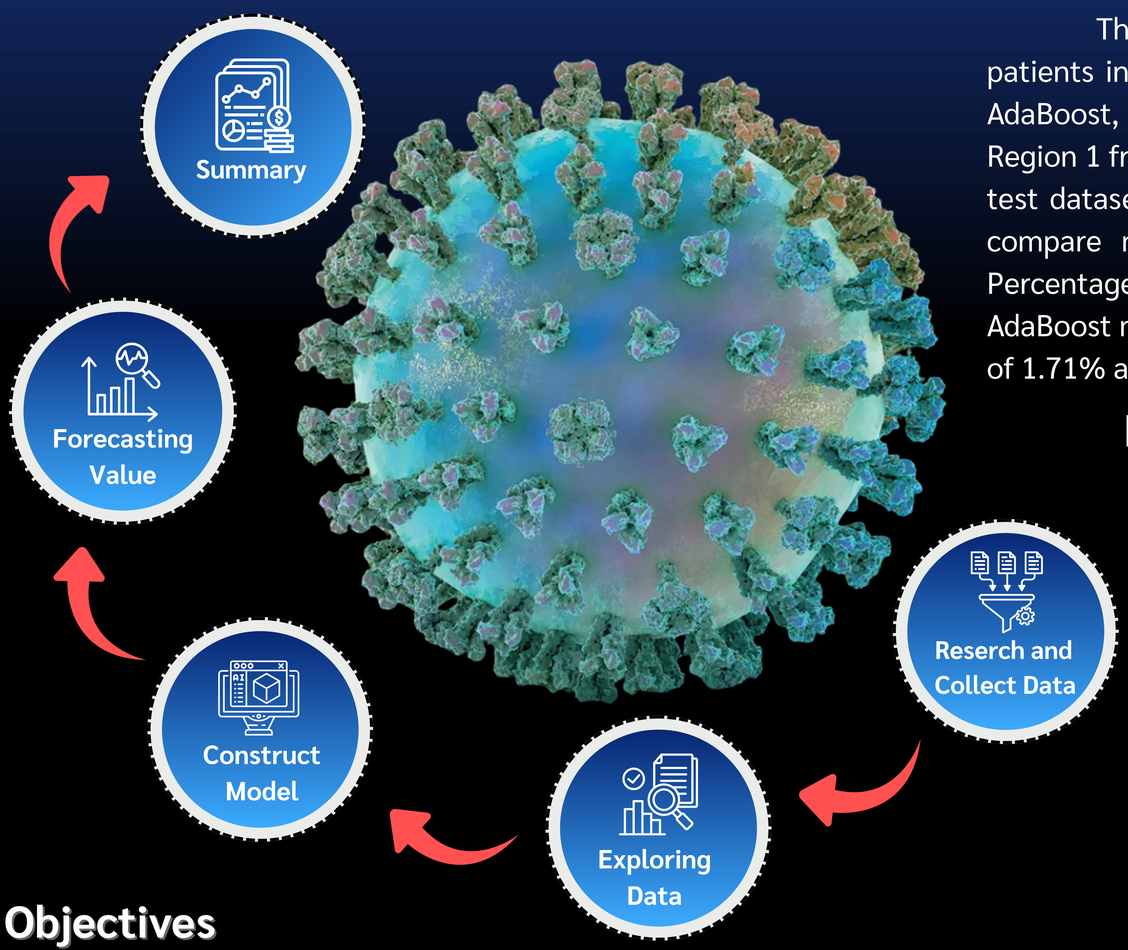
Forecasting the Monthly Cumulative Number of Influenza Patients in Health Region 1

Author: Chutipapha Pattaraprasong **Advisors:** Acting Capt. Dr. Chalermrat Nontapa and Assistant Professor Dr.Parichart Pattarapanitchai

Department of Statistics, Faculty of Science, Chiang Mai University, Chiang Mai, 50200, Thailand

E-mail: chutipapha_p@cmu.ac.th

Methodology

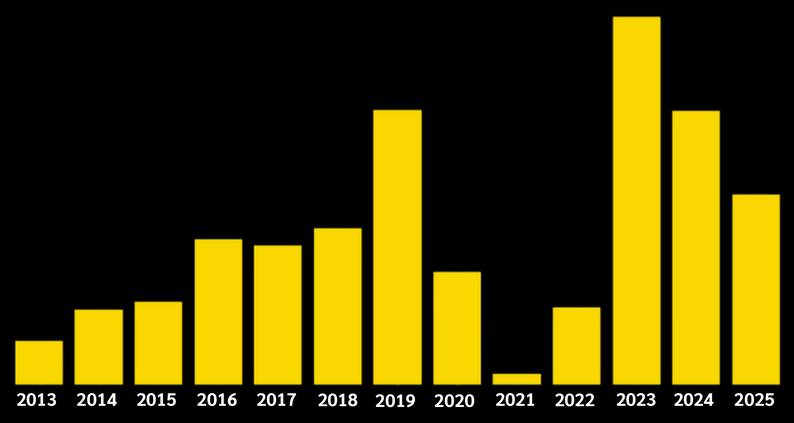


Abstract

This study aimed to develop forecasting models for the monthly cumulative number of influenza patients in Health Region 1 and to compare the performance of four forecasting models: Random Forest, AdaBoost, XGBoost, and LightGBM. Monthly data on the cumulative number of influenza patients in Health Region 1 from January 2013 to March 2024 (135 months) were used to construct the forecasting models. The test dataset consisted of data from April 2024 to March 2025 (12 months) and was used to evaluate and compare model accuracy. The evaluation criteria for forecasting error included the Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R-Squared). The results showed that the AdaBoost model was the most suitable forecasting model, as it achieved the lowest MAPE on the training set of 1.71% and the lowest MAPE on the test set of 4.36%, as well as the highest R-Squared value of 99.98%.

Introduction

Influenza is a rapidly spreading respiratory infection, especially during the rainy and winter seasons. Thailand continues to face this issue, with some years seeing over 100,000 cases and more than ten deaths. Health Region 1 is one of the areas affected by the outbreak, with the highest number of cases occurring during the rainy season, from June to October.



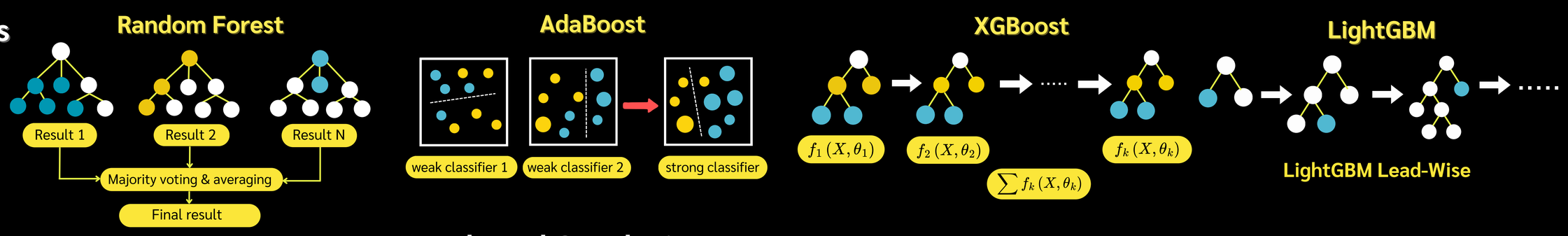
Time Series plot of Influenza cases in Health Region 1 from 2013 to 2025

Objectives

1. To create and compare models for forecasting the monthly cumulative number of influenza patients in Health Region 1, using four forecasting models: Random Forest, AdaBoost, XGBoost, and LightGBM.

2. To select the best model for forecasting the monthly cumulative number of influenza patients in Health Region 1.

Models



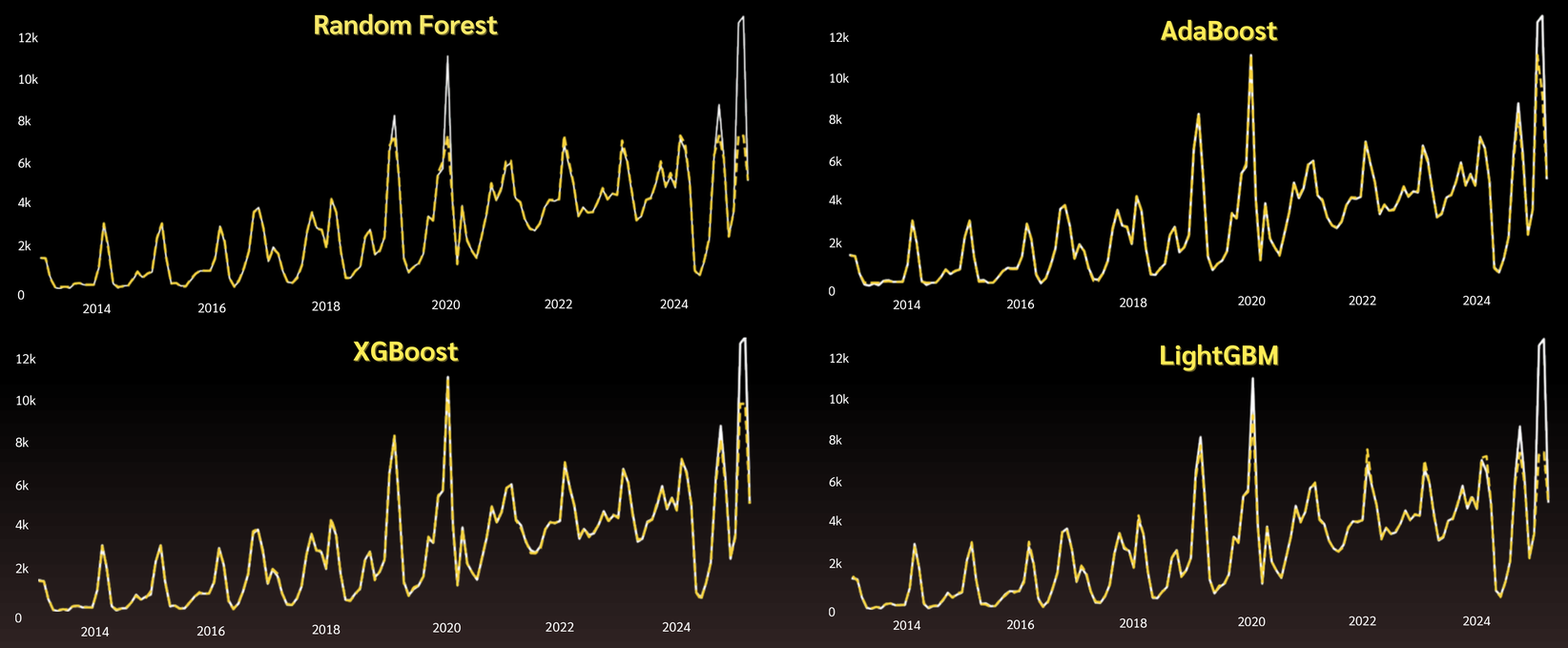
Evaluation Criteria

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

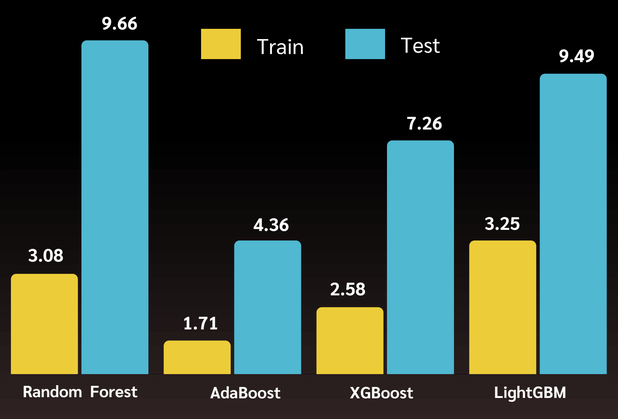
$$R^2\ (\%) = \left(1 - \frac{\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}\right) \times 100$$

Result and Conclusion

The results show that AdaBoost is the best model for forecasting the monthly cumulative number of influenza patients in Health Region 1, with the lowest MAPE in the training set of 1.71% and MAPE in the test set of 4.36%. The R-Squared value is 99.98%.



Discussion



The study found that the AdaBoost model is suitable for forecasting the monthly accumulated influenza cases in Health Region 1. These findings align with the research of Kedir Hussein Abegaz et al. (2023), which stated that the AdaBoost model is efficient and accurate, with the ability to handle data complexity effectively. Additionally, Prophet is suitable for use in research that deals with missing data over the long term. This is consistent with the findings of Blake VanBerlo et al. (2021), which stated that Prophet has a high potential for forecasting even with partial missing data.

We can experiment with various Deep Neural Network techniques, such as Long Short-Term Memory (LSTM) and Transformer (TF), by applying them to problems or large datasets with diverse features. Additionally, we can incorporate Exogenous Variables that impact the data and adjust the data frequency from monthly more records to weekly records in order to enhance forecasting performance.

This research supports Sustainable Development Goal 3 (Good Health and Well-Being) of the United Nations by focusing on health promotion and disease prevention, thereby contributing to the sustainable improvement of people's quality of life.

References

[1] Abegaz, K. H., & Etikan, İ. (2023). Boosting the performance of artificial intelligence-driven models in predicting COVID-19 mortality in Ethiopia. Diagnostics, 13(4), 658.

[2] VanBerlo, B., Ross, M. A., & Hsia, D. (2021). Univariate long-term municipal water demand forecasting. arXiv preprint arXiv:2105.08486.