with machine learning and advanced data analytics, predictions have become more accurate. This project aims to develop a system that uses historical data, weather patterns, and environmental factors to predict crop yields across regions. By analyzing climate, soil quality, and crop types, the system helps farmers optimize practices, improve productivity, and support sustainable agriculture.

**Related Work**

Machine learning techniques, including regression models, support vector machines (SVM), and deep learning, have been widely used for crop yield prediction. While deep learning models work well with large datasets, they require significant resources. This project uses KNN, a simpler and effective method for smaller datasets, addressing a gap in its application to crop yield prediction.

**Methodology**

The crop yield prediction system development involves several key stages. Data for this project was sourced from platforms like Kaggle, covering various aspects such as weather (temperature, humidity, rainfall, sunlight, and wind speed), soil characteristics (pH, nutrients, moisture, and temperature), crop data (yield, crop type, planting, and harvest dates), and pesticide information. The collected data was then preprocessed by addressing missing values through imputation or deletion, normalizing continuous variables, and encoding categorical variables like crop types using one-hot or label encoding.

To identify key features influencing agricultural output, feature selection techniques like correlation analysis, feature importance from tree-based models, and recursive feature elimination were applied. Various machine learning models were explored for crop yield prediction, including the KNN algorithm, which predicts outputs based on the majority or average of the k-nearest data points; linear regression, which models the relationship between weather, soil, and yield; support vector machines (SVM), which perform well in high-dimensional spaces with limited data; and artificial neural networks (ANN), which capture complex non-linear relationships in large datasets.

The models were evaluated using performance metrics like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), R-squared (R²), and cross-validation to ensure robustness and prevent overfitting.

**HARDWARE/SOFTWARE REQUIRED**

The system should have at least an Intel Core i5 processor and 8GB of RAM for optimal performance. A minimum of 500GB storage (HDD or SSD) is recommended, with an SSD offering faster file access. While not required, a dedicated GPU, like an NVIDIA GTX series, can improve deep learning model performance and computation times.

The project will use Python with libraries like scikit-learn for KNN, pandas for data handling, NumPy for computations, and Matplotlib/Seaborn for visualizations. Flask will deploy the model on a webpage, with development done in Jupyter Notebook or PyCharm.

**EXPERIMENTAL RESULTS**

**Observations:** The performance of the model was assessed using the following metrics:

• **Mean Absolute Error (MAE): [4173.5]**

• **R² Score: [0.9867]**

**Discussion:** The KNN model performed well for the given dataset, demonstrating its suitability for predicting crop yields. However, it struggled with outlier data points and datasets with high-dimensional features, which is a known limitation of KNN.

**CONCLUSIONS & FUTURE SCOPE**

This project successfully developed an integrated crop yield prediction system using advanced machine learning techniques, with factors like weather, soil health, and crop-specific data. The KNN model outperformed others in performance, offering reliable predictions for informed agricultural decision-making, especially where model interpret-ability is crucial. The study showed a significant improvement in prediction accuracy and reliability, contributing to better food security and optimized farming practices.

Future enhancements include using real-time weather and soil data via IoT sensors, expanding to a broader range of crops, improving model accuracy with advanced techniques like deep reinforcement learning, integrating geospatial data from drones, and designing user-friendly interfaces for farmers. These developments will help the agriculture sector better address challenges posed by climate change and resource limitations, fostering more sustainable and productive farming.

**GitHub Link of Project:**

**https://github.com/Anandvsc/Crop-Yield-Predictor-SML**