**BSSE2301 SOFTWARE ENGINEERING MINI PROJECT 2025**

**Assignment Five(05): (10marks)**

You have been hired as a data scientist by AgriSmart AI, a company specializing in AI-driven precision agriculture. The company collects data from farms using sensors and drones to monitor soil moisture, pH levels, temperature, humidity, and crop conditions to optimize irrigation and fertilization practices. You have received a dataset named climate\_action\_data.csv **containing 913 records** across 10 variables, but it contains several issues including missing values, incorrect data formats, erroneous entries labeled as 'error', and duplicate records. Your task is to load this dataset into a Pandas DataFrame and conduct a thorough data inspection to identify the structure, missing values, and data types. Clean the dataset by removing duplicate records, replacing erroneous values with NaN, and handling missing data appropriately by either filling or dropping values based on logical reasoning. After cleaning, **perform an exploratory data analysis** by generating descriptive statistics, visualizing the distribution of numeric variables using histograms, and creating a correlation heatmap to examine the relationships between soil and environmental variables. From your analysis, identify which variables most influence fertilizer recommendations, determine the crop type with the highest average soil moisture, and suggest appropriate irrigation adjustments for crops experiencing average temperatures above 30°C. Finally, export your clean dataset as cleaned\_precision\_agriculture\_data.csv and document your insights and recommendations based on the analysis. (10 Marks)