Rainfall Prediction in Bangladesh

Mohaiminul Islam (1905018)
Tanveer Rahman (1905025)

Problem Definition

- We have various weather related information collected from 35 stations of our country. The information ranges from 1948-2022.
- From this dataset, we need to use various machine learning algorithms to predict the rainfall and analyze the predictions.
- Comparing to other countries, less work is done in our country because of the unavailability of relevant datasets.
- Few previous works :
    - Comparative analysis of different rainfall prediction models: A case study of Aligarh City, India
    - Precipitation Forecasting in Northern Bangladesh Using a Hybrid Machine Learning Model
    - Standardization Of Rainfall Prediction In Bangladesh Using Machine Learning Approach
    - Effectiveness of Ensemble Machine Learning Algorithms in Weather Forecasting of Bangladesh

Dataset Preparation

- 65 years weather dataset from Kaggle was taken as the basis. This dataset ranges from 1948 to 2013.
- The recent data from 2014 to 2022 was taken from Bangladesh Agriculture Research Council.
- Combining this two we get a dataset of shape `(24868, 10)`.
- The 10 features are:
  - Station, Year, Month, Max Temp, Min Temp,
  - Rainfall(mm), Humidity (%), Wind Speed (m/s), Cloud Coverage (Octs)
  - Sunshine (hrs)
- The Station column contains the data of 35 stations in total.

Proposed Solution

We will compare the results between different machine learning models.

Traditional ML Models

- Linear Regression
- Random Forest (Ensemble Learning)

Deep Learning Models

- RNN
- LSTM
- CNN

Performance metrics

Mean Squared Error (MSE)

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

R-squared Test

Proportion of variance in the dependent variable explained by independent variable. Higher R-square means better fit.