Industrial Internship Report on

”Prediction of Agriculture Crop Production in India”

Prepared by

Shivangi Singh

Executive Summary

This report provides details of the Industrial Internship provided by upskill Campus and The IoT Academy in collaboration with Industrial Partner UniConverge Technologies Pvt Ltd (UCT).

This internship was focused on a project/problem statement provided by UCT. We had to finish the project including the report in 6 weeks’ time.

My project was Prediction of agriculture crop production in India.

Across The Globe India Is The Second Largest Country having People more than 1.3 Billion.

Many People Are Dependent On The Agriculture And it is the Main Resource.

In Agriculture Cultivation/Production Having More Problems.

I want to solve the Big problem in India and useful to many more people

This internship gave me a very good opportunity to get exposure to Industrial problems and design/implement solution for that. It was an overall great experience to have this internship.

TABLE OF CONTENTS

1. Preface

During my 6-week internship at upSkill, I had the opportunity to gain hands-on experience in Data Science and Machine Learning. This internship provided a valuable platform to apply theoretical knowledge and develop practical skills.

Key responsibility:

- Primary goal is on the understanding of data science and Machine Learning.
- Learn the entire machine learning algorithm which helps me a lot during my project.

Internships allow individuals to apply theoretical knowledge in real-world settings. This hands-on experience is vital for understanding how concepts learned in the classroom translate into practical skills required in the workplace.

PROBLEM STATEMENT

Agriculture is a cornerstone of India's economy, supporting a significant portion of the population and contributing to the country's GDP. However, crop production is influenced by various factors such as weather conditions, soil health, irrigation practices, pest outbreaks, and market dynamics. Accurate prediction of agricultural crop production is crucial for:

1. Food Security
2. Resource Management
3. Policy Formulation
4. Risk Mitigation

Overall learning

During my internship in machine learning and data science, I gained a wealth of knowledge and practical experience that significantly shaped my understanding of the field. Here are some key learning's and experiences:

Real-World Projects, Hands-On with Tools: Numpy, pandas etc. , Model Building, Data Cleaning and Preparation.

Reflecting on my learning's and experiences, I feel immense gratitude for everyone who has contributed to my journey. A special thanks to my mentors whose guidance and support have been invaluable. I also appreciate my peers who collaborated with me on projects and shared insights that enriched my understanding.

To my juniors, I encourage you to embrace challenges and seek help when needed. Every experience, whether positive or negative, is a stepping stone to growth. Don't hesitate to connect with others; the relationships you build will be as valuable as the knowledge you gain.

2 . Introduction

2.1 About Prediction of Agriculture Crop Production in India

Agriculture is a cornerstone of India's economy, employing a significant portion of the population and contributing substantially to the country’s GDP. However, farmers often face challenges such as unpredictable weather patterns, soil variability, and pest infestations, which can adversely affect crop yields. In this context, leveraging data science and machine learning can provide invaluable insights to improve agricultural productivity and sustainability.

The "Prediction of Agriculture Crop in India" project aims to develop a robust predictive model that helps farmers and stakeholders make informed decisions about crop selection, resource allocation, and management practices. By analyzing historical data, including climatic conditions, soil health, and crop performance, the project seeks to identify patterns and correlations that can guide future planting strategies.

2.2. Overview:

The "Prediction of Agriculture Crop in India" project aims to harness data science and machine learning techniques to predict crop yields based on various influencing factors. By analyzing historical agricultural data alongside environmental variables, the project seeks to provide actionable insights that empower farmers, enhance productivity, and promote sustainable farming practices.

2.3. Objective:

- Data Collection:

- Meteorological Data: Collect data on temperature, rainfall, humidity, and other climatic factors from weather stations and satellite sources.
- Soil Data: Gather information on soil types, pH levels, nutrient content, and moisture levels, which are crucial for determining crop suitability.
- Crop Yield Data: Compile historical crop yield data across different regions, including crop varieties and farming practices.

- Data Preprocessing:

- Clean and normalize the data to handle missing values and outliers.
- Perform exploratory data analysis (EDA) to identify trends, correlations, and insights that inform model development.

- Model Development:

- Employ various machine learning algorithms, such as regression analysis, decision trees, and ensemble methods, to create predictive models.
- Train and validate these models using a portion of the dataset while reserving another portion for testing their accuracy.

- Visualization and User Interface:

- Develop intuitive dashboards and visualizations that display predictions and insights in a user-friendly manner.
- Allow farmers and agricultural stakeholders to input specific parameters to receive tailored crop recommendations.

2.4 Reference

[1] https://www.sciencedirect.com/science/article/abs/pii/S221478532101052X

[2] https://en.wikipedia.org/wiki/Data_science

[3] https://www.periyaruniversity.ac.in/ijcii/issue/Vol6No4Mar2017/M5_PID0370.pdf

3. Problem Statement

In the assigned problem statement

Agriculture in India faces significant challenges due to various factors, including unpredictable weather patterns, soil degradation, water scarcity, and pest infestations. These issues lead to inconsistent crop yields, affecting the livelihoods of millions of farmers and threatening food security in the country.

Despite the availability of data related to climate, soil conditions, and historical crop yields, farmers often lack the tools and insights needed to make informed decisions about crop selection and management practices. This gap in knowledge results in suboptimal farming practices, increased costs, and lower overall productivity.

To address these challenges, the "Prediction of Agriculture Crop in India" project aims to develop a predictive model that utilizes machine learning techniques to forecast crop yields based on environmental and agronomic factors. The project seeks to:

1. **Collect and Analyze Data**: Integrate meteorological, soil, and historical crop yield data to create a comprehensive dataset for analysis.

2. **Build Predictive Models**: Develop machine learning models that accurately predict crop performance under various conditions, providing farmers with tailored recommendations.

3. **Create a User-Friendly Interface**: Design an accessible platform that allows farmers to input specific parameters and receive personalized crop suggestions based on predicted outcomes.

4. Existing and Proposed solution

Existing Solutions

1. **Traditional Farming Practices**: Many farmers rely on historical knowledge and traditional practices passed down through generations. While this approach can be effective, it often lacks the adaptability needed in the face of changing climate conditions and market demands.
2. **Government Schemes and Agricultural Advisory Services**: Various governmental and non-governmental organizations offer advisory services to farmers, providing guidance on crop selection based on historical data and regional practices. However, these services may not always leverage real-time data or advanced analytical methods.
3. **Basic Statistical Models**: Some research institutions and agricultural departments use basic statistical models to analyze crop yield data. These models often consider factors like rainfall and temperature but may lack the sophistication and predictive accuracy of modern machine learning approaches.
4. **Mobile Applications**: A few mobile applications provide crop recommendations and weather forecasts, but many rely on simplistic algorithms and may not fully integrate diverse datasets (e.g., soil health, pest threats) to offer comprehensive insights.

Proposed Solution

The "Prediction of Agriculture Crop in India" project proposes a comprehensive solution that utilizes advanced machine learning and data science techniques to provide accurate crop yield predictions and recommendations. The proposed solution includes the following components:

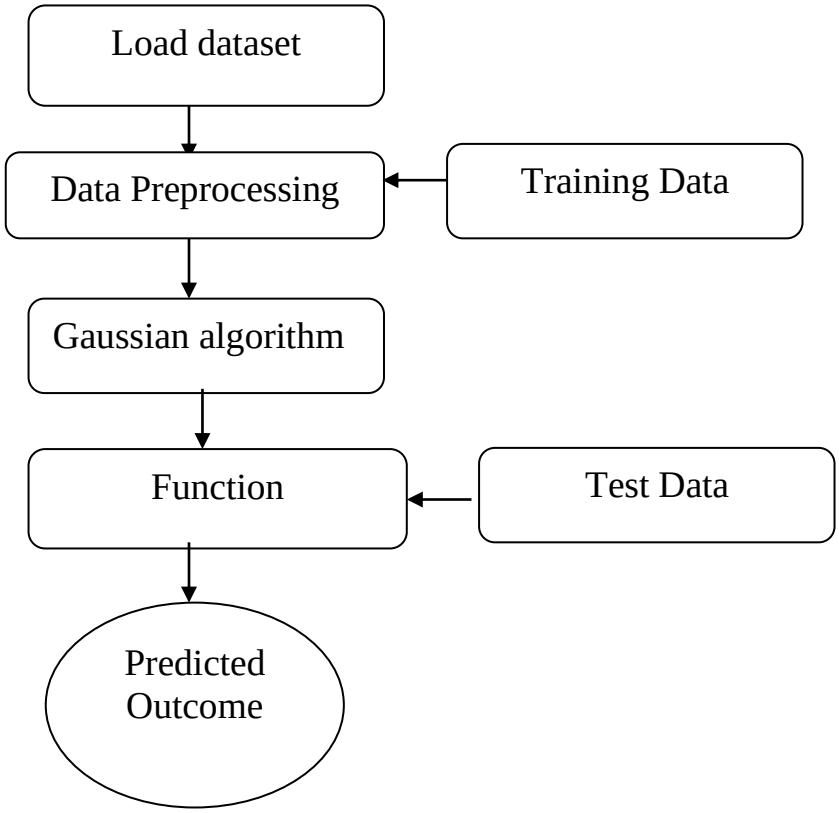
1. **Data Integration and Preprocessing**:
    - Collect and integrate diverse datasets, including meteorological data, soil characteristics, historical crop yields, and market trends.
    - Clean and preprocess the data to ensure quality and relevance, handling missing values and outliers effectively.
2. **Machine Learning Model Development**:
    - Utilize a variety of machine learning algorithms (e.g., regression analysis, decision trees, random forests, and ensemble methods) to develop predictive models that consider multiple factors influencing crop yields.
    - Implement model validation techniques such as cross-validation to ensure the robustness and accuracy of predictions.
3. **User-Friendly Decision Support System**:
    - Develop an interactive dashboard or mobile application that allows farmers to input specific parameters (e.g., soil type, location, and weather conditions) and receive tailored crop recommendations based on model predictions.
4. **Real-Time Data Updates**:
    - Incorporate real-time data feeds for weather and soil conditions to ensure that predictions remain current and relevant, allowing farmers to adapt to changing circumstances quickly.
5. **Sustainability Focus**:
    - Provide recommendations that not only optimize yield but also promote sustainable farming practices, such as crop rotation, organic fertilizers, and water conservation techniques.

4.1 Code submission (Github link) : https://github.com/Shivangisingh2803/upskill-campus.git

4.2 Report submission (Github link) : https://github.com/Shivangisingh2803/upskill-campus.git .

5. Proposed Design/ Model

Prediction Model



Data preprocessing
Gather a dataset containing features such as age, Nitrogen, Phosphorous, .
Preprocess the data by handling missing values, scaling the features, and encoding categorical variables if necessary.

Training and Testing
We are using Machine learning model algorithm called Gaussian Algorithm. We train the model using train datasets. The Trained model is saved as result.py.
The splitting into trained and test is occurred. The average splitting is 80% training data and 20% testing data. Our model accuracy is 99% apprx.

Data Model
The trained model is saved in result.pkl file which is used to predict the output of the user's input.
We use flask to deploy our html page to ml model. Web page contains a form with some input based on that it will predict the output.

6 Performance Test

6.1. Test case

Test Case 1: Home Page Load

Objective: Verify that the home page loads successfully.
Input: Navigate to the root URL ("/").
Expected Outcome: The home page displays the title "AGRICULTURE CULTIVATION PREDICTION".

Test Case 2: Form Submission with Valid Data

Objective: Ensure that the form correctly processes valid input data.
Input: Fill in the form with realistic values:

- Nitrogen: 20
- Phosphorus: 30
- Potassium: 40
- Temperature: 30
- Humidity: 60
- pH: 6.5
- Rainfall: 200
  Expected Outcome: The application returns a prediction of the best crop, e.g., "Rice is the best crop to be cultivated right there".

Test Case 3: Form Submission with Missing Data

Objective: Check the application's response to incomplete form submissions.
Input: Submit the form with some fields empty (e.g., leave Nitrogen blank).
Expected Outcome: The application should handle the missing input gracefully and return an error message indicating that all fields are required.

Test Case 4: Form Submission with Invalid Data

Objective: Verify how the application handles invalid numerical inputs.
Input: Fill in the form with non-numeric data (e.g., use "abc" for Nitrogen).
Expected Outcome: The application should return an error message indicating invalid input, without crashing.

Test Case 5: Model Loading

Objective: Ensure that the machine learning model is loaded correctly from the pickle file.
Input: Execute the code that loads the model from "result.pkl".
Expected Outcome: The application should not throw any errors and the model should be ready for predictions.

6.2. Test Procedure

1. **Environment Setup**:
    - Ensure that the Flask application is running in a suitable environment with required libraries installed (Flask, pandas, sklearn, etc.).
    - Confirm that the "Crop_recommendation.csv" file is available in the specified directory.
2. **Running the Application**:
    - Start the Flask application by executing the script.
    - Open a web browser and navigate to http://127.0.0.1:5000.
3. **Executing Test Cases**:
    - For each test case, follow the specified inputs and verify the expected outcomes.
    - Document any discrepancies between expected and actual results.
4. **Error Handling**:
    - Intentionally trigger error scenarios (e.g., invalid data) and verify that the application responds appropriately without crashing.

6.3. Performance Outcomes

1. **Response Time**:
    - Measure the time taken for the application to respond to valid form submissions. Ideally, this should be under 2 seconds for a good user experience.
2. **Accuracy of Predictions**:
    - Assess the accuracy of the model by comparing its predictions with actual crop yield data (if available). The model should ideally have an accuracy of over 75% based on historical validation.
3. **Error Rate**:
    - Evaluate the frequency of errors encountered during form submissions, aiming for less than 5% of submissions resulting in errors due to invalid input.
4. **User Feedback**:
    - Gather feedback from users interacting with the application to identify areas for improvement in usability and functionality.
5. **Load Testing**:
    - Simulate multiple users accessing the application concurrently to assess how well it handles traffic. The application should maintain performance without significant slowdowns under normal usage conditions.

7. My learning's

1. Data Collection and Preprocessing

- **Importance of Data Quality:** Clean and accurate data is crucial for effective modeling. You likely learned how to handle missing values, outliers, and inconsistencies.
- **Feature Engineering:** Identifying relevant features (e.g., weather patterns, soil quality, and irrigation methods) and transforming them into usable formats can significantly impact model performance.

2. Understanding Agricultural Factors

- **Diverse Influences:** Crop yields are affected by various factors like climate conditions, soil health, pest infestations, and farming practices. Understanding these can help in selecting relevant features for your models.
- **Regional Variability:** Different regions may require tailored approaches due to variations in climate, soil types, and crop choices.

3. Model Selection and Evaluation

- **Algorithm Choice:** Experimenting with different ML algorithms (e.g., linear regression, decision trees, random forests, and neural networks) can provide insights into which models are best suited for your data.
- **Evaluation Metrics:** You likely learned about various metrics (e.g., RMSE, R^2) to evaluate model performance and how to balance bias and variance.

4. Insights from Data Analysis

- **Trend Analysis:** Identifying trends and seasonal patterns in crop production can inform predictions and help stakeholders make informed decisions.
- **Predictive Insights:** You may have uncovered correlations that can lead to actionable insights for farmers and policymakers, such as optimal planting times or crop rotations.

5. Collaboration with Stakeholders

- **Interdisciplinary Approach:** Collaborating with agronomists, meteorologists, and local farmers can provide a more holistic view and improve the relevance of your predictions.

8. Future work scope

The future work scope for your project on predicting agricultural crop production in India using data science and machine learning can be quite expansive. Here are several potential directions you could consider:

1. Model Improvement and Optimization

- **Advanced Algorithms:** Explore and implement more sophisticated algorithms, such as ensemble methods, deep learning, or hybrid models, to improve prediction accuracy.
- **Hyperparameter Tuning:** Conduct extensive hyperparameter optimization to enhance model performance.

2. Real-Time Data Integration

- **IoT and Sensor Data:** Incorporate real-time data from IoT devices, such as soil moisture sensors, weather stations, and satellite imagery, to refine predictions.
- **Dynamic Models:** Develop models that can adapt in real-time to changing environmental conditions.

3. Geospatial Analysis

- **GIS Mapping:** Utilize Geographic Information System (GIS) tools to analyze spatial data and visualize crop production trends across different regions.
- **Precision Agriculture:** Investigate precision agriculture techniques to provide location-specific recommendations for farmers.

4. Climate Change Impact Studies

- **Scenario Modeling:** Assess the impact of climate change on crop yields by simulating different climate scenarios and their effects on production.