Analysis of Factors Influencing Paddy Yield
DataSet Summary

STA 4320 - Fall 2025
Gerardo Gutierrez, Blaine Behen, Anh Tran

Our dataset is a Paddy Dataset sourced from UCI's Machine Learning repository: https://archive.ics.uci.edu/dataset/1186/paddy+dataset.

The *Paddy Dataset* contains **2,789 instances** with **45 features**. It includes agronomic and environmental variables related to rice (paddy) cultivation, covering **farm characteristics** (such as land area, soil type, and variety), **management practices** (like seed rate, fertilizer and pesticide applications, and nursery details), and **weather conditions** (including rainfall, temperature, wind speed and direction, and relative humidity across different growth stages).

The **response variable**, **"Paddy yield (in Kg)"**, measures the total yield obtained per plot or farm.

The goal of this analysis is to identify and evaluate the **factors that are most significant in explaining variations in paddy yield**.