การพยากรณ์จำนวนผู้ป่วยสะสมโรคไข้หวัดใหญ่รายเดือนในเขตสุขภาพที่ 1 โดยใช้การเรียนรู้ของเครื่อง

FORECASTING THE MONTHLY CUMULATIVE NUMBER OF INFLUENZA PATIENTS IN HEALTH REGION 1 USING MACHINE LEARNING

ชุติปภา ภัทรประสงค์[1*], เฉลิมรัช นนทะภา[1], ปาริชาต ภัทรพานิชชัย[1] และ ชูสกุล พิริยะ[2]

Chutipapha Pattaraprasong[1*], Chalermrat Nontapa[1], Parichart Pattarapanitchai[1] and Choosakun Piriya[2]

[1] *ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่*

[2] *สำนักงานป้องกันควบคุมโรคที่ 1 เชียงใหม่*

* ผู้นิพนธ์ประสานงาน: ชุติปภา ภัทรประสงค์ อีเมล: chutipapha_p@cmu.ac.th

บทคัดย่อ:

การศึกษาในครั้งนี้มีวัตถุประสงค์เพื่อสร้างตัวแบบการพยากรณ์จำนวนผู้ป่วยสะสมโรคไข้หวัดใหญ่รายเดือนในเขตสุขภาพที่ 1 โดยใช้การเรียนรู้ของเครื่อง และเปรียบเทียบตัวแบบการพยากรณ์ทั้ง 4 ตัวแบบ คือ Random Forest, AdaBoost, XGBoost และ LightGBM โดยใช้ข้อมูลรายเดือนของจำนวนผู้ป่วยสะสมโรคไข้หวัดใหญ่รายเดือนในเขตสุขภาพที่ 1 ตั้งแต่เดือนมกราคม พ.ศ. 2556 ถึงเดือนมีนาคม พ.ศ. 2567 จำนวน 135 เดือนเพื่อใช้ในการสร้างตัวแบบพยากรณ์ และข้อมูลชุดทดสอบตั้งแต่เดือนเมษายน พ.ศ. 2567 ถึง มีนาคม พ.ศ. 2568 จำนวน 12 เดือน เพื่อเปรียบเทียบค่าความถูกต้องโดยเกณฑ์ที่ใช้เปรียบเทียบความคลาดเคลื่อนของการพยากรณ์ คือ ค่าเฉลี่ยสัมบูรณ์ของเปอร์เซ็นความคลาดเคลื่อน (MAPE) และค่าสัมประสิทธิ์การตัดสินใจ (R-Squared) ผลการศึกษาพบว่า ตัวแบบที่เหมาะสมที่สุด คือ ตัวแบบ AdaBoost โดยมีค่า MAPE ของชุดฝึกสอนต่ำที่สุด เท่ากับ 1.71% และ MAPE ข้อมูลชุดทดสอบต่ำที่สุด เท่ากับ 4.36% มีค่า R-Squared มากที่สุดเท่ากับ 99.98%

คำสำคัญ : โรคไข้หวัดใหญ่, การพยากรณ์, การเรียนรู้ของเครื่อง, อนุกรมเวลา

Abstract:

This study aimed to develop forecasting models for the monthly cumulative number of influenza patients in Health Region 1 Using Machine Learning 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%.

Keywords: Influenza, Forecasting, Machine Learning, Time Series