

การเปรียบเทียบการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ ประเภท 1 ระหว่างตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้ Comparison of the Prediction of the First Class Automobile Insurance Renewal between Random Forest Model and Adaptive Boosting Model

กมล บุชบา*, มथนาลัย สุตสายสาคร, ภัณฑิรา เตชะพรสิน และสรลพร โกมลทองทิพย์
สาขาวิชาคณิตศาสตร์และสถิติ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยธรรมศาสตร์

Kamon Budsaba*, Mathanalai Sutsaisakhon, Pantira Techapornsin
and Saranporn Komolthonhthip

Department of Mathematics and Statistics, Faculty of Science and Technology, Thammasat University

Received: February 2, 2021 ; Accepted: May 28, 2021

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ระหว่างตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้ และศึกษาอัตราส่วนที่เหมาะสมในการแบ่งข้อมูลชุดฝึกฝนและชุดทดสอบ ข้อมูลที่ใช้ในการวิจัยเป็นข้อมูลการต่ออายุกรมธรรม์ประกันภัยรถยนต์ส่วนบุคคลประเภท 1 ของบริษัทประกันภัยแห่งหนึ่งในประเทศไทย พ.ศ. 2561 และ พ.ศ. 2562 จำนวนทั้งสิ้น 10,000 ชุด ประกอบด้วยจำนวนตัวแปรที่สนใจ 8 ตัวแปร ตัวแปรตอบสนองที่สนใจในการศึกษา คือ สถานะการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 แบ่งเป็นการต่อหรือไม่ต่ออายุกรมธรรม์ ตัวแปรอิสระ ได้แก่ เพศของผู้เอาประกัน, เบี้ยประกันภัยรถยนต์, อายุรถยนต์ที่รับประกันภัย, จำนวนการเคลมใน 1 ปี, จำนวนความเสียหายที่เกิดขึ้นใน 1 ปี, เบี้ยประกันส่วนลด และส่วนเพิ่มเบี้ยประกันเมื่อเกิดความเสียหายเกินขั้นต่ำ ใช้โปรแกรม RStudio Version 1.2.5033 แพคเกจ rattle ในการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 จากผลการวิจัยพบว่าตัวแบบบูสต์ติงปรับได้เมื่อการแบ่งชุดข้อมูลชุดฝึกฝนและชุดทดสอบในแต่ละอัตราส่วน 85:15 มีความเหมาะสมมากที่สุดโดยมีค่าความแม่นยำร้อยละ 66.90 และค่าประสิทธิภาพโดยรวมร้อยละ 78.44 มากกว่าตัวแบบป่าสุ่มเล็กน้อย โดยตัวแปรอิสระที่มีความสำคัญในการทำนายการต่ออายุกรมธรรม์ลำดับแรกคือเบี้ยประกันภัยรถยนต์ และ จำนวนความเสียหายที่เกิดขึ้นใน 1 ปี รองลงมาคือเบี้ยประกันส่วนลด

คำสำคัญ : ต้นไม้ตัดสินใจ, วิธีการแบ็กกิง, วิธีการบูสต์ติง, เมทริกซ์ความสับสน, ความแม่นยำ

Abstract

This paper aims to compare the prediction of the first class automobile insurance renewal between the random forest model and adaptive boosting model, and to study the partition proportion between the training and testing dataset. Secondary data were obtained from an insurance company in Thailand from 2018 to 2019 with 10,000 datasets and 8 interested variables. The response variable in this study is the renewal status of each policy, e.g., renew or not renew. The independent variables are the gender of the policyholder, car premium, car year, number of claims in one year, number of incurred losses in one year, car premium discount, and car premium surcharge if the incurred loss excess the policy limit by using RStudio version 1.2.5033 program with rattle package for prediction of the first class automobile insurance renewal. The result showed that the adaptive boosting model with a proportion of training and testing dataset as 85:15 is an appropriate model with 66.90 % accuracy and 78.44 % of overall efficiency. This adaptive boosting model is a little better than the random forest model. The most important independent variables for insurance policy renewal are car premium and the number of claims in one year, followed by the car premium discount.

Keywords: decision trees; bagging, boosting; confusion matrix; accuracy

1. บทนำ

ประกันภัยรถยนต์เป็นหนึ่งในประกันภัยวินาศภัยที่คุ้มครองภัยจากอุบัติเหตุทางรถยนต์ รวมถึงทรัพย์สินของผู้เอาประกันและคู่กรณีไม่ว่าผู้เอาประกันจะเป็นผู้เสียหายหรือผู้ก่อเหตุ หากอยู่ในความคุ้มครองที่บริษัทประกันภัยระบุไว้บริษัทจะชดเชยค่าเสียหายโดยเรียกจำนวนเงินก่อนนี้ว่าค่าสินไหมทดแทนซึ่งจะจ่ายตามค่าเสียหายจริงแต่ไม่เกินจำนวนที่ระบุไว้ตามกรมธรรม์ประกันภัย การประกันภัยรถยนต์ถือเป็นเครื่องมือหลักประกันทางการเงินคือการโอนความเสี่ยงของผู้เอาประกันภัยไปสู่ผู้รับประกันภัย

ปัจจุบันรถยนต์เข้ามามีบทบาทสำคัญในการเดินทางในชีวิตประจำวัน ทำให้มีจำนวนรถยนต์บนท้องถนนมากขึ้น ซึ่งจะเห็นได้จากจำนวนรถจดทะเบียนสะสมในปี พ.ศ. 2561 ที่มีจำนวน

36,124,623 คัน เมื่อเทียบกับปี พ.ศ. 2562 ที่มีจำนวน 36,196,629 คัน [1] ซึ่งการที่รถยนต์มีจำนวนเพิ่มมากขึ้น ส่งผลให้มีจำนวนการเกิดอุบัติเหตุทางรถยนต์เพิ่มสูงขึ้นทุกปี [2] ดังนั้นการประกันภัยรถยนต์จึงมีความจำเป็นต่อผู้ใช้รถยนต์เพื่อเป็นการแบ่งเบาภาระค่าใช้จ่ายหากเกิดอุบัติเหตุขึ้น

ประกันภัยรถยนต์ในประเทศไทยประกอบไปด้วย ประกันภัยรถยนต์ภาคบังคับ (พ.ร.บ.) ตามกฎหมายรถที่จดทะเบียนต้องทำทุกคัน และประกันภัยรถยนต์ภาคสมัครใจ เป็นประกันที่สามารถทำหรือไม่ทำก็ได้ขึ้นอยู่กับความสมัครใจ โดยแบ่งออกเป็น 5 ประเภท ประกอบด้วย ประกันภัยรถยนต์ชั้น 1 ประกันภัยรถยนต์ชั้น 2 ประกันภัยรถยนต์ชั้น 2+ ประกันภัยรถยนต์ชั้น 3 ประกันภัยรถยนต์ชั้น 3+ ซึ่งการประกันรถยนต์ชั้น

1 หรือ ประกันภัยรถยนต์ประเภท 1 เป็นประกันภัยรถยนต์ที่ให้ความคุ้มครองครอบคลุมเกือบทุกกรณี ที่กฎหมายกำหนดมากกว่าการประกันรถยนต์ประเภทอื่น ๆ

การประกันภัยรถยนต์ภาคสมัครใจเป็นการประกันที่มีความหลากหลายรูปแบบความคุ้มครองได้รับความสนใจจากผู้บริโภคจำนวนมาก โดยเอาประกันสามารถเลือกซื้อได้ตามความเหมาะสมพิจารณาจากลักษณะการใช้งาน อายุรถยนต์ ประเภทรถยนต์ หากพิจารณาจากความคุ้มครองและผลประโยชน์สูงสุด ประกันภัยรถยนต์ชั้น 1 หรือ ประกันรถยนต์ประเภท 1 เป็นประกันรถยนต์ที่ให้ความคุ้มครองที่ครอบคลุมเกือบทุกกรณีตามที่กฎหมายกำหนด ซึ่งจะคุ้มครองทุกสถานการณ์ ทั้งการเกิดอุบัติเหตุที่มีคู่กรณีหรือไม่มีคู่กรณี ไม่ว่าเราจะเป็นผู้กระทำหรือผู้เสียหาย รวมถึงภัยจากไฟไหม้ สูญหาย โจรกรรม ภัยทางธรรมชาติ และน้ำท่วม การประกันภัยรถยนต์ชั้น 1 หรือ การประกันภัยรถยนต์ประเภท 1 จึงเป็นการประกันภัยที่น่าสนใจกับผู้เอาประกันภัยเป็นจำนวนมากเพื่อได้รับความคุ้มครองที่เต็มรูปแบบ

การทำประกันภัยรถยนต์ให้ความคุ้มครองเพียง 1 ปี ดังนั้นจึงจำเป็นต้องมีการต่ออายุกรมธรรม์ทุกปี การต่ออายุกรมธรรม์ประกันภัยถือว่าเป็นผลิตภัณฑ์ที่ได้รับความสนใจมาโดยตลอด ส่งผลให้มูลค่าตลาดประกันภัยมีแนวโน้มที่จะเติบโตขึ้น การประกันภัยรถยนต์เป็นผลิตภัณฑ์ที่มีอยู่ในหลากหลายบริษัท แต่ก็ยังมีบางบริษัทประกันภัยประสบปัญหาฐานะทางการเงินที่ไม่มั่นคงขาด

สภาพคล่องรุนแรง และค้างสินไหมทดแทนจำนวนมากอย่างต่อเนื่อง เนื่องจากจัดสรรงบประมาณเงินสำรองที่ไม่เพียงพอ แต่ก็มีบริษัทอีกมากที่สามารถดำรงกิจการให้มีเงินทุนสำรอง โดยมีประกันภัยหลากหลายรูปแบบกรมธรรม์ที่สามารถตอบสนองความต้องการของผู้เอาประกันภัย ทำให้กลุ่มผู้เอาประกันภัยรายเดิมมีความเชื่อมั่นต่อบริษัทประกันภัย รวมถึงทำให้เกิดกลุ่มผู้เอาประกันภัยรายใหม่เพิ่มขึ้น จึงทำให้บริษัทสามารถดำเนินธุรกิจมาได้ยาวนานและไม่ขาดสภาพคล่องทางการเงิน ส่งผลให้บริษัทประกันภัยมีเงินทุนสำรองอย่างมั่นคง ซึ่งทำให้คณะผู้วิจัยเล็งเห็นถึงความสำคัญของการสำรองเงินประกันภัยของบริษัทประกันภัยเพื่อให้ดำรงเงินกองทุนให้เพียงพอในการรองรับความเสี่ยงที่จะเกิดขึ้นในอนาคตอย่างเหมาะสมตามที่กฎหมายกำหนด สำหรับเงินสำรองที่กำหนดไว้คือให้วางร้อยละ 25 ของรายได้ที่ยังไม่ตกเป็นรายได้ของบริษัท ดังนั้นผู้วิจัยจึงสนใจและต้องการทำนายนการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 เพื่อให้บริษัทประกันภัยสามารถจัดสรรงบประมาณเงินสำรองประกันภัยได้อย่างแม่นยำ

การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบการทำนายนการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ระหว่างตัวแบบป่าสุ่มและตัวแบบบูสต์ ดิงปรับได้ และศึกษาสัดส่วนที่เหมาะสมในการแบ่งข้อมูลชุดฝึกฝนและชุดทดสอบเพื่อหาตัวแบบที่มีความแม่นยำและเหมาะสมที่สุด รวมทั้งพิจารณาลำดับความสำคัญของตัวแปรอิสระในการทำนายนการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1

2. วิธีการ

2.1 ขอบเขตการวิจัย

2.1.1 ด้านข้อมูล

ข้อมูลที่ใช้ในการวิจัยเป็นสถานะการต่ออายุกรมธรรม์ประกันภัยรถยนต์ส่วนบุคคลประเภท 1 และตัวแปรที่คาดว่าจะมีผลต่อการต่ออายุกรมธรรม์ประกันภัยรถยนต์ส่วนบุคคลประเภท 1 ของบริษัทประกันภัยแห่งหนึ่งในประเทศไทยปี พ.ศ. 2561 และ พ.ศ. 2562 จำนวนทั้งหมด 10,000 ชุด

2.1.2 ด้านตัวแบบและวิธีการที่ใช้

2.1.2.1 ตัวแบบป่าสุ่ม (Random Forest Model)

ตัวแบบป่าสุ่มพัฒนาจากวิธีการแบ็กกิง (bagging) ซึ่งเป็นการสร้างต้นไม้ตัดสินใจหลาย ๆ ต้นโดยใช้ชุดตัวอย่างซ้ำด้วยหลักการบูตสแตรป์ (bootstrap) จากข้อมูลชุดฝึกฝนเพียงชุดเดียวโดยการสร้างต้นไม้ใช้หลักการแบ่งต้นไม้ด้วยตัวอย่างสุ่มขนาด m จากตัวแปรอิสระทั้งหมด p ตัว (โดยทั่วไปจะให้ m เท่ากับรากที่สองของ p) เพื่อเป็นเกณฑ์ในการแบ่งต้นไม้ ซึ่งเมื่อนำต้นไม้ตัดสินใจมาทำนายร่วมกันโดยใช้ค่าเฉลี่ยรวมจากผลการทำนายจากต้นไม้แต่ละต้นของในกรณีนี้ที่ตัวแปรตอบสนองเป็นข้อมูลเชิงปริมาณหรือใช้ฐานนิยมจากผลการทำนายจากต้นไม้แต่ละต้นในกรณีนี้ที่ตัวแปรตอบสนองเป็นข้อมูลเชิงคุณภาพ ทำให้ได้ตัวแบบที่มีความแม่นยำมากกว่าการทำนายโดยใช้ต้นไม้ตัดสินใจเพียงต้นเดียว [3]

2.1.2.2 ตัวแบบบูสต์ติงปรับได้ (Adaptive Boosting Model)

ตัวแบบบูสต์ติงปรับได้เป็นวิธีการที่พัฒนาจากวิธีบูสต์ติง (boosting) ซึ่งต่างจากวิธีการแบ็กกิงที่ต้นไม้ที่สร้างขึ้นหลาย ๆ ต้นนั้นเป็นอิสระกัน แต่วิธีบูสต์ติงต้นไม้ต้นใหม่จะสร้างขึ้นโดยอาศัยสารสนเทศจากต้นไม้ต้นเดิม นั่นคือใช้ตัวแบบต้นไม้

ตัดสินใจปัจจุบันก่อนจากนั้นจึงสร้างต้นไม้ตัดสินใจต้นใหม่โดยใช้ส่วนเหลือ (residuals) ของตัวแบบต้นไม้ตัดสินใจปัจจุบัน จากนั้นจึงเพิ่มต้นไม้ใหม่ในนี้ในตัวแบบเพื่อให้ได้ส่วนเหลือใหม่ที่เป็นปัจจุบันมากขึ้น ขั้นตอนวิธีในลักษณะนี้จะดำเนินต่อไปเรื่อย ๆ ให้ต้นไม้ตัดสินใจเรียนรู้อย่างซ้ำ ๆ โดยมีการปรับพารามิเตอร์เช่น จำนวนต้นไม้ จำนวนการแบ่งต้นไม้ แบบค่อยเป็นค่อยไปจนกว่าจะได้จำนวนต้นไม้ตัดสินใจที่มีความเหมาะสมที่สุด [4]

2.1.2.3 เกณฑ์การประเมินประสิทธิภาพของตัวแบบ

ใช้เมทริกซ์ความสับสน (confusion matrix) ที่แสดงต่อไปนี้เป็นพื้นฐานในการหาค่าความแม่นยำ (accuracy) และประสิทธิภาพโดยรวม

Table 1 confusion matrix

ผลลัพธ์การ ทำนาย	ผลลัพธ์ที่เกิดขึ้นจริง	
	บวก	ลบ
บวก	จำนวนบวก จริง (TP)	จำนวนบวก เท็จ (FP)
ลบ	จำนวนลบเท็จ (FN)	จำนวนลบจริง (TN)

ค่าความแม่นยำ = $(TP+TN)/(TP+TN+FP+FN)$

ค่าประสิทธิภาพโดยรวม = $2(\text{ค่าความเที่ยง})/(\text{ค่าการเรียกคืน})$

โดยที่ ค่าความเที่ยง = $TP/(TP+FP)$

ค่าการเรียกคืน = $TP/(TP+FN)$

2.1.2.4 เกณฑ์การวัดความสำคัญ

สำคัญของตัวแปรอิสระในตัวแบบ สำหรับตัวแบบป่าสุ่ม ใช้ค่าความแม่นยำลดลงเฉลี่ย (mean decrease accuracy) ซึ่งเป็นร้อยละของความแม่นยำในการทำนายที่ลดลงเมื่อนำตัวแปรอิสระนั้นออกจากตัวแบบ

สำหรับตัวแบบบูสต์ดิงปรับได้ ใช้ค่าความถี่ (frequency) ในการสุ่มเรียกใช้ตัวแปรอิสระนั้นเพื่อปรับค่าถ่วงน้ำหนักเพื่อให้ได้ผลการทำนายที่ดีขึ้นเมื่อกำหนดค่าให้โปรแกรมทำการสุ่มตัวแปรอิสระมา 50 ครั้ง

2.1.3. ด้านการคำนวณ

2.1.3.1 ใช้โปรแกรม SPSS Version 23 เพื่อคาดหมายในเบื้องต้นเกี่ยวกับตัว

แปรอิสระที่ส่งผลต่อการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1

2.1.3.2 ใช้โปรแกรม RStudio Version 1.2.5033 โดยใช้แพ็คเกจ rattle ในการสร้างตัวแบบป่าสุ่มและตัวแบบบูสต์ดิงปรับได้ในการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1

Table 1 Proportion of first class automobile insurance renewal policy classified by qualitative independent variables.

Independent variables		Status of first class automobile insurance policy			χ^2
		Renewal policy	Not renewal policy	Total	
Gender	male	3,279 (63.03%)	1,923 (36.97%)	5,202 (100%)	0.149 P-value = 0.699
	female	2,997 (62.66%)	1,786 (37.34%)	4,783 (100%)	
Car year	1-4 years	1,772 (62.88%)	1,046 (37.12%)	2,818 (100%)	20.138 P-value < 0.0001
	5-6 years	1,865 (59.99%)	1,244 (40.01%)	3,109 (100%)	
	7-8 years	1,066 (64.14%)	596 (35.86%)	1,662 (100%)	
	9 years and over	1,573 (65.65%)	823 (34.35%)	2,396 (100%)	
No. of claim	No claim	4,497 (67.44%)	2,171 (32.56%)	6,668 (100%)	180.907 P-value < 0.0001
	Have a claim	1,779 (53.63%)	1,538 (46.27%)	3,317 (100%)	

3. ผลการวิจัยและวิจารณ์ผล

2.2 วิธีการดำเนินงานวิจัย

2.2.1 รวบรวมข้อมูลอุบัติเหตุภูมิของการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 จากบริษัทประกันภัยแห่งหนึ่ง ในพ.ศ. 2561 และ พ.ศ. 2562

2.2.2 ตรวจสอบข้อมูลเบื้องต้น เช่น หาสถิติเชิงพรรณนาและสถิติทดสอบด้วยวิธีไคกำลังสองและการทดสอบที โดยใช้โปรแกรม SPSS

2.2.3 จัดกลุ่มชุดข้อมูลใหม่ของตัวแปรอิสระ เช่น ตัวแปรอายุรถยนต์ที่รับประกันภัยและตัวแปรจำนวนการเคลมใน 1 ปี

2.2.4 แบ่งข้อมูลชุดฝึกฝนและชุดทดสอบระหว่างอัตราส่วนต่าง ๆ ได้แก่ 85:15, 80:20, 75:25, 70:30, 65:35, 60:40, 55:45 และ 50:50

2.2.5 สร้างตัวแบบป่าสุ่มและตัวแบบบูสต์ดิงปรับได้โดยใช้โปรแกรม R studio แพคเกจ rattle

2.2.6 เปรียบเทียบประสิทธิภาพผลลัพธ์ของการทำนายทั้ง 2 ตัวแบบ โดยใช้เกณฑ์ประสิทธิภาพโดยรวมและค่าความแม่นยำในการเปรียบเทียบ

2.2.7 สรุปผลการวิจัย

3.1 ผลการวิเคราะห์สถิติเชิงพรรณนาและสถิติทดสอบ

3.1.1 จากการตรวจสอบข้อมูลเบื้องต้น พบว่าตัวแปรจำนวนความเสียหายที่เกิดขึ้นใน 1 ปี ของบางรายมีค่าไม่ถูกต้องและมีความสัมพันธ์ไม่ตรงกับความเป็นจริงจำนวน 15 ราย ผู้วิจัยจึงได้ตัดชุดข้อมูลนี้ออกไป ทำให้มีข้อมูลที่นำไปวิเคราะห์ในขั้นตอนต่อไปจำนวน 9,985 ชุด

3.1.2 ตัวแปรอิสระที่เป็นข้อมูลเชิงคุณภาพ มีทั้งหมด 3 ตัวแปร ได้แก่ เพศ (Gender),

อายุของรถยนต์ (Car Year) และจำนวนการเคลม (No of Claim) เมื่อนำมาจำแนกตามสถานภาพการต่ออายุกรมธรรม์และคำนวณสัดส่วนของจำนวนกรมธรรม์ที่ต่ออายุรถยนต์ประเภท 1 (ตารางที่ 1) พบว่า เพศชายและเพศหญิงมีสัดส่วนการต่ออายุกรมธรรม์รถยนต์ประเภท 1 ใกล้เคียงกัน กล่าวคือ ร้อยละ 63.03 และ ร้อยละ 62.66 ตามลำดับรถยนต์ที่มีอายุ 9 ปีขึ้นไปมีสัดส่วนการต่ออายุกรมธรรม์รถยนต์ประเภท 1 สูงสุด รองลงมาคือรถยนต์ที่มีอายุ 7-8 ปี กล่าวคือร้อยละ 65.65 และ ร้อยละ 64.14 ตามลำดับ ผู้ที่ไม่เคยเคลมเลยมีสัดส่วนการต่ออายุร้อยละ 67.44 มากกว่าผู้ที่เคยเคลม (ร้อยละ 53.63) เมื่อทดสอบสมมติฐานเกี่ยวกับความสัมพันธ์ระหว่างตัวแปรอิสระเชิงคุณภาพ (เพศ, อายุของรถยนต์ และจำนวนการเคลม) กับการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ด้วยสถิติไคกำลังสองพบว่า อายุของรถยนต์และจำนวนการเคลมมีความสัมพันธ์กับการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 อย่างมีนัยสำคัญยิ่ง (P-value น้อยกว่า 0.0001)

3.1.3 ตัวแปรอิสระที่เป็นตัวแปรเชิงปริมาณ มีทั้งหมด 4 ตัวแปร คือ เบี้ยประกันภัยรถยนต์ (Premium), จำนวนความเสียหายที่เกิดขึ้นใน 1 ปี (Incurred Loss), เบี้ยประกันภัยส่วนลด (NCB) และส่วนเพิ่มเบี้ยประกันภัยเมื่อเกิดความเสียหายเกินขั้นต่ำ (Surcharge) เมื่อนำมาคำนวณสถิติเชิงพรรณนาและทดสอบสมมติฐานความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรเชิงปริมาณแต่ละตัวด้วยสถิติทดสอบที (t-test) ระหว่างผู้ที่ต่ออายุที่ไม่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 (ตารางที่ 2) พบว่าเบี้ยประกันภัยรถยนต์โดยเฉลี่ยของผู้ที่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 เท่ากับ 13,007.04 บาทต่ำกว่าเบี้ยประกันภัยรถยนต์โดยเฉลี่ยของผู้ที่ไม่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 (13,823.65 บาท) อย่างมีนัยสำคัญยิ่ง (P-value < 0.0001)

Table 2 Mean, standard deviation and t-test for each quantitative independent variable between first class automobile insurance renewal policy and nonrenewal policy

Independent variables	Status of first class automobile insurance policy	Means	Standard deviation	t-test
Premium (baht)	renewal policy (n = 6,276)	13,007.04	6,416.22	-5.897 (P-value < 0.0001)
	not renewal policy (n = 3,709)	13,823.65	6,841.57	
Incurred loss (baht)	renewal policy (n = 6,276)	6,005.98	31,304.99	-9.389 (P-value < 0.0001)
	not renewal policy (n = 3,709)	14,178.41	47,232.85	
NCB (baht)	renewal policy (n = 6,276)	7,808.15	4,386.54	4.294 (P-value < 0.0001)
	not renewal policy (n = 3,709)	7,421.13	4,330.49	
Surcharge (baht)	renewal policy (n = 6,276)	1.14	64.18	-2.744 (P-value = 0.006)
	not renewal policy (n = 3,709)	12.66	250.97	

จำนวนความเสียหายที่เกิดขึ้นใน 1 ปีโดยเฉลี่ยของผู้ที่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 เท่ากับ 6,005.98 บาทน้อยกว่าจำนวนความเสียหายที่เกิดขึ้นใน 1 ปีโดยเฉลี่ยของผู้ที่ไม่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 (14,178.41 บาท) อย่างมีนัยสำคัญยิ่ง (P-value < 0.0001)

เบี้ยประกันภัยส่วนลดโดยเฉลี่ยของผู้ที่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 เท่ากับ 7,808.15 บาทสูงกว่าเบี้ยประกันภัยส่วนลดโดยเฉลี่ยของผู้ที่

ไม่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 (7,421.13 บาท) อย่างมีนัยสำคัญยิ่ง (P-value < 0.0001)

ส่วนเพิ่มเบี้ยประกันภัยเมื่อเกิดความเสียหายเกินขั้นต่ำโดยเฉลี่ยของผู้ที่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 เท่ากับ 1.14 บาทต่ำกว่าส่วนเพิ่มเบี้ยประกันภัยเมื่อเกิดความเสียหายเกินขั้นต่ำโดยเฉลี่ยของผู้ที่ไม่ต่ออายุกรมธรรม์รถยนต์ประเภท 1 (12.66 บาท) อย่างมีนัยสำคัญยิ่ง (P-value = 0.006)

จากผลการทดสอบความสัมพันธ์ระหว่างตัวแปรอิสระเชิงคุณภาพกับการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ด้วยสถิติไคกำลังสองและการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของตัวแปรเชิงปริมาณด้วยการทดสอบที่ ผู้วิจัยคาดว่าตัวแปรอิสระที่นำจะมีส่วนช่วยในการทำนายการต่ออายุกรมธรรม์รถยนต์ประเภท 1 ของบริษัทประกันภัยแห่งนี้ ได้แก่

- อายุของรถยนต์ (Car Year)
- จำนวนการเคลม (No of Claim)
- เบี้ยประกันภัยรถยนต์ (Premium)
- จำนวนความเสียหายที่เกิดขึ้นใน 1 ปี (Incurred Loss)
- เบี้ยประกันส่วนลด (NCB)
- ส่วนเพิ่มเบี้ยเมื่อเกิดความเสียหายเกินขั้นต่ำ (Surcharge)

3.2 ผลการทดสอบตัวแบบ

จากการพิจารณาค่าข้อมูลดิบโดยละเอียดพบว่าตัวแปรส่วนเพิ่มเบี้ยประกันเมื่อเกิดความเสียหายเกินขั้นต่ำมีจำนวนกรมธรรม์ที่ไม่มีการเสียเบี้ยประกันภัยเพิ่ม (ค่าเป็นศูนย์) มากกว่า ร้อยละ 90 ผู้วิจัยจึงได้ทำการตัดตัวแปรนี้จากการสร้างตัวแบบ ทำให้เหลือตัวแปรอิสระที่นำมาสร้างตัวแบบจำนวน 5 ตัวแปรคือ

- อายุของรถยนต์ (Car Year)
- จำนวนการเคลม (No of Claim)
- เบี้ยประกันภัยรถยนต์ (Premium)
- จำนวนความเสียหายที่เกิดขึ้นใน 1 ปี (Incurred Loss)
- เบี้ยประกันส่วนลด (NCB)

ตารางที่ 3 แสดงการเปรียบเทียบการประเมินประสิทธิภาพเมื่อแบ่งข้อมูลชุดฝึกฝนและชุดทดสอบด้วยสัดส่วนต่าง ๆ และระหว่างตัวแบบป่าสุ่มกับตัวแบบบูสต์ติงปรับได้ โดยใช้ค่าความ

แม่นยำและค่าประสิทธิภาพโดยรวม สำหรับข้อมูลที่ได้จากบริษัทประกันภัยแห่งนี้จะเห็นว่าเมื่อสัดส่วนการแบ่งข้อมูลชุดฝึกฝนและชุดทดสอบเท่ากับ 85:15 ให้ค่าความแม่นยำสูงสุดคือ ร้อยละ 65.97 และร้อยละ 66.90 สำหรับตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้ตามลำดับ โดยมีค่าประสิทธิภาพโดยรวมคือ ร้อยละ 77.54 และร้อยละ 78.44 สำหรับตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้ตามลำดับ และเมื่อเปรียบเทียบระหว่างตัวแบบป่าสุ่มกับตัวแบบบูสต์ติงปรับได้จะเห็นว่า ตัวแบบบูสต์ติงปรับได้มีทั้งค่าความแม่นยำและค่าประสิทธิภาพโดยรวมสูงกว่าตัวแบบป่าสุ่ม อย่างไรก็ตามค่าความแม่นยำและค่าประสิทธิภาพโดยรวมของตัวแบบบูสต์ติงปรับได้สูงกว่าค่าความแม่นยำและค่าประสิทธิภาพโดยรวมของตัวแบบป่าสุ่มเล็กน้อยตารางที่ 4 แสดงความสำคัญของตัวแปรอิสระในตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้ สำหรับตัวแบบป่าสุ่ม ตัวแปรเบี้ยประกันส่วนลด (NCB) มีค่าความแม่นยำลดลงเฉลี่ยมากที่สุดเท่ากับ 40.62 หมายความว่าตัวแปรเบี้ยประกันส่วนลดเป็นตัวแปรอิสระที่มีความสำคัญมากที่สุดของตัวแบบ เมื่อนำตัวแปรนี้ออกจากตัวแบบจะทำให้ให้ตัวแบบป่าสุ่มมีความแม่นยำลดลงถึงร้อยละ 40.62 รองลงมาได้แก่ตัวแปรจำนวนความเสียหายที่เกิดขึ้นใน 1 ปี (Incurred Loss) และเบี้ยประกันภัยรถยนต์ (Premium) มีค่าความแม่นยำลดลงเฉลี่ยเท่ากับ 37.62 และ 36.18 ตามลำดับ

สำหรับตัวแบบบูสต์ติงปรับได้ ตัวแปรเบี้ยประกันภัยรถยนต์ (Premium) มีค่าความถี่ในการเรียกใช้เพื่อปรับค่าถ่วงน้ำหนักเพื่อให้ผลการทำนายที่ดีขึ้นมากที่สุดเท่ากับ 35 ครั้งจาก 50 ครั้ง รองลงมาได้แก่ ตัวแปรจำนวนความเสียหายที่เกิดขึ้นใน 1 ปี (Incurred Loss) และตัวแปรเบี้ยประกันส่วนลด (NCB) มีค่าความถี่ในการเรียกใช้ 34 และ 29 ครั้งตามลำดับ

Table 3 Accuracy and overall efficiency of random forest model and adaptive boosting model classified by proportion of training and testing dataset

Proportion of training and testing dataset	Accuracy (%)		Overall efficiency (%)	
	Random forest model	Adaptive boosting model	Random forest model	Adaptive boosting model
85:15	65.97	66.90	77.54	78.44
80:20	64.80	66.47	76.53	77.98
75:25	64.66	65.57	76.39	77.40
70:30	64.60	65.80	76.18	77.35
65:35	64.40	65.87	75.98	77.43
60:40	64.70	65.83	76.20	77.23
55:45	65.50	66.33	76.86	77.64
50:50	65.60	66.27	76.94	77.64

Table 4 Variable importance ranking for random forest model with mean decrease accuracy in parentheses and adaptive boosting model with its' frequency in parentheses

Variable important rank	Random forest model	Adaptive boosting model
1.	NCB (40.62)	Premium (35)
2.	Incurred loss (37.62)	Incurred loss (34)
3	Premium (36.18)	NCB (29)
4	No. of claim (27.04)	Car year (14)
5	Car year (23.08)	No. of claim (1)

ให้สังเกตว่าตัวแปรจำนวนความเสียหายที่เกิดขึ้นใน 1 ปีมีค่าความถี่ในการเรียกใช้ต่ำกว่าตัวแปรเบี้ยประกันภัยรถยนต์เพียงเล็กน้อย

4. สรุป

งานวิจัยครั้งนี้ใช้ข้อมูลทุติยภูมิจากบริษัทประกันภัยแห่งหนึ่งในพ.ศ. 2561 และ พ.ศ. 2562 จำนวน 10,000 ชุด และมีจำนวนตัวแปรทั้งหมด 8 ตัวแปร ตัวแปรตอบสนองที่สนใจในการศึกษา คือ สถานะการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 แบ่งเป็นการต่อหรือไม่ต่ออายุกรมธรรม์ตัวแปรอิสระ ได้แก่ เพศของผู้เอาประกัน, เบี้ยประกันภัยรถยนต์, อายุรถยนต์ที่รับประกันภัย, จำนวนการเคลมใน 1 ปี, จำนวนความเสียหายที่เกิดขึ้นใน 1 ปี, เบี้ยประกันส่วนลด และส่วนเพิ่มเบี้ยประกันเมื่อเกิดความเสียหายเกินขั้นต่ำ ใช้โปรแกรม RStudio Version 1.2.5033 แพคเกจ rattle ในการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1

จากการตรวจสอบข้อมูลเบื้องต้น พบว่าตัวแปรจำนวนความเสียหายที่เกิดขึ้นใน 1 ปี ของบางรายมีค่าไม่ถูกต้องและมีความสัมพันธ์ไม่ตรงกับความ เป็นจริงจำนวน 15 ราย ผู้วิจัยจึงได้ตัดชุดข้อมูลนี้ออกไป ทำให้มีข้อมูลที่นำไปวิเคราะห์ในขั้นต่อไปจำนวน 9,985 ชุด

ในการเปรียบเทียบประสิทธิภาพระหว่างตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้เมื่อสัดส่วนในการแบ่งข้อมูลชุดฝึกฝนและชุดทดสอบเป็น 85:15, 80:20, 75:25, 70:30, 65:35, 60:40, 55:45 และ 50:50 พบว่าตัวแบบป่าสุ่มและตัวแบบบูสต์ติงปรับได้เมื่อแบ่งชุดข้อมูลฝึกฝนและชุดทดสอบในสัดส่วน 85:15 มีค่าความแม่นยำ และค่าประสิทธิภาพโดยรวมมากกว่าการแบ่งชุดข้อมูลฝึกฝนและชุดทดสอบด้วยสัดส่วนอื่น ๆ กล่าวคือมีค่าความแม่นยำร้อยละ 66.90 และค่าประสิทธิภาพโดยรวมร้อยละ 78.44 ดังนั้น

ผู้วิจัยจึงใช้ตัวแบบนี้เป็นหลักในการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 โดยได้ลองใช้ตัวแบบป่าสุ่มเมื่อแบ่งชุดข้อมูลฝึกฝนและชุดทดสอบในสัดส่วน 85:15 เช่นเดียวกันในการทำนายการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 เพื่อจะได้พิจารณาว่าตัวแปรที่สำคัญในการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ของบริษัทประกันภัยแห่งนี้เป็นตัวแปรชุดเดียวกันหรือไม่

เมื่อใช้ตัวแบบบูสต์ติงปรับได้ในการทำนายการต่ออายุกรมธรรม์ประกันภัยรถยนต์ประเภท 1 ของบริษัทประกันภัยแห่งนี้ พบว่าตัวแปรอิสระที่มีความสำคัญลำดับแรกคือตัวแปรเบี้ยประกันภัยรถยนต์ และ ตัวแปรจำนวนความเสียหายที่เกิดขึ้นใน 1 ปี ลำดับต่อไปคือตัวแปรเบี้ยประกันส่วนลดตัวแปรอิสระสามตัวนี้สอดคล้องกับการใช้ตัวแบบป่าสุ่มในการพิจารณาความสำคัญของตัวแปรอิสระ แม้ว่าลำดับที่ของความสำคัญตัวแปรทั้งสามตัวจะต่างกันเล็กน้อยเมื่อใช้ตัวแบบป่าสุ่ม

5. กิตติกรรมประกาศ

งานวิจัยนี้มีความสมบูรณ์ยิ่งขึ้นโดยได้รับความอนุเคราะห์จาก ดร.ชัยณรงค์ เกษามูล อาจารย์ประจำสาขาวิชาคณิตศาสตร์และสถิติ มหาวิทยาลัยธรรมศาสตร์ และดร.อรรวรรณ สุภาพิ่ง นักวิจัยหน่วยระบอดวิทยาคลินิก คณะแพทยศาสตร์ศิริราชพยาบาล มหาวิทยาลัยมหิดล

6. References

Department of Land Transport, 2020. The number of vehicles registered collectors, Available Source: <https://www.dlt.go.th/>, February 12, 2020. (in Thai)

Thai SRC, 2017, The information of car accident, Available Source: <http://www.thaisrc.com>, February 12, 2020. (in Thai)

Laisue Panyadee, Poom Jewprasertpan and Nutthaphong Pengplian, 2020, A Comparative Study of Partition Ratio between Training to Testing Dataset and Study Factors Affecting the Renewal of a

Hospital Benefit Rider in Random Forest Model, Senior Project, Actuarial Science program, Department of Mathematics and Statistics, 64 p. (in Thai)

Sirawich Jaichuen, 2019. Adaboost Algorithm, Available Source: <https://www.sirawichjaichuen.medium.com/adaboost-algorithm-cfe6b58e60fa>, October 26, 2020. (in Thai)