AI-Driven Agricultural Yield Prediction

A Project Work Synopsis

Submitted in the partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN
ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by:

21BCS6027 Chennupati Pavan Sanjay
21BCS6034 Thiramdas Karthik
21BCS6110 Guttula Venkata Surya
21BCS6087 Nimmala Manideep Reddy

Under the Supervision of:

Ms.Tanvi

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,
PUNJAB

November, 2024

Abstract

This research presents an innovative approach to crop yield prediction utilizing the K-Nearest Neighbor (KNN) model integrated with multiple and diverse agricultural data sources. The project's success is underscored by its potential to revolutionize farming practices, providing farmers with valuable insights and informed decision-making tools. By embracing the principles of precision farming, the model assists farmers in optimizing inputs, minimizing environmental impact, and enhancing overall efficiency. Leveraging machine learning techniques, the KNN model analyses average rainfall per year, average temperature, etc. to forecast crop yields with high accuracy. The system aims to empower farmers with actionable insights, enabling them to make informed decisions regarding resource allocation, crop management, and risk mitigation.

Keywords: K-Nearest Neighbor, Artificial Intelligence, algorithm, prediction, Explainable Artificial Intelligence

Table of Contents

1. INTRODUCTION

1.1 Problem Definition

Nowadays, farmers encounter several significant challenges that impede their agricultural operations. One critical issue revolves around resource allocation. In the absence of accurate predictions, farmers struggle to allocate resources like water, fertilizers, and pesticides efficiently. This often leads to either excessive use, resulting in unnecessary costs, or insufficient application, leading to reduced crop yields. Furthermore, the inability to anticipate risks, such as adverse weather conditions, pests, or diseases, leaves farmers vulnerable to crop losses. Without proactive measures in place, these unforeseen challenges can significantly impact farmers' livelihoods and agricultural productivity. Additionally, unpredictable market conditions present another obstacle. Without knowledge of expected crop yields, farmers find it challenging to plan their harvesting and marketing strategies effectively. This uncertainty can result in surplus or shortage situations, leading to price fluctuations and decreased profitability. Financial instability is another concern, as the lack of crop yield predictions makes it difficult for farmers to secure loans, purchase crop insurance, or manage financial risks associated with agricultural production. Moreover, the absence of accurate predictions limits farmers' ability to make informed decisions about crop rotations, variety selection, and planting schedules, leading to suboptimal outcomes in terms of productivity, profitability, and sustainability. Lastly, inefficient resource use and lack of planning without crop yield predictions can exacerbate environmental degradation, contributing to soil degradation, water pollution, and habitat destruction.

Overall, implementing crop yield prediction models is crucial for addressing these challenges and empowering farmers with the foresight and insights needed to optimize their agricultural practices effectively.

1.2 Problem Overview

The project on crop yield prediction using the K-Nearest Neighbors (KNN) algorithm is designed to address the pressing need for accurate and reliable methods of estimating crop yields. With agriculture facing increasing challenges from climate variability, resource constraints, and market fluctuations, having predictive models that can forecast crop yields is essential for farmers to optimize their practices and make informed decisions. The project begins with collecting historical data on crop yields and relevant agricultural factors, followed by preprocessing to ensure data quality and integrity. Key features that influence crop yields, such as weather conditions, soil quality, and crop types, are identified and incorporated into the model through feature engineering. The KNN algorithm is then implemented for regression to train the predictive model, with careful consideration given to hyperparameter tuning and model evaluation techniques. Importantly, the project emphasizes interpretability and explainability, with efforts made to generate saliency maps or feature importance scores to help users understand the factors driving crop yield predictions. The ultimate goal of the project is to deploy the trained model into production environments, making it accessible to stakeholders such as farmers, agricultural planners, and policymakers. Through stakeholder engagement and education, the project aims to empower users with actionable insights and contribute to the advancement of sustainable farming practices. Overall, the project represents a significant step towards improving agricultural decision-making processes and ensuring the resilience and sustainability of food production systems.

1.3 Hardware Specification

1. most operations can be performed on standard computers.
2. High-performance computing power may be needed for training complex AI models.
3. CPU GPU and RAM, Internet Connectivity

1.4 Software Specification

1. Python libraries like scikit-learn, NumPy, and TensorFlow will be used for data analysis, KNN model development, and XAI implementation.
2. Libraries like Matplotlib and Seaborn will be used to create clear and informative saliency maps and charts for the dashboards.
3. IDE- Google Colab, Jupyter Notebook

2. LITERATURE SURVEY

2.1 Existing System

The traditional methods for predicting crop yields often rely on manual analysis of historical data and expert knowledge. These methods can be time-consuming, labor-intensive, and prone to errors. Furthermore, they may not adequately capture the complex
relationships between various factors influencing crop yield

2.2 Proposed System

The articles presented explore various methods and techniques for utilizing different machine learning and deep learning approaches to address the limitations of traditional methods for crop yield prediction and smart farming practices.

2.3 Literature Review Summary

Year and Citatio n	Article/ Author	Tools/ Software	Technique	Source	Evaluation Parameter
2023	Sharm a et al.	Regression and Deep learning	Yield Prediction	IEEE Access	R-Squred, RMSE
2023	Bagheri et al.	Graph Neural Networks (GCN)	Yield Prediction	arXiv	Mean Squared Eror(MSE)
2023	Albert et al.	KNN	Hydroponic System Monitoring and Nutrition Control	IEEE Access	Accuracy, Precision, Recall, F1score
2023	Islam et al.	Deep Learning	Crop Disease Prediction	Journal of Agriculture ND Food Research	Accuracy, Precision, Recall, F1score

2023	Sharma et al.	Machine Learning	Crop Recommendation	IGI Global	Accuracy, Precision, Recall, F1score
2023	Shreya et al.	Blockchain and IOT	Secure Smart Farming Framework	Sustainable Computing: Informatics and Systems	Security, Scalability, Energy Consumption
2023	Kwaghtyo and Eke	Various	Smart Farming Prediction Models SUrvey	Artificial Intelligence (AI)	Survey Article

.

3. PROBLEM FORMULATION

The objective of this project is to develop a predictive model using the K-Nearest Neighbours (KNN) regression algorithm to estimate crop yields based on various agricultural factors. The dataset used for this analysis is sourced from "yield_df.csv" and contains information on crop yields, agricultural practices, and environmental conditions across different regions.

1. Scope Definition:

The scope of the problem centres around farmers' limited understanding of how AI models predict crop yields. This lack of comprehension hampers their trust in the models and, consequently, their ability to optimize agricultural practices for enhanced yield.

2. Specific Questions:

- How can we improve farmers' understanding of AI models used in predicting crop yields?
- What specific challenges do farmers face in comprehending and trusting AI-driven yield predictions?
- Which methodologies and tools can be employed to address farmers' concerns and enhance their trust in AI models?

3. Assessment Methods:

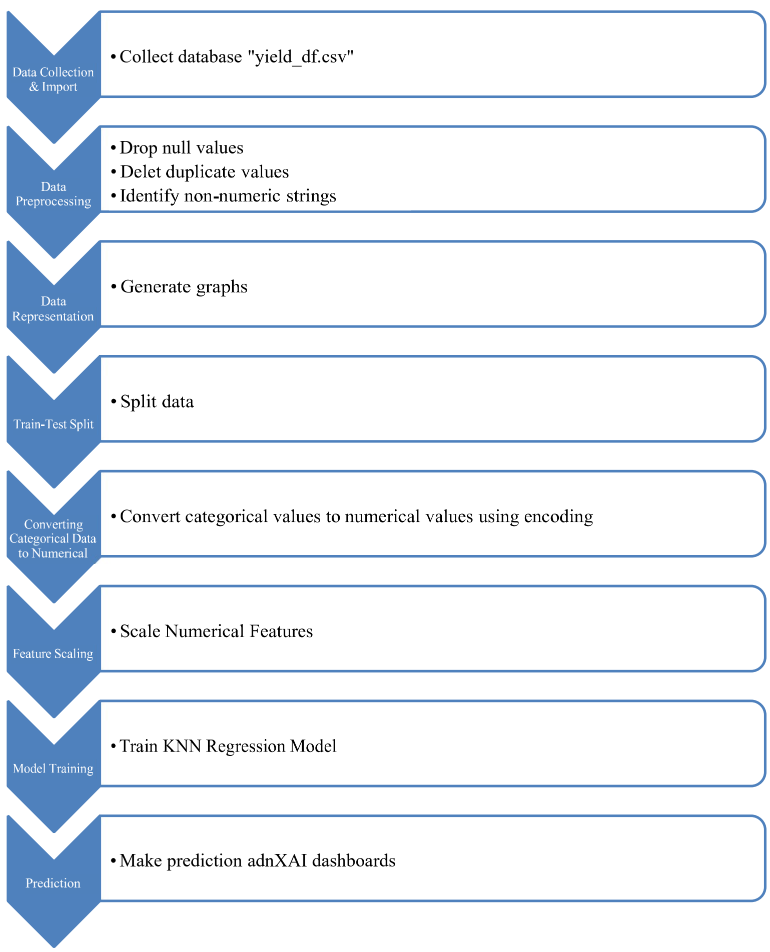
The assessment will involve:

- Analysing the existing level of understanding among farmers regarding AI-driven crop yield predictions.
- Identifying challenges through surveys, interviews, or focus group discussions with farmers.

- Evaluating the effectiveness of proposed solutions using metrics like improved comprehension, increased trust, and positive changes in agricultural practices.

Problem Statement:

While AI models can predict crop yields with increasing accuracy, their opaque nature presents a challenge for farmers. Lack of understanding hinders trust and adoption of AI driven recommendations, limiting potential benefits for agricultural efficiency and sustainability. Proposed Solution: This project aims to bridge the gap between AI and farmers by developing explainable crop yield prediction models and interactive dashboards.

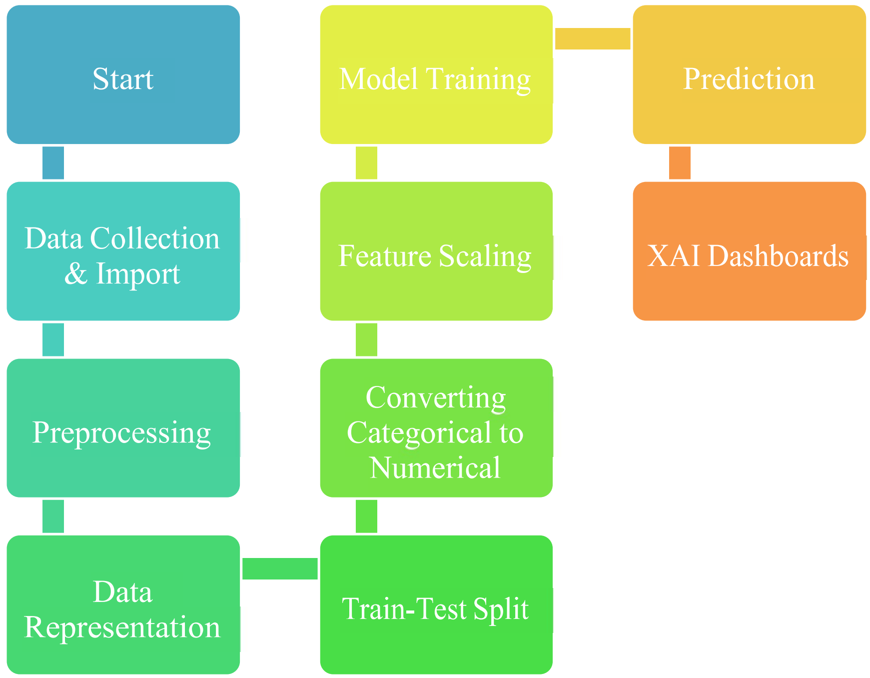


4. OBJECTIVES

1. Develop KNN models with embedded saliency maps: These maps will visually highlight the critical features, like rainfall or soil nutrients, that most influence specific yield predictions.
2. Build interactive XAI dashboards: Farmers can input their field data and receive AI-powered yield predictions alongside clear explanations in the form of saliency maps, charts, and text summaries.
3. Evaluate the impact of explain ability: Measure the increase in farmer trust, understanding, and adoption of AI-driven recommendations through controlled trials and farmer interviews.

This project focuses on explainable KNN models and dashboards for a specific crop or region. Further research could expand to other AI models and crops. User testing and evaluation will involve a limited group of farmers initially, aiming for wider adoption later.

4. METHODOLOGY



1. Data Collection and Import:

- Collected the dataset "yield_df.csv" containing information on crop yields and agricultural factors.
- Imported the dataset into the programming environment using appropriate libraries such as pandas.

2. Data Pre-processing:

- Identified and dropped null values from the dataset to ensure data integrity and completeness.
- Removed duplicate entries to eliminate redundancy and bias

  in the analysis.
- Identified rows where the values in the column

'average_rain_fall_mm_per_year' are not numeric strings and considered these rows for removal to maintain consistency in data type.

3. Data Representation:

- Utilized Seaborn library to create graphical representations of various aspects of the dataset:
- Generated a frequency distribution plot of crop yields against different areas (Frequency vs. Area) to understand regional yield patterns.
- Created a yield per country graph to visualize the variation in crop yields across different regions.
- Produced a frequency distribution plot of crop types grown (Frequency vs. Item) to identify common crops.
- Developed a scatter plot depicting the relationship between crop yield and crop type (Yield vs. Item) to explore yield variations based on crop type.

4. Train-Test Split:

- Employed train_test_split function from scikit-learn to split the dataset into training and testing sets with a ratio of 0.8 for training data.

5. Converting Categorical to Numerical:

- Utilized techniques such as one-hot encoding or label encoding to convert categorical variables such as "Area" and "Item" into numerical representations suitable for model training.

6. Feature Scaling:

- Applied standard scaling using StandardScaler from scikitlearn to scale the values of numerical features including 'Year', 'average_rain_fall_mm_per_year', 'pesticides_tonnes', and 'avg_temp'. This step ensures that all features contribute equally to the model training process and prevents bias due to differences in feature scales.

7. Model Training:

- Utilized the KNeighborsRegressor class from scikit-learn to train the KNN regression model on the training data. This algorithm learns the relationships between input features and crop yields based on the nearest neighbors' information.

8. Prediction:

- Utilized the trained KNN regression model to predict crop yields on the testing dataset. This step evaluates the model's performance in estimating crop yields based on the input features.

9. XAI dashboards:

- Performing interactive XAI dashboards, allowing farmers to visualize key factors impacting yield predictions and adjust their decisions accordingly, fostering trust and sustainable agricultural practices.

By following these methodologies, we aim to develop a robust crop yield prediction model using the KNN algorithm, which can provide valuable insights for agricultural decision-making and optimization.

6. EXPERIMENTAL SETUP

Enhancing Understanding of AI-Driven Crop Yield Predictions

1. Data Preparation:

- Collect and preprocess a diverse dataset containing historical crop yield data and relevant environmental factors.

2. Model Training:

- Train the KNN model with the training dataset, experimenting with different values of hyperparameter k and exploring distance metrics and weighting schemes.

3. Model Evaluation:

- Assess KNN model performance using regression metrics (MAE, MSE, RMSE, R-squared).
- Compare against baseline models or alternative regression algorithms.

4. Interpretability and Explainability:

- Generate saliency maps to understand feature importance and visualize model predictions for crop yield insights.

5. Parameter Tuning:

- Conduct hyperparameter tuning using grid search or random search with cross-validation for optimal model configuration.

6. Deployment and Integration:

- Deploy the trained KNN model into a production environment.
- Integrate the model into decision support systems for real-time use by farmers or agricultural advisors.

7. Documentation and Reporting:

- Document experimental setup, data preprocessing, model configuration, and evaluation metrics.
- Prepare a comprehensive report summarizing results, insights, and recommendations for further deployment or improvements.

8. XAI Dashboard Development:

  - Develop an interactive XAI dashboard that allows farmers to visualize saliency maps and understand the impact of various factors on yield predictions.
  - Include features for zooming, filtering, and exploring different scenarios.

7.CONCLUSION

The integration of the K-Nearest Neighbors (KNN) model within the Crop Yield Prediction AI project marks a significant milestone in agricultural innovation. This initiative produces precise crop production projections by combining data from many sources, giving farmers useful information for maximising resource use and raising productivity.

The project gives farmers the ability to make well-informed decisions that promote efficiency and sustainability in agriculture by utilising machine learning. With accurate crop yield forecasts at their disposal, farmers can reduce input costs, manage risks, and increase harvests, which will eventually boost the agricultural industry's resilience and profitability.

The project's ability to predict crop yields with precision contributes to sustainability efforts by enabling more efficient use of resources such as water, fertilizers, and pesticides. This reduction in resource wastage not only benefits the environment but also enhances the economic viability of farming operations.

Optimizing crop production and reducing uncertainties associated with yield fluctuations, the project can bolster food security initiatives at regional, national, and global levels. Access to reliable crop yield predictions allows policymakers to implement targeted interventions, allocate resources effectively, and mitigate potential food shortages or price volatility.

In conclusion, the KNN model-powered Crop Yield Prediction AI study provides evidence of the revolutionary potential of technology in contemporary farming methods. Its successful implementation promises to revolutionize crop management strategies, enabling farmers to achieve higher yields while minimizing environmental impact and contributing to global food security.

8. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

CHAPTER 1: INTRODUCTION

- Farmers struggle with resource allocation due to inaccurate predictions, leading to wasteful or insufficient use of resources like water and fertilizers, impacting costs and yields.
- Unforeseen risks such as adverse weather and pests threaten farmers' livelihoods and productivity, highlighting the need for proactive measures.
- Unpredictable market conditions complicate planning, resulting in fluctuating profitability and financial instability for farmers.
- Without reliable yield predictions, farmers face challenges in securing loans, purchasing crop insurance, and managing financial risks.
- The absence of accurate predictions hampers decision-making on crop rotations, variety selection, and planting schedules, contributing to environmental degradation.

CHAPTER 2: LITERATURE REVIEW

In our research, we analyzed recent studies in agricultural technology:

- Sharma et al. (2023) utilized Regression and Deep Learning for Yield Prediction, focusing on R-Squared and RMSE in their IEEE Access article.
- Bagheri et al. (2023) applied Graph Neural Networks (GCN) for Yield Prediction, with MSE as their evaluation metric, published on arXiv.
- Albert et al. (2023) used KNN for Hydroponic System Monitoring, assessing accuracy, precision, recall, and F1-score in IEEE Access.
- Islam et al. (2023) employed Deep Learning for Crop Disease Prediction, evaluating accuracy, precision, recall, and F1-score in their study in the Journal of Agriculture and Food Research.
- Sharma et al. (2023) applied Machine Learning for Crop Recommendation, focusing on accuracy, precision, recall, and F1-score in their IGI Global publication.
- Shreya et al. (2023) presented a Secure Smart Farming Framework using Blockchain and IoT, addressing security, scalability, and energy consumption in Sustainable Computing: Informatics and Systems.
- Kwaghtyo and Eke (2023) conducted a Survey Article on Smart Farming Prediction Models, covering various AI techniques in agriculture.

CHAPTER 3: OBJECTIVE

- Development of KNN models with embedded saliency maps: These maps will visually highlight crucial features influencing specific yield predictions, such as rainfall or soil nutrients.

- Creation of interactive XAI dashboards: Farmers can input their field data and receive AI-powered yield predictions, accompanied by clear explanations in the form of saliency maps, charts, and text summaries.
- Evaluation of explainability impact: Measure the increase in farmer trust, understanding, and adoption of AI-driven recommendations through controlled trials and farmer interviews.
- Focus on explainable KNN models and dashboards for a specific crop or region initially, with potential expansion to other AI models and crops in further research.
- User testing and evaluation involving a limited group of farmers initially, with the goal of achieving wider adoption later on.

CHAPTER 4: METHODOLOGIES

1. Data Collection and Preprocessing:

   - Collected and imported dataset "yield_df.csv" using pandas.
   - Dropped null values, removed duplicates, and ensured consistency in data types.

2. Data Representation:

   - Utilized Seaborn library to create graphical representations.
   - Generated frequency distribution plots for crop yields, yield per country, crop types, and relationships between yield and crop type.

3. Train-Test Split and Feature Engineering:

   - Split dataset using train_test_split function.

- Converted categorical variables to numerical representations and performed feature scaling.

4. Model Training and Prediction:

- Employed KNeighborsRegressor for model training.
- Utilized trained model to predict crop yields on the testing dataset.

5. XAI Dashboards:

- Implementing interactive XAI dashboards for farmers to visualize key factors influencing yield predictions.

Through these steps, we aim to develop a robust crop yield prediction model using the KNN algorithm, facilitating informed agricultural decision-making and fostering sustainability.

CHAPTER 5: EXPERIMENTAL SETUP

1. Data Preparation:

- Collect and preprocess diverse dataset including crop yield data and environmental factors.

2. Model Training:

- Train KNN model with varied hyperparameter values, exploring distance metrics and weighting schemes.

3. Model Evaluation:

- Assess KNN model performance using regression metrics, comparing against baseline models or alternative algorithms.

4. Interpretability and Explainability:

- Generate saliency maps to understand feature importance and visualize model predictions for crop yield insights.

5. Parameter Tuning:

- Conduct hyperparameter tuning using grid search or random search with cross-validation.

6. Deployment and Integration:

- Deploy trained KNN model into production environment and integrate into decision support systems for real-time use.

7. Documentation and Reporting:

- Document experimental setup, preprocessing, model configuration, and evaluation metrics. Prepare comprehensive report summarizing results and recommendations.

8. XAI Dashboard Development:

- Develop interactive dashboard allowing farmers to visualize saliency maps and understand factors affecting yield predictions. Include features for exploration and scenario analysis.

CHAPTER 6: CONCLUSION AND FUTURE SCOPE

- Integration of KNN model in Crop Yield Prediction AI project signifies a significant advancement in agriculture, offering precise crop production projections by amalgamating data from various sources.
- Farmers gain access to well-informed decisions, promoting efficiency and sustainability in agriculture through machine learning-based predictions.
- Accurate crop yield forecasts aid in reducing input costs, managing risks, and increasing harvests, thereby enhancing the agricultural industry's resilience and profitability.
- The project contributes to sustainability efforts by enabling more efficient use of resources like water, fertilizers, and pesticides, benefiting both the environment and economic viability of farming operations.
- By optimizing crop production and reducing yield fluctuations, the project supports food security initiatives at regional, national, and global levels.
- Reliable crop yield predictions facilitate policymakers in implementing targeted interventions, allocating resources effectively, and mitigating potential food shortages or price volatility.
- Overall, the successful implementation of the KNN model-powered Crop Yield Prediction AI study promises to revolutionize crop management strategies, enabling higher yields while minimizing environmental impact and contributing to global food security.

REFERENCES

1. Sharma, P., Dadheech, P., Aneja, N., & Aneja, S. (2023). Predicting Agriculture Yields Based on Machine Learning Using Regression and Deep Learning. IEEE Access.
2. Bagheri, S., Cheung, G., & Eadie, T. (2023). Graph Sparsification for GCN Towards Optimal Crop Yield Predictions. arXiv preprint arXiv:2306.01725.
3. Albert, M. C., Hans, H., Karteja, H., & Widianto, M. H. (2023, February). Development of Hydroponic IoT-based Monitoring System and Automatic Nutrition Control using KNN. In 2023 International Conference on Computer Science, Information Technology and Engineering (ICCoSITE) (pp. 974-979). IEEE.
4. Islam, M. M., Adil, M. A. A., Talukder, M. A., Ahamed, M. K. U., Uddin, M. A., Hasan, M. K., ... & Debnath, S. K. (2023). DeepCrop: Deep learning-based crop disease prediction with web application. Journal of Agriculture and Food Research, 14, 100764.
5. Sharma, P., Dadheech, P., & Senthil, A. S. K. (2023). AI-Enabled Crop Recommendation System Based on Soil and Weather Patterns. In Artificial Intelligence Tools and Technologies for Smart Farming and Agriculture Practices (pp. 184-199). IGI Global.

6. Shreya, S., Chatterjee, K., & Singh, A. (2023). BFSF: A secure IoT based framework for smart farming using blockchain. Sustainable Computing: Informatics and Systems, 40, 100917.
7. Kwaghtyo, D. K., & Eke, C. I. (2023). Smart farming prediction models for precision agriculture: a comprehensive survey. Artificial Intelligence Review, 56(6), 5729-5772.