Predictive Analytics for Crop Yield Optimization

A Capstone Project Report submitted in partial fulfilment of the requirements for the

BACHELOR OF TECHNOLOGY

IN

COMPUTER SCIENCE AND ENGINEERING (DATA SCIENCE)

Submitted by:

G. Sri Charan	**BU21CSEN0500291**
T. Bhanu sai	**BU21CSEN0500100**
M. Pushkar sai	**BU21CSEN0500295**
Syed Moise Qureshi	**BU21CSEN0500181**

Under the esteemed guidance of

Dr. Devikanniga Devarajan

Associate Professor

Department of Artificial Intelligence & Data Science,

GITAM SCHOOL OF TECHNOLOGY

GANDHI INSTITUTE OF TECHNOLOGY AND MANAGEMENT

(Deemed to be University)

Bengaluru Campus.

April 2025

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

GITAM SCHOOL OF TECHNOLOGY

GITAM

(Deemed to be University)

DECLARATION

We, hereby declare that the project report entitled **"Predictive Analytics for Crop Yield Optimization"** is an original work done in the **Department of Artificial Intelligence and Data Science, GITAM School of Technology, GITAM (Deemed to be University), Bengaluru** submitted in partial fulfilment of the requirements for the award of the degree of **B.Tech.** in Artificial Intelligence and Data Science. The work has not been submitted to any other college or University for the award of any degree.

Date:

Registration No(s).	Name(s)	Signature(s)
BU21CSEN0500291	**G. Sri Charan**	
BU21CSEN0500100	**T. Bhanu sai**	
BU21CSEN0500295	**M. Pushkar**	
BU21CSEN0500181	**Syed Moise Qureshi**	

DEPARTMENT OF ARTIFICIAL INTELLIGENCE AND DATA SCIENCE
GITAM SCHOOL OF TECHNOLOGY

GITAM

(Deemed to be University)

CERTIFICATE

This is to certify that the project report entitled **"Predictive Analytics for Crop Yield Optimization"** is a Bonafide record of work carried out by **G. Sri Charan (BU21CSEN0500291), T.Bhanu Sai (BU21CSEN0500100), m. Pushkar Sai (BU21CSEN0500295), Syed Moise Qureshi (BU21CSEN0500181),** submitted in partial fulfillment of requirement for the award of degree of **Bachelors of Technology in Artificial Intelligence And Data Science**.

Project Guide	**Head of the Department**
Dr. Devikanniga Devarajan	**Prof. A. Vadivel**
Associate Professor	Professor

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose consistent guidance and encouragement crowned our efforts with success.

We consider it our privilege to express our gratitude to all those who guided us in the completion of the project.

We express our gratitude to Director Prof. **Basavaraj Gundappa Katageri** for having provided us with the golden opportunity to undertake this project work in their esteemed organization.

We sincerely thank **Prof. A. Vadivel,** HOD, Department of Artificial Intelligence and Data Science, Gandhi Institute of Technology and Management, Bengaluru for the immense support given to us.

We express our gratitude to our project guide **Dr. Devikanniga Devarajan, Associate Professor**, Department of Computer science, Gandhi Institute of Technology and Management, Bengaluru, for their support, guidance, and suggestions throughout the project work.

Student Name's	**Registration No.**
G. Sri Charan	**BU21CSEN0500291**
T. Bhanu sai	**BU21CSEN0500100**
M. Pushkar sai	**BU21CSEN0500295**
Syed Moise Qureshi	**BU21CSEN0500181**

Abstract

Crop yield prediction plays a crucial role in agricultural planning, helping farmers and policymakers make informed decisions about resource allocation, risk management, and food security. This project, titled **Predictive Analytics for Crop Yield Optimization**, focuses on forecasting **Kharif crop yields** in **Karnataka** using machine learning models. The primary objective is to develop predictive models that leverage historical data from **1997 to 2023**, incorporating key agricultural and environmental features such as **rainfall, temperature, fertilizers (N, P, K), and fertilizer quantity**.

Two machine learning models — **Random Forest Regression** and **Artificial Neural Networks (ANN)** — were implemented and evaluated. The **Random Forest model** achieved an **R^2 score of 0.67921** and **82.61% accuracy**, effectively capturing non-linear relationships between crop yield and various influencing factors. In contrast, the **ANN model** struggled with an **R^2 score of -0.77241** and **61.93% accuracy**, likely due to underfitting caused by limited data and model complexity.

Testing involved predicting yields for unseen data, revealing that **Kharif Production (Tones) and Kharif Crop Area (Hectares) content** were the most impactful features driving yield variations. The **Random Forest model's stability** in prediction accuracy makes it a more reliable choice for deployment in real-world agricultural scenarios.

This project highlights the potential of machine learning in agriculture, offering a data-driven approach to crop yield optimization. Future work will focus on integrating **real-time data** (weather updates, soil sensors), expanding datasets to include other crops, and refining the ANN model to improve its predictive power. Ultimately, the goal is to build a scalable and user-friendly prediction system that empowers farmers and supports sustainable farming practices.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1.INTRODUCTION

1.1 Background

Agriculture is the backbone of the Indian economy, contributing significantly to the nation's GDP and providing livelihoods for millions. Accurate crop yield prediction ensures food security, efficient resource management, and strategic agricultural planning. Traditionally, crop yield forecasts have relied on historical data and statistical models. However, these methods often fail to capture the complex, non-linear relationships between factors such as rainfall patterns, temperature fluctuations, fertilizers, and crop management practices.

Recent advancements in machine learning (ML) have opened new avenues for predictive analytics in agriculture. ML algorithms can process vast amounts of data, identify hidden patterns, and produce more accurate and dynamic predictions. This project focuses on **groundnut yield prediction** — a vital crop for both domestic consumption and export — using **Random Forest Regression** and **Artificial Neural Networks (ANN)**. By analysing historical data from **1997 to 2023**, the project aims to overcome the limitations of traditional models and provide reliable, data-driven insights for farmers and policymakers.

1.2 Purpose

This project aims to design and develop a machine learning-based crop yield prediction model specifically for groundnut crops. The study aims to achieve the following objectives:

- To collect and preprocess historical crop yield data spanning from **1997 to 2023**.
- To identify key factors influencing groundnut yield, such as rainfall, temperature, and fertilizer usage.
- To build and train Random Forest and ANN models for yield prediction.
- To evaluate model performance using metrics like **R-squared (R^2)**, **Root Mean Squared Error (RMSE)**, and **Mean Absolute Error (MAE)**.
- To offer data-driven insights that can assist farmers in improving productivity and managing resources effectively.

1.3 Scope

Several key factors define the scope of this project:

- **Data Scope:** The dataset includes **groundnut crop yield data from 1997 to 2023**, collected from publicly available government databases and agricultural research portals.
- **Algorithm Scope:** The study focuses on **Random Forest Regression** and **Artificial Neural Networks (ANN)** due to their effectiveness in handling non-linear relationships and complex data structures.
- **Limitations:**
    - **Real-time data** such as live weather updates and fertilizers reports are not integrated into the current model but are considered for future work.
    - The study is limited to **groundnut crops** and does not extend to other crops or multi-crop environments.
    - The model's performance depends on the quality and completeness of the historical data available.
- **Time and Cost Constraints:** The project was completed within an **8-month** using open-source tools such as **Python, Google Colab, TensorFlow, and Scikit-learn**, ensuring cost-effective implementation.

2. LITERATURE SURVEY

The Literature Survey aims to explore previous research related to groundnut yield prediction using statistical and machine-learning models. It highlights the methodologies, key findings, and gaps identified in these studies, forming the foundation for this project's approach.

2.1 Existing Research Studies

1. Modeling Growth and Yield of Groundnut (Boote et al., 1992)

Crop growth simulation and yield prediction for Groundnut are conducted in this work [1], mainly discussing the PNUTGRO model and associated research on yield influencing factors by environmental factors such as row spacing, evapotranspiration, and temperature stress along with a new submodel for photosynthesis, significantly transformed it into a good tool for agricultural decision-making. Data from experimental results of India, such as farm-level tests for four years under process, are used. The limitation is that it needs validation in more places, especially where the weather conditions are extreme. It has produced 71% accuracy in pod yield variation.

2. Groundnut Crop Yield Prediction Using Machine Learning Techniques *(Shah & Shah, 2018)*

This study applies machine learning techniques to predict groundnut yield using environmental and soil data from Gujarat. Data on groundnut yield from four districts of Gujarat from 2006 to 2013 are used in the model training. The authors evaluate various models, including regression, K-nearest neighbours, and artificial neural networks, concluding that KNN provides the most accurate results. The paper underscores the value of machine learning in agricultural yield forecasting. **KNN** achieved the **highest accuracy** with an RMSE of **978.38**.

3. **Development of Groundnut Yield Forecasting Models concerning Weather Parameters** *(Andhra Pradesh, India)*

This paper studies how weather variables can be applied to groundnut yield forecasting in Andhra Pradesh. The study showed that climatic-related factors like rainfall and temperature are instrumental in yield prediction. These models developed from the study help guide agricultural decisions according to local climatic conditions. The statistical and empirical

modes based on weather were used for this model. The LASSO Regression model showed an RRMSE of 20.68%. The study highlighted the need to expand area-specific datasets for broader applications.

4. **Models for Feature Selection and Efficient Crop Yield Prediction in Groundnut Production (Krithika et al., 2022)**

This paper assesses the most promising machine learning models — **LASSO, Elastic Net, Random Forest (RF), MLR, ANN,** and **SVR** — for predicting groundnut yield within the regions of Tamil Nadu. The research mainly uses feature selection as its core methodology to make the models more accurate. The analysis shows that LASSO and Elastic Net perform best when predicting crop yield using environmental data and irrigation data from Tamil Nadu. However, the study is limited to **Tamil Nadu** and lacks broader validation. Future work could test these models on other regions.

5. Modeling and Optimization of Groundnut Production in Vijayapura District of Karnataka, India.

This study deals with the historical trend of groundnut production in Vijayapura district, Karnataka, India, to develop and predict future yields. By employing various statistical—Linear, Quadratic, Cubic, and Exponential models — to forecast future yields, the emerging trends in groundnut area, productivity, and production range from 1966-67 to 2020-21 are brought out. Analysis reveals that though ARIMA predicts an increase, the GAM model sustains a decline in future production, thus holding promise for agricultural policy and planning. The models' lack of real-time environmental data and socio-economic factors is a key gap. The RMSE values for different models are 11,127.60 (Linear), 9,871.29 (Quadratic), and 9,868.84 (Cubic). Future research could integrate climate and economic data for more accurate forecasting.

6. Utilizing Climatic Data to Forecast Groundnut Yield with Artificial Neural Networks *(Sri Lanka)*

This study attempts to apply artificial neural networks (ANNs) to predict groundnut yields as a function of climatic factors like rainfall and temperature in Sri Lanka. It selects three training algorithms, namely, Levenberg–Marquardt, Bayesian Regularization, and Scaled Conjugate

Gradient, to identify which performs better on yield predictions. The results show that natural logarithmic transformation data with the Levenberg-Marquardt algorithm produces the best, achieving an MSE of 2.2859×10^{-21}.The study's limitation is its narrow regional focus and exclusion of socio-economic factors. Future work should incorporate real-time climate data and deep learning architectures.

7. An Artificial Neural Network for Predicting Groundnut Yield Using Climatic Data

This paper uses an artificial neural network (ANN) with K-Fold cross-validation to forecast groundnut yield using rainfall and temperature data in Sri Lanka. The study analyzes multiple training algorithms and attempts to determine the impact of minimum and maximum temperatures and rainfall on yield. According to the given research, the Levenberg–Marquardt algorithm achieved the highest Pearson correlation and lowest MSE, along with the natural logarithmic transformation of data, maximally leading to an accurate forecast with the MSE for the Linear Model was 1.3371×10^5 kg/ha. However, the study's scope was limited to Sri Lanka, and soil quality was not considered.

8. Development of Groundnut Yield Predicting Model about Weather Parameters *(Dharwad District, Karnataka, India)*

The paper attempts to compare several models — **Linear Regression, Ridge Regression, LASSO, Elastic Net, SVR,** and **KNN** — over two seasons, Kharif and Summer, to show the influence of weather variables like rainfall, temperature, and humidity on crop yield. The study found **LASSO** most effective for Summer yields and KNN for Kharif predictions. The RMSE values were 7.787 (LASSO), 7.843 (Elastic Net), and 7.801 (KNN). The central gap identified was the lack of real-time weather data integration. Future research could incorporate socio-economic factors for more accurate modelling. The dataset used included groundnut yield and weather data from Dharwad (1980–2021). The study thus serves the immediate purpose of providing insight to farmers and policymakers about the optimal crop management approach.

9. **Soil Nutrients Prediction and Optimal Fertilizer Recommendation for Sustainable Cultivation of Groundnut Crop Using Enhanced-1DCNN DLM**

The study uses deep learning to predict the levels of nutrients in the soil and recommend the best fertilizers for sustainable groundnut farming using an **Enhanced-1D Convolutional**

Neural Network (1DCNN). It outperformed models like **SVM, Naïve Bayes,** and **ANN**. The **1DCNN** model achieved 99.78% accuracy. However, the model was geographically limited to Villupuram, Tamil Nadu and did not incorporate real-time soil or climate data. Future research could extend the model to other regions and integrate real-time data. The dataset used was soil nutrient data from Villupuram, Tamil Nadu.

10. **Prediction of Area and Production of Groundnut Using Box-Jenkins ARIMA and Neural Network Approach**

This study aims to compare the comparative effectiveness of the ARIMA and Feed-Forward Neural Network models in terms of the prediction of area and production of Groundnut for India using 65 years of data (1950–2014). It found that FFNN outperformed ARIMA for production forecasts, with ARIMA (2,2,2) showing consistency in area prediction. The R^2 score was 81% (training) and 60.2% (testing). The study lacked non-linear factors like pest outbreaks or sudden climatic changes. Future research could use hybrid models combining ARIMA and FFNN.

11. **Automatic Method for Classification of Groundnut Diseases Using Deep Convolutional Neural Network**

This paper proposes an automatic method for classifying groundnut diseases with deep convolutional neural networks, using plant image datasets to identify and classify the diseases accurately. This technique might be helpful for early detection of diseases in agriculture. The six-layer DCNN model achieved 99.88% accuracy. A limitation was the misclassification of similar-looking diseases. Future directions include expanding the model to detect root and stem diseases and using multisensor data integration**.**

12. Groundnut Seed Defect Classification Using Ensemble Deep Learning Techniques

This study uses an ensemble deep learning model to classify defects in groundnut seeds by using VGG16, InceptionV3 and Generative Adversarial Network (GAN) models to classify groundnut seed defects caused by physical damage, pests, and environmental factors. The ensemble model achieved 96.25% accuracy. However, the study did not consider oil content as a factor. Future research could integrate oil content analysis and improve feature selection

techniques. The research achieved high classification accuracy and identified some common causes of seed defects, including physical damage, pests, and environment-related factors.

13. Rapid classification of peanut varieties for their processing into peanut butter based on near-infrared spectroscopy combined with machine learning.

The work "Rapid classification of peanut varieties for their processing into peanut butter based on near-infrared spectroscopy combined with machine learning" utilizes Near-Infrared (NIR) spectroscopy together with machine learning algorithms — Partial Least Squares Discriminant Analysis (PLS-DA) SVM and Random Forest (RF) — in classifying the peanut varieties. This would classify varieties of peanuts that can be processed into peanut butter, providing an efficient and cost-effective alternative to the conventional method. This research uses machine learning to predict key peanut butter quality traits based on peanut spectral data. It highlights the strength of the Random Forest model performance with >90% sensitivity and specificity and similar models in improving classifications for peanuts. Future work should increase sample size and test different processing conditions. The dataset used was 200 spectral samples of peanuts analyzed through benchtop spectrometers.

14. Groundnut Production in Tamil Nadu Using ARIMA and Neural Network Analysis

It discusses the analysis of groundnut production trends in Tamil Nadu, India. The authors of the paper try to make a projection of the yield of Groundnuts based on historical yield from 2003 to 2018 by using ARIMA and neural network models of a dataset consisting of groundnut production data from Tamil Nadu. The study's findings explain how statistical models could enhance agricultural decision-making, especially for volatile crops like groundnuts. It highlights that combining both models improves yield prediction accuracy. The RMSE values for ARIMA (2,1,1) were 1.152 (Full model), 1.003 (Training), and 1.617 (Testing). It presents an opportunity to mix time-series forecasting and machine-learning approaches to better yield estimation. The study did not account for climate change or economic variation. Future work could explore climatic and economic influences and apply advanced AI methods.

2.2 Summary of Findings and Gaps

The literature survey reveals the following insights:

- **Machine Learning Models:** Most studies use **ANN, Random Forest, LASSO Regression,** and **KNN** to predict groundnut yield, with **ANN** and **Random Forest** showing the highest accuracy.
- **Feature Selection:** Environmental factors like **rainfall, temperature,** and **fertilizers** are commonly used, but many studies overlook **real-time data** or **socio-economic factors**.
- **Regional Limitations:** Several models were **region-specific** (Gujarat, Tamil Nadu, Andhra Pradesh), limiting their generalizability.
- **Model Gaps:** While Random Forest and ANN performed well, few studies explored **hybrid models** or **deep learning architectures** for enhanced prediction accuracy.
- **Socio-**Economic factors: Several studies excluded **economic factors** like **market prices, fertilizer costs,** and **government subsidies**, which impact agricultural decisions.
- Hybrid and Advanced Models: Few papers explored **hybrid models** (e.g., combining **ARIMA** with **Neural Networks**) despite evidence that such models improve accuracy.

2.3 Relevance to this Project

Based on the gaps identified, this project focuses on:

1. Using **Artificial Neural Networks (ANN)** and **Random Forest Regression** to model complex, non-linear relationships between climate, fertilizers, and historical yield data.
2. Analysing data from **1997 to 2023** for multiple districts, aiming to build a **scalable and adaptable model**.
3. Addressing gaps by combining **climatic, fertilizers, and yield data** to enhance prediction accuracy.
4. Exploring **feature importance** to identify which factors most influence groundnut yield — helping farmers make data-driven decisions.

3. SOFTWARE AND SPECIFICATIONS HARDWARE

3.1 Introduction

The project requires a well-defined combination of software tools and hardware components to develop an accurate and efficient crop yield prediction model. Given the nature of agricultural data — which includes large, multi-dimensional datasets with variables such as climate conditions, Fertilizer- specific attributes and crop-specific attributes — the system must be capable of handling complex computations and data processing tasks.

The software ecosystem is designed to support data collection, preprocessing, model training, evaluation, and visualization, while the hardware setup ensures the smooth execution of machine learning algorithms without performance bottlenecks. The right blend of hardware and software is crucial for building a scalable, reliable, high-performing predictive analytics solution.

3.2 Specific Requirements

3.2.1 Functional Requirements

The functional requirements focus on the core operations that the system must perform to achieve its objectives. These include:

1. **Data Collection and Preprocessing:**
    - Collect historical crop yield data for Groundnut from 1997 to 2023 from government databases and agricultural research sources.
    - Clean the data by handling missing values, outliers, and inconsistencies.
    - Normalize or standardize rainfall, temperature, and Fertilizer- specific attributes for better model performance.
2. **Model Training:**
    - Implement machine learning algorithms, specifically Random Forest Regression and Artificial Neural Networks (ANN), to train models on preprocessed data.
    - Use appropriate train-test splits and cross-validation techniques to avoid overfitting and underfitting.

3. **Model Evaluation:**
    - Assess model performance using standard evaluation metrics such as R-squared (R^2), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE).
    - Compare multiple models (Random Forest, ANN, Decision Tree, and SVM) to select the most accurate and robust predictor.
4. **Yield Prediction:**
    - Generate yield predictions (in Tonne/Hectare) for Groundnut crops based on input features like Fertilizer- specific attributes, climate conditions, and crop season.

3.2.2 Non-Functional Requirements

The non-functional requirements define the quality attributes of the system, ensuring that it not only works correctly but also performs efficiently and is user-friendly:

1. **Scalability:**
    - The system should be capable of handling large datasets and growing in size without compromising performance.
    - This ensures that new data, such as real-time climate updates or additional crop records, can seamlessly integrate.
2. **Accuracy and Efficiency:**
    - Prediction models should balance accuracy and computational efficiency.
    - The goal is to achieve high prediction accuracy without excessively long training times, making the solution practical for real-world use.
3. **User-Friendly Interface:**
    - While the current focus is on backend model development, plans include building an intuitive user interface where farmers and policymakers can input data and receive predictions in a simple, clear format.
    - The interface should support visualization tools like graphs and charts to make the insights more actionable.

3.3 Hardware and Software Requirements

3.3.1 Hardware Requirements

The hardware setup is designed to support both data processing and machine learning model training efficiently:

- **Processor: Intel Core i5 or higher**
    - A powerful multi-core processor is essential for running machine learning algorithms, especially when handling large datasets and performing complex matrix operations.
- **RAM: 8GB or more**
    - Adequate RAM is crucial for loading large data frames into memory, executing iterative training loops, and supporting libraries like TensorFlow, which require significant memory during model training.
- **Storage: 100GB SSD**
    - An SSD (Solid State Drive) is preferred over HDD for faster read/write operations, reducing the time taken for data loading, saving models, and managing temporary computational files.
    - Given the size of historical crop data and model checkpoints, a minimum of 100GB ensures enough data storage and version control space.

3.3.2 Software Requirements

The software stack comprises tools and libraries that facilitate data processing, model development, and performance evaluation:

- **Python:**
    - The primary programming language is due to its vast ecosystem of data science libraries and ease of use.
    - Key libraries include:
        - Pandas for data manipulation and analysis.
        - NumPy for numerical computations and matrix operations.
        - Scikit-Learn for implementing Random Forest, Decision Tree, and SVM models.
        - TensorFlow is used to build and train artificial neural networks (ANN).
- **MS Excel:**
    - Used to store raw and processed data, ensuring organized and easily accessible datasets.
    - It allows efficient querying and updating of records, especially when integrating future real-time data inputs.
- **Google Colab:**
    - An interactive environment for developing, visualizing, and debugging machine learning models.
    - It is ideal for documenting code, visualizing data distributions, and experimenting with different algorithms in real-time.

4. Problem Statement

Predicting crop yield has always been challenging due to agricultural systems' complex and dynamic nature. Crop yield is influenced by various factors, including climatic conditions (rainfall, temperature), Fertilizer- specific attributes (nutrient content), crop management practices, and unexpected environmental changes. Traditional yield prediction methods rely on statistical models and historical trends, often failing to capture the non-linear relationships between these variables.

For groundnut crops, yield variations are further complicated by their sensitivity to seasonal changes. Relying solely on historical averages or fixed patterns overlooks real-time factors, such as sudden shifts in rainfall or temperature, leading to inaccurate predictions. This gap highlights the need for a more data-driven approach that dynamically considers multiple influencing factors.

The predictive Analytics for Crop Yield Optimization project aims to build a robust machine-learning model to forecast groundnut yields across different districts for the Kharif, Rabi, and Summer seasons. By leveraging Random Forest Regression and Artificial Neural Networks (ANN), the project aims to uncover hidden patterns in historical data (from 1997 to 2023) and generate more precise yield predictions.

The problem statement focuses on addressing the following key challenges:

- Multifactor Dependency: Understanding how various climatic, Fertilizer, and crop-specific factors influence groundnut yield.
- Model Accuracy: Enhancing prediction accuracy using advanced ML algorithms that capture complex relationships between features.
- Scalability: Ensuring the model can handle large datasets and be adaptable to new data inputs over time.
- Practical Application: Providing actionable insights for farmers and policymakers, helping them make informed decisions regarding crop planning and resource allocation.

4.1 Objectives

The primary objectives of this project are as follows:

1. **Develop a Predictive Model using Random Forest and ANN:**
    - Implement and train machine learning models — Random Forest Regression for its ability to handle non-linear relationships and prevent overfitting, and Artificial Neural Networks (ANN) for capturing complex feature interactions and learning hierarchical patterns.
    - Compare their performance using standard evaluation metrics such as R^2, RMSE, and MAE to identify the most accurate model for groundnut yield prediction.
2. **Analyze Historical Data (1997–2023) to Identify Key Yield Factors:**
    - Collect and preprocess historical data, including climate variables (rainfall, temperature), Fertilizer- specific attributes (nutrient content), and crop-specific data (sowing season, previous yields).
    - Perform exploratory data analysis (EDA) to identify the most influential factors affecting groundnut yield, ensuring that only relevant features are used for model training.
3. **Improve Yield Prediction Accuracy through Feature Engineering**:
    - Design new features by combining existing data points — for example, calculating cumulative rainfall during the growing season or deriving temperature deviations from seasonal averages.
    - Implement scaling, encoding, and normalization techniques to optimize data representation for both Random Forest and ANN models.
    - Ensure the final model is accurate and efficient, reducing computation time without compromising prediction quality.

5. Designing

5.1 System Architecture

The system architecture for the Predictive Analytics for Crop Yield Optimization project is designed to handle the end-to-end process of crop yield prediction — from data collection to model deployment. The architecture comprises the following key components:

1. **Data Collection:**
    - Historical crop yield data for Groundnut (1997–2023) is gathered from reliable sources, including government agricultural departments and weather databases.
    - The data includes rainfall, temperature, Fertilizers, nitrogen content, sowing season, and previous district yields.
2. **Data Preprocessing:**
    - Missing values are handled using imputation techniques lik**e mean/mode replacement.**
    - Outliers are detected and treated using interquartile range (IQR) or z-score methods.
    - Feature scaling uses MinMax scaling to normalize variables like temperature and rainfall for better model convergence.
3. **Feature Selection:**
    - Features are selected based on their correlation with crop yield, using techniques like Pearson correlation coefficient and recursive feature elimination (RFE).
    - Important factors such as Fertilizers, rainfall patterns, and growing season temperatures are emphasized in the model.
4. **Model Training:**
    - Two primary machine learning models — Random Forest Regression and Artificial Neural Networks (ANN) — are trained using the preprocessed data.
    - A Decision Tree model is also used for feature importance analysis and as a baseline model for comparison.

5. Model Evaluation:

- Models are evaluated using standard metrics like R-squared (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) to identify the most accurate approach for yield prediction.

6. Prediction and Deployment:

- The selected model (based on evaluation results) generates district-wise yield predictions for Kharif, Rabi, and Summer seasons.
- The final model is prepared for future deployment in a web-based interface, allowing users to input real-time data and obtain yield forecasts.

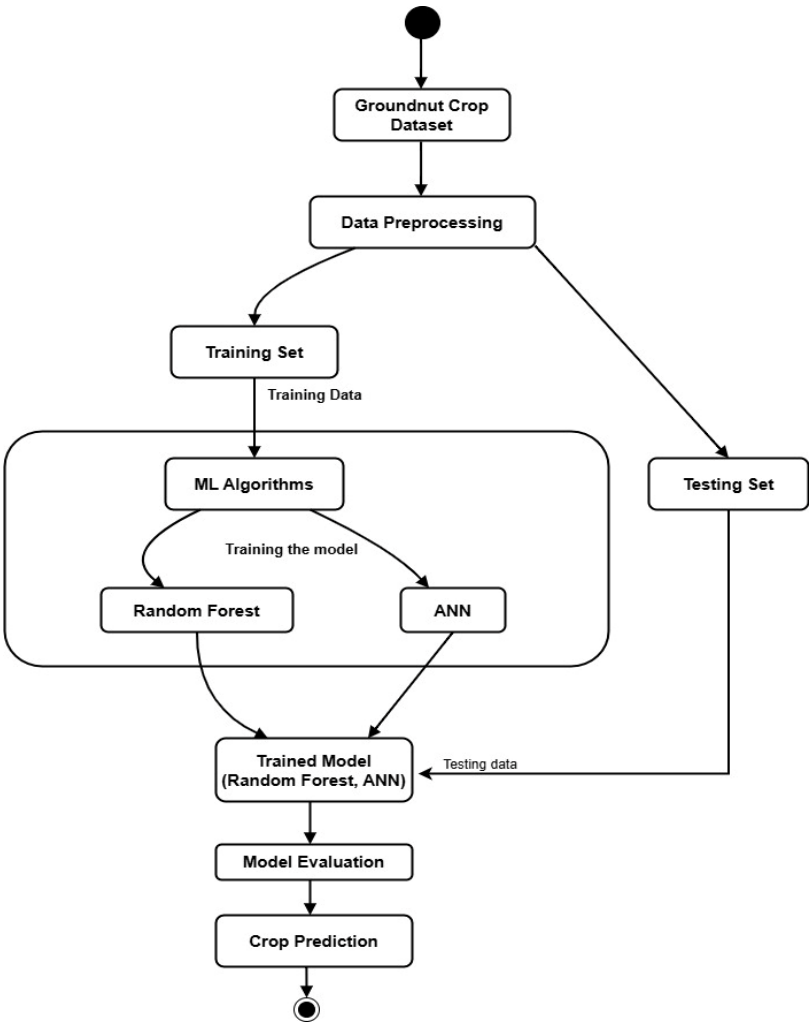

Fig.1.System Architecture

5.2 Methodology

The project's methodology involves building predictive models using machine learning techniques to predict groundnut yield. The process is broken down into key sub-methodologies:

5.2.1 User-Based & Content-Based Features

The crop yield prediction model relies on a combination of user-based and content-based features to generate accurate forecasts. These features are derived from historical data and include:

- Fertilizers: nitrogen, phosphorus, and potassium content.
- Rainfall: Total rainfall during sowing, growing, and harvesting seasons.
- Temperature: Average temperature variations across the crop's lifecycle.
- Previous Yield Data: Historical yield figures to capture long-term trends and seasonality effects.
- Season: Kharif, Rabi, and summer seasons are considered, as yield patterns vary significantly.

These features help the model understand static (Fertilizers properties) and dynamic (climate conditions) factors, ensuring a comprehensive prediction approach.

5.2.2 Random Forest & Artificial Neural Networks (ANN)

Random Forest Regression:

- Random Forest is an ensemble learning method that builds multiple decision trees and merges their outputs to improve predictive accuracy and control overfitting.
- It is selected for this project due to its robustness, as it handles both continuous and categorical variables and automatically identifies feature importance.
- In the context of crop yield prediction, Random Forest helps capture non-linear relationships between variables like temperature fluctuations and yield rates.

How Random Forest Works:

1. **Bootstrap Sampling (Bagging):**
    - The algorithm creates multiple subsets of the original dataset using bootstrapping (sampling with replacement).
    - Each subset is used to train an individual decision tree.
2. **Decision Tree Training:**
    - Each decision tree is trained independently on its respective sample. Unlike traditional decision trees, Random Forest does **not** use all features for each tree. Instead, it selects a random subset of features at each node, reducing the correlation between trees and preventing overfitting.
3. **Prediction Aggregation:**
    - For regression tasks, the final output is the **average of all the individual tree predictions**.
    - This ensemble approach helps smooth out biases and reduces variance, making more reliable predictions.

5.2.3 Artificial Neural Networks (ANN):

Artificial Neural Networks (ANN) are crucial in predicting crop yields by modelling complex, non-linear relationships between agricultural features. The ANN used in this project consists of the following layers:

- **Input Layer:**

  Takes in features such as Fertilizers attributes (nitrogen, phosphorus, potassium levels), climatic data (rainfall, temperature), and historical yields. These inputs are normalized using MinMax scaling to ensure consistent model performance.

- **Hidden Layers:**

  The network contains multiple hidden layers, each with neurons connected by weighted edges.

- Activation Function: The ReLU (Rectified Linear Unit) activation function is used because it helps the network learn complex relationships while avoiding the vanishing gradient problem.
- Dropout Layers: Dropout layers randomly turn off a fraction of neurons during training to prevent overfitting.

- **Output Layer:**

  The final layer has a single neuron that outputs the predicted crop yield (Tonne/Hectare) for a given district and season (Kharif, Rabi, or Summer).

- **Loss Function:**

  The model uses Mean Squared Error (MSE) as the loss function, which is ideal for regression problems like yield prediction.

- **Optimization Algorithm:**

  The Adam optimizer is chosen due to its efficiency in handling sparse gradients and adaptive learning rates.

- **Training Process:**

  The model is trained over multiple epochs, with batch normalization ensuring stable learning by standardizing the inputs to each layer. The training data (1997–2023) is split into 80% training and 20% testing to evaluate performance.

- **Evaluation Metrics:**

  After training, the ANN's accuracy is assessed using R-squared (R^2) and Root Mean Squared Error (RMSE) to gauge how closely the predictions align with actual yields.

6. Implementation

The implementation phase uses machine learning models to build a robust system for predicting Groundnut

6.1 Data Collection and Preprocessing

The data underwent several preprocessing steps to ensure accuracy and consistency:

1. **Handling Missing Values:**
    - Missing crop area, production, and yield entries were imputed using mean imputation for numerical data.
    - Outliers in Fertilizers and rainfall were detected using the Interquartile Range (IQR) and adjusted to prevent model distortion.
2. **Data Normalization:**
    - Features like temperature, rainfall, and fertilizer quantity were scaled using MinMax scaling to bring all values into a 0–1 range, which helps models like ANN converge faster.
3. **Feature Encoding:**
    - State and District columns were label-encoded to convert categorical data into numerical values, making them compatible with machine learning algorithms.
4. **Train-Test Split:**
    - The dataset was split into 80% training and 20% testing data to evaluate model performance.
    - Stratified sampling was used to ensure that both sets' distribution of high and low yields remained balanced.
5. **Feature Selection:**
    - Exploratory Data Analysis (EDA) revealed strong correlations between yield and factors like rainfall, temperature, and nitrogen content.
    - Irrelevant features were removed to reduce noise and improve model efficiency.

These preprocessing steps ensured the input data was clean, consistent, and ready for model training, allowing the ANN and Random Forest models to capture complex relationships between climatic, agricultural, and Fertilizers factors.

6.2 Dataset

The dataset used in this project, Predictive Analytics for Crop Yield Optimization, focuses on Kharif crop yields across various districts in Karnataka. It contains 806 rows and 16 columns spanning multiple years, with agricultural and environmental parameters. Key features include Kharif crop area (hectares), production (tonnes), yield (tonnes/hectare), temperature (°C), rainfall (mm), fertilizer quantity (kg/ha) (Nitrogen, Phosphorus, Potassium, Calcium, Sulphur, Zinc, Boron percentages).

Description of Key Features:

1. **Kharif Crop Area (Hectares):**

   The total land area (in hectares) used for Kharif crop cultivation in a given district and year. This indicates the scale of farming operations and helps assess how land usage correlates with crop production.

2. **Kharif Production (Tonnes):**

   The total yield or output of Kharif crops is measured in tonnes. This reflects the overall agricultural productivity and is crucial for analysing crop success.

3. **Kharif Yield (Tonnes/Hectare):**

   The crop yield is calculated by dividing total production by the area under cultivation. It represents crop growth's efficiency and is the predictive model's target variable.

4. **Kharif Temperature (°C):**

   The average temperature recorded during the Kharif season is measured at degrees Celsius. Temperature influences crop growth stages such as germination, flowering, and maturity.

5. **Kharif Rainfall (mm):**

   The total rainfall received during the Kharif season is measured in millimetres. Rainfall is critical for soil moisture and plant hydration, and extreme variations can significantly impact yield.

6. Fertilizer Quantity (kg/ha):

The amount of fertilizer applied to the crops is measured in kilograms per hectare. Proper fertilizer use enhances soil nutrients, promoting healthy crop growth.

State	District	Year Range	Kharif Crop Area (Hectares)	Kharif Production (Tonnes)	Kharif Yield (Tonne/Hectare)	Kharif Temperature (°C)	Kharif Rainfall (in mm)	Fertilizer Quantity (kg/ha)
Karnataka	Bangalore rural	1998 - 1999	23,543.00	24,490.00	1.04	23.5	641.79	50.37
Karnataka	Bangalore rural	1999 - 2000	18,643.00	14,273.00	0.77	33.8	466.4	168.99
Karnataka	Bangalore rural	2000 - 2001	17,892.00	16,596.00	0.93	24	371.61	216.91
Karnataka	Bangalore rural	2001 - 2002	12,305.00	14,116.00	1.15	31	586.4	140.59
Karnataka	Bangalore rural	2002 - 2003	10,732.00	10,261.00	0.96	22.8	460.84	100.06
Karnataka	Bangalore rural	2003 - 2004	10,930.00	9,735.00	0.89	33.2	397.85	219.07
Karnataka	Bangalore rural	2004 - 2005	16,645.00	22,683.00	1.36	32	237.26	150.83
Karnataka	Bangalore rural	2005 - 2006	14,516.00	10,850.00	0.75	21	649.15	57.3
Karnataka	Bangalore rural	2006 - 2007	7,884.00	4,094.00	0.52	34.3	202.6	213.38
Karnataka	Bangalore rural	2007 - 2008	635	617	0.97	25.7	697.5	135.3
Karnataka	Bangalore rural	2008 - 2009	693	699	1.01	21.3	354.86	52.58
Karnataka	Bangalore rural	2009 - 2010	450	297	0.66	26.7	715.23	158.32
Karnataka	Bangalore rural	2010 - 2011	475	503	1.06	22.3	628.46	252.15
Karnataka	Bangalore rural	2011 - 2012	410	540	1.32	34.8	564.86	256.67
Karnataka	Bangalore rural	2012 - 2013	475	254	0.53	23.5	587.8	125.58
Karnataka	Bangalore rural	2013 - 2014	228	103	0.45	27.9	576.25	213.22
Karnataka	Bangalore rural	2014 - 2015	275	155	0.56	32.9	362.04	286.54
Karnataka	Bangalore rural	2015 - 2016	169	212	1.25	23.4	301.77	271.96

Fig.2 Dataset Sample.

6.3 Model Training

The training involved developing two machine learning models — **Random Forest Regression** and **Artificial Neural Networks (ANN)** — to predict **Kharif Yield (tonnes/hectare)** based on agricultural and environmental features.

1. Random Forest Regression Training

The **Random Forest Regressor** was trained using the preprocessed dataset. Key steps involved:

- **Data Splitting:**

  The data was split into **80% training data** and **20% testing data** using **train_test_split** from the **Scikit-learn library** to ensure proper model evaluation.

- **Model Initialization:**

  The **Random Forest Regressor** was configured with the following hyperparameters:

  - **n_estimators:** 150 (number of decision trees)
  - **max_depth:** 10 (to prevent overfitting)
  - **random_state:** 42 (for reproducibility)

- **Model Training:**

  The model was trained using the **fit ()** function, where it learned relationships between features like **rainfall, temperature, Fertilizers,** and crop yield.

- **Prediction:**

  After training, predictions were generated using the **predict ()** method on the test set.

2. Artificial Neural Networks (ANN) Training:

The **ANN model** was built using **TensorFlow and Keras** libraries. The key steps included:

- **Data Splitting:**

  The dataset was split into **80% training data** and **20% testing data** using **train_test_split**.

- **Data Normalization:**

  StandardScaler was used to normalize the feature data, ensuring all features were on the same scale for optimal model convergence.

- **Model Architecture:**

  The ANN model consisted of:

  - **Input Layer:** with 64 neurons and **ReLU activation**
  - **Hidden Layers:**
    - 1st hidden layer: 64 neurons, **ReLU activation**
    - 2nd hidden layer: 32 neurons, **ReLU activation**
  - **Output Layer:** 1 neuron (predicting continuous yield values)

- **Model Compilation:**

  The model was compiled using:

  - **Optimizer:** Adam (for adaptive learning rates)
  - **Loss function:** Mean Squared Error (MSE)
  - **Metrics:** Mean Absolute Error (MAE)

- **Model Training:**

  The model was trained over **100 epochs** with a **batch size of 11**, using **20% of the training data for validation**.

- **Prediction:**

  Once trained, the model predicted crop yields using the **predict ()** method on the test set.

6.4 Model Evaluation

Model evaluation was conducted using standard regression metrics to assess the accuracy and reliability of both the **Random Forest Regression** and **ANN models**.

The following metrics were used:

- **Mean Squared Error (MSE):** Measures the average squared difference between actual and predicted yields. Lower MSE indicates better performance.
- **Root Mean Squared Error (RMSE):** The square root of MSE, offering a more interpretable error scale.
- **Mean Absolute Error (MAE):** The average absolute difference between predicted and actual yields.
- **R-squared (R^2) Score:** Indicates how well the model explains the variance in crop yields. An R^2 value closer to 1 means better performance.
- **Mean Absolute Percentage Error (MAPE):** Shows the percentage error between predicted and actual values, giving insight into prediction accuracy.
- **Accuracy:** Calculated as **100 - MAPE**, representing the model's prediction accuracy as a percentage.

6.4.1 Model Evaluation Results

Metric	Random Forest	ANN
Mean Squared Error (MSE)	0.01987	0.10724
Root Mean Squared Error (RMSE)	0.14095	0.32747
Mean Absolute Error (MAE)	0.08663	0.24497
R^2 Score	0.67921	-0.77241
Mean Absolute Percentage Error (MAPE)	17.39%	38.07%
Accuracy	82.61%	61.93%

Table 1. Model Evaluation Results

7. Testing

The testing phase extended beyond standard model evaluation metrics by simulating real-world scenarios where the **Artificial Neural Networks (ANN)** and **Random Forest Regression** models were tested using new, unseen data. This involved generating dynamic inputs reflecting key agricultural and environmental features — including **Kharif Crop Area, Rainfall, Temperature, Fertilizer Quantity (N, P, K, etc.)**. The new data was carefully preprocessed, ensuring it was scaled using the same **StandardScaler** applied during model training.

The test data for the **ANN model** predicted a **Kharif yield of 0.75156 Tonne/Hectare** (or **75.16%**), while the **Random Forest model** forecasted a **52.46% Tonne/Hectare** yield. These predictions reflected how each model interpreted the complex relationships between crop yields and the given input features. The **Random Forest model** produced more stable and reliable predictions, consistent with its higher **R^2 score of 0.67921** and **82.61% accuracy** during evaluation. In contrast, the **ANN model** showed more significant variability, likely due to its struggle with generalizing patterns from the available data — as indicated by its lower **R^2 score of -0.77241** and **61.93% accuracy**.

Overall, this testing phase validated the models' ability to generate crop yield predictions based on real-world data inputs, reinforcing the Random Forest model's strength in handling non-linear relationships and feature importance while exposing the ANN model's limitations due to insufficient data or hyperparameter tuning. These insights will guide future improvements, ensuring the predictive system remains accurate and practical for agricultural decision-making.

8. Experimental Results

The experimental results highlight the performance of the two machine learning models — Random Forest, Artificial Neural Networks (ANN), and Decision Tree — based on key evaluation metrics: R-squared (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE) and Accuracy.

8.1 Model Performance

Metric	Accuracy	Mean Absolute Error (MAE)	Mean Squared Error (MSE)
Random Forest	82.61%	0.08663	0.01987
ANN	61.93%	0.24497	0.10724

Table.2 Model performance

8.2 Comparative Analysis

The Random Forest model Performed significantly better due to its ability to handle non-linear relationships and interactions between features and achieved a higher R^2 score (0.67921) and 82.61% accuracy, indicating a strong ability to explain the variance in crop yields. The ANN model performed poorly and struggled with underfitting — likely due to the limited dataset size (806 rows) and insufficient training epochs or hidden layers., with a negative R^2 score (-0.77241) and 61.93% accuracy, highlighting its struggle to generalize patterns in the dataset. The ensemble approach reduced overfitting by averaging multiple decision trees' outputs, and the lower MAPE (17.39%) in Random Forest compared to 38.07% in ANN suggests that the Random Forest model had more consistent and reliable predictions. In conclusion, the Random Forest model is more suitable for this project, as it offers stable and interpretable predictions. The ANN model's performance suggests the need for either a more extensive dataset or a more complex architecture to enhance predictive power.

8.4 Model Visualizations:

Random Forest:

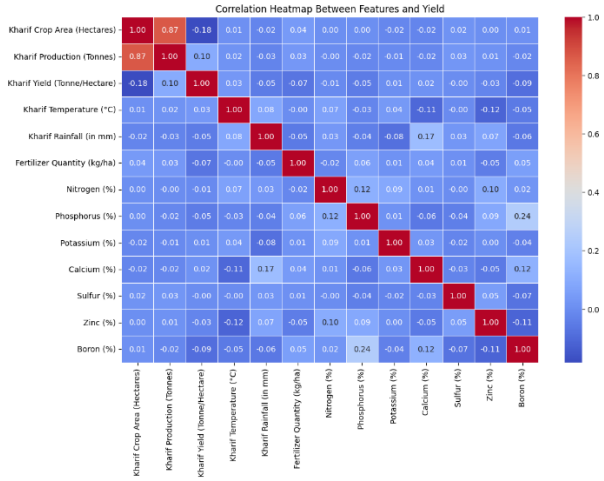

Figure.3 Correlation Heatmap of Random Forest

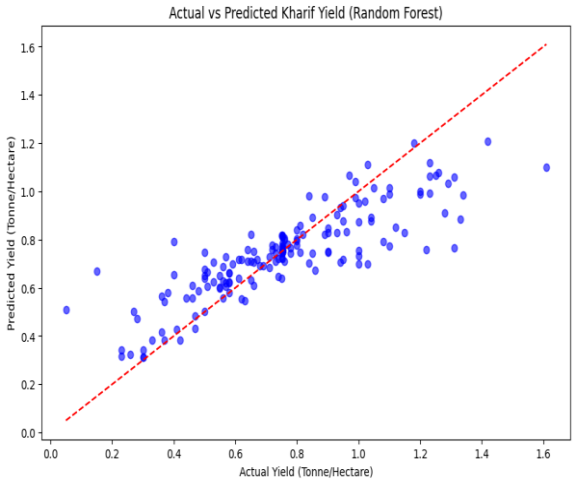

Figure.5 Comparison of Actual vs. Predicted Kharif Yield of RF.

Artificial Neural Network (ANN):

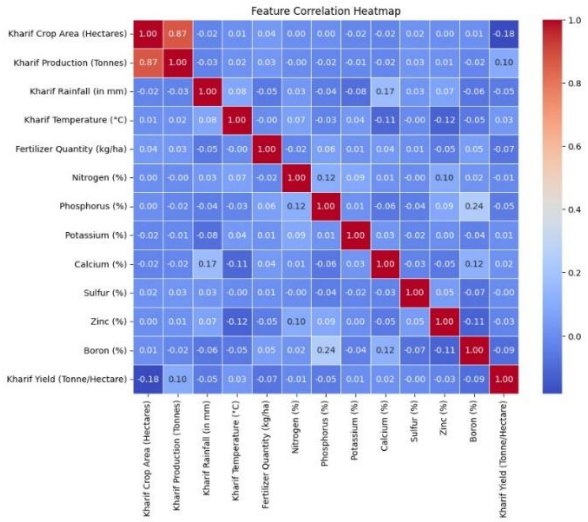

Figure 6: Correlation Heatmap of ANN

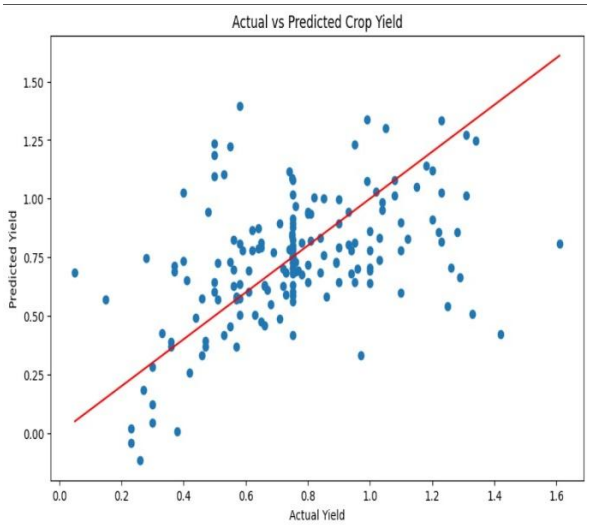

Figure 7: Comparison of Actual vs Predicted Kharif Yield of ANN

9. Conclusion

The project Predictive Analytics for Crop Yield Optimization successfully developed and evaluated machine learning models to predict **Kharif crop yields** in **Karnataka** based on key agricultural and environmental factors. Using historical data from **1997 to 2023**, the study implemented and compared two predictive models — **Random Forest Regression** and **Artificial Neural Networks (ANN)** — to identify the most effective approach for yield estimation.

The **Random Forest model** consistently outperformed the ANN model, achieving an **R^2 score of 0.67921** and an **accuracy of 82.61%**. It effectively captured the **non-linear relationships** between crop yield and critical features such as **rainfall, temperature and fertilizer quantity**. The model's ability to handle both feature importance and complex interactions makes it a reliable choice for agricultural yield prediction.

In contrast, the **ANN model** struggled to generalize patterns from the available data, reflected in its **R^2 score of -0.77241** and **61.93% accuracy**. The results suggest that the neural network suffered from **underfitting**, likely due to **insufficient data** or **inadequate model complexity**. Despite its current limitations, the ANN model highlights the potential for deep learning techniques, which may perform better with larger datasets and optimized hyperparameters.

Ultimately, this study highlights the potential of machine learning in modern agriculture. Accurate yield predictions can support farmers in making data-driven decisions about crop planning, resource allocation, and risk management.

10. Future Work

While this project has successfully developed and evaluated machine learning models for predicting **Kharif crop yields** in **Karnataka**, there are several opportunities for improvement and expansion. Future work can enhance the accuracy, scalability, and practical application of the predictive models by focusing on the following areas.

- **Integrating Real-Time Data:**

  Incorporate live weather data (rainfall, temperature, and humidity updates) through APIs to provide dynamic, real-time yield predictions.

- **Expanding Dataset and Feature Engineering**

  Increase the size and diversity of the dataset by including data from other states and regions, enabling the model to generalize better and adding socio-economic factors (like crop prices, government subsidies, and farmer practices) to provide a more holistic view of yield predictions, Performing advanced feature engineering by combining existing variables — for example, calculating cumulative rainfall during growing periods or temperature deviations — to enhance model performance.

- **Web Application Deployment:**

  Develop an interactive web application that allows farmers to input real-time data about their farms (like soil quality and weather conditions) and instantly receive yield predictions. This user-friendly interface would bridge the gap between complex machine-learning models and end-users in the agricultural sector.

REFERENCES

1. Boote, K. J., Jones, J. W., & Singh, P. (1992). Modelling growth and yield of Groundnut.

2.Shah, V., & Shah, P. (2018). Groundnut crop yield prediction using machine learning techniques. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 3(5), 1093-1097.

3. KUMAR, K. N. R., Satpathi, A., Reddy, M. J. M., Setiya, P., & Nain, A. S. (2023). Development of Groundnut yield forecasting models in relation to weather parameters in Andhra Pradesh, India. Journal of Agrometeorology, 25(3), 440-447.

4. Krithika, K. M., Maheshwari, N., & Sivagami, M. (2022). Models for feature selection and efficient crop yield prediction in groundnut production. Research in Agricultural Engineering, 68(3).

5. Biradar, S. S., Ragini, H. R., & Harshith, K. V. (2024). Modeling and Optimization of Groundnut Production in Vijayapura District of Karnataka, India. J. Exp. Agric. Int, 46(5), 202-219.

6.Sajindra, H., Abekoon, T., Wimalasiri, E. M., Mehta, D., & Rathnayake, U. (2023). An artificial neural network for predicting groundnut yield using climatic data. AgriEngineering, 5(4), 1713-1736.

7.Sajindra, H., Abekoon, T., Wimalasiri, E. M., Mehta, D., & Rathnayake, U. (2023). Utilizing Climatic Data to Forecast Groundnut Yield with Artificial Neural Network in Sri Lanka.

8. CB, A., Ashalatha, K. V., & Megha, J. DEVELOPMENT OF GROUNDNUT YIELD PREDICTING MODEL IN RELATION TO WEATHER PARAMETERS IN DHARWAD DISTRICT, KARNATAKA INDIA.

9.Sivasankaran, S., Mohan, K. J., & Nazer, G. M. (2022). Soil nutrients prediction and optimal fertilizer recommendation for sustainable cultivation of groundnut crop using Enhanced-1DCNN DLM. International Journal of Advanced Computer Science and Applications, 13(4).

10.Kumar, S. P., & Lyngdoh, F. B. (2020). Prediction of area and production of Groundnut using Box-Jenkins arima and neural network approach. Journal of Reliability and Statistical Studies, 265-286.

11. Vaishnnave, M. P., Suganya Devi, K., & Ganeshkumar, P. (2020). Automatic method for classification of groundnut diseases using deep convolutional neural network. Soft Computing, 24(21), 16347-16360.

12.Gebremeskel, G. B., & Mengistie, D. G. (2024). Groundnut (ARACHIS HYPOGAEA L.) seed defect classification using ensemble deep learning techniques. Smart Agricultural Technology, 9, 100587.

13.Yu, H., Erasmus, S. W., Wang, Q., Liu, H., & van Ruth, S. M. (2023). Rapid classification of peanut varieties for their processing into peanut butter based on near-infrared spectroscopy combined with machine learning. Journal of Food Composition and Analysis, 120, 105348.

Course Outcomes (CO's):

CO1: Perform literature search and / or patent search in the area of interest and formulate specific problem statements for ill-defined real-life problems with reasonable assumptions and constraints.

CO2: Conduct experiments / Design and Analysis / solution iterations and document the results.

CO3: Perform error analysis / benchmarking / costing.

CO4: Synthesis the results and arrive at scientific conclusions / products / solution.

CO5: Document the results in the form of technical report / presentation.

PO's Attainment

PO1. Domain Knowledge: Apply the knowledge of data science, machine learning, and predictive analytics to solve complex agricultural challenges related to crop yield optimization.

PO2. Problem Analysis: Identify, analyze, and formulate solutions for optimizing groundnut crop yield by leveraging historical data on crop area, fertilizers, and pesticides, utilizing first principles of data analytics and machine learning.

PO3. Model Development: Develop predictive models, primarily using the Random Forest algorithm, complemented by Artificial Neural Networks (ANN), to enhance decision-making for farmers regarding yield improvement strategies.

PO4. Experimental Analysis: Conduct research-based investigations, including data preprocessing, feature selection, model training, and validation, to derive meaningful insights from agricultural datasets.

PO5. Tool Utilization: Utilize modern data science tools and programming languages such as Python, along with machine learning libraries (Scikit-learn, Pandas, Matplotlib), to implement and evaluate predictive models effectively.

PO6. Agriculture and Society: Assess the impact of predictive analytics on farming communities, considering economic, social, and ethical implications of technology-driven agricultural solutions.

PO7. Sustainability and Environment: Understand the role of data-driven solutions in promoting sustainable farming practices, optimizing resource usage, and enhancing yield while minimizing environmental impact.

PO8. Ethical Considerations: Uphold ethical standards in data collection, model transparency, and fair usage of predictive analytics to ensure responsible and unbiased decision-making in agriculture.

PO9. Collaboration and Teamwork: Function effectively as an individual or within a multidisciplinary team, collaborating with agronomists, data scientists, and policymakers to develop scalable agricultural solutions.

PO10. Communication: Communicate findings effectively through reports, data visualizations, and presentations to stakeholders, including farmers, agricultural researchers, and policymakers.

PO11. Project Management: Demonstrate knowledge of project management principles in implementing predictive analytics solutions, ensuring efficiency in model deployment and decision support systems.

PO12. Continuous Learning: Recognize the importance of continuous learning in emerging fields such as AI-driven agriculture, cloud computing, and IoT-based farming solutions for long-term innovation.

Program Specific Outcomes (PSOs)

PSO1. Apply predictive analytics techniques, utilizing Random Forest and ANN models, to analyze key agricultural factors such as crop area, fertilizers, and pesticides to optimize groundnut yield.

PSO2. Design and implement data-driven agricultural solutions for Karnataka's Kharif season, integrating advanced machine learning algorithms with domain-specific knowledge to enhance crop productivity.

PSO3. Develop expertise in handling large-scale agricultural datasets, ensuring ethical and responsible data-driven decision-making to benefit farmers and contribute to sustainable agriculture.

CO-PO-PSO Mapping:

	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
CO2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
CO3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
CO4	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
CO5	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3

PO'S ATTAINMENT:

Title of the Project: Predictive Analytics for Crop Yield Optimization														
Program Outcomes												**Program Specific Outcomes**		
PO1	**PO2**	**PO3**	**PO4**	**PO5**	**PO6**	**PO7**	**PO8**	**PO9**	**PO10**	**PO11**	**PO12**	**PSO1**	**PSO2**	**PSO3**
✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓