A Comparative of Time Series Model of Dengue Hemorrhagic Fever Cases for Monthly Dengue Cases in Chiang Mai Province

Author : Thanakorn Rueansit **Advisors :** Dr.Chalermrat Nontapa and Dr.Parichart Pattarapanitchai
Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand
E-mail : Thanakorn_rue@cmu.ac.th

Abstract

This study aimed to study models for forecasting the number of Dengue Hemorrhagic Fever (DHF) cases in Chiang Mai and compare four forecasting models, namely SARIMA, Holt-Winters, LightGBM, and XGBoost. The dataset is a monthly dataset of the number of Dengue Hemorrhagic Fever cases in Chiang Mai province from January 2011 to May 2024 (161 months). This dataset was divided into two sets: the first training dataset from January 2011 to May 2023 (149 months), to construct the forecasting model; and the second testing dataset from June 2023 to May 2024 (12 months), to compare the accuracy of each model. The performance of each model was compared by the Mean Absolute Percentage Error (MAPE), the coefficient of determination (R-Squared), and the Reduce of Error Rate (RER). The results show that XGBoost is the best model for forecasting the number of Dengue Hemorrhagic Fever, with the lowest MAPE of the training set of 0.078% and MAPE of the test set of 6.809%, R-Squared of 85.90%, and RER of 53.103%.

Introduction

Dengue Hemorrhagic Fever (DHF) is caused by a virus that is transmitted to humans via **female Aedes mosquitoes.** DHF is common in tropical and subtropical locations, largely affecting Asian and Latin American countries. Chiang Mai Province is one of the areas hit by the outbreak. The pattern is every 3 to 5 years during the rainy season, when a major dengue-endemic outbreak develops, with the top 3 months being in the decreasing order of July, June, and August.
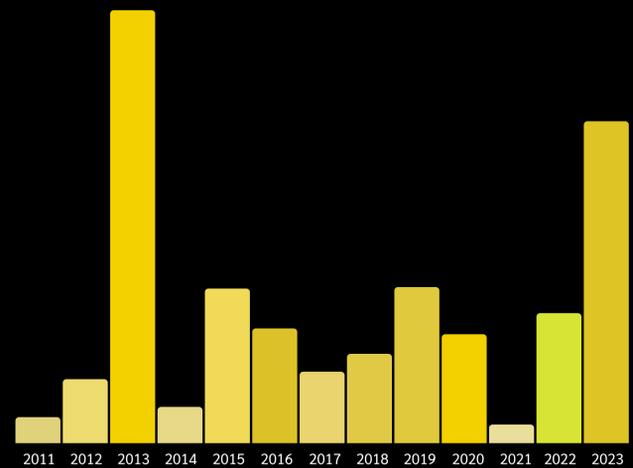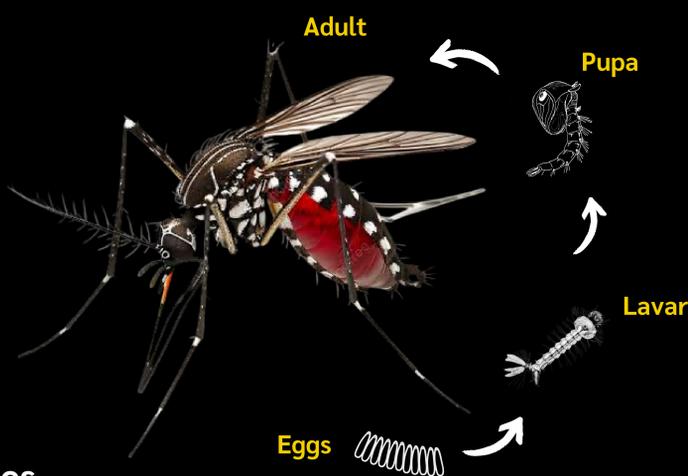
Time Series plot of DHF cases in Chiang Mai from 2011 to 2023

Objectives

1 Study and compare four models of forecasting Dengue Hemorrhagic Fever (DHF) cases in Chiang Mai province. SARIMA, Holt-Winters, LightGBM, and XGBoost.

2 To compare the forecasting performance of SARIMA, Holt-Winters, LightGBM, and XGBoost models in predicting time series data on dengue fever cases in Chiang Mai province.
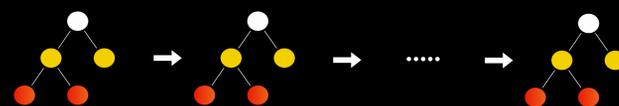
Methodology

Reserch and Collect Data ➤ Exploring Data ➤ Construct Model ➤ Forecasting Value ➤ Summary
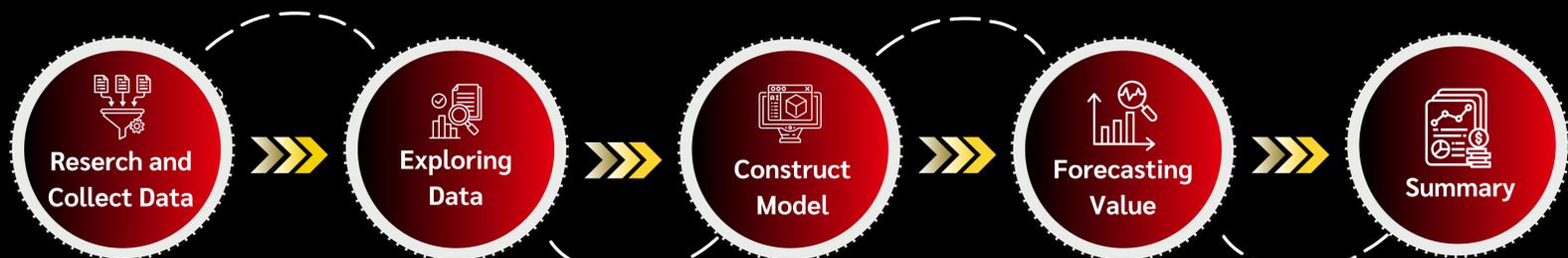
Models

Holt - Winters

$$Y_{t+m} = \left[\beta_{0,t} + m \cdot \beta_{1,t}\right] \times S_{t+m-s} + \varepsilon_t$$
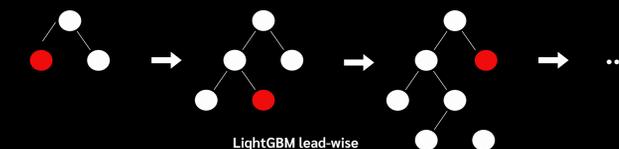
SARIMA

$$\left(1 - \sum_{i=1}^{p}\emptyset_i B^i\right)\left(1 - \sum_{k=1}^{P}\Theta_k B^{ks}\right)Z_t = \left(1 - \sum_{j=1}^{q}\theta_j B^j\right)\left(1 - \sum_{l=1}^{Q}\Theta_l B^{ls}\right)\varepsilon_t$$

where, $Z_t = (1-B)^d(1-B^s)^D Y_t$
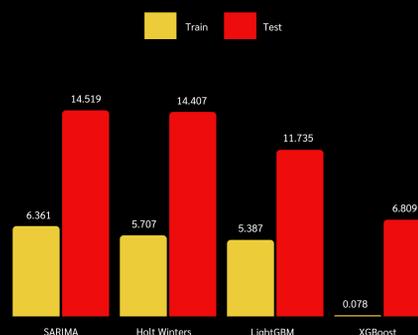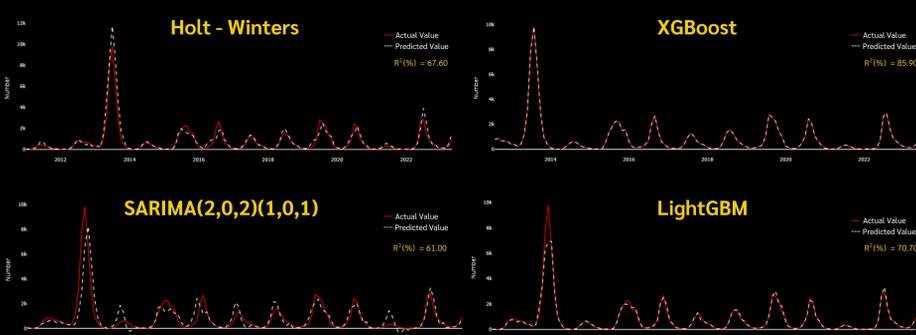
XGBoost

LightGBM

LightGBM lead-wise

Evaluation Criteria

$$MAPE\ (\%) = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Y_t - \hat{Y}_t}{Y_t}\right| \times 100$$

$$R^2(\%) = \left(1 - \frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}\right) \times 100$$

Result and Conclusion

The results show that XGBoost is the best model for forecasting the number of Dengue Hemorrhagic Fever, with the lowest MAPE of the training set of 0.078% and MAPE of the test set of 6.809%, R-Squared of 85.90%, and RER of 53.102%.

Holt - Winters — Actual Value / Predicted Value / $R^2(\%) = 67.60$

XGBoost — Actual Value / Predicted Value / $R^2(\%) = 85.90$

SARIMA(2,0,2)(1,0,1) — Actual Value / Predicted Value / $R^2(\%) = 61.00$

LightGBM — Actual Value / Predicted Value / $R^2(\%) = 70.70$

Train / Test

Model	Train	Test
SARIMA	6.361	14.519
Holt Winters	5.707	14.407
LightGBM	5.387	11.735
XGBoost	0.078	6.809

Discussion

From the study, it was found that the XGBoost model suitable for forecasting Dengue Hemorrhagic Fever (DHF) cases in Chiang Mai province. These findings align with the research of (Na Tian et al. 2024), which stated that XGBoost is an efficient and practical model that offers high performance while requiring relatively short training time.

Suggestion

We can try to use the variety of Deep Neural Network such as Long Short-Term Memory (LSTM), Transformer (TF), Gated Recurrent Units (GRUs) by applying this technique to some other problems and big data sets with various numbers of features or we can add the Exogenous Variables in this model.

References

[1] Tian, N., Zheng, J. X., Li, L. H., Xue, J. B., Xia, S., Lv, S., & Zhou, X. N. (2024). Precision prediction for dengue fever in Singapore: A machine learning approach incorporating meteorological data. Tropical Medicine and Infectious Disease, 9(4), 72.

[2] Sintupachee, S., Ruksachol, O., Pollar, M., & Promprao, S. (2023). Development of the univariate time series model for forecasting Dengue Hemorrhagic Fever cases in Nakhon Si Thammarat. Malaysian Journal of Public Health Medicine, 23(2), 63-72.