



การประยุกต์ใช้การเรียนรู้ของเครื่องในการตรวจจับการเรียกร้องค่าสินไหมทดแทน ประกันภัยรถยนต์ที่ไม่มีการออกสำรวจภัย

Applying Machine Learning to Detect Automobile Insurance Claims without Surveys

กิตติศักดิ์ จังพานิช¹ ศุภเบญญา สีวันนา² วรรณพร เชาวน์ชวานิล² กุลภัสสรณ์ ชีวาลักษณ์ลิขิต²
และ ธีระวัฒน์ สิมมาจันทร์^{2*}

¹คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ พื้นที่สาละวิน นครปฐม 73170

² สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ปทุมธานี 12120

Kittisak Jangphanish¹, Supabenya Siwanna², Wannapon Jaova-javanil²,
Kulphassorn Chivalakkhanalikhit² and Teerawat Simmachan^{2*}

¹Faculty of Science and Technology, Rajamangala University of Technology Rattanakosin,
Salaya Campus, Nakhon Pathom 73170

² Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University,
Pathum Thani 12120

*Corresponding author: teerawat@mathstat.sci.tu.ac.th

Received: 25 April 2023/ Revised: 4 June 2023/ Accepted: 7 June 2023

บทคัดย่อ

การประกันภัยรถยนต์มีการฉ้อฉลมากที่สุดเนื่องจากการเรียกร้องค่าสินไหมทดแทนที่ไม่มีการออกสำรวจภัย หรือเรียกว่า “เคลมแห้ง” ถือเป็นการอ้างสิทธิ์ที่จะนำไปสู่การฉ้อฉลและเกิดจากผู้เอาประกันภัยไม่ได้แจ้งรายละเอียดการเกิดเหตุที่ตรงกับความเป็นจริง การตรวจสอบการฉ้อฉลด้วยวิธีการแบบดั้งเดิมนั้นยากต่อการบริหารจัดการ งานวิจัยนี้ใช้การเรียนรู้ของเครื่องในการแก้ปัญหาข้างต้น ตัวแปรตามที่ใช้ในการศึกษาคือการเรียกร้องค่าสินไหมทดแทน (เคลมแห้ง, เคลมสด) วัตถุประสงค์ของการวิจัยประกอบด้วย 1) ค้นหาปัจจัยที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทน 2) เปรียบเทียบประสิทธิภาพการทำนายของวิธีการคัดเลือกตัวแปรอิสระ 3 วิธี คือ ไม่มีการคัดเลือกตัวแปรอิสระ วิธีการถดถอยลอจิสติกที่ละชั้น และวิธีต้นไม้ตัดสินใจ 3) เปรียบเทียบประสิทธิภาพการทำนายของ 3 อัลกอริทึม คือ วิธีนาอูฟเบย์ วิธีป่าสุ่ม และวิธีบูสต์ติงปรับได้ และ 4) เปรียบเทียบประสิทธิภาพการทำนายของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระ ข้อมูลที่ใช้คือรายงานการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ประเภท 1 จากบริษัทประกันภัยแห่งหนึ่งในประเทศไทย ปี 2564 ผลการศึกษาพบว่าลักษณะการเกิดเหตุ ภูมิภาคที่เกิดเหตุ อายุผู้เอาประกันภัย ประเภทการซ่อม ทุนประกันภัย และจำนวนเงินที่จ่ายรวมทั้งเคลม เป็นปัจจัยที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทน การคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุด วิธีป่าสุ่ม วิธีบูสต์ติงปรับได้ และวิธีนาอูฟเบย์เป็น

อัลกอริทึมที่มีประสิทธิภาพสูงสุดตามลำดับ และวิธีป่าสุ่มร่วมกับวิธีการคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุด

คำสำคัญ: วิธีนาอิวเบย์, วิธีบูสต์ติงปรับได้, วิธีป่าสุ่ม, การคัดเลือกตัวแปร

Abstract

Automobile insurance was the most fraudulent because claims without surveys, also known as “dry claims,” led to fraud and arose from the insured failing to provide details of the incident match with reality. Traditional fraud detection methods were difficult to administer. This research utilized machine learning to solve the above problems. The dependent variable used in the study was compensation claims (dry or fresh claims). The research objectives consisted of 1) searching for factors influencing compensation claims; 2) comparing the predictive performance of the three variable selection methods, no selection of independent variables, Stepwise logistic regression, and decision tree; 3) comparing the predictive performance of the three algorithms, namely Naïve Bayes, random forest, and adaptive boosting; and 4) comparing the predictive performance of the algorithms together with variable selection methods. The claim report of the Type I automobile insurance from an anonymous insurance company in Thailand in 2021 was employed in this study. The study found that the cause of the incident, scene region, insured age, type of repair, sum insured, and amount paid influenced the claims. The most effective variable selection method was the decision tree. Random forest, adaptive boosting, and Naïve Bayes was the most efficient algorithm, respectively. The random forest and the independent variable selection method with the decision tree method were the most efficient.

Keywords: Naïve Bayes, Adaptive boosting, Random Forest, Variable selection

บทนำ

อุตสาหกรรมประกันภัยเป็นอุตสาหกรรมที่เติบโตอย่างรวดเร็วและมีการจัดการข้อมูลจำนวนมาก การฉ้อฉลที่เกี่ยวข้องกับการทำประกันภัยเป็นปัญหาหลักในอุตสาหกรรม การฉ้อฉลประกันภัยแบ่งเป็น 2 ประเภท ได้แก่ Soft Insurance Fraud และ Hard Insurance Fraud [1, 2] Soft Insurance Fraud เป็นการฉ้อฉลที่ไม่ได้มีการวางแผนไว้ล่วงหน้า แต่สถานการณ์พาไปให้กระทำ ส่วน Hard Insurance Fraud เป็นการฉ้อฉลที่มีแผนการหรือวิธีการที่อาจจะมีบุคคลภายในบริษัทผู้รับประกันภัยร่วมมือช่วยบุคคลภายนอกให้ได้รับค่าสินไหมทดแทนจากบริษัทประกันภัย การฉ้อฉลรูปแบบนี้เกิดขึ้นได้น้อยกว่าแบบแรกแต่มูลค่าความเสียหายต่อหนึ่งเหตุการณ์มากกว่าหลายเท่าตัว [1] การประกันภัยรถยนต์เป็นหนึ่งในประเภทที่มีการฉ้อฉลมากที่สุด [3, 4] ประมาณร้อยละ 21–36 ของการเคลมประกันรถยนต์ถูกสงสัยว่าเป็นการฉ้อฉล แต่มีเพียงน้อยกว่าร้อยละ 3 เท่านั้นที่ถูกตรวจจับการฉ้อฉลตามกฎหมาย [5] การเรียกร้องค่าสินไหมทดแทนที่ฉ้อฉลจำนวนมากไม่เพียงแต่ส่งผลกระทบต่อบริษัทประกันภัยเท่านั้นแต่ยังรวมถึงผู้เอาประกันภัยที่จริงใจ [6] กล่าวคือผู้เอาประกันภัยที่จริงใจจะได้รับภาระเบี้ยประกันภัยที่สูงขึ้นเนื่องจากธุรกิจประกันภัยเป็นธุรกิจที่มีการระดมทุนจากประชาชนโดยการเก็บเบี้ยประกันภัยจากผู้เอาประกันภัยหลายรายมาไว้เป็นกองทุนและยังส่งผลกระทบต่อการบริหารจัดการและความมั่นคงทางการเงินของบริษัทประกันภัยอีกด้วย หากบริษัทประกันภัยมีระบบตรวจจับการฉ้อฉลที่มีประสิทธิภาพจะช่วยลดความสูญเสียดังกล่าวได้ การตรวจจับการฉ้อฉลประกันภัยโดยทั่วไปจะใช้วิธีการตรวจสอบด้วย



ตนเอง (manual inspection method) ซึ่งต้องอาศัยผู้เชี่ยวชาญหรือวิธีการเฉพาะของแต่ละบริษัท [2] วิธีการแบบดั้งเดิมนี้ยากต่อการบริหารจัดการทั้งด้านเวลาและทรัพยากรบุคคล ดังนั้นการตรวจจับการฉ้อฉลประกันภัยจึงเป็นโจทย์ที่ท้าทาย

การเรียกร้องค่าสินไหมทดแทนที่ไม่มีการออกสำรวจภัยคือ Soft Insurance Fraud ประเภทหนึ่ง โดยพนักงานไม่จำเป็นต้องออกไปตรวจสอบ ณ จุดเกิดเหตุหรือที่เรียกว่า “เคลมแห้ง” ถือเป็นการอ้างสิทธิ์ที่จะนำไปสู่การฉ้อฉลเนื่องจากผู้เอาประกันภัยอาจไม่ได้แจ้งรายละเอียดการเกิดเหตุที่ตรงกับความเป็นจริง เป็นเรื่องราวแรงที่แสดงถึงช่องว่างของระบบประกันภัยที่ทำให้ผู้เอาประกันภัยสามารถกระทำการฉ้อฉลบริษัทประกันภัยได้หรืออาจเป็นช่องทางหนึ่งของอาชญากรรมที่เป็นอันตรายต่อธุรกิจประกันภัย ดังนั้นการสร้างตัวแบบเชิงทำนายเพื่อตรวจจับการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ที่ไม่มีการออกสำรวจภัยจึงเป็นสิ่งที่สำคัญยิ่งต่อธุรกิจประกันภัย

จากการตรวจสอบวรรณกรรมที่เกี่ยวกับการทำนายการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ที่ไม่มีการออกสำรวจภัย พบว่ายังไม่ปรากฏการศึกษาในเรื่องดังกล่าวโดยตรง อย่างไรก็ตามมีการศึกษาเกี่ยวกับการฉ้อฉลประกันภัยจำนวนมาก ตัวอย่างเช่น กฤษณา ศักติวรพงศ์ [7] ได้ศึกษาเชิงเปรียบเทียบของกฎหมายว่าด้วยการฉ้อฉลในการใช้สิทธิเรียกร้องค่าสินไหมทดแทนของประเทศอังกฤษและออสเตรเลียเพื่อหามาตรการทางแพ่งที่เหมาะสมกับการฉ้อฉลการเรียกร้องค่าสินไหมทดแทนตามหลักสุจริตอย่างยิ่ง พรวานภัส จันทรวดี [8] ได้ศึกษาการฉ้อฉลประกันภัย กรณีศึกษาคนเดินเท้าแก่ล้งก่อให้เกิดอุบัติเหตุ เพื่อวิเคราะห์ถึงสภาพปัญหาและหาแนวทางที่เหมาะสมเชิงป้องกันในมุมมองทางด้านนิติศาสตร์และเศรษฐศาสตร์ เมื่อไม่นานนี้มีมีการประยุกต์ใช้การเรียนรู้ของเครื่อง (Machine Learning) ในการตรวจจับการฉ้อฉลประกันภัยอย่างแพร่หลาย เช่น Wang และ Xu [9] ได้ศึกษาการตรวจจับการฉ้อฉลประกันภัยรถยนต์โดยใช้การเรียนรู้เชิงลึก (Deep Learning) พบว่าวิธีการเรียนรู้เชิงลึกมีประสิทธิภาพดีกว่าวิธีการวิเคราะห์การถดถอยลอจิสติก (Logistic regression) Kowshalya และ Nandhini [5] ได้ศึกษาการทำนายการฉ้อฉลการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ พบว่าวิธีป่าสุ่ม (Random Forest) ให้ประสิทธิภาพในการทำนายสูงที่สุด Simmachan และคณะ [10] ได้นำเสนอกัลยสูตรในการทำนายการฉ้อฉลการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ พบว่าการคัดเลือกตัวแปรด้วยวิธีแบบไม่อิงพารามิเตอร์มีประสิทธิภาพในการทำนายดีกว่าวิธีแบบอิงพารามิเตอร์ วิธีป่าสุ่มและวิธีบูสต์ติงปรับได้ (Adaptive Boosting) เป็นอัลกอริทึมที่มีประสิทธิภาพในการทำนายสูง Abdelhadi และคณะ [11] ได้ศึกษาการทำนายการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ พบว่าวิธีเอ็กซ์ตรีมกราดิเอนท์บูสต์ติง (Extreme Gradient Boosting: XGBoost) และวิธีต้นไม้ตัดสินใจ (Decision tree) เป็นอัลกอริทึมที่มีประสิทธิภาพในการทำนายสูง

จากปัญหาดังกล่าวข้างต้น การค้นหาวิธีการตรวจจับการฉ้อฉลประกันภัยที่มีประสิทธิภาพเป็นงานที่ท้าทายและมีความสำคัญยิ่งต่อการดำเนินธุรกิจประกันภัย ผู้วิจัยจึงนำการเรียนรู้ของเครื่องซึ่งเป็นวิธีการหนึ่งที่มีประสิทธิภาพมาใช้ในการทำนายการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ที่ไม่มีการออกสำรวจภัย โดยมีวัตถุประสงค์ ดังนี้ 1) เพื่อหาปัจจัยที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ 2) เพื่อเปรียบเทียบประสิทธิภาพของการคัดเลือกตัวแปรอิสระระหว่างตัวแบบที่ไม่มีการคัดเลือกตัวแปรอิสระ (Full model), วิธีการถดถอยลอจิสติกทีละขั้น (Stepwise logistic regression) และวิธีต้นไม้ตัดสินใจ (Decision tree) ในการทำนายการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ 3) เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนายการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ 3 วิธี ได้แก่ วิธีนาอิวเบย์ (Naïve Bayes), วิธีป่าสุ่ม (Random Forest) และวิธีบูสต์ติงปรับได้ (Adaptive Boosting) จากการศึกษาวรรณกรรมข้างต้นพบว่าวิธีป่าสุ่มและวิธีบูสต์ติงปรับได้เป็นวิธีการที่มีประสิทธิภาพ ส่วนวิธีนาอิวเบย์ถูกเลือกมาศึกษาในฐานะอัลกอริทึมควบคุม (Control algorithm) เนื่องจากเป็นหนึ่งในวิธีการที่เก่าแก่ที่สุด [10] และ 4) เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระในการทำนายการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ การศึกษานี้อาจใช้เป็นแนวทางในการบริหารจัดการความเสี่ยงของบริษัทประกันภัยในการรับมือกับผู้เอาประกันกลุ่มเสี่ยงได้



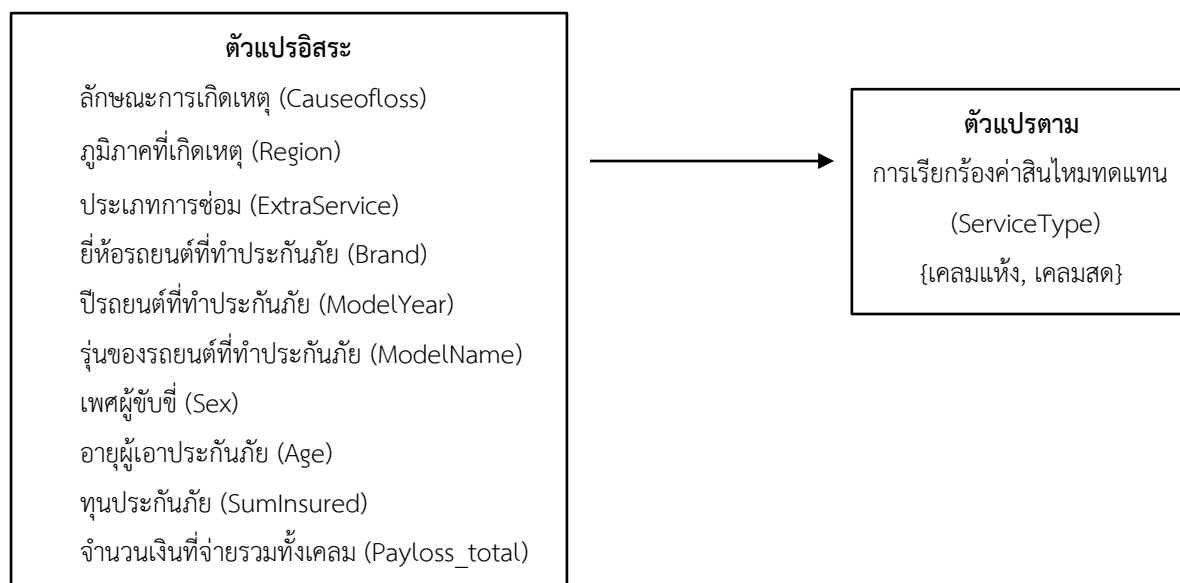
วิธีดำเนินการวิจัย

1. การเตรียมข้อมูลและตัวแปรที่ใช้ในการศึกษา

วัตถุประสงค์หลักของงานวิจัยนี้คือการนำเสนอวิธีการสำหรับการตรวจจับการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ที่ไม่มีการออกสำรวจภัย (การอ้างสิทธิ์ที่จะนำไปสู่การฉ้อฉล) โดยทั่วไปข้อมูลการเคลมของบริษัทประกันภัยเป็นความลับ ในงานวิจัยนี้ใช้ข้อมูลรายงานการรับแจ้งการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ภาคสมัครใจประเภท 1 ของบริษัทประกันภัยแห่งหนึ่งในประเทศไทย ปี 2564 มีการเรียกร้องค่าสินไหมทดแทนทั้งหมด 167,780 กรณี ทำการตรวจสอบข้อมูลสูญหายและการซ้ำซ้อนของข้อมูล พบว่ามีข้อมูลสูญหายจำนวนมาก จากนั้นจึงตัดข้อมูลที่มีปัญหาออกจากการวิเคราะห์และเหลือข้อมูลที่สมบูรณ์สำหรับการวิเคราะห์ 89,965 กรณี ตัวแปรตามที่ใช้ในการศึกษาคือการเรียกร้องค่าสินไหมทดแทน มีค่าเป็นการเรียกร้องค่าสินไหมทดแทนที่ไม่มีการออกสำรวจภัย (เคลมแห้ง) 60,139 กรณี (ร้อยละ 66.85) และการเรียกร้องค่าสินไหมทดแทนที่มีการออกสำรวจภัย (เคลมสด) 29,826 กรณี (ร้อยละ 33.15) สำหรับตัวแปรอิสระมีทั้งหมด 10 ตัวแปร ประกอบด้วยตัวแปรเชิงคุณภาพ 7 ตัวแปร และตัวแปรเชิงปริมาณ 3 ตัวแปร โดยมีรายละเอียดดังตารางที่ 1 ภาพที่ 1 แสดงกรอบแนวคิดการวิจัย ซึ่งเป็นการแสดงความสัมพันธ์ระหว่างตัวแปรอิสระและตัวแปรตาม ก่อนทำการวิเคราะห์ข้อมูลในขั้นตอนต่อไป ตัวแปรเชิงคุณภาพจะถูกแปลงเป็นตัวแปรหุ่น (Dummy variable) และตัวแปรเชิงปริมาณจะถูกแปลงเป็นค่ามาตรฐาน (Z-score Normalization) เนื่องจากอัลกอริทึมการเรียนรู้ของเครื่องส่วนใหญ่ใช้ระยะทางแบบยูคลิด (Euclidean distance) ระหว่างจุดข้อมูลสองจุด ดังนั้นอัลกอริทึมอาจทำงานไม่มีประสิทธิภาพหากไม่มีการแปลงข้อมูล [12] อีกเหตุผลหนึ่งที่ทำให้ต้องมีการแปลงข้อมูลก่อนนำไปสร้างแบบจำลองคืออัลกอริทึมจะทำงานได้เร็วกว่ากรณีไม่มีการแปลงข้อมูล [13]

ตารางที่ 1 ตัวแปรอิสระและคำอธิบาย

ชื่อตัวแปร	ค่าของตัวแปร
ลักษณะการเกิดเหตุ (Cause of loss)	{มีคู่กรณี, ไม่มีคู่กรณี, ภัยจากสิ่งแวดล้อม, เคลมรอบคัน, อื่น ๆ}
ภูมิภาคที่เกิดเหตุ (Region)	{ภาคเหนือ, ภาคกลาง, ภาคตะวันออก, ภาคตะวันออกเฉียงเหนือ, ภาคใต้, ภาคใต้ชายแดน}
ประเภทการซ่อม (Extra Service)	{ซ่อมอู่, ซ่อมข้าง}
ยี่ห้อรถยนต์ที่ทำประกันภัย (Brand)	{BMW, CHEVROLET, FORD, HONDA, ISUZU, MAZDA, MERCEDES BENZ, MG, MITSUBISHI, NISSAN, SUZUKI, TOYOTA, Other Brands}
ปีรถยนต์ที่ทำประกันภัย (Model Year)	{2002-2006, 2007-2011, 2012-2016, 2017-2021, Other Model Years}
รุ่นของรถยนต์ที่ทำประกันภัย (Model Name)	{BMW 320 D, CHEVROLET CAPTIVA, FORD RANGER, HONDA CITY, ISUZU D-MAX, MAZDA 2, MERCEDES-BENZ E200, MG ZS, MITSUBISHI ATTRAGE, NISSAN ALMERA, TOYOTA YARIS, SUZUKI SWIFT, Other Model Names}
เพศผู้ขับขี่ (Sex)	{เพศหญิง, ชาย}
อายุผู้เอาประกันภัย (Age)	{18-95 ปี}
ทุนประกันภัย (Sum Insured)	{50,000- 8,200,000 บาท}
จำนวนเงินที่จ่ายรวมทั้งเคลม (Payloss_total)	{0-3,977,191 บาท}



ภาพที่ 1 กรอบแนวคิดการวิจัย

2. การคัดเลือกตัวแปรอิสระ

วัตถุประสงค์ข้อที่ 2) คือการเปรียบเทียบประสิทธิภาพการทำนายการเรียกร้องค่าสินไหมทดแทนของการคัดเลือกตัวแปรอิสระ 3 วิธี กล่าวคือหลังจากการคัดเลือกตัวแปรจะได้ตัวแปรอิสระมา 3 ชุด และนำไปใช้สร้างตัวแบบเชิงทำนาย เราต้องการทราบว่าตัวแปรอิสระชุดใดให้ประสิทธิภาพสูงสุดและการคัดเลือกตัวแปรอิสระกับการไม่คัดเลือกตัวแปรอิสระให้ผลเป็นอย่างไร การคัดเลือกตัวแปรอิสระจะใช้ข้อมูลทั้งหมดในการวิเคราะห์ข้อมูลโดยใช้โปรแกรม R Studio เวอร์ชัน 1.4.1717 มีรายละเอียดของวิธีการคัดเลือกตัวแปรอิสระ 3 วิธี ดังนี้

2.1 ตัวแบบที่ไม่มีการคัดเลือกตัวแปรอิสระ (Full model)

วิธีการสร้างตัวแบบที่ไม่ใช้การคัดเลือกตัวแปรอิสระ ในที่นี้ใช้ตัวย่อเป็น Full คือการใช้ตัวแปรอิสระทั้งหมดในการสร้างตัวแบบเชิงทำนายโดยไม่สนใจลำดับความสำคัญของตัวแปรแต่ละตัวหรือไม่สนใจความมีนัยสำคัญของตัวแปรแต่ละตัว

2.2 วิธีการถดถอยลอจิสติกทีละขั้น (Stepwise logistic regression)

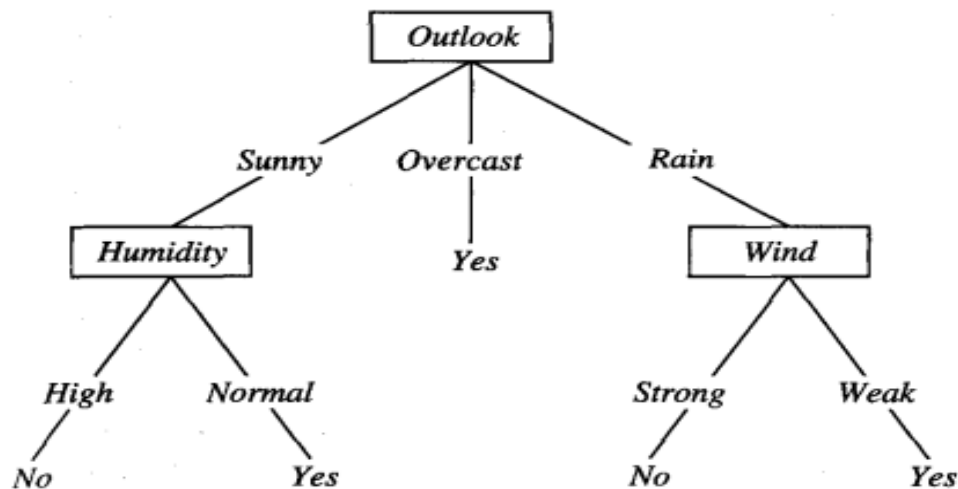
เนื่องจากงานวิจัยนี้ใช้โปรแกรม R Studio ในการวิเคราะห์ข้อมูล เราจึงอธิบายหลักการควบคู่ไปกับการใช้โปรแกรม ดังนี้ ขั้นตอนแรก สร้างตัวแบบการถดถอยลอจิสติกโดยใช้ตัวแปรอิสระทุกตัวและใช้ฟังก์ชัน `glm(..., family = binomial)` ใน `library(MASS)` จากนั้นใช้ฟังก์ชัน `stepAIC(..., direction = "both")` ในการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยลอจิสติกทีละขั้น ในที่นี้ใช้ตัวย่อเป็น Step โดยอาศัยเกณฑ์สารสนเทศของอะกิเกะ (Akaike's Information Criterion: AIC) ในขั้นตอนสุดท้ายจะได้ตัวแปรอิสระที่มีความสำคัญในตัวแบบและจะนำตัวแปรชุดดังกล่าวไปสร้างตัวแบบเชิงทำนายในขั้นตอนต่อไป หลักการทำงานของวิธีการถดถอยลอจิสติกทีละขั้น [6, 10, 14] คือใช้การคัดเลือกตัวแปรแบบไปข้างหน้า (Forward Selection) และการคัดเลือกตัวแปรแบบถดถอยหลัง (Back Elimination) ร่วมกัน โดยเริ่มจากการนำตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามเข้าตัวแบบทีละ

หนึ่งตัวแปร จากนั้นทำการคัดเลือกตัวแปรอิสระที่มีความสัมพันธ์กับตัวแปรตามเข้าตัวแบบเป็นตัวแปรตัวที่สอง แล้วใช้วิธีการคัดเลือกตัวแปรแบบถอยหลังตรวจสอบว่าตัวแปรที่เข้าไปอยู่ในตัวแบบมีตัวแปรตัวใดควรคัดออกจากตัวแบบหรือไม่ การคัดตัวแปรเข้าหรือออกขึ้นอยู่กับค่า AIC ซึ่งตัวแบบในขั้นตอนสุดท้ายจะได้ชุดของตัวแปรอิสระที่ทำให้ได้ค่า AIC ต่ำที่สุด

2.3 วิธีต้นไม้ตัดสินใจ (Decision tree)

วิธีต้นไม้ตัดสินใจ ในที่นี้ใช้ตัวย่อเป็น DT เป็นแบบจำลองทางคณิตศาสตร์ที่ใช้ในการแบ่งกลุ่มปริภูมิของตัวแปรทำนายหรือตัวแปรอิสระออกเป็นอาณาบริเวณเพื่อใช้ในการทำนายค่าของตัวแปรตาม เกณฑ์ในการแบ่งปริภูมิของตัวแปรทำนายสรุปได้ในรูปแบบของต้นไม้ วิธีต้นไม้ตัดสินใจสามารถใช้ได้ทั้งตัวแปรตามเชิงปริมาณซึ่งเรียกว่า “ต้นไม้การถดถอย (Regression tree)” และตัวแปรตามเชิงคุณภาพหรือตัวแปรจำแนกประเภท (Categorical variable) ซึ่งเรียกว่า “ต้นไม้จำแนก (Classification tree)” [15] ในงานวิจัยนี้ใช้วิธีต้นไม้ตัดสินใจในรูปแบบของต้นไม้จำแนก ภาพที่ 2 แสดงตัวอย่างของต้นไม้จำแนกสำหรับการจำแนกสภาพอากาศที่เหมาะสมกับการเล่นเทนนิส

สำหรับต้นไม้จำแนก ความสำคัญของตัวแปร (Variable importance) คือการวัดความสำคัญสัมพัทธ์หรือการมีส่วนร่วมของตัวแปรอิสระแต่ละตัวในการสร้างตัวแบบเชิงทำนายที่แม่นยำ ช่วยให้เราเข้าใจว่าตัวแปรใดมีอิทธิพลมากกว่าในกระบวนการสร้างต้นไม้จำแนก วิธีหนึ่งที่ใช้กันทั่วไปในการคำนวณความสำคัญของตัวแปรคือดัชนีจินี (Gini index) ซึ่งเป็นตัววัดความไม่บริสุทธิ์หรือความผิดปกติภายในโหนด (Node) ของต้นไม้จำแนกและวัดว่าตัวแปรใดแยกข้อมูลออกเป็นหมวดหมู่เป้าหมาย (Target class) ได้ดีเพียงใด [14] ในงานวิจัยนี้ใช้ฟังก์ชัน `fit()` ใน `library(tidymodels)` การสร้างต้นไม้ตัดสินใจ และใช้ฟังก์ชัน `vi()` ใน `library(vip)` เพื่อคำนวณค่า Variable importance ของตัวแปรอิสระ ตัวแปรใดที่ให้ค่า Variable importance สูงจะถูกนำไปสร้างตัวแบบเชิงทำนายในขั้นตอนต่อไป ในที่นี้จะถือว่าตัวแปรมีความสำคัญเมื่อมีค่า Variable importance มากกว่าหรือเท่ากับ 1000



ภาพที่ 2 ตัวอย่างโครงสร้างต้นไม้ตัดสินใจสำหรับการจำแนกสภาพอากาศที่เหมาะสมกับการเล่นเทนนิส [15]



3. อัลกอริทึมที่ใช้ในการสร้างตัวแบบเชิงทำนาย

3.1 วิธีนาอืฟเบส์ (Naïve Bayes)

วิธีนาอืฟเบส์เป็นวิธีการเรียนรู้โดยใช้หลักการของความน่าจะเป็นซึ่งมีพื้นฐานจากทฤษฎีของเบส์ (Bayes theorem) ในที่นี้ใช้ตัวย่อเป็น NB โดยมีจุดมุ่งหมายเพื่อต้องการสร้างตัวแบบหรือวิธีเชิงทำนายที่อยู่ในรูปของความน่าจะเป็น ซึ่งเป็นค่าที่บันทึกได้จากการสังเกต จากนั้นนำตัวแบบมาพิจารณาสมมติฐานใดถูกต้องที่สุดโดยอาศัยความน่าจะเป็นเข้ามาช่วยเมื่อกล่าวถึงความน่าจะเป็นของเหตุการณ์ที่เกิดขึ้น (A) ถ้ามีเหตุการณ์อีกเหตุการณ์หนึ่งเกิดมาแล้ว (B) สามารถเขียนให้อยู่ในรูปความสัมพันธ์ [11, 16] ดังสมการที่ (1)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (1)$$

$P(A|B)$ คือ ความน่าจะเป็นที่เหตุการณ์ A จะเกิดขึ้นเมื่อเหตุการณ์ B เกิดขึ้นมาแล้ว

$P(B|A)$ คือ ความน่าจะเป็นที่เหตุการณ์ B จะเกิดขึ้นเมื่อเหตุการณ์ A เกิดขึ้นมาแล้ว

$P(A)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ A

$P(B)$ คือ ความน่าจะเป็นที่จะเกิดเหตุการณ์ B

การใช้สมการที่ (1) ในการทำนายว่าการเรียกร้องค่าสินไหมทดแทนจะเป็นเคลมแห้งหรือเคลมจะอาศัยความน่าจะเป็นแบบเบย์หรือที่เรียกว่า “ความน่าจะเป็นภายหลัง (posterior class probability)” สามารถเขียนได้ดังนี้

$$P(Y = y_j | X = x) = P(y_j | x) = \prod_{i=1}^n \frac{P(x_i | y_j)P(y_j)}{P(x)} \quad (2)$$

ตัวแปร Y คือ ตัวแปรตามที่มีค่าเป็น y_j ในที่นี้ $j = 1, 2$ เนื่องจากมี 2 หมวดหมู่ (class) นั่นคือ {เคลมแห้ง, เคลมสด}

ตัวแปร $X = \{X_1, X_2, \dots, X_k\}$ คือ เวกเตอร์ของตัวแปรอิสระ k ตัว จากสมการที่ (2) จะสังเกตได้ว่า $P(x)$ เป็นค่าคงที่กับทุก ๆ หมวดหมู่ และไม่กระทบต่อการคำนวณความน่าจะเป็น ดังนั้นในการพิจารณาการจำแนกประเภทข้อมูลจึงพิจารณาแค่เพียงค่า $P(x_i | y_j)P(y_j)$ และได้ความสัมพันธ์ ดังนี้

$$\begin{aligned} P(y_1 | x) &\approx \prod_{i=1}^k P(x_i | y_1)P(y_1) = P(x_1 | y_1) \times P(x_2 | y_1) \times \dots \times P(x_k | y_1)P(y_1) \\ P(y_2 | x) &\approx \prod_{i=1}^k P(x_i | y_2)P(y_2) = P(x_1 | y_2) \times P(x_2 | y_2) \times \dots \times P(x_k | y_2)P(y_2) \end{aligned} \quad (3)$$

$P(y_1)$ คือ ความน่าจะเป็นที่จะเป็นหมวดหมู่ที่ 1 หรือความน่าจะเป็นของเหตุการณ์ที่เป็นเคลมแห้ง

$P(y_2)$ คือ ความน่าจะเป็นที่จะเป็นหมวดหมู่ที่ 2 หรือความน่าจะเป็นของเหตุการณ์ที่เป็นเคลมสด

$P(x_i | y_j)$ คือ ความน่าจะเป็นที่จะเป็นตัวแปร $X = x_i$ เมื่อทราบว่ามีมาจากหมวดหมู่ y_j

$P(y_j | x)$ คือ ความน่าจะเป็นที่จะเป็นหมวดหมู่ y_j เมื่อกำหนดค่าของตัวแปรอิสระ $X = x$

ตัวแปรตามจะมีค่าเป็นหมวดหมู่ใดจะพิจารณาจากค่า $P(y_j | x)$ ของแต่ละหมวดหมู่ หากหมวดหมู่ใดให้ค่าความน่าจะเป็นสูงที่สุด จะทำนายว่าเป็นหมวดหมู่นั้น ดังสมการที่ (4) และใช้ฟังก์ชัน naiveBayes() ในการสร้างตัวแบบเชิงทำนายด้วยวิธี naiveBayes

$$\hat{y} = \max \{P(y_1 | x), P(y_2 | x)\} \quad (4)$$

3.2 วิธีป่าสุ่ม (Random Forest)

วิธีป่าสุ่ม ในที่นี้ใช้ตัวย่อเป็น RF เป็นอัลกอริทึมเชิงทำนายที่เกิดจากการรวมต้นไม้ตัดสินใจหลายๆ ต้น ซึ่งมาจากการสุ่มตัวอย่างแบบคืนที่หลาย ๆ ครั้ง ด้วยเทคนิคที่เรียกว่า Bagging หรือ Bootstrap Aggregation สามารถใช้ได้ทั้งตัวแปรตามเชิงปริมาณและเชิงคุณภาพ โดยผลลัพธ์การทำนายของวิธีป่าสุ่มในงานด้านการจำแนกประเภท (Classification) คือผลการทำนายที่มีจำนวนซ้ำกันมากที่สุด (Majority vote) ซึ่งวิธีการนี้จะทำให้ผลการทำนายมีความแม่นยำมากกว่าการใช้วิธีต้นไม้ตัดสินใจเพียงหนึ่งต้น และใช้ฟังก์ชัน randomForest() ในการสร้างตัวแบบเชิงทำนายด้วยวิธีป่าสุ่ม

3.3 วิธีบูสต์ติงปรับตัว (Adaptive Boosting)

วิธีบูสต์ติงปรับตัว ในที่นี้ใช้ตัวย่อเป็น AB เป็นอัลกอริทึมเชิงทำนายที่สร้างจากต้นไม้ตัดสินใจเป็นพื้นฐาน (Tree base) เช่นเดียวกับวิธีป่าสุ่ม แต่จะต่างกันที่วิธีบูสต์ติงปรับตัวมีแนวคิดในการพัฒนาอัลกอริทึมที่มีประสิทธิภาพต่ำ (Weak classifier) ให้มีประสิทธิภาพสูงขึ้นโดยการปรับค่าน้ำหนักของค่าสังเกตที่ทำนายผิด ใช้อัลกอริทึมที่มีประสิทธิภาพต่ำทำการจำแนกประเภทข้อมูลซ้ำ ๆ ไปเรื่อย ๆ ซึ่งการทำซ้ำแต่ละครั้งนั้นจะใช้เพื่อปรับค่าน้ำหนักของค่าสังเกตที่ทำนายผิด จนกระทั่งได้ตัวแบบที่ไม่มีคลาดเคลื่อน (Error) หรือมีความคลาดเคลื่อนตามที่กำหนด และใช้ฟังก์ชัน ada() ในการสร้างตัวแบบเชิงทำนายด้วยวิธีบูสต์ติงปรับตัว

4. การแบ่งชุดข้อมูลในการประเมินประสิทธิภาพตัวแบบเชิงทำนาย

งานวิจัยนี้ใช้วิธีการตรวจสอบไขว้ (Cross-Validation Test) เพื่อวัดประสิทธิภาพของตัวแบบเชิงทำนาย โดยแบ่งข้อมูลทั้งหมดออกเป็น K ส่วนเท่า ๆ กัน ซึ่งเรียกว่า “K-Fold Cross Validation” ในที่นี้จะใช้ $K = 10$ เนื่องจากตัวอย่างที่ใช้ในการศึกษามีขนาดใหญ่ [6] หากแบ่งข้อมูลเป็น 10 ส่วน (Fold) จะได้ขนาดตัวอย่างของแต่ละส่วนประมาณ 9,000 หลักการทำงานของวิธีนี้จะสุ่มแบ่งข้อมูลทั้งหมดออกเป็น 10 ส่วนเท่า ๆ กัน โดยใช้ข้อมูล 9 ส่วน (ข้อมูลชุดฝึกสอน หรือ Training set) เพื่อสร้างตัวแบบเชิงทำนายจากการประกอบกัน (Combination) ของตัวแปรอิสระทั้ง 3 ชุด และอัลกอริทึมทั้ง 3 และใช้ข้อมูล 1 ส่วน (ชุดทดสอบ หรือ Test set) เพื่อประเมินประสิทธิภาพการทำนาย และดำเนินการในลักษณะแบบนี้จำนวน 10 ครั้ง โดยในแต่ละครั้งจะเปลี่ยนข้อมูลชุดฝึกสอนและข้อมูลชุดทดสอบไปเรื่อย ๆ จนครบ 10 ครั้ง จากนั้นเก็บค่าเกณฑ์วัดประสิทธิภาพของตัวแบบเชิงทำนายแต่ละครั้งไว้แล้วนำมาพิจารณาค่าเฉลี่ยทั้ง 10 ครั้ง

5. เกณฑ์วัดประสิทธิภาพตัวแบบเชิงทำนาย

เกณฑ์วัดประสิทธิภาพการทำนายคำนวณโดยอาศัยเมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) แสดงดังตารางที่ 2 เป็นการเปรียบเทียบผลลัพธ์ของการทำนาย (Predicted) กับสิ่งที่เกิดขึ้นจริง (Actually) สัญลักษณ์ต่าง ๆ ในตารางที่ 2 แบ่งออกเป็น 4 กรณี [2, 6, 10] สามารถอธิบายได้ดังนี้



ตารางที่ 2 เมตริกซ์วัดประสิทธิภาพ (Confusion Matrix) ขนาด 2X2

Predicted	Actually	
	Positive (เคลมแห้ง)	Negative (เคลมสด)
Positive (เคลมแห้ง)	True Positives (TP)	False Positives (FP)
Negative (เคลมสด)	False Negatives (FN)	True Negatives (TN)

True Positive (TP) หมายถึง แบบจำลองสามารถทำนายผลลัพธ์ว่าเป็นเคลมแห้งและตรงกับสิ่งที่เกิดขึ้นจริงคือเป็นเคลมแห้ง

True Negative (TN) หมายถึง แบบจำลองสามารถทำนายผลลัพธ์ว่าเป็นเคลมสดและตรงกับสิ่งที่เกิดขึ้นจริงคือเป็นเคลมสด

False Positive (FP) หมายถึง แบบจำลองสามารถทำนายผลลัพธ์ว่าเป็นเคลมแห้งแต่สิ่งที่เกิดขึ้นจริงคือเป็นเคลมสด

False Negative (FN) หมายถึง แบบจำลองสามารถทำนายผลลัพธ์ว่าเป็นเคลมสดแต่สิ่งที่เกิดขึ้นจริงคือเป็นเคลมแห้ง

งานวิจัยนี้มุ่งเน้นการตรวจจัดการเรียกร้องค่าสินไหมทดแทนที่ไม่มีการออกสำรวจภัย (เคลมแห้ง) ซึ่งงานวิจัยส่วนใหญ่ที่ศึกษาเกี่ยวกับการตรวจจัดการข้อผิดพลาดจะให้ความสนใจในหมวดหมู่ที่เป็นการผิดพลาดมากกว่าหมวดหมู่ที่ไม่ใช่การผิดพลาด [6, 10, 17] ดังนั้นเกณฑ์วัดประสิทธิภาพการทำนายในงานวิจัยนี้จะให้ความสนใจกับความสามารถในการทำนายหมวดหมู่ที่เป็นเคลมแห้ง การอธิบายความหมายของแต่ละเกณฑ์สามารถอธิบายได้ด้วยหลักการคำนวณความน่าจะเป็น ซึ่งจะมีค่าอยู่ในช่วง $[0, 1]$ ตัวแบบเชิงทำนายใดที่ให้ค่าเกณฑ์วัดประสิทธิภาพสูง (ค่าเข้าใกล้ 1) หมายถึงตัวแบบที่มีความน่าจะเป็นในการทำนายถูกต้องสูง โดยมีรายละเอียดดังนี้

5.1 ค่าความแม่นยำ (Accuracy) เป็นการวัดความถูกต้องของตัวแบบเชิงทำนายในภาพรวม สามารถคำนวณได้ดังสมการที่ (5)

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \quad (5)$$

5.2 ค่าความไว (Sensitivity) หรือเรียกอีกชื่อหนึ่งว่าค่าความระลึก (Recall) เป็นการวัดความถูกต้องของตัวแบบเชิงทำนายในสถานการณ์ที่ทำนายถูกว่าเป็นเคลมแห้งและตรงกับสิ่งที่เกิดขึ้นจริงคือเป็นเคลมแห้ง สามารถคำนวณได้ดังสมการที่ (6)

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad (6)$$

5.3 ค่าความเที่ยง (Precision) เป็นการวัดความถูกต้องของตัวแบบเชิงทำนายในสถานการณ์ที่ทำนายว่าเป็นเคลมแห้งและสนใจเหตุการณ์ที่ทำนายถูกว่าเป็นเคลมแห้ง สามารถคำนวณได้ดังสมการที่ (7)

$$\text{Precision} = \frac{TP}{TP + FP} \quad (7)$$



5.4 ค่าคะแนนเอฟ (F-score) เป็นค่าเฉลี่ยฮาร์โมนิก (Harmonic mean) ของค่าความเที่ยงและค่าความไว สามารถคำนวณได้ดังสมการที่ (8) ดังนั้นในงานวิจัยนี้จะให้ความสำคัญกับค่าคะแนนเอฟมากที่สุดเนื่องจากการถ่วงดุลของสองเกณฑ์คือค่าความเที่ยงและค่าความไว

$$F\text{-score} = \frac{2TP}{2TP + FP + FN} \quad (8)$$

ผลการวิจัย

1. ปัจจัยที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์

จากตารางที่ 3 พบว่าตัวแปรอิสระที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์โดยใช้วิธีการถดถอยลอจิสติกทีละขั้น ได้แก่ ลักษณะการเกิดเหตุ (ประเภทอื่น ๆ, ภัยจากสิ่งแวดล้อม, มีคู่กรณี และไม่มีคู่กรณี) ภูมิภาคที่เกิดเหตุ (ภาคตะวันออก, ภาคตะวันออกเฉียงเหนือ, ภาคใต้, ภาคใต้ชายแดน และภาคเหนือ) ประเภทการซ่อม ทุนประกันภัยและจำนวนเงินที่จ่ายรวมทั้งหมด เนื่องจากในการวิเคราะห์การถดถอยลอจิสติกจะต้องแปลงตัวแปรอิสระเชิงคุณภาพให้เป็นตัวแปรหุ่น ส่งผลให้มีจำนวนตัวแปรอิสระเพิ่มขึ้นเท่ากับ จำนวนหมวดหมู่ - 1 ตามหลักของการสร้างตัวแปรหุ่น เช่น ตัวแปรลักษณะการเกิดเหตุเดิมมี 5 หมวดหมู่ ดังนั้นจะมีตัวแปรหุ่นทั้งหมด 4 ตัวแปร ดังตารางที่ 3 ตัวแปรอื่น ๆ จะถูกสร้างด้วยวิธีการเดียวกัน

ตารางที่ 3 ผลการคัดเลือกตัวแปรอิสระโดยใช้วิธีการถดถอยลอจิสติกทีละขั้น

ชื่อตัวแปร	คำอธิบาย	ค่าพี (P-value)
Constant	ค่าคงที่	$2.00 \times 10^{-16} ***$
Causeofloss_อื่น ๆ	ลักษณะการเกิดเหตุประเภทอื่น ๆ	$8.49 \times 10^{-5} ***$
Causeofloss_ภัยจากสิ่งแวดล้อม	ลักษณะการเกิดเหตุประเภทภัยจากสิ่งแวดล้อม	$2.00 \times 10^{-16} ***$
Causeofloss_มีคู่กรณี	ลักษณะการเกิดเหตุประเภทมีคู่กรณี	$2.00 \times 10^{-16} ***$
Causeofloss_ไม่มีคู่กรณี	ลักษณะการเกิดเหตุประเภทไม่มีคู่กรณี	$6.51 \times 10^{-14} ***$
Region_East	ภูมิภาคที่เกิดเหตุเป็นภาคตะวันออก	$1.60 \times 10^{-11} ***$
Region_NorthEast	ภูมิภาคที่เกิดเหตุเป็นภาคตะวันออกเฉียงเหนือ	$2.00 \times 10^{-16} ***$
Region_South	ภูมิภาคที่เกิดเหตุเป็นภาคใต้	$1.19 \times 10^{-7} ***$
Region_Southborder	ภูมิภาคที่เกิดเหตุเป็นภาคใต้ชายแดน	$1.68 \times 10^{-3} **$
Region_North	ภูมิภาคที่เกิดเหตุเป็นภาคเหนือ	$2.00 \times 10^{-16} ***$
ExtraService_Y	ประเภทการซ่อม	$2.39 \times 10^{-6} ***$
SumInsured	ทุนประกันภัย	$2.68 \times 10^{-6} ***$
PayLoss_total	จำนวนเงินที่จ่ายรวมทั้งหมด	$2.00 \times 10^{-16} ***$

หมายเหตุ * ตัวแปรอิสระที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.05, ** ตัวแปรอิสระที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.01 และ *** ตัวแปรอิสระที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทน ที่ระดับนัยสำคัญ 0.001



ตารางที่ 4 ผลการคัดเลือกตัวแปรอิสระโดยวิธีต้นไม้ตัดสินใจ

ชื่อตัวแปร	คำอธิบาย	Variable Importance
Causeofloss	ลักษณะการเกิดเหตุ	18,158
PayLoss_total	จำนวนเงินที่จ่ายรวมทั้งหมด	1,256
Age	อายุผู้เอาประกันภัย	1,084

จากตารางที่ 4 พบว่าตัวแปรอิสระที่มีความสำคัญต่อการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์โดยใช้วิธีต้นไม้ตัดสินใจ เรียงลำดับตามความสำคัญ ได้แก่ ลักษณะการเกิดเหตุ จำนวนเงินที่จ่ายรวมทั้งหมด และอายุผู้เอาประกันภัย ซึ่งมีค่าความสำคัญของตัวแปรเท่ากับ 18,158, 1,256 และ 1,084 ตามลำดับ สำหรับตัวแปรอื่น ๆ ให้ค่าความสำคัญของตัวแปรเช่นเดียวกัน แต่ไม่ผ่านเงื่อนไขที่ตั้งไว้ (มีค่ามากกว่าหรือเท่ากับ 1,000) เนื่องจากสำหรับข้อมูลชุดนี้ โปรแกรม R Studio จะแสดงแผนภูมิแท่งของค่าความสำคัญของตัวแปรที่มีค่ามากกว่าหรือเท่ากับ 1000 เท่านั้น จึงใช้เป็นเกณฑ์ในการพิจารณา

2. การเปรียบเทียบประสิทธิภาพของการคัดเลือกตัวแปรอิสระ

จากตารางที่ 5 เมื่อพิจารณาแยกแต่ละอัลกอริทึมและแต่ละเกณฑ์ พบว่าสำหรับวิธีนาอิวเบส วิธีการคัดเลือกตัวแปรอิสระที่มีประสิทธิภาพสูงสุดคือวิธีต้นไม้ตัดสินใจในเกณฑ์ ค่าความแม่นยำ (ร้อยละ 86.78) ค่าความไว (ร้อยละ 94.55) และค่าคะแนนเอฟ (ร้อยละ 89.59) ในขณะที่วิธีการถดถอยลอจิสติกทีละขั้นมีประสิทธิภาพสูงสุดสำหรับเกณฑ์ค่าความเที่ยง (ร้อยละ 85.63) สำหรับวิธีป่าสุ่ม วิธีการคัดเลือกตัวแปรอิสระที่มีประสิทธิภาพสูงสุดคือวิธีต้นไม้ตัดสินใจในทุกเกณฑ์ ได้แก่ ค่าความแม่นยำ (ร้อยละ 86.85) ค่าความไว (ร้อยละ 94.51) ค่าความเที่ยง (ร้อยละ 85.28) และค่าคะแนนเอฟ (ร้อยละ 89.66) สำหรับวิธีบูสต์ติ่งปรับได้ วิธีการคัดเลือกตัวแปรอิสระที่มีประสิทธิภาพสูงสุดคือวิธีต้นไม้ตัดสินใจในทุกเกณฑ์ ได้แก่ ค่าความแม่นยำ (ร้อยละ 86.82) ค่าความไว (ร้อยละ 94.35) ค่าความเที่ยง (ร้อยละ 85.40) และค่าคะแนนเอฟ (ร้อยละ 89.65)

ตารางที่ 5 ผลการเปรียบเทียบประสิทธิภาพของการคัดเลือกตัวแปรอิสระ

อัลกอริทึม	การคัดเลือกตัวแปร	ค่าความแม่นยำ	ค่าความไว	ค่าความเที่ยง	ค่าคะแนนเอฟ
NB	Full	0.8540	0.9219	0.8540	0.8867
	Step	0.8603	0.9292	0.8563	0.8612
	DT	0.8678	0.9455	0.8512	0.8959
RF	Full	0.8595	0.9329	0.8502	0.8899
	Step	0.8659	0.9437	0.8501	0.8944
	DT	0.8685	0.9451	0.8528	0.8966
AB	Full	0.8603	0.9383	0.8515	0.8902
	Step	0.8649	0.9411	0.8512	0.8939
	DT	0.8682	0.9435	0.8540	0.8965

หมายเหตุ ตัวหนา หมายถึง การคัดเลือกตัวแปรอิสระที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาแยกแต่ละอัลกอริทึมและแต่ละเกณฑ์



3. การเปรียบเทียบประสิทธิภาพของอัลกอริทึม

จากตารางที่ 6 พบว่าอัลกอริทึมมีประสิทธิภาพแตกต่างกันไปแต่ละสถานการณ์ เมื่อพิจารณาแยกแต่ละวิธีการคัดเลือกตัวแปรอิสระและแต่ละเกณฑ์ พบว่าสำหรับตัวแบบที่ไม่มีการคัดเลือกตัวแปร อัลกอริทึมที่มีประสิทธิภาพสูงสุดคือวิธีวิบูล์สตริงปรับได้ในเกณฑ์ ค่าความแม่นยำ (ร้อยละ 86.03) ค่าความไว (ร้อยละ 93.83) และค่าคะแนนเอฟ (ร้อยละ 89.02) ในขณะที่วิธีวินาอีฟเป็นอัลกอริทึมที่มีประสิทธิภาพสูงสุดสำหรับเกณฑ์ค่าความเที่ยง (ร้อยละ 85.40) สำหรับตัวแบบที่คัดเลือกตัวแปรด้วยวิธีการถดถอยลอจิสติกที่ละชั้น อัลกอริทึมที่มีประสิทธิภาพสูงสุดคือวิธีวิบูล์สตริงปรับได้ในเกณฑ์ ค่าความแม่นยำ (ร้อยละ 86.59) ค่าความไว (ร้อยละ 94.37) และค่าคะแนนเอฟ (ร้อยละ 89.44) ในขณะที่วิธีวินาอีฟเป็นอัลกอริทึมที่มีประสิทธิภาพสูงสุดสำหรับเกณฑ์ค่าความเที่ยง (ร้อยละ 85.63) สำหรับตัวแบบที่คัดเลือกตัวแปรด้วยวิธีต้นไม้ตัดสินใจ อัลกอริทึมที่มีประสิทธิภาพสูงสุดคือวิธีวิบูล์สตริงปรับได้ในเกณฑ์ ค่าความแม่นยำ (ร้อยละ 86.85) และค่าคะแนนเอฟ (ร้อยละ 89.66) ในขณะที่วิธีวินาอีฟเป็นอัลกอริทึมที่มีประสิทธิภาพสูงสุดสำหรับเกณฑ์ค่าความไว (ร้อยละ 94.55) และวิธีวิบูล์สตริงปรับได้เป็นอัลกอริทึมที่มีประสิทธิภาพสูงสุดสำหรับเกณฑ์ค่าความเที่ยง (ร้อยละ 85.40) หากพิจารณาเฉพาะเกณฑ์ค่าคะแนนเอฟ พบว่าวิธีวิบูล์สตริงปรับได้มีประสิทธิภาพสูงสุดสำหรับตัวแบบที่ไม่มีการคัดเลือกตัวแปร ในขณะที่วิธีวิบูล์สตริงปรับได้มีประสิทธิภาพสูงสุดสำหรับตัวแบบที่มีการคัดเลือกตัวแปรอิสระทั้ง 2 วิธี

ตารางที่ 6 ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึม

การคัดเลือกตัวแปร	อัลกอริทึม	ค่าความแม่นยำ	ค่าความไว	ค่าความเที่ยง	ค่าคะแนนเอฟ
Full	NB	0.8540	0.9219	0.8540	0.8867
	RF	0.8595	0.9329	0.8502	0.8899
	AB	0.8603	0.9383	0.8515	0.8902
Step	NB	0.8603	0.9292	0.8563	0.8612
	RF	0.8659	0.9437	0.8501	0.8944
	AB	0.8649	0.9411	0.8512	0.8939
DT	NB	0.8678	0.9455	0.8512	0.8959
	RF	0.8685	0.9451	0.8528	0.8966
	AB	0.8682	0.9435	0.8540	0.8965

หมายเหตุ ตัวหนา หมายถึง อัลกอริทึมที่มีประสิทธิภาพสูงสุดเมื่อพิจารณาแยกแต่ละวิธีการคัดเลือกตัวแปรอิสระและแต่ละเกณฑ์

4. การเปรียบเทียบประสิทธิภาพของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระ

ตารางที่ 7 แสดงผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระ โดย NB-Full หมายถึงวิธีวินาอีฟร่วมกับตัวแปรอิสระทุกตัว, RF-Full หมายถึงวิธีวิบูล์สตริงร่วมกับตัวแปรอิสระทุกตัว, AB-Full หมายถึงวิธีวิบูล์สตริงปรับได้ร่วมกับตัวแปรอิสระทุกตัว, NB-Step หมายถึงวิธีวินาอีฟกับการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยลอจิสติกที่ละชั้น, RF-Step หมายถึงวิธีวิบูล์สตริงปรับได้กับการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยลอจิสติกที่ละชั้น, AB-Step หมายถึงวิธีวิบูล์สตริงปรับได้ร่วมกับการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยลอจิสติกที่ละชั้น, NB-DT หมายถึงวิธีวินาอีฟกับการคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจ, RF-DT หมายถึงวิธีวิบูล์สตริงปรับได้ร่วมกับการคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจ และ



ตารางที่ 7 ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระ

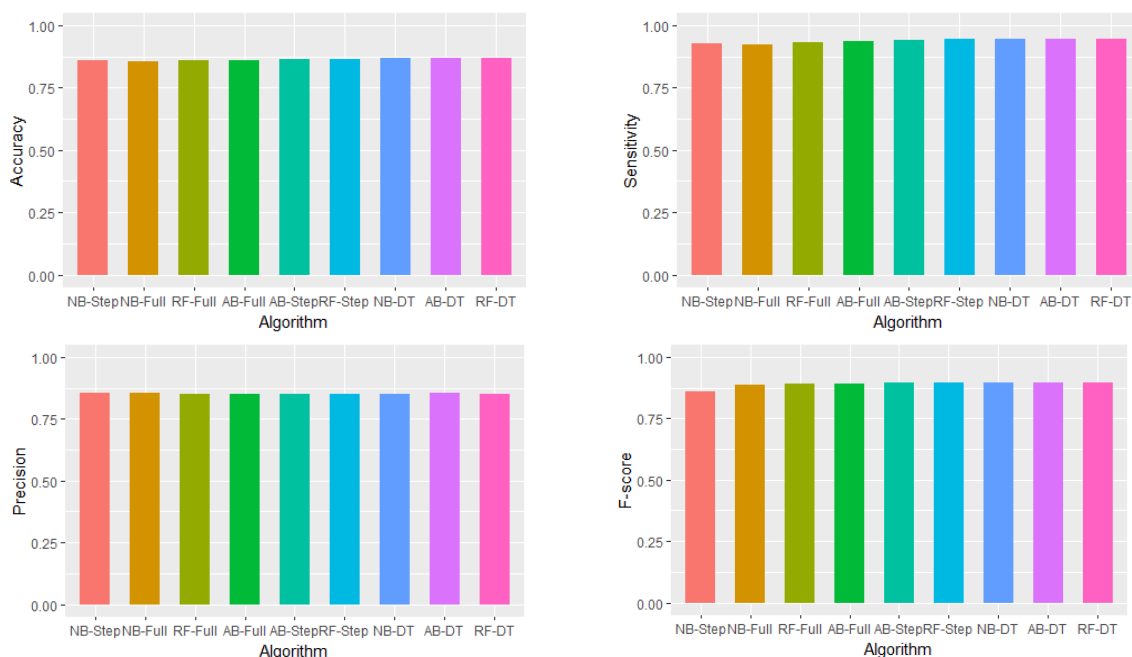
อัลกอริทึม-การคัดเลือกตัวแปร	ค่าความแม่นยำ	ค่าความไว	ค่าความเที่ยง	ค่าคะแนนเอฟ
NB-Full	0.8540 (9)	0.9219 (9)	0.8540 (2.5)	0.8867 (8)
RF-Full	0.8595 (8)	0.9329 (7)	0.8502 (8)	0.8899 (7)
AB-Full	0.8603 (6.5)	0.9383 (6)	0.8515 (5)	0.8902 (6)
NB-Step	0.8603 (6.5)	0.9292 (8)	0.8563 (1)	0.8612 (9)
RF-Step	0.8659 (4)	0.9437 (3)	0.8501 (9)	0.8944 (4)
AB-Step	0.8649 (5)	0.9411 (5)	0.8512 (6.5)	0.8939 (5)
NB-DT	0.8678 (3)	0.9455 (1)	0.8512 (6.5)	0.8959 (3)
RF-DT	0.8685 (1)	0.9451 (2)	0.8528 (4)	0.8966 (1)
AB-DT	0.8682 (2)	0.9435 (4)	0.8540 (2.5)	0.8965 (2)

หมายเหตุ ตัวหนา หมายถึง อัลกอริทึมที่มีประสิทธิภาพสูงสุดในแต่ละเกณฑ์ (พิจารณาในแนวคอลัมน์)

ตัวเลขที่อยู่ในวงเล็บแสดงการจัดอันดับ (Rank) ของอัลกอริทึมในแต่ละเกณฑ์ โดยที่ (1) หมายถึงมีประสิทธิภาพสูงสุด

AB-DT หมายถึงวิธีสุดดิ่งปรับได้ร่วมกับการคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจ และภาพที่ 3 นำเสนอผลลัพธ์ด้วยแผนภูมิแท่ง ซึ่งแต่ละแท่งแสดงถึงอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระที่แตกต่างกัน ผลการศึกษาพบว่าอัลกอริทึมทั้งหมดมีความถูกต้องในการทำนายสูงมากกว่าร้อยละ 85 ในทุกเกณฑ์วัดประสิทธิภาพ

สำหรับเกณฑ์ค่าความแม่นยำ อัลกอริทึมที่ร่วมกับการคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุด 3 อันดับแรก โดยอัลกอริทึม RF-DT มีประสิทธิภาพสูงสุด คิดเป็นร้อยละ 86.85 อันดับ 2 คืออัลกอริทึม AB-DT คิดเป็นร้อยละ 86.82 และอันดับ 3 คืออัลกอริทึม NB-DT คิดเป็นร้อยละ 86.78



ภาพที่ 3 ผลการเปรียบเทียบประสิทธิภาพของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระ

สำหรับเกณฑ์ค่าความไว อัลกอริทึมที่ร่วมกับการคัดเลือกตัวแปรอิสระด้วยวิธีต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุด 2 อันดับแรก คืออัลกอริทึม NB-DT คิดเป็นร้อยละ 94.55 และอัลกอริทึม RF-DT คิดเป็นร้อยละ 94.51 ตามลำดับ และอันดับ 3 คือ อัลกอริทึม RF-Step คิดเป็นร้อยละ 94.37

สำหรับเกณฑ์ค่าความเที่ยง วิธีนาอ็พเบสมีประสิทธิภาพสูงสุด 2 อันดับแรก โดยอัลกอริทึม NB-Step มีประสิทธิภาพสูงสุด คิดเป็นร้อยละ 85.63 และอันดับ 2 ร่วมคือ อัลกอริทึม NB-Full และ AB-DT คิดเป็นร้อยละ 85.40

สำหรับเกณฑ์ค่าคะแนนเอฟ ให้ผลลัพธ์เช่นเดียวกับเกณฑ์ค่าความแม่นยำใน 3 อันดับแรก โดยอัลกอริทึม RF-DT มีประสิทธิภาพสูงสุด คิดเป็นร้อยละ 89.66 อันดับ 2 คืออัลกอริทึม AB-DT คิดเป็นร้อยละ 89.65 และอันดับ 3 คืออัลกอริทึม NB-DT คิดเป็นร้อยละ 89.59 อย่างไรก็ตามเมื่อพิจารณาแผนภูมิแท่งในภาพที่ 3 พบว่าประสิทธิภาพของอัลกอริทึมในแต่ละเกณฑ์ค่อนข้างใกล้เคียงกัน

อภิปรายและสรุปผลการวิจัย

จากการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยลอจิสติกทีละชั้นและวิธีต้นไม้ตัดสินใจ พบว่าปัจจัยที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ ได้แก่ 1) ลักษณะการเกิดเหตุประเภทต่าง ๆ ซึ่งสอดคล้องกับงานวิจัย [6, 10, 18, 19] ที่อธิบายว่าค่าสินไหมทดแทนเมื่อเกิดอุบัติเหตุต้องสอดคล้องกับความหลากหลายของรูปแบบประกันภัย 2) ภูมิภาคที่เกิดเหตุ ซึ่งสอดคล้องกับงานวิจัย [6, 10] 3) อายุผู้เอาประกันภัย ซึ่งสอดคล้องกับงานวิจัย [6, 10] โดยพบว่าตัวแปรด้านประชากรศาสตร์ เช่น อายุ การศึกษา สถานภาพ และอาชีพ ส่งผลต่อรายการการฉ้อฉลประกันภัยรถยนต์ 4) ทุนประกันภัย ซึ่งสอดคล้องกับงานวิจัย [10] ปัจจัยที่ค้นพบใหม่สำหรับงานวิจัยนี้คือ 5) ประเภทการซ่อม และ 6) จำนวนเงินที่จ่ายรวมทั้งหมด ซึ่งสามารถนำไปใช้เป็นปัจจัยร่วมในการสร้างตัวแบบเชิงทำนายการเรียกร้องค่าสินไหมทดแทนในอนาคต อย่างไรก็ตาม งานวิจัยนี้มีตัวแปรอิสระเพียง 10 ตัวแปรเท่านั้น ดังนั้นการค้นหาตัวแปรอิสระอื่นๆ ที่มีอิทธิพลต่อการเรียกร้องค่าสินไหมทดแทนประกันภัยรถยนต์ ยังคงต้องดำเนินการต่อไปเพื่อประสิทธิภาพในการทำนายให้สูงขึ้น

ตัวแบบเชิงทำนายที่ทำงานร่วมกับการคัดเลือกตัวแปรอิสระมีประสิทธิภาพสูงกว่าตัวแบบที่ไม่มีการคัดเลือกตัวแปรอิสระ โดยพบว่าวิธีต้นไม้ตัดสินใจมีประสิทธิภาพสูงสุด รองลงมาคือวิธีการถดถอยลอจิสติกทีละชั้น ผลการศึกษานี้สอดคล้องกับงานวิจัย [10] ที่กล่าวว่า การคัดเลือกตัวแปรอิสระด้วยวิธีแบบไม่อิงพารามิเตอร์มีประสิทธิภาพสูงกว่าวิธีแบบอิงพารามิเตอร์ ผลการศึกษานี้ชี้ให้เห็นว่าการคัดเลือกตัวแปรอิสระเพื่อสร้างตัวแบบเชิงทำนายการเรียกร้องค่าสินไหมทดแทนที่ไม่มีการออกสำรวจภัยมีความสำคัญยิ่ง การศึกษาในอนาคตควรศึกษาวิธีการคัดเลือกตัวแปรอิสระด้วยวิธีการอื่น ๆ เช่น ใช้การถดถอย LASSO และวิธีป่าสุ่ม [6] และใช้ค่า SHAP [20] เป็นต้น

ประสิทธิภาพในการทำนายของอัลกอริทึมแตกต่างกันไปในแต่ละสถานการณ์ แต่หากพิจารณาแยกแต่ละวิธีการคัดเลือกตัวแปรอิสระ สามารถสรุปในภาพรวมได้ว่า วิธีบูสต์ติงปรับได้มีประสิทธิภาพสูงสุดสำหรับตัวแบบที่ไม่มีการคัดเลือกตัวแปร ผลการศึกษานี้สอดคล้องกับงานวิจัย [18] ที่ระบุว่าวิธีบูสต์ติงปรับได้มีประสิทธิภาพสูงสุดในการทำนายการต่อกรมธรรม์ประกันภัย วิธีป่าสุ่มมีประสิทธิภาพสูงสุดสำหรับตัวแบบที่คัดเลือกตัวแปรด้วยวิธีการถดถอย ลอจิสติกทีละชั้นและวิธีต้นไม้ตัดสินใจ ผลการศึกษานี้สอดคล้องกับงานวิจัย [10, 19] ที่ระบุว่าวิธีป่าสุ่มเป็นวิธีที่มีประสิทธิภาพสูงสุด จะเห็นว่าทั้ง 2 อัลกอริทึมยังคงเป็นวิธีเชิงทำนายที่มีประสิทธิภาพ ส่วนวิธีนาอ็พเบสที่เป็นอัลกอริทึมควมคุมมีประสิทธิภาพสูงสุดในบางสถานการณ์ เช่น เมื่อไม่มีการคัดเลือกตัวแปรอิสระและการคัดเลือกตัวแปรอิสระด้วยวิธีการถดถอยลอจิสติกทีละชั้น สำหรับเกณฑ์ค่าความเที่ยง การศึกษาใน



อนาคตควรพิจารณาวิธีเชิงทำนายอื่น ๆ เช่น วิธีการเกรเดียนต์ (Gradient Boosting) และวิธีเพื่อนบ้านใกล้ที่สุด (K-Nearest Neighbor: KNN) [20] และวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) [6]

เมื่อพิจารณาประสิทธิภาพในการทำนายของอัลกอริทึมร่วมกับวิธีการคัดเลือกตัวแปรอิสระ พบว่าวิธีป่าสุ่มร่วมกับวิธีต้นไม้ตัดสินใจ (RF-DT) มีประสิทธิภาพสูงสุด ผลสรุปที่ได้จากงานวิจัยนี้ใช้เกณฑ์พิจารณาด้วยค่าคะแนนเอฟเป็นหลักเนื่องจากเป็นค่าถ่วงดุลระหว่างค่าความไวและค่าความเที่ยงจึงบอกสารสนเทศได้ครอบคลุมมากกว่าการพิจารณาเกณฑ์ค่าใดค่าหนึ่ง สำหรับเกณฑ์ค่าความแม่นยำเป็นค่าที่บอกความถูกต้องในภาพรวมและไม่สามารถบอกประสิทธิภาพการทำนายแยกตามหมวดหมู่ได้ เป็นที่น่าสนใจว่าอัลกอริทึมทั้งหมดในงานวิจัยนี้ให้ค่าความถูกต้องในการทำนายสำหรับทุกเกณฑ์สูงมากกว่าร้อยละ 85

กิตติกรรมประกาศ

ผู้วิจัยขอขอบคุณโครงการปริญญาตรีภาคพิเศษหลักสูตรวิทยาการประกันภัย สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์ ที่สนับสนุนทุนวิจัยและขอขอบคุณคณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยเทคโนโลยีราชมงคลรัตนโกสินทร์ พื้นที่สาธิต ที่สนับสนุนเวลาให้นักวิจัยทำวิจัยลุล่วงไปด้วยดี

เอกสารอ้างอิง

1. ชูลีกร แด้ศภาพพงษ์. บทบาทของการใช้ข้อมูลเชิงวิเคราะห์ (Data Analytics) ในการตรวจสอบการฉ้อฉลในประกันภัย. วารสารประกันภัย 2558;127:9-14.
2. Belhadji EB, Dionne G, Tarkhani F. A Model for the Detection of Insurance Fraud. Geneva Pap Risk Insur Issues Pract 2000;25(4):517-38.
3. Roy R, George KT. Detecting insurance claims fraud using machine learning techniques. In: Proceeding of the 2017 International Conference on Circuit Power and Computing Technologies (ICCPCT), April 20-21, 2017, Kollam, India. Institute of Electrical and Electronics Engineers; 2017. p. 1-6.
4. สำนักงานคณะกรรมการกำกับและส่งเสริมการประกอบธุรกิจประกันภัย. สรุปผลแบบสำรวจความคิดเห็นต่อสถานะความเสี่ยงและทิศทางการดำเนินธุรกิจประกันภัยปี 2565 (TIPeR). [อินเทอร์เน็ต]. 2565. [เข้าถึงเมื่อ 30 ธ.ค. 2565]. เข้าถึงได้จาก: <https://www.oic.or.th/th/consumer/92762>
5. Kowshalya G, Nandhini M. Predicting Fraudulent Claims in Automobile Insurance. In: Proceedings of the 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018), April 20-21, 2018, Coimbatore, India. Institute of Electrical and Electronics Engineers; 2018. p. 1338-43.
6. Moon H, Pu Y, Ceglia C. A Predictive Modeling for Detecting Fraudulent Automobile Insurance Claims. Theor Econ Lett 2019;9:1886-900.
7. กฤษณา ศักติวรพงษ์. การฉ้อฉลในการใช้สิทธิเรียกร้องค่าสินไหมทดแทนในธุรกิจประกันวินาศภัย. วิทยานิพนธ์นิติศาสตรมหาบัณฑิต สาขาวิชานิติศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย. กรุงเทพฯ; 2551.



8. พรานวนภัส จันทรวดี. นิติเศรษฐศาสตร์ว่าด้วยการฉ้อฉลประกันภัย: กรณีศึกษา คนเดินเท้าแกล้งก่ออุบัติเหตุเพื่อเรียกร้องค่าสินไหมทดแทน. การค้นคว้าอิสระศิลปศาสตรมหาบัณฑิต สาขานิติเศรษฐศาสตร์การค้าระหว่างประเทศ, คณะนิติศาสตร์และคณะเศรษฐศาสตร์ มหาวิทยาลัยธรรมศาสตร์. ปทุมธานี; 2559.
9. Wang Y, Xu W. Leveraging deep learning with LDA-based text analytics to detect automobile insurance fraud. Decis Support Syst 2018;105(1):87-95.
10. Simmachan T, Manopa W, Neamhom P, Poothong A, Phaphan W. Detecting Fraudulent Claims in Automobile Insurance Policies by Data Mining Techniques. Thailand Stat 2023;(in press).
11. Abdelhadi S, Elbahnasy K, Abdelsalam MA. Proposed model to predict auto insurance claims using machine learning techniques. J Theor Appl Inf Technol 2020;98(22):3428-37.
12. Aksoy S, Haralick RM. Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognit Lett 2001;22(5):563-82.
13. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proceedings of the 32nd International conference on machine learning, July 7-9, 2015, Lille, France. PMLR; 2015. p. 448-56.
14. Mauša G, Galinac Grbac T, Bašić BD. Multivariate logistic regression prediction of fault-proneness in software modules. In: 2012 Proceedings of the 35th International Convention MIPRO, May 21-25, 2012, Opatija, Croatia. Institute of Electrical and Electronics Engineers; 2012. p. 698-703.
15. Tom M. Mitchell: Machine Learning. Burr Ridge 1997;45(37):870-7.
16. Patil TR, Sherekar S. Performance analysis of Naive Bayes and J48 classification algorithm for data classification. Int J Comput Sci Appl 2013;6(2):256-61.
17. ชาลีสา จิตบุญญาพินิจ, ปราณี มณีรัตน์, นิเวศ จิระวิจิตชัย การพัฒนาแบบจำลองการวิเคราะห์ความรู้สึกบนสื่อสังคมออนไลน์ไทยโดยใช้เทคนิคการเรียนรู้เชิงลึก. ว วิทย เทคโนโลยี หัวเฉียวเฉลิมพระเกียรติ 2565;8(2),8-18.
18. กมล บุชบา, มหนาลัย สุดสายสาคร, ภัณฑิรา เตชะพรสิน, สรภัพร โกมลทองทิพย์. การเปรียบเทียบการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ ประเภท 1 ระหว่างตัวแบบป่าสุ่มและตัวแบบบูสต์ติ่งปรับได้. TSTJ 2564;10(2):124-34.
19. สุภัตสรสา สมเจตนา, จาริ ทองคำ. การเปรียบเทียบประสิทธิภาพเทคนิคเหมืองข้อมูลในการสร้างแบบจำลองเพื่อจำแนกความคิดเห็นของผู้ปกครองต่อการใช้สมาร์ทโฟนของบุตร. ววท มอว 2564;23(1):21-30.
20. Vanishkorn B, Supanich W. Crash Severity Classification Prediction and Factors Affecting Analysis of Highway Accidents. In: Proceedings of the 9th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA 2022), September 28-29, 2022, Tokoname, Japan. Institute of Electrical and Electronics Engineers; 2022. p. 1-6.