Predictive Analytics Lab

Submitted To: Dr. Achala Shakya

Submitted By:

Name	SAP ID	Branch
Siddharth Joshi	500107461	CSE AIML(Hons.)
Rishabh Verma	500108272	CSE AIML(Hons.)

Predictive Analytics Report (2024)		
SNo.	**TOPIC**	**PageNo.**
1.	Abstract	1
2.	Introduction	1
3.	Objective	2
4.	Literature Review	2
5.	Methodology	3
6.	Conclusion	5
7.	References	7

Agricultural Crop Price and Health Detection

Abstract

The agricultural industry faces challenges related to crop price fluctuations and diseases affecting crop health. Predicting crop prices and identifying diseases early can significantly improve farmers' decision-making. This project aims to develop a machine learning-based system that predicts crop prices using historical data and detects crop diseases from images. The system integrates image processing and regression models to provide a dual output—price prediction and disease identification. The proposed approach will help farmers estimate market conditions and evaluate crop health, reducing economic losses and increasing productivity. The project uses both structured data (CSV) for price prediction and unstructured data (images) for disease detection. A convolutional neural network (CNN) is employed for image-based disease classification, while a regression model is used for predicting crop prices.

Introduction

Agriculture is a vital industry for the global economy, especially in developing countries where it serves as the backbone for many livelihoods. However, agricultural productivity is constantly affected by unpredictable crop prices and diseases that can significantly reduce yields. Timely and accurate predictions regarding crop prices and early detection of diseases can empower farmers to make better decisions.

Historically, farmers have relied on their intuition and local market trends to estimate prices, but with advancements in data science and machine learning, it is now possible to predict these prices more accurately by analyzing historical trends. Similarly, crop diseases, traditionally identified through visual inspection, can now be automatically detected through machine learning models that process images of crops.

This project explores the development of a machine learning-based system for agricultural price prediction and disease detection. It aims to use historical crop price data and image-based disease datasets to provide farmers with a comprehensive tool for making informed decisions about crop sales and disease management.

Objectives

The primary objectives of this project are:

1. **Price Prediction:**
    - To predict the future prices of selected crops based on historical data using machine learning models.
2. **Disease Detection:**
    - To classify crops as healthy or diseased by analyzing images of crop leaves or fruits using Convolutional Neural Networks (CNNs).
3. **Integration of Models:**
    - To create a unified system that takes an image of a crop and provides both price predictions and disease diagnosis.

Literature Review

1. Price Prediction in Agriculture

Machine learning models, such as regression models, have been increasingly used in agriculture to predict crop prices. Research shows that regression algorithms such as **Linear Regression**, **Random Forest**, and **XGBoost** have been successfully applied to historical price data to make future predictions. Studies highlight the importance of feature engineering, such as considering factors like **weather conditions, economic indicators, and regional demand** for improving the accuracy of price prediction models (Shah et al., 2019).

Recent advancements in **time-series forecasting models**, such as **Long Short-Term Memory (LSTM)** networks, have demonstrated significant potential for predicting crop prices by learning temporal patterns from historical data (Patel et al., 2021).

2. Disease Detection in Crops

Convolutional Neural Networks (CNNs) are widely used in computer vision tasks, including crop disease detection. Research in this domain has shown that CNN models trained on large datasets of crop images can successfully detect diseases with high accuracy. For instance, models such as **ResNet, InceptionNet, and MobileNet** have been widely employed to identify plant diseases (Ferentinos, 2018).

Studies have shown that **data augmentation** and **transfer learning** can improve CNN models' accuracy when limited labeled data is available (Sun et al., 2020). Transfer learning involves fine-tuning pre-trained models on large datasets (such as ImageNet) for specific tasks like plant disease detection.

Methodology

The methodology of this project involves two key tasks: **price prediction** and **disease detection**. The system integrates these models into a unified framework to provide predictions based on both data types.

1. Dataset Collection

- **Price Prediction Dataset (CSV):** Historical price data for 3-4 crops will be collected, including features such as date, region, crop type, and price. This dataset will be preprocessed to handle missing values, outliers, and normalizations.
- **Disease Detection Dataset (Images):** Image datasets containing healthy and diseased crops will be collected from public sources (such as PlantVillage). The dataset will include images labeled as "Healthy," "Disease A," "Disease B," etc.

2. Data Preprocessing

- **For Price Data (CSV):**
  - Clean the dataset by handling missing values and removing outliers.
  - Normalize or standardize numerical values, such as prices.
  - Feature engineering: Convert the date into separate features like year, month, and day to account for seasonal variations.
  - Split the data into training, validation, and test sets.
- **For Image Data:**
  - Images will be resized to a fixed dimension.
  - Data augmentation techniques such as random cropping, rotation, and flipping will be applied to increase the diversity of training data.
  - Normalize pixel values for better model training.

3. Model Building

- **Price Prediction Model:**
  - Regression algorithms will be used to predict the crop price.
  - **Linear Regression** is used for price prediction.
  - The performance of the models will be evaluated using metrics such as **Mean Squared Error (MSE)** and **R-squared**.
- **Disease Detection Model:**
  - A **Convolutional Neural Network (CNN)** will be built to classify crop images into healthy or diseased categories.
  - Pre-trained CNN models **MobileNet** will be fine-tuned on the dataset to improve accuracy.

4. Model Integration

- A unified system will be developed where the user can input a crop image, and the system will predict both the price and the health status of the crop.
- The models will run in parallel, with one predicting the price (based on crop type) and the other predicting the disease status (based on the image).

5. System Design

Frontend (Streamlit):

1. **Health Prediction**:
    a. Users can upload an image of the crop's leaves, which Streamlit processes to determine if the crop is healthy or unhealthy.
    b. The model for disease detection runs directly within Streamlit, classifying the uploaded image.
2. **Price Prediction**:
    a. Users select the crop type and enter **Min Price** and **Max Price** per quintal for price prediction.
    b. Streamlit takes these inputs and runs the prediction model to output an estimated price range or modal price for the crop.
3. **User Interface**:
    a. A simple, user-friendly layout where:
        i. An **image uploader** component allows users to upload a leaf image.
        ii. **Dropdown** and **input fields** for selecting crop type and setting price ranges.
        iii. **Results Display**: Predicted health status and price are displayed on the same interface.

Model Integration:

- Both health prediction and price prediction models are embedded within the Streamlit app, allowing for direct processing without needing a separate backend.

6. Testing and Validation

- The system will be tested on unseen data to ensure robustness.
- Cross-validation will be used to tune hyperparameters for both price prediction and disease detection models.

7. Deployment

- The final system will be deployed using streamlit.
- A user interface will be created to allow users to interact with the system by uploading images and viewing predictions.

Conclusion

This project proposes a machine learning-based approach to help farmers with two critical tasks: predicting the price of crops and detecting diseases in crops. By using both structured (CSV) and unstructured (image) data, the system aims to provide a comprehensive solution that can improve decision-making in the agricultural sector. The integration of price prediction and disease detection into a single system makes it highly applicable for real-world usage.

References

Ferentinos, K. P. (2018). Deep learning models for plant disease detection and diagnosis. *Computers and Electronics in Agriculture, 145*, 311–318.

Patel, D., Singh, J., & Sharma, M. (2021). Application of LSTM in time-series forecasting for agricultural price prediction. *International Journal of Advanced Research in Artificial Intelligence, 10*(2), 45-54.

Shah, R., Kumar, M., & Desai, A. (2019). Machine learning techniques for crop price prediction using historical data. *Agricultural Informatics Journal, 15*(3), 98–107.

Sun, Y., Qian, J., & Zhu, H. (2020). Enhancing CNN performance in crop disease detection through data augmentation and transfer learning. *Journal of Plant Science and Technology, 18*(1), 62–70.

Xu, L., Zhang, X., & Wang, R. (2022). CNN-based image processing for crop freshness detection. *Journal of Food Quality, 46*, 293–302.