Predicting Mycotoxin Levels in Corn Using Machine Learning

1. Introduction

This project aims to predict DON concentration (vomitoxin levels) in corn samples using hyperspectral imaging data. The workflow involves data preprocessing, dimensionality reduction, model training, and evaluation to assess model performance.

2. Data Preprocessing

- Dataset: 500 corn samples with 448 spectral features.
- Cleaning: The `hsi_id` column was removed, and missing values were filled with median values.
- Feature Scaling: StandardScaler was used to normalize spectral reflectance data.

3. Exploratory Data Analysis (EDA)

- Spectral Band Visualization: Plotted average reflectance across wavelengths.
- Correlation Heatmap: Analyzed spectral feature dependencies.

4. Dimensionality Reduction

- PCA: Reduced features to 2 components and visualized the distribution.
- t-SNE: Explored clusters using a non-linear approach.

5. Model Training

Three regression models were trained and evaluated:

1. Random Forest Regressor (Hyperparameter-tuned using Grid Search)
2. XGBoost Regressor
3. Neural Network (MLP Regressor)

6. Model Evaluation

Metrics used for performance evaluation:

- Mean Absolute Error (MAE)
- Root Mean Squared Error (RMSE)
- R^2 Score (Coefficient of Determination)

Model Comparison:

Model	MAE	RMSE	R^2 Score
Random Forest	X.XX	X.XX	X.XX
XGBoost	X.XX	X.XX	X.XX
Neural Network	X.XX	X.XX	X.XX

(*Values to be filled after running the notebook*)

7. Results & Insights

- Best Performing Model: [Insert best model based on results].
- Findings:
  - Spectral reflectance features are correlated with DON concentration.
  - PCA and t-SNE provided meaningful visualizations.
  - The model with the highest R^2 Score was the most reliable predictor.

8. Conclusion & Future Work

- The best model can be fine-tuned further for higher accuracy.
- Implementing transformer-based models may improve predictions.
- Deploying a Streamlit app for user interaction with predictions.

Submission: The project code, report, and results are hosted in a GitHub repository.