Abstract No: 128 ***ICT, Mathematics, and Statistics***

UNDERSTANDING RAINFALL DYNAMICS IN KEGALLE, SRI LANKA: INSIGHTS FROM CIRCULAR AND LINEAR ANALYSIS

H.W.N.H. Dissanayaka*, P. Wijekoon and S.G.J. Senarathne

Department of Statistics and Computer Science, Faculty of Science,
University of Peradeniya, Sri Lanka
**nimsarad@sci.pdn.ac.lk*

Rainfall is a crucial component of the Earth's water cycle, providing essential moisture for life. Traditional linear regression models often struggle to capture the complex interactions between meteorological variables, especially circular variables like wind direction. This study focused on circular regression models to overcome these drawbacks and enhance rainfall forecasts. The dataset obtained from Bogala Graphite Lanka PLC, covered hourly data from January 1, 2022, to January 1, 2023, in Kegalle. It contained observations of rainfall, wind speed, wind direction, and temperature. Summary statistics were calculated for the numerical variables, and various plots, including boxplots, hydro plots, and circular histograms, were used to identify trends. Additional predictor variables were created to avoid predicting rainfall occurrence for the same hour by introducing lagged values for wind speed, wind direction, and temperature. First, several Circular-Linear Logistic Regression Models (CLLRMs) were fitted to the entire dataset, with an 80% training index dividing the data into training and testing sets. Then, the best-fitted model was selected based on the performance statistics. To account for the seasonal variability of Sri Lanka's monsoon patterns, the dataset was divided into subsets corresponding to the first inter-monsoon, southwest monsoon, second inter-monsoon, and northeast monsoon. A Linear-Ridge Logistic Regression Model was employed for the first inter-monsoon dataset to address multicollinearity among the predictor variables, while a standard Linear Logistic Regression Model was constructed for the northeast monsoon dataset, as circular variables were found to be statistically insignificant. For the other two seasons, CLLRMs were fitted. The Spearman's rank correlation coefficient revealed a weak negative correlation between rainfall and temperature, with no significant correlation to wind speed. The seasonal models exhibited high accuracy (80%-90%) and specificity (85%-90%), demonstrating notable sensitivity compared to CLLRM fitted for the entire dataset, with temperature having a greater impact on rainfall than wind speed and direction. Future research may focus on developing CLLRMs and Multinomial Models to understand rainfall patterns more accurately, aiming to predict both rainy and non-rainy conditions and distinguish between light, moderate, and heavy rainfall.

Keywords: Correlation, Multicollinearity, Ridge regression, Sensitivity, Specificity