DengAI: Predicting Disease Spread

Group 15

140105N - I.D.D.G. de Silva

140120F - I.N. Dharmadasa

140474R - M.K.S. Pramuditha

140605X - W.H.N. Sudharaka

Introduction

- DengAI:Prediction Disease Spread is a competition organized by DrivenData platform.
- Focused on predicting the number of dengue fever cases reported each week in two cities; San Juan and Iquitos.

Methodology

- Time series method
  - Analysis of time series data in the dataset
  - Machine learning models used
    - Random Forest, Gradient Boost, KNN, SVM, Multi Layer Perceptron

Methodology Contd.

- Regression method
  - Feature selection
  - Use of different window sizes
  - Machine learning models used
    - Linear Regression, Random Forest, Gradient Boost, KNN, SVM, Multi Layer Perceptron

Methodology Contd.

- Ensemble method
  - Error propagation of time series method was higher than that of the regression method.
  - Error propagations of Linear Regression, KNN, SVM and MLP were higher than that of Gradient Boost and Random Forest.
  - Ensembling of Random Forest and Gradient Boost

Results and Analysis

Model	Mean Absolute Error (MAE)
Time Series Model	20.9375
Gradient Boosting Regression Model	19.1859
Random Forest Regression Model	20.8669
Ensemble Model	16.8870

Results and Analysis Contd.

- MAE produced by time series method was higher than the regression method
- MAE obtained by the ensemble method was much less than the MAEs produced by random forest and gradient boost models
    - Ensemble model counterbalanced the error propagation of those two models.

Thank You