Mycotoxin Prediction in Corn

Introduction

Mycotoxin contamination in corn is a major agricultural issue. This project develops a machine learning model to predict deoxynivalenol (DON) levels based on spectral features using hyperspectral imaging data.

Data Preprocessing

Steps Taken:

- Loaded raw hyperspectral data from MLE-Assignment.csv.

- Handled missing values using mean/mode imputation.

- Applied SHAP feature selection to retain the 24 most relevant features.

- Used RobustScaler for numerical stability.

Dimensionality Reduction & Insights

To remove redundant features:

- SHAP analysis retained 24 key spectral bands.

- PCA (optional) captured 98% variance in top 10 components.

Key Findings:

- Features 78, 123, 136 had the highest impact.

- Features 371, 73, 384 were less influential and removed.

Model Selection & Training

Tested models: Random Forest, XGBoost, Stacked Models.

Final Model: Stacked Regression (Random Forest + XGBoost)
- Hyperparameter tuning used Optuna.

- Stacking reduced bias and variance.

Model Evaluation

Results:

- R^2 Score: -0.269 (Needs improvement)

- RMSE: 885.023 (High error)

Mycotoxin Prediction in Corn

Observations:

- Model struggles with small DON levels.

- High DON values (>1000) are underpredicted.

- Feature noise and outliers impact performance.

Suggested Improvements

1. Enhance feature engineering with spectral indices.

2. Try alternative models (LightGBM, CNNs).

3. Apply log transformation for outliers.

4. Increase training data to improve robustness.

Next Steps:

- Improve stacked model performance with ensemble methods.

- Deploy model using Flask/FastAPI & Docker.