

ST-16

วิธีการทำนายด้วยการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจสำหรับการซื้อแบบประกันชีวิต
A Prediction Method with Machine Learning Based on Decision Tree for
Life Insurance Purchase

นฤมล เพชรแท้¹, เตชต์ฐสิณป์ เพี้ยซ้าย² และณัฐพร เห็นเจริญเลิศ³
Narumon Phetthae¹, Tejtasin Phiasai² and Nuttaporn Hencharoenlert³
^{1,2,3}สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช
อีเมล: Narumon.phe@stou.ac.th

บทคัดย่อ

บทความนี้เสนอวิธีการทำนายด้วยการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจสำหรับการซื้อแบบประกันชีวิต เพื่อพัฒนาวิธีการทำนายเพื่อจำแนกประเภทการซื้อแบบประกันชีวิต และเพื่อประเมินประสิทธิภาพของวิธีการทำนายเพื่อจำแนกประเภทการซื้อแบบประกันชีวิต โดยพิจารณาจากตัวแปรอิสระ 4 ตัว ได้แก่ เพศ ชั้นอาชีพ ฐานเงินเดือน และสถานภาพ โดยผลการทดลองจากลูกค้าทั้งหมดที่เลือกทำประกันภัยสามารถสรุปได้ว่า วิธีนาอีฟเบย์ (Naïve Bayes) มีค่าเท่ากับร้อยละความถูกต้อง 78.64 วิธีซัพพอร์ตเวกเตอร์แมทชีน (Support Vector Machine) มีค่าร้อยละความถูกต้อง 80.96 วิธีโครงข่ายประสาทเทียม (Artificial Neural Network) มีค่าร้อยละความถูกต้อง 81.77 และวิธีต้นไม้ตัดสินใจ (Decision Tree) มีค่าร้อยละความถูกต้อง 81.81 โดยพบว่าเทคนิคการทำนายเพื่อจำแนกประเภท วิธีการเรียนรู้ของเครื่องแบบต้นไม้ตัดสินใจเป็นวิธีที่มีประสิทธิภาพมากที่สุดในการจำแนกประเภทของการซื้อแบบประกันชีวิต โดยให้ค่าความถูกต้องร้อยละ 81.81

คำหลัก: การทำนาย, แบบประกัน, ประกันชีวิต

Abstract

This paper proposes a prediction method with machine learning based on a decision tree for life insurance purchases to develop the predictive method for the classification of the life insurance purchase and to evaluate the effectiveness of the prediction method for the classification of the life insurance purchase. The proposed prediction method considered 4 independent variables, such as gender, occupational level, salary base, and status. The experimental results for all customers who chose to purchase the life insurance showed that the method of Naïve Bayes received an accuracy of 78.64 percent, the method of support vector machine received the accuracy of 80.96 percent, the method of the artificial neural network received the accuracy of 81.77 percent, and the method of decision tree received the accuracy of 81.81 percent. For all results, the



prediction method with machine learning based on decision tree for the classification of the life insurance purchase received the maximum effectiveness with the accuracy of 81.81.

Keywords: Predicting, Insurance, Life Insurance

บทนำ

การซื้อแบบประกันชีวิตในปัจจุบันถือเป็นเรื่องที่สำคัญ และอัตราการทำประกันชีวิตในปัจจุบันมีแนวโน้มที่จะเพิ่มขึ้นเรื่อย ๆ ไม่ว่าจะด้วยวิธีการเสนอขายแบบใดก็ตาม แต่เนื่องจากตัวผลิตภัณฑ์ประกันชีวิตในปัจจุบันมีมากมายหลายรูปแบบให้เลือกซื้อ ซึ่งแต่ละแบบจะมีลักษณะการคุ้มครองและผลประโยชน์ต่างกันไป เช่น แบบตลอดชีพ แบบชั่วระยะเวลา แบบสะสมทรัพย์ และแบบควบการลงทุน แบบประกันแต่ละแบบ แต่ละประเภท มีข้อดี-ข้อเสียที่แตกต่างกัน ทั้งนี้ผู้คนทั่วไปไม่ได้มีความรู้และไม่ได้ทำการศึกษารูปแบบประกันแต่ละประเภทนี้ก่อนตัดสินใจเลือกซื้อ ทำให้ผู้คนไม่ทราบว่าตัวเองเหมาะกับแบบประกันประเภทใด ควรเลือกซื้อและลงรายละเอียดแบบใดจึงจะเหมาะสม และเกิดประโยชน์สูงสุด

คุณ รุจิรา (2555) ใช้เทคนิคต้นไม้ตัดสินใจเป็นแนวทางในการตัดสินใจเลือกใช้แพคเกจ อินเทอร์เน็ตมือถือจากพฤติกรรมของผู้ใช้บริการแต่ละคน โดยสร้างตัวแบบจากอัลกอริทึมต้นไม้ตัดสินใจ คือ ID3 และ C4.5 (148) เพื่อตรวจสอบความถูกต้องของตัวแบบทั้ง 2 สมการที่ได้สร้างขึ้นมา ซึ่งผลที่ได้คือตัวแบบที่ถูกสร้างจากอัลกอริทึมต้นไม้ตัดสินใจ ID3 มีค่าความถูกต้องมากกว่า C4.5 (148)

คุณ ธนธร (2559) ใช้สถิติการวิเคราะห์ความถดถอยลอจิสติกอนอกนาม (Multinomial Logistic Regression) ในการศึกษาวิเคราะห์ปัจจัยด้านองค์ประกอบเว็บไซต์ร้านค้าออนไลน์ที่ส่งผลต่อ พฤติกรรมการทำธุรกรรมในธุรกิจซื้อขายออนไลน์ของนักศึกษาระดับปริญญาตรี มหาวิทยาลัยใน จังหวัดนครปฐม โดยสร้างตัวแบบ 3 สมการ ได้แก่ ตัวแบบพยากรณ์ความถี่ในการทำธุรกรรมใน ธุรกิจซื้อขายออนไลน์, ตัวแบบพยากรณ์พฤติกรรมการชำระเงินในการทำธุรกรรมใน ธุรกิจซื้อขายออนไลน์ และตัวแบบพยากรณ์พฤติกรรมการประเมินสินค้าที่ซื้อในธุรกิจซื้อขาย ออนไลน์ โดยได้ประสิทธิภาพในการพยากรณ์ความถูกต้องเป็นร้อยละ 42.50, 50.50 และ 44.30 ตามลำดับ

คุณจามรี และ คุณสมศรี (2560) สร้างตัวแบบเพื่อพยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ใน กระบวนการทดสอบความเชื่อมั่นก่อนการส่งมอบแก่ผู้บริโภค ตัวแบบที่ใช้ในการพยากรณ์คุณภาพของผลิตภัณฑ์ ฮาร์ดดิสก์ที่พิจารณา มี 2 แบบ คือ การวิเคราะห์การถดถอยลอจิสติกและโครงข่ายประสาทเทียม จากงานวิจัยพบว่าตัวแบบที่ได้จากโครงข่ายประสาทเทียมให้ ความถูกต้องในการ พยากรณ์คุณภาพของผลิตภัณฑ์ฮาร์ดดิสก์ได้ดีกว่าตัวแบบการถดถอยลอจิสติก

คุณวิชชกานต์ และคณะ (2560) ได้ทำการสร้างและเปรียบเทียบตัวแบบสัญญาณเตือนภัย ล่วงหน้าทางการเงินของบริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย ใช้ วิธีการวิเคราะห์การถดถอยลอจิสติกกับวิธี โครงข่ายประสาทเทียม ในการวิเคราะห์ข้อมูลจาก การศึกษาด้วยวิธีการวิเคราะห์การถดถอยลอจิสติก พบว่าวิธี โครงข่ายประสาทเทียมมีประสิทธิภาพในการจำแนกประเภทดีกว่าวิธีการวิเคราะห์การ ถดถอยลอจิสติก

คุณปานทิพย์ พุ่มพุด (2561) ได้ศึกษาเรื่องการทำนายผลการรีวิวกัดในโครงการโอเพนซอร์สบนกิตฮับ โดยการ วิเคราะห์ ข้อมูลบนกิตฮับ กำหนดตัวแปรที่คาดว่าจะมีผลกระทบกับการรีวิวกัด รวมถึงการหา ความสัมพันธ์ที่เกิดขึ้นบ่อยจากกฎ



ความสัมพันธ์ จากนั้นทำการสร้างแบบจำลองการทำนายการรีวิว โค้ดบนกิตฮับด้วยวิธีการทางสถิติโลจิสติก และหลักการเรียนรู้ด้วยเครื่อง การวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น ผลการทดลองแสดงให้เห็นว่าแบบจำลองที่ใช้ในการทำนายผลการโดยวิเคราะห์โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น มีเปอร์เซ็นต์ความแม่นยำมากกว่าการวิเคราะห์โลจิสติก

ชนิตา หางแก้ว (2563) ได้ศึกษาเรื่องการทำนายการเลือกประเภทประกันภัยของลูกค้า มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อการตัดสินใจการเลือกทำประเภทประกันภัยของลูกค้า และ เปรียบเทียบประสิทธิภาพเทคนิคการวิเคราะห์จำแนกกลุ่ม 4 วิธี คือ วิธีการวิเคราะห์การถดถอยโลจิสติกพหุ วิธีการวิเคราะห์จำแนกกลุ่ม วิธีต้นไม้ตัดสินใจ และวิธี โครงข่ายประสาทเทียม โดยพิจารณาจากตัวแปรอิสระ 6 ตัว ผลการวิจัยสรุปได้การเปรียบเทียบประสิทธิภาพเทคนิคการวิเคราะห์จำแนก กลุ่มทั้ง 4 วิธี พบว่าวิธีโครงข่ายประสาทเทียมเป็นวิธีที่มีประสิทธิภาพในการจำแนกกลุ่มมากที่สุด

คุณ วีระพันธ์ พานิชย์ (2563) ได้ศึกษาเรื่องการใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา โดนทำการทดสอบและเปรียบเทียบประสิทธิภาพโมเดลระหว่างวิธี Decision Tree, Naive Bayes , Neural Network และ Support Vector Machine ผลการวิจัยพบว่าประสิทธิภาพทำนายด้วยวิธี Neural Network มีค่าความถูกต้อง (Accuracy) มากที่สุด

Weerasinghe and Wijegunasekara (2016) ได้ทำการสร้างตัวแบบทำนายค่าสินไหม ทดแทนของประกันรถยนต์ โดยทำการเปรียบเทียบประสิทธิภาพในการทำนายระหว่างวิธี โครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ และวิธีการวิเคราะห์ การถดถอยโลจิสติกอนเคนนาม พบว่าวิธี โครงข่ายประสาทเทียมให้ประสิทธิภาพดีที่สุด

จากประเด็นงานวิจัยที่ได้ศึกษาทั้งหมดพบว่าการเปรียบเทียบวิธีการทำนายที่คล้ายคลึงกัน แต่ท้ายที่สุดผลลัพธ์ของงานวิจัยออกมาแตกต่างกัน อาจจะเป็นด้วยความเหมาะสมและสอดคล้องกันของข้อมูลและวิธีการทำนาย ดังนั้นผู้เขียนจึงต้องการจะทดสอบเปรียบเทียบเพื่อหาวิธีการทำนายที่เหมาะสมกับข้อมูลของแบบประกัน และ ให้ค่าความถูกต้องแม่นยำที่สูงขึ้น

โดยบทความนี้ได้แบ่งการอธิบายเป็น 5 ส่วนได้แก่ ส่วนวัตถุประสงค์ ส่วนวิธีดำเนินการวิจัยส่วนผลการวิจัย ส่วนสรุปและอภิปรายผลการวิจัย และส่วนข้อเสนอแนะ

วัตถุประสงค์

1. เพื่อประยุกต์วิธีการทำนายสำหรับการจำแนกการซื้อแบบประกันชีวิต
2. เพื่อประเมินประสิทธิภาพของวิธีการทำนายเพื่อจำแนกการซื้อแบบประกันชีวิต

วิธีดำเนินการวิจัย

งานวิจัยนี้ผู้วิจัยได้ใช้กระบวนการทำเหมืองข้อมูล (Data Mining) โดยใช้เทคนิค Machine Learning เพื่อการทำนายการเลือกซื้อแบบประกันชีวิต ซึ่งการศึกษาวิธีการทำนายด้วยการเรียนรู้ของเครื่องสำหรับแนะนำการซื้อแบบประกันชีวิตนั้น จะเป็นข้อมูลให้แก่บริษัทประกันได้ใช้ประโยชน์ในการเลือกผลิตภัณฑ์ประกันชีวิตในแต่ละประเภทได้อย่างมีประสิทธิภาพ รวมถึงทราบปัจจัยสำคัญที่มีอิทธิพลต่อการตัดสินใจเลือกทำแบบประกันชีวิตของลูกค้า

1. จัดเตรียมข้อมูล

โดยมีตัวแปรอิสระ 4 ตัว คือ เพศ แบ่งเป็น 2 กลุ่ม คือเพศชายและเพศหญิง ชั้นอาชีพ แบ่งเป็น 4 กลุ่มคือ ชั้นอาชีพ 1,2,3 และ 4 ฐานเงินเดือน แบ่งเป็น 4กลุ่ม 15,000-30,000 30,000-40,000 40,000-80,000 และ80,000 ขึ้นไป สถานภาพ แบ่งเป็น 3 กลุ่ม โสด สมรส และหม้าย/หย่า ตัวแปรตาม 4 ตัว คือ แบบชั่วคราวระยะเวลายาว แบบตลอดชีพ แบบสะสมทรัพย์ และแบบควบการลงทุน

2. การวิเคราะห์ข้อมูล

ทำการวิเคราะห์ข้อมูล โดยวิธีการวิเคราะห์ข้อมูลที่ใช้ในงานวิจัยนี้ประกอบด้วย 4 วิธี คือ

2.1 วิธีการวิเคราะห์หน้าอ็อปเบย์ ขั้นตอนการสร้างตัวแบบมีดังนี้

2.1.1 นำข้อมูลที่ได้ทำการจัดเตรียมไว้เป็นชุดข้อมูลฝึกฝนเพื่อให้ระบบใช้เรียนรู้มาจัดเก็บลงฐานข้อมูลเพื่อสร้างตัวแบบโดยใช้โปรแกรม Weka

2.1.2 นำเข้าข้อมูลที่ได้จัดเตรียมไว้ในขั้นตอนก่อนหน้ามาสร้างการแบ่งประเภทโดยใช้อัลกอริทึมที่ได้เลือกไว้คือ Naïve Baye ในโปรแกรม Weka

2.1.3 ทำการทดสอบประสิทธิภาพความถูกต้องในการจำแนกประเภทข้อมูลของตัวแบบ โดยใช้ค่าความแม่นยำ

2.1.4 เปรียบเทียบประสิทธิภาพความถูกต้องในการวิเคราะห์หรือจำแนกประเภทข้อมูลของแต่ละวิธี โดยใช้ค่าความแม่นยำเป็นเครื่องมือในการวัดประสิทธิภาพ

2.2 วิธีการวิเคราะห์ซัพพอร์ตเวกเตอร์แมทซิน ขั้นตอนการสร้างตัวแบบมีดังนี้

2.2.1 นำข้อมูลที่ได้ทำการจัดเตรียมไว้มาจัดเก็บลงฐานข้อมูลเพื่อเตรียมสร้างตัวแบบในโปรแกรมเช่นเดียวกับในวิธีการก่อนหน้า

2.2.2 สร้างตัวแบบจากชุดข้อมูลฝึกฝน ที่ได้เลือกไว้คือ SMO ในโปรแกรม Weka

2.2.3 ทำการทดสอบประสิทธิภาพความถูกต้องในการจำแนกประเภทข้อมูลของตัวแบบโดยใช้ค่าความแม่นยำ

2.3 วิธีการวิเคราะห์ต้นไม้ตัดสินใจ ขั้นตอนการสร้างตัวแบบมีดังนี้

2.3.1 นำข้อมูลที่ได้ทำการจัดเตรียมไว้เป็นชุดข้อมูลฝึกฝนเพื่อให้ระบบใช้เรียนรู้มาจัดเก็บ ลงฐานข้อมูลเพื่อสร้างตัวแบบโดยใช้โปรแกรม Weka

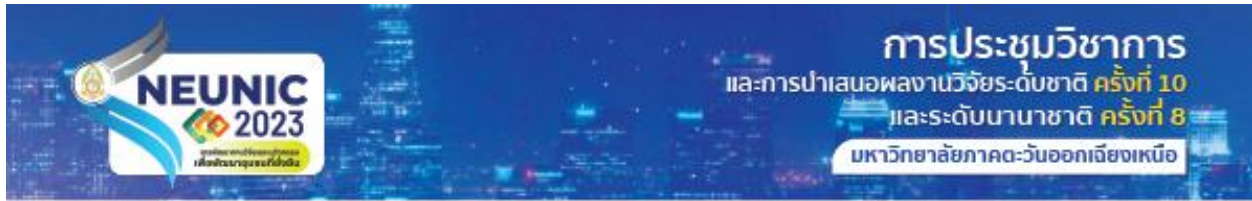
2.3.2 นำเข้าข้อมูลที่ได้จัดเตรียมไว้ในขั้นตอนก่อนหน้ามาสร้างการแบ่งประเภทโดยใช้ อัลกอริทึมที่ได้เลือกไว้คือ C4.5(J48) ในโปรแกรม Weka ซึ่งผลลัพธ์ที่ได้เมื่อโปรแกรมได้ทำการสร้างรูปแบบคือจะแสดงรูปของต้นไม้ (Tree) ที่ได้จากการประมวลผล

2.3.3 ทำการทดสอบประสิทธิภาพความถูกต้องในการจำแนกประเภทข้อมูลของตัวแบบ โดยใช้ค่าความแม่นยำ

2.4 วิธีโครงข่ายประสาทเทียม ขั้นตอนการสร้างตัวแบบมีดังนี้

2.4.1 นำข้อมูลที่ได้ทำการจัดเตรียมไว้มาจัดเก็บลงฐานข้อมูลเพื่อเตรียมสร้างตัวแบบในโปรแกรมเช่นเดียวกับในวิธีการก่อนหน้า

2.4.2 สร้างตัวแบบจากชุดข้อมูลฝึกฝน โดยสร้างโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบหลายชั้น (Multilayer Perceptron) และใช้อัลกอริทึมแบบแพร่กลับ (Back Propagation) เพื่อหาค่าน้ำหนักที่ดีที่สุดของเส้นเชื่อมโยงแต่ละเส้น



2.4.3 ทำการทดสอบประสิทธิภาพความถูกต้องในการจำแนกประเภท ข้อมูลของตัวแบบ โดยใช้ค่าความถูกต้อง

2.4.4 เปรียบเทียบประสิทธิภาพความถูกต้องในการวิเคราะห์หรือจำแนกประเภทข้อมูลของแต่ละวิธี โดยใช้ค่าความถูกต้องเป็นเครื่องมือในการวัดประสิทธิภาพ

ผลการวิจัย

ผลการวิจัย ผู้วิจัยเสนอผลการวิเคราะห์ข้อมูล แบ่งเป็น 5 ตอน ดังนี้ ตอนที่ 1 ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีนาอิวเบย์ (Naïve Bayes) ตอนที่ 2 ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ตอนที่ 3 ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) ตอนที่ 4 ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีโครงข่ายประสาทเทียม (Artificial Neural Network) ตอนที่ 5 การเปรียบเทียบประสิทธิภาพการทำนายการเลือกซื้อประกันชีวิต ทั้ง 4 วิธี

1. ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีนาอิวเบย์ (Naïve Bayes)

ในขั้นตอนการเตรียมและคัดเลือกข้อมูลที่จะนำมาใช้นั้น เนื่องจากข้อมูลที่เราได้ทำการรวบรวมมานั้นได้จัดเก็บไว้ในรูปแบบที่ไม่สามารถนำมาใช้ในการทดสอบวิเคราะห์ด้วยเครื่องมือการทำเหมืองข้อมูลได้ ดังนั้น ทำการโอนย้ายข้อมูล (Data Transfer) โดยการโอนข้อมูลเดิมที่ถูกจัดเก็บอยู่ในรูปแบบไฟล์ .XLSX ไปเป็นฐานข้อมูลในรูปแบบไฟล์ .CSV เพื่อให้สามารถนำไปใช้งานในโปรแกรม Weka ได้และจากข้อมูลที่เราได้ทำการรวบรวมมาเพื่อวิเคราะห์ในการเลือกทำประเภทประกันภัยของลูกค้าประกัน มีทั้งหมด 5 แอทริบิวท์ ซึ่งมีความสำคัญต่อการวิเคราะห์ทั้งหมด จากการนำข้อมูลทั้งหมดมาใช้สร้างตัวแบบนาอิวเบย์ในโปรแกรม Weka ผลจากการทดสอบประสิทธิภาพความถูกต้องในการจำแนกประเภทข้อมูลของตัวแบบได้ผลการวิเคราะห์ ดังตารางที่ 1

ตารางที่ 1 ผลการวิเคราะห์ที่แสดงค่าความถูกต้องของวิธีนาอิวเบย์โดยใช้โปรแกรม Weka

การวิเคราะห์ด้วยวิธีนาอิวเบย์	ผลการวิเคราะห์
ค่าความถูกต้องของการทำนาย	78.64%
ค่าการผิดพลาดของการทำนาย	21.36%
ค่าสถิติแคปปา	0.6867
รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย	0.2766

ในการวัดประสิทธิภาพของตัวแบบสามารถพิจารณาผลต่าง ๆ ได้ดังนี้ ค่าความถูกต้องหรือค่าความถูกต้องของการทำนายการเลือกทำประเภทประกันภัยของลูกค้ามีค่าเท่ากับร้อยละ 78.64 ค่าความผิดพลาดของการทำนายการเลือกทำประเภทประกันภัยของลูกค้ามีค่าเท่ากับร้อยละ 21.36 สถิติแคปปา (Kappa Statistic) หรือ ค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติ ความสอดคล้องของการทำนายเลือกทำประเภทประกันภัยของลูกค้า มีค่าเท่ากับ 0.6867 ซึ่งค่าสถิติแคปปา จะมีค่าอยู่ในช่วง 0.000 -1.000 นั่นคือ ข้อมูลค่อนข้างมีความสอดคล้องกัน รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error :



RMSE) ของการเลือกทำประเภทร่วมกันภัยของลูกค้า มีค่าเท่ากับ 0.2766 ซึ่งเป็นค่าที่ใช้ตรวจสอบความเคลื่อนไหวในการทำนายยังมีค่าน้อยยิ่งดี และจากค่าที่ได้แสดงว่า ตัวแบบที่ได้นั้นมีความถูกต้องดี

2. ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

ในขั้นตอนการเตรียมและคัดเลือกข้อมูลที่จะนำมาใช้นั้น จะนำข้อมูลที่ได้ทำการเตรียมและคัดเลือกไว้ในขั้นตอนวิธีนาอ็อปเบย์ก่อนหน้า มาใช้ในการวิเคราะห์ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีนนี้ จากการนำข้อมูลทั้งหมดที่มาสกัดตัวแบบในโปรแกรม Weka โดยใช้อัลกอริทึมชนิด SMO ซึ่งได้ผลการวิเคราะห์ ดังตารางที่ 2

ตารางที่ 2 ผลการวิเคราะห์แสดงค่าความถูกต้องของวิธีซัพพอร์ตเวกเตอร์แมชชีนโดยใช้โปรแกรม Weka

การวิเคราะห์ด้วยวิธีซัพพอร์ตเวกเตอร์แมชชีน	ผลการวิเคราะห์
ค่าความถูกต้องของการทำนาย	80.96%
ค่าการผิดพลาดของการทำนาย	19.04%
ค่าสถิติแคปปา	0.723
รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย	0.3397

ในการวัดประสิทธิภาพของตัวแบบสามารถพิจารณาผลต่าง ๆ ได้ดังนี้ ค่าความถูกต้องหรือค่าความถูกต้องของการทำนายการเลือกทำประเภทร่วมกันภัยของลูกค้ามีค่าเท่ากับร้อยละ 80.96 ค่าความผิดพลาดของการทำนายการเลือกทำประเภทร่วมกันภัยของลูกค้ามีค่าเท่ากับร้อยละ 19.04 สถิติแคปปา (Kappa Statistic) หรือ ค่าสัมประสิทธิ์ตัวชี้วัดทางสถิติความสอดคล้องของการทำนายเลือกทำประเภทร่วมกันภัยของลูกค้า มีค่าเท่ากับ 0.723 ซึ่งค่าสถิติแคปปา จะมีค่าอยู่ในช่วง 0.000 -1.000 นั่นคือ ข้อมูลค่อนข้างมีความสอดคล้องกัน รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) ของการเลือกทำประเภทร่วมกันภัยของลูกค้า มีค่าเท่ากับ 0.3397 ซึ่งเป็นค่าที่ใช้ตรวจสอบความเคลื่อนไหวในการทำนายยังมีค่าน้อยยิ่งดี และจากค่าที่ได้แสดงว่า ตัวแบบที่ได้นั้นมีความถูกต้องดี

3. ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีต้นไม้ตัดสินใจ

ในขั้นตอนการเตรียมและคัดเลือกข้อมูลที่จะนำมาใช้นั้น จะนำข้อมูลที่ได้ทำการเตรียมและคัดเลือกไว้ในขั้นตอนก่อนหน้า มาใช้สร้างตัวแบบต้นไม้ตัดสินใจด้วยอัลกอริทึม C4.5 (J48) ในโปรแกรม Weka ผลจากการทดสอบประสิทธิภาพความถูกต้องในการจำแนกประเภทข้อมูลของตัวแบบได้ผลการวิเคราะห์ ดังตารางที่ 3



ตารางที่ 3 ผลการวิเคราะห์แสดงค่าความถูกต้องของวิธีต้นไม้ตัดสินใจโดยใช้โปรแกรม Weka

การวิเคราะห์ด้วยวิธีต้นไม้ตัดสินใจ	ผลการวิเคราะห์
ค่าความถูกต้องของการทำนาย	81.81%
ค่าการผิดพลาดของการทำนาย	18.19%
ค่าสถิติแคปปา	0.7299
รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย	0.2632

ในการวัดประสิทธิภาพของตัวแบบสามารถพิจารณาผลต่าง ๆ ได้ดังนี้ ค่าความถูกต้องหรือค่าความถูกต้องของการทำนายการเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับร้อยละ 81.81 ค่าความผิดพลาดของการทำนายการเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับร้อยละ 18.19 สถิติแคปปา (Kappa Statistic) หรือ ค่าสัมประสิทธิ์ตัววัดทางสถิติ ความสอดคล้องของการทำนายเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับ 0.7299 ซึ่งค่าสถิติแคปปา จะมีค่าอยู่ในช่วง 0.000 -1.000 นั่นคือ ข้อมูลก่อนข้างมีความสอดคล้องกัน รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) ของการเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับ 0.2632 ซึ่งเป็นค่าที่ใช้ตรวจสอบความเคลื่อนในการทำนายยังมีค่าน้อยยิ่งดี และจากค่าที่ได้แสดงว่า ตัวแบบที่ได้ั้นมีความถูกต้องดี

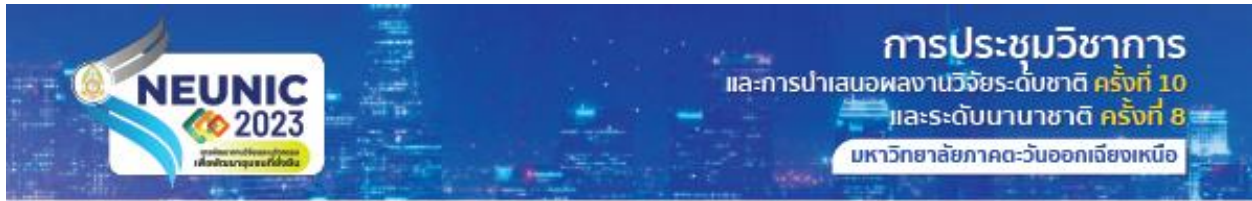
4. ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีโครงข่ายประสาทเทียม

ในขั้นตอนการเตรียมและคัดเลือกข้อมูลที่จะนำมาใช้นั้น จะนำข้อมูลที่ได้ทำการเตรียมและคัดเลือกไว้ในขั้นตอนวิธีต้นไม้ตัดสินใจก่อนหน้านี้ มาใช้ในการวิเคราะห์ด้วยวิธีโครงข่ายประสาทเทียมนี้เลย จากการนำข้อมูลทั้งหมดที่มาสสร้างตัวแบบในโปรแกรม Weka โดยใช้อัลกอริทึมชนิดเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron) ซึ่งจากการทำการทดลองหลาย ๆ ครั้ง พบว่าการกำหนดค่าอัตราการเรียนรู้ (Learning Rate) เป็น 0.3 ค่าโมเมนตัม (Momentum) เป็น 0.2 จำนวนรอบการสอน (Training Time) 500 รอบ และใช้โครงข่ายประสาทเทียมที่มีชั้นปกปิด (Hidden Layer) 3 ชั้น ซึ่งได้ผลการวิเคราะห์ ดังตารางที่ 4

ตารางที่ 4 ผลการวิเคราะห์แสดงค่าความถูกต้องของวิธีโครงข่ายประสาทเทียม

การวิเคราะห์ด้วยวิธีโครงข่ายประสาทเทียม	ผลการวิเคราะห์
ค่าความถูกต้องของการทำนาย	81.77%
ค่าการผิดพลาดของการทำนาย	18.23%
ค่าสถิติแคปปา	0.7297
รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย	0.2631

ในการวัดประสิทธิภาพของตัวแบบสามารถพิจารณาผลต่าง ๆ ได้ดังนี้ ค่าความถูกต้องหรือค่าความถูกต้องของการทำนายการเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับร้อยละ 81.77 ค่าความผิดพลาดของการทำนายการเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับร้อยละ 18.23 สถิติแคปปา (Kappa Statistic) หรือ ค่าสัมประสิทธิ์ตัววัดทางสถิติ ความสอดคล้องของการทำนายเลือกทำประเภทประกันภัยของลูกคามีค่าเท่ากับ 0.7297 ซึ่งค่าสถิติแคปปา จะมีค่าอยู่ในช่วง



0.000 -1.000 นั่นคือ ข้อมูลค่อนข้างมีความสอดคล้องกัน รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย (Root Mean Squared Error : RMSE) ของการเลือกทำประเภทประกันภัยของลูกค้า มีค่าเท่ากับ 0.2631 ซึ่งเป็นค่าที่ใช้ตรวจสอบความคลื่อนในการทำนายยังมีค่าน้อยยิ่งดี และจากค่าที่ได้แสดงว่า ตัวแบบที่ได้นั้นมีความถูกต้องดี

5. ผลการวิเคราะห์ประสิทธิภาพการจำแนกประเภทของตัวแปรตามด้วยวิธีนาอิวเบย์ (Naive Bayes)

จากเทคนิคการจำแนกประเภทของตัวแปรตามทั้ง 4 วิธี พบว่า วิธีการวิเคราะห์ต้นไม้ตัดสินใจเป็นวิธีที่มีประสิทธิภาพมากที่สุดในการจำแนกประเภท โดยให้ค่าความถูกต้องร้อยละ 81.81 รองลงมา คือ วิธีการวิเคราะห์โครงข่ายประสาทเทียม วิธีการวิเคราะห์ซัพพอร์ตเวกเตอร์แมชชีน และวิธีการวิเคราะห์นาอิวเบย์ ตามลำดับ ซึ่งผลวิเคราะห์เป็นไปดังตารางที่ 5

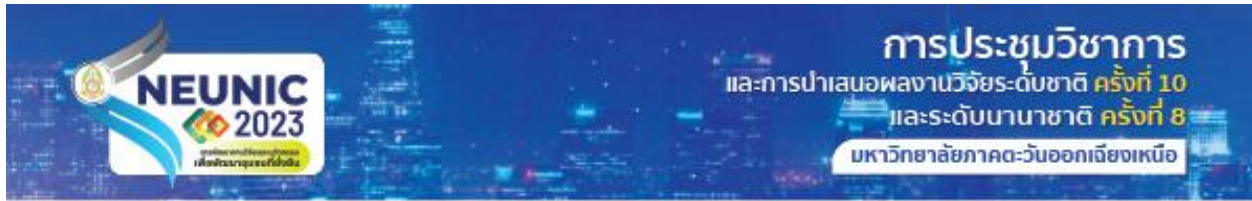
ตารางที่ 5 การเปรียบเทียบค่าความถูกต้อง (Accuracy) ของการทำนาย

เทคนิคการจำแนกประเภทของตัวแปรตาม	ค่าความถูกต้อง (Accuracy)
นาอิวเบย์	78.64%
ซัพพอร์ตเวกเตอร์แมชชีน	80.96%
ต้นไม้ตัดสินใจ	81.81%
โครงข่ายประสาทเทียม	81.77%

สรุปและอภิปรายผลการวิจัย

ในการเปรียบเทียบประสิทธิภาพในการทำนายการเลือกซื้อประเภทประกันของลูกค้า ทั้ง 4 วิธี พบว่า วิธีต้นไม้ตัดสินใจให้ค่าความถูกต้องในการทำนายการเลือกทำประเภทประกันชีวิตของลูกค้าได้ถูกต้องร้อยละ 81.81 รองลงมาคือวิธีโครงข่ายประสาทเทียม (Artificial Neural Network) ให้ค่าความถูกต้องในการทำนายถูกต้องร้อยละ 81.77 และวิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) วิธีนาอิวเบย์ (Naive Bayes) ให้ค่าความถูกต้องร้อยละ 80.96 และ 78.64 ตามลำดับ

ดังนั้น วิธีต้นไม้ตัดสินใจ มีประสิทธิภาพในการทำนายการเลือกซื้อ ประเภทประกันภัยของลูกค้ามากที่สุด เมื่อเปรียบเทียบกับงานวิจัยของ Weerasinghe and Wijegunasekara (2016) ที่ได้ทำการสร้างตัวแบบทำนายค่าสินไหมทดแทนของประกันรถยนต์ โดยทำการเปรียบเทียบประสิทธิภาพในการทำนายระหว่างวิธีโครงข่ายประสาทเทียม วิธีต้นไม้ตัดสินใจ และวิธีการวิเคราะห์การถดถอยลอจิสติกอเนกนาม พบว่าวิธี โครงข่ายประสาทเทียมให้ประสิทธิภาพดีที่สุด โดยให้ค่าความถูกต้องร้อยละ 61.71 รองลงมาคือวิธีต้นไม้ตัดสินใจที่ให้ค่าความถูกต้องร้อยละ 57.05 และสุดท้ายคือวิธีการวิเคราะห์การถดถอยลอจิสติกอเนกนาม ที่มีค่าความถูกต้องร้อยละ 52.39 และจากงานวิจัยของ ชนิตา หางแก้ว (2563) ได้ศึกษาเรื่องการทำนายการเลือกประเภทประกันภัยของลูกค้า มีวัตถุประสงค์เพื่อศึกษาปัจจัยที่มีผลต่อการตัดสินใจการเลือกทำประเภทประกันภัยของลูกค้าโดยพิจารณาจากตัวแปรอิสระ 6 ตัว ได้แก่ เพศ อายุ รายได้ต่อเดือน วุฒิกการศึกษา สถานภาพครอบครัว และอาชีพ ผลการวิจัยสรุปได้ว่า การเปรียบเทียบประสิทธิภาพเทคนิคการวิเคราะห์จำแนก กลุ่มทั้ง 4 วิธี พบว่า วิธีโครงข่ายประสาทเทียมเป็นวิธีที่มีประสิทธิภาพในการจำแนกกลุ่มมากที่สุด โดยให้ค่าความถูกต้องร้อยละ 85.75 รองลงมาคือ วิธีต้นไม้ตัดสินใจ วิธีการวิเคราะห์การถดถอยลอจิสติกพหุ และวิธีการวิเคราะห์จำแนกกลุ่ม โดยให้ค่าความถูกต้องร้อยละ 76.50, 70.00 และ 61.80



ตามลำดับ ทั้งนี้ในการวิจัยในครั้งถัดไป ควรเพิ่ม Attribute ที่เกี่ยวข้องกับลูกค้าที่เลือกซื้อประกันให้มากขึ้น เช่น จำนวนบุตร อายุ และ เพิ่มข้อมูลในการเรียนรู้ของเครื่องให้มากขึ้น โดยกระจายตัวอย่างของข้อมูล

ข้อเสนอแนะ

สามารถนำผลการคำนวณที่ได้ไปใช้เป็นต้นแบบเพื่อสร้างระบบการทำนายเพื่อเลือกซื้อประกันชีวิตและอาจพิจารณาข้อมูลอื่น ๆ ที่เกี่ยวข้องกับการเลือกซื้อประกันที่กำหนดเป็นตัวแปรอิสระเพื่อเพิ่มประสิทธิภาพการทำนาย

กิตติกรรมประกาศ

บทความวิจัยนี้สำเร็จลุล่วงได้ด้วยความกรุณาเป็นอย่างยิ่งจาก ดร. เตชรัฐสินธุ์ เพ็ญชัย อาจารย์ที่ปรึกษาที่ได้กรุณาให้คำแนะนำและติดตามการวิจัยครั้งนี้ตลอดมา ผู้วิจัยขอขอบคุณในความกรุณาของท่านเป็นอย่างยิ่ง ขอขอบคุณครอบครัวของผู้วิจัยที่ได้ให้การสนับสนุนมาโดยตลอด และท้ายที่สุดขอขอบคุณตนเองที่ยังมีแรงใจและพยายามจนบทความวิจัยนี้สำเร็จลุล่วงด้วยดี

เอกสารอ้างอิง

- จามรี ชูบัวทอง และสมศรี บัณฑิตวิไล. (2560) การพัฒนาตัวแบบเพื่อพยากรณ์คุณภาพผลิตภัณฑ์ฮาร์ดดิสก์ด้วยการถดถอยโลจิสติกส์และโครงข่ายประสาทเทียมโดยใช้การวิเคราะห์เหมืองข้อมูล, *วารสารวิทยาศาสตร์และเทคโนโลยี*, 25(1), 1-13.
- ชนิตา หางแก้ว. (2563) .การทำนายการตัดสินใจสำหรับการเลือกทำประเภทประกันภัยของลูกค้า. (วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยเกษตรศาสตร์).
- ธนธร โลจันดา. (2559) .ปัจจัยที่ส่งผลต่อพฤติกรรมการทำธุรกรรมในธุรกิจซื้อขายออนไลน์ของนักศึกษาระดับปริญญาตรีในมหาวิทยาลัยเขตจังหวัดนครปฐมด้วยเทคนิคการวัดพฤติกรรมเชิงเส้นแบบหลายกลุ่ม. (วิทยานิพนธ์ปริญญาโท, มหาวิทยาลัยศิลปากร)
- ปานทิพย์ พุ่มพุ่ม. (2561) การทำนายผลการรีวิวโค้ดในโครงการโอเพนซอร์สบนกิตฮับ. (วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต) จุฬาลงกรณ์มหาวิทยาลัย ,กรุงเทพมหานคร).
- รุจิรา ธรรมสมบัติ. (2555). ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือโดยใช้ต้นไม้ตัดสินใจ. หน่วยงานวิจัยวิทยาลัยราชพฤกษ์ นนทบุรี,
- วิชชานันต์ เมธาวิริยะกุล, ปานฉัตร อาการักษ์ และวัฒนา ยืนยง.(2560) ตัวแบบสัญญาณเตือนภัยล่วงหน้าทางการเงิน การเปรียบเทียบระหว่างวิธีการถดถอยโลจิสติกส์และโครงข่ายประสาทเทียมของบริษัทจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทย, *วารสารวิชาชีพบัญชี*, 13(39), 36-57
- วีระพันธ์ พานิชย์. (2564) .การประยุกต์ใช้ Machine Learning ทำนายผลการเรียนวิชา Web Database ของนิสิตสาขาเทคโนโลยีการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยบูรพา. (วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต, มหาวิทยาลัยธุรกิจบัณฑิตย์).
- Weerasinghe KPMLP, Wijegunasekara M. (2559). A comparative study of data mining algorithms in the prediction of auto insurance claims, *Eur Int J Sci Technol*, 5(1), 47-54.