

## การทำนายผลลัพธ์ของประกันภัยรถยนต์โดยใช้ตัวแบบการเรียนรู้ของเครื่อง

### PREDICTIVE MODELING FOR CAR INSURANCE OUTCOMES USING MACHINE LEARNING

วชิรพล ไชยชนต์<sup>1\*</sup>, โรจันสุต เครือทอง<sup>1</sup>, ธีรภัทร์ น้อยตัน<sup>1</sup>, เก็จมณี ก้างอนตา<sup>1</sup>, ณัฐนันรี ประดิษฐ์สุวรรณ<sup>1</sup>, พิมพัลลัญช์ จอมมงคล<sup>1</sup> และ เฉลิมรัช นนทะภา<sup>2</sup>

Wachiraphon Phaichayon<sup>1\*</sup>, Rotjanut Kruetong<sup>1</sup>, Theerapat Noiton<sup>1</sup>, Ketmanee Kangonta<sup>1</sup>, Natnaree Pradisuwon<sup>1</sup>, Pimvalan Jommongkol<sup>1</sup> and Chalermrat Nontapa<sup>2</sup>

<sup>1</sup> โครงการจัดตั้งศูนย์วิจัยวิทยาการข้อมูล คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

<sup>2</sup> ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยเชียงใหม่

\* ผู้นิพนธ์ประสานงาน: วชิรพล ไชยชนต์ อีเมล: wachiraphon\_p@cmu.ac.th

#### บทคัดย่อ:

งานวิจัยนี้มีวัตถุประสงค์ในการเปรียบเทียบประสิทธิภาพของตัวแบบที่ใช้ในการพยากรณ์โอกาสที่ผู้ถือกรมธรรม์จะยื่นคำร้องขอสินไหมทดแทน 6 เดือนข้างหน้าหรือไม่ โดยชุดข้อมูลประกอบด้วยประชากร 58,592 ราย และตัวแปรสำคัญ 7 ตัวแปร ซึ่งข้อมูลที่นำมาศึกษาเป็นชุดข้อมูลจาก Kaggle ชุดข้อมูลมีชื่อว่า Car Insurance Claim Prediction การพยากรณ์โอกาสที่ผู้ถือกรมธรรม์จะยื่นคำร้องขอสินไหมทดแทน 6 เดือนข้างหน้าหรือไม่ ใช้ตัวแบบ 4 ตัว ได้แก่ LightGBM, XGBoost, CatBoost และ Random Forest เพื่อเปรียบเทียบประสิทธิภาพการจำแนกประเภทข้อมูลด้วยค่าความถูกต้อง (Accuracy) ค่าประสิทธิภาพโดยรวม (F<sub>1</sub>-Score) ค่าความแม่นยำ (Precision) และค่าความระลึก (Recall) และคัดเลือกตัวแบบที่เหมาะสมที่สุดเพื่อใช้ในการทำนายผลการวิจัยพบว่าในตัวแบบ LightGBM เป็นตัวแบบที่เหมาะสมที่สุดในการทำนายโอกาสที่ผู้ถือกรมธรรม์จะยื่นคำร้องขอสินไหมทดแทน 6 เดือนข้างหน้า โดยมีค่าความถูกต้องร้อยละ 94.51 ประสิทธิภาพโดยรวม ร้อยละ 94.24 ค่าความแม่นยำ ร้อยละ 99.28 และค่าความระลึก ร้อยละ 89.68

**คำสำคัญ:** การเรียนรู้ของเครื่อง, สินไหมทดแทน, รถยนต์

#### Abstract:

This research aims to evaluate and compare the performance of machine learning models in predicting the probability of insurance claim submissions within a six-month horizon. The study utilizes the 'Car Insurance Claim Prediction' dataset from Kaggle, comprising 58,592 records and 7 key predictive variables. To identify the most effective classification approach, four prominent algorithms—LightGBM, XGBoost, CatBoost, and Random Forest—were implemented and benchmarked. The models were rigorously assessed based on four performance metrics: Accuracy, F1-Score, Precision, and Recall. The experimental results indicate that LightGBM emerged as the most optimal model for forecasting claim probabilities. It achieved superior performance with an Accuracy of 94.51%, an F1-Score of 94.24%, a Precision of 99.28%, and a Recall of 89.68%, outperforming the alternative models. These findings provide a robust framework for insurers to enhance their decision-making processes and optimize future operational strategies through high-precision risk assessment.

**Keywords:** Machine Learning, LightGBM, Insurance Claim Prediction, Risk Assessment, Vehicle Attributes, Binary Classification