DENGAI: PREDICTING DISEASE SPREAD

CS4622 - Machine Learning
Group 06

Team Members

- 140062D – G.P.P.D. Bandara
- 140066T – K.P.T.K. Bandara
- 140506U – R.A.T.K. Ranatunga
- 140640A – B.K.D.V. Vimarshana

GitHub Link

- https://github.com/vibs94/ML_Project_DengAI

Introduction

- Research is based on relationship between weather and vegetation in San Juan and Iquitos, and total dengue cases reported in those areas.
- The relationship and interaction between diseases and climatic changes are very complicated. So, modeling of such relationships and interactions has been identified as a difficult problem in many research studies.

Introduction

- More researches show that depending on the weather conditions and climate changes, dengue transmission patterns varied between years.
- The focus of this research study was to explore the different climatic variables and recognize short and long-term predictors and build a model for prediction by ensembling the results from regression methods

Methodology

- Two methods used for dengue prediction; **regression method** and **time series method**
- For regression method, two training models were created for the regression method by changing the window size for smoothing, feature selection and method of imputation of missing values.

Methodology - Regression Model 1

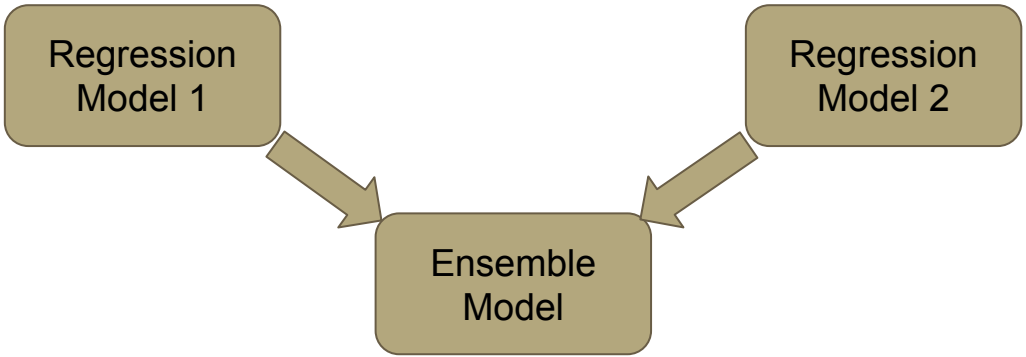
- Regression model 1 is created for the Random forest, Extra trees, and Multilayer Perceptron, and got results separately.
- Final results were calculated by ensembling results

Methodology - Regression Model 2

- Regression model 2 is created for Linear Regression, Random Forest, Gradient Boost, KNN, SVM and MLP with the different preprocessing compared to the first method.
- After cross-validation for all six models, the error propagation of Linear Regression, KNN, SVM and MLP is considerably high.
- Output values of Random Forest and Gradient Boost ensembled.

Methodology - Regression Model

- Final regression model was created by ensembling the results of two regression models

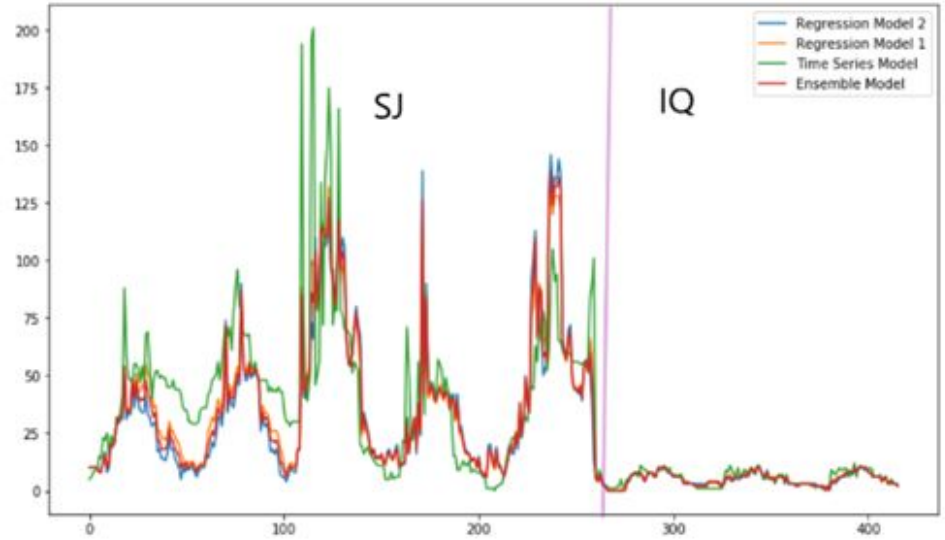


Methodology - Time Series Model

- Monthly Trend was predicted by giving month of the record and total number of cases recorded.
- Residual values for each week was calculated by deducting monthly trend from actual value.
- Residual values and trend was predicted separately by two separate models and predicted values of those two models were added to get the final predicted total cases.

Results

- Following image indicates the predicted values of all four models.



Results

- Following is the accuracies obtained from four model

Model	Tested Accuracy (MAE)
Regression Model 1	21.2188
Regression Model 2	21.0168
Time Series Model	24.8486
Ensemble Model	19.4856

Analysis

- Regression model 1 and regression model 2 have achieved superior performance of 21.2188 and 21.0168 MAE respectively and Ensemble Model got 19.4856
- Regression model 1 and Regression model 2 the data preprocessing has done in different methods.
- If there is some negative impact because of these operations that were neutralized by ensembling these two models.

Analysis

- It can be observed that time series analysis gives better results when there is a dengue outbreak.
- Overall Regression Model has abled to provide better results than Time Series Model.

Conclusion

- Regression method and time series method used for generating the dengue prediction model.
- Two regression models were created changing the methods of the imputation of missing values, smoothing window size and feature selection.

Conclusion

- Accuracy of these two models is low when considered separately. But accuracy will increase more when constructing a new model by ensembling the two models.
- Reason for this is, error from one model is fixed from the second model.

Conclusion

- Another model created using time series analysis.
- By comparing the result models from the regression and time series methods, it is found that the accuracy of time series method is considerably low.
- **Ensembling Regression Model** is used to predict the dengue cases on long-term and short-term climate changes and geographical changes.

Thank You!