

การพยากรณ์จำนวนผู้ป่วยสะสมโรคปอดอักเสบรายเดือนโดยใช้การเรียนรู้ของเครื่อง ในเขตสุขภาพที่ 1 FORECAST THE NUMBER OF PNEUMONIA USING MACHINE LEARNING IN HEALTH REGION 1

กฤษณะ ปาอุดร¹, อธิเทพ สุขทรัพย์ศรี^{1,*}, อาชากรินทร์ ปินตา¹ และ เฉลิมรัช นนทะภา¹

Kitsana Paaudon¹, Athitthep Suksuprsri^{1,*}, Archakrin Pinta¹ and Chalermrat Nontapa¹

¹ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

*ผู้นิพนธ์ประสานงาน: อธิเทพ สุขทรัพย์ศรี อีเมล: athitthep_k@cmu.ac.th

บทคัดย่อ:

การศึกษานี้มีวัตถุประสงค์เพื่อสร้างตัวแบบการพยากรณ์จำนวนผู้ป่วยโรคปอดอักเสบรายเดือนในเขตสุขภาพที่ 1 และเปรียบเทียบตัวแบบการพยากรณ์ทั้ง 3 ตัวแบบ คือ Random Forest, AdaBoost และ XGBoost โดยใช้ข้อมูลรายเดือนของจำนวนผู้ป่วยโรคปอดอักเสบรายเดือนในเขตสุขภาพที่ 1 ตั้งแต่เดือน มกราคม พ.ศ. 2556 ถึงเดือน มีนาคม พ.ศ. 2568 รวมทั้งสิ้น 147 เดือน โดยแบ่งข้อมูลออกเป็น 2 ชุด ได้แก่ ข้อมูลชุดฝึกสอนตั้งแต่เดือน มกราคม พ.ศ. 2556 ถึง เดือนมีนาคม พ.ศ. 2567 จำนวน 135 เดือน เพื่อใช้ในการสร้างตัวแบบพยากรณ์ และข้อมูลชุดทดสอบตั้งแต่เดือน เมษายน พ.ศ. 2567 ถึง มีนาคม พ.ศ. 2568 จำนวน 12 เดือน เพื่อเปรียบเทียบค่าความถูกต้องโดยเกณฑ์ที่ใช้เปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ คือ ค่าเฉลี่ยสัมบูรณ์ของเปอร์เซ็นต์ความคลาดเคลื่อน (MAPE) และค่าสัมประสิทธิ์การตัดสินใจ (R-Squared) ผลการศึกษาพบว่า ตัวแบบที่เหมาะสมที่สุด คือ ตัวแบบ XGBoost โดยมีค่า MAPE ของชุดฝึกสอนต่ำที่สุดเท่ากับ 5.65% และ MAPE ของชุดทดสอบต่ำที่สุดเท่ากับ 7.71% มีค่า R-Squared มากที่สุดเท่ากับ 96.08%

คำสำคัญ : โรคปอดอักเสบ, การพยากรณ์, การเรียนรู้ของเครื่อง, อนุกรมเวลา

Abstract:

This study aimed to develop and compare forecasting models for predicting the monthly number of pneumonia cases in Health Region 1. Three machine learning models Random Forest, AdaBoost, and XGBoost were evaluated. The dataset consisted of monthly pneumonia case counts from January 2013 to March 2025, spanning 147 months. The data was split into two sets: a training set from January 2013 to March 2024 (135 months) used to build the forecasting models, and a testing set from April 2024 to March 2025 (12 months) used to evaluate their predictive accuracy. Model performance was assessed using Mean Absolute Percentage Error (MAPE) and the coefficient of determination (R-squared). Results indicated that the XGBoost model outperformed the others, achieving the lowest MAPE values of 5.65% on the training set and 7.71% on the testing set, along with the highest R-squared value of 96.08%.

Keywords : Pneumonia, Forecasting, Machine Learning, Time Series