Objective

To develop a robust ML/DL pipeline for predicting DON (vomitoxin) concentration in corn samples using hyperspectral reflectance data.

1. **Data Preprocessing**

- Dataset: 500 corn samples with 448 spectral features and DON concentration as target.

- Preprocessing Steps:

- Missing value inspection (None found).

- Normalization using StandardScaler for spectral features.

- Visualization: Average reflectance plot and heatmap of first 20 samples.

2. **Dimensionality Reduction**

- Technique: Principal Component Analysis (PCA)

- Key Insights:

- PCA reduced features to ~95% variance explained using fewer components.

- 2D PCA scatter plot shows partial separation of samples.

3. **Model Training**

- Dataset split: 80% training / 20% testing.

- Models Used:

- Random Forest

- XGBoost

- MLP Neural Network (with target scaling)

- Optimization Techniques:

- Early stopping for MLP

4. **Model Evaluation**

- Metrics: MAE, RMSE, R^2 Score

- Visualization: Actual vs Predicted scatter plots

Model	MAE	RMSE	R^2 Score
Random Forest	1798.10	3751.89	0.9496
XGBoost	1805.82	4221.16	0.9363
MLP Neural Nerwork	2061.37	3991.98	0.9430

Conclusion

This project demonstrated a robust pipeline to preprocess, visualize, reduce, model, and evaluate hyperspectral data using both traditional ML and feedforward neural models.