**AI Enhanced Agricultural Forecasting: Dimensionality Reduction and Similarity Metrics for Predicting Crop Yields in Sparse Data Microclimates**

Leif Huender, Mary Everett

Department of Computer Science, University of Idaho

**Abstract**

This study explores and develops new methodologies for predicting agricultural outcomes, such as crop yields, in microclimates characterized by sparse meteorological data. Specifically, it focuses on reducing the dimensionality in time series data as a preprocessing step to generate simpler and more explainable forecast models. Dimensionality reduction helps in managing large data sets by simplifying the information into more manageable forms without significant loss of information. We explore and utilize various 'similarity' metrics, including Kullback-Leibler Divergence, Euclidean Distance, Manhattan Distance, Cosine Similarity, Pearson Correlation, and Spearman Rank Correlation. These metrics help in identifying patterns and relationships across different microclimate features for both locations and seasons (time and space dimensions). We analyze continuous, temporally aligned data streams from two distinct geographic locations to assess the similarity of various weather features like temperature, humidity, cloud cover and many more. This involves comparing long-term weather patterns to identify common traits that might influence crop yields. Additionally, we examine seasonal blocks of meteorological data across different seasons within the same geographic region and separate geographic regions. Analyzing data in blocks helps in understanding how seasonal variations impact agricultural outcomes. This information is leveraged as input for multiple machine learning techniques, ranging from small classical models to advanced approaches like Riemannian learning and transformer models. These models are adept at handling complex, high-dimensional data and extracting meaningful predictions. The study employs high-dimensional temporal datasets from four geographic regions in New Mexico (Otero, Sierra, Doña Ana, and Chaves) as input data, with pecan crop yields as the ultimate outcome of interest. Although still in the preliminary stages, early models suggest a strong predictive link between similar microclimates and agricultural outcomes like crop yield. Moving forward, we propose future avenues of research to refine these predictive models. We also propose to explore a new Mixture of Experts architecture, which combines insights from various specialized models to provide more precise and localized predictions across different agricultural regions. This research, still in its early stages, holds promise for improving forecasting practices in agriculture.