CornCount: A Phenologically-Aware Machine Learning Framework for Early-Season Yield Forecasting

Prakash, Anika (School: Eastlake High School)

In 2024, corn losses due to weather and fire events exceeded $3.85 billion in the U.S. alone. Corn is critical for livestock feed, fuel, and trade, but yield variability causes supply instability and price volatility. Traditional forecasting methods and machine learning approaches have limited accuracy earlier in the growing season or during anomaly weather years. This project addressed these limitations by creating a machine learning model that considers the corn plant's phenological stage in parallel with environmental data to capture more biological trends and their relation to the final yield. First, almost one million data points regarding climate, vegetation, soil, and historical yield were compiled, spanning over one thousand counties and 24 years. This data trained a neural network, XGBoost, and LightGBM ensemble model (CornCount) that could predict county-level yield at the V6, V12, R1, and R3 phenological stages. CornCount achieved a mean absolute error (MAE) of 18.43 bushels/acre at the V6 stage, outperforming the traditional Yield Component method weeks before it became available. By the R3 stage, it reached an R^2 of 0.80 and MAE of 13.56 bushels/acre. Notably, the model was robust against climate anomalies, retaining a MAE of 7.90 bushels/acre during the 2023 Illinois flash droughts. Permutation importance analysis confirmed the model's validity, as it showed the model understood how the impact of key stressors changes throughout the season. CornCount's success shows how phenologically-aware machine learning can potentially supplement existing forecasting methods to provide earlier accuracy, aiding financial risk management and food security efforts.

Awards Won: