# **README: Hyperspectral Imaging Mycotoxin Prediction**

## **Project Overview**

This project predicts **mycotoxin (vomitoxin) levels in corn** using **hyperspectral imaging data**. Various machine learning models were tested, and **XGBoost** was selected as the best-performing model.

## **1️⃣ Installation Instructions**

### **🔹 Step 1: Clone the Repository**

bashCopy codegit clone https://github.com/KashishK30/ImagoAI\_Hyperspectral\_Prediction.git  
cd ImagoAI\_Hyperspectral\_Prediction

### **🔹 Step 2: Create a Virtual Environment (Optional but Recommended)**

bashCopy codepython -m venv env  
source env/bin/activate # For Mac/Linux  
env\Scripts\activate # For Windows

### **🔹 Step 3: Install Dependencies**

bashCopy codepip install -r requirements.txt

## **2️⃣ Running the Jupyter Notebook**

To explore data preprocessing, dimensionality reduction, and model training:

bashCopy codejupyter notebook

Then open **ImagoAI\_internship\_assignment.ipynb** and run the cells step by step.

## **3️⃣ Running the Streamlit App**

The web app allows users to upload new data and make predictions using the trained XGBoost model.

### **🔹 Run Streamlit**

bashCopy codestreamlit run app.py

📌 **Note:** If running in Google Colab, use **localtunnel** or **ngrok** for external access.

## **4️⃣ Model Training Pipeline**

The pipeline includes:  
✅ **Preprocessing:** SNV Normalization, Savitzky-Golay Smoothing  
✅ **Dimensionality Reduction:** PCA (95% variance), Successive Projections Algorithm (SPA)  
✅ **Model Comparison:** SVM, Random Forest, CNN, XGBoost  
✅ **Final Model:** XGBoost (Best R² = 0.75)

## **5️⃣ Expected Outputs**

📌 **Training & Model Evaluation:**

* The best model’s performances:
* XGBoost → MAE: 2558.92, RMSE: 8382.82, R²: 0.75
* XGBoost (Optuna) → MAE: 1974.12, RMSE: 6791.12, R²: 0.84

📌 **Prediction Output (from app.py):**

* Users can upload spectral data, and the model will predict **mycotoxin levels**.

## **6️⃣ Notes & Future Work**

💡 **Next Steps:**

* Improve feature engineering using domain-specific spectral techniques.
* Test hybrid models (CNN feature extraction + XGBoost prediction).
* Collect more hyperspectral samples or use data augmentation.

📌 **For questions or contributions, open an issue on GitHub!** 🚀