การเปรียบเทียบประสิทธิภาพการแปลงข้อมูลที่มีผลต่อการจำแนกข้อมูลของการอนุมัติสินเชื่อของลูกค้าธนาคารด้วยเทคนิคเหมืองข้อมูล

ชุติมา สุขสมัย[1], ไก้รุ่ง เฮงพระพรหม[2*] และ เกล้ากัลยา ศิลาจันทร์[3]

[1,2]สาขาวิชาวิทยาการข้อมูล คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม
[3]สาขาเทคโนโลยีคอมพิวเตอร์ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยราชภัฏนครปฐม
*kairung2011.heng@gmail.com

บทคัดย่อ

งานวิจัยนี้มีวัตุประสงค์ 1) เพื่อศึกษาผลกระทบจากการแปลงข้อมูลให้เป็นมาตรฐานกับการจำแนกข้อมูลอนุมัติสินเชื่อของลูกค้า 2) เพื่อศึกษาประสิทธิภาพการจำแนกข้อมูล 3 เทคนิค ได้แก่ เพื่อนบ้านใกล้สุด (K-Nearest Neighbor) ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม (Neural Network) การอนุมัติสินเชื่อของลูกค้าธนาคาร โดยในงานวิจัยนี้ได้ศึกษากับข้อมูลการอนุมัติสินเชื่อของลูกค้าธนาคาร

เทคนิคการแปลงข้อมูให้เป็นมาตรฐานในรูปแบบ Min-Max, Z-score และ Mean มีผลต่อการจำแนกข้อมูลด้วยเทคนิคเพื่อนบ้านใกล้สุด (K-Nearest Neighbor) สูงสุดเนื่องจากให้ประสิทธิภาพในการจำแนกข้อมูลสูงสุด เทคนิคที่ให้ประสิทธิภาพในการจำแนกข้อมูรองลงมา คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และสุดท้าย คือ เทคนิคโครงข่ายประสาทเทียม (Neural Network) ผลการทดลองพบว่า เทคนิค KNN-MinMax, KNN-Mean และ KNN-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.63 % ส่วนเทคนิคที่ให้ประสิทธิภาพรองลงมา คือ DT-Mean ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.53 % รองลงมา คือ DT-MinMax และ DT-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.43 % ถัดมาเทคนิค NN-MinMax, NN-Mean, NN-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 77.45 %

คำสำคัญ: การแปลงข้อมูล การจำแนกประเภท โครงข่ายประสาทเทียม ต้นไม้ตัดสินใจ เพื่อนบ้านใกล้สุด k ตัว

A Comparison of Normalization Data Transformation Efficiency Affecting with Bank Customer Credit Data Classification using Data Mining Techniques

Chutima Suksamai[1], Kairung Hengpraprohm[2] and Klaokanlaya Silachan[3]

[1,2]Data Science Department, Faculty of Science and Technology, Nakhon Pathom Rajabhat University

[3]Technologycomputer Department, Faculty of Science and Technology, Nakhon Pathom Rajabhat University

* kairung2011.heng@gmail.com

Abstract

The objectives of this research 1) to compare the effects of data transformation using 3 methods namely Min-Max, Z-score and Mean 2) to study the efficiency of data classification of 3 techniques namely K-Nearest Neighbor, Decision Tree and Neural Network. This research studied credit approval data of bank customers. The results showed that the Knn-MinMax, Knn-Mean and Knn-Z-score techniques yielded 80.63% classification efficiency, while the second most efficient technique, DT-Mean, yielded the second 80.53% classification efficiency. Next, DT-MinMax and DT-Z-scores had provided a classification efficiency of 80.43%. Finaly, the NN-MinMax, NN-Mean, NN-Z-score techniques provide a classification efficiency of 77.45%.

Therefore, the techniques of normalization data transformation into Min-Max, Z-score and Mean. They have highest effect on the K-Nearest Neighbor classification technique because they provide the highest classification efficiency. The second most effective technique for normalization data transformation was Decision Tree technique and the last one was the Neural Network technique.

Keywords: Transformation, Classification, Neural Network, Decision Tree, K-nearest neighbor

1. บทนำ

สินเชื่อโครงการธนาคารประชาชนเป็นโครงการที่จัดตั้งขึ้นตามนโยบายของรัฐบาลเพื่อกระจาย โอกาสในการเข้าถึงแหล่งเงินทุนของลูกค้ารายย่อย ซึ่งมีวัตถุประสงค์เพื่อตอบสนอง นโยบาย ของ รัฐบาลและสอดคล้องกับวิสัยทัศน์ของธนาคาร คือเป็นสถาบันการเงินที่มั่นคงเพื่อการออมการ พัฒนา เศรษฐกิจและสังคมของประเทศโดยเฉพาะเศรษฐกิจฐานราก เพื่อพื้นฐานและความมั่นคง แก่ ผู้ประกอบการรายย่อยและประชาชนทั่วไป

ในการพิจารณาเพื่อขออนุมัติสินเชื่อแต่ละประเภทนั้น ธนาคารได้อาศัยนโยบายและ หลักเกณฑ์ ของผู้ใช้สินเชื่อแต่ละรายโดยทั่วไปแล้วจะมีปัจจัยหลักที่ใช้ประกอบการพิจารณา คือ นโยบายการให้ สินเชื่อของสินเชื่อแต่ละประเภท คุณลักษณะและความสามารถในการชำระ หนี้ของผู้ขอสินเชื่อ คุณลักษณะและความ น่าเชื่อถือของผู้ขอสินเชื่อ ความสามารถ ในการชำระหนี้คืนได้ตามระยะเวลาที่กำหนด เงินทุน สินทรัพย์ หรือเงินฝากของผู้ขอสินเชื่อ ผู้ค้าประกันหรือหลักประกันที่ผู้ขอสินเชื่อนำมาจำนำหรือจำนองเพื่อให้สถาบันการเงินมีความ มั่นใจและลดความเสี่ยง และปัจจัยภายนอกที่มีผลกระทบต่อ รายได้ของผู้ขอสินเชื่อ

การให้สินเชื่อทุกประเภทนั้นย่อมมีความเสี่ยงจากการที่ไม่ได้รับชำระเงินคืนหรือ เกิดหนี้ที่ไม่ก่อรายได้ขึ้น การวิเคราะห์เพื่ออนุมัติเงินกู้ที่ มีประสิทธิภาพ และดูแนวโน้มความสามารถในการชำระหนี้ของ ลูกค้าในอนาคตและครอบคลุมไป จนถึงการวิเคราะห์เพื่อพิจารณาประนอมหนี้ให้แก่ลูกค้า การบริหารสินเชื่อที่ดีต้องอาศัยข้อมูล ทางการบัญชีเป็นเครื่องมือที่สำคัญในการวิเคราะห์ฐานะ การเงิน ผลการดำเนินงาน สภาพคล่องของ กิจการ ประเมินผลการดำเนินทั้งในอดีตและปัจจุบัน รวมทั้งการวางแผนใน อนาคต เพื่อใช้ ประกอบการวิเคราะห์อนุมัติสินเชื่อ ติดตามฐานะของลูกหนี้เพื่อดูแนวโน้มความสามารถในการชำระหนี้ว่าอาจจะเกิดการค้างชำระได้หรือไม่ และช่วย ในการตัดสินใจพิจารณาอนุมัติการแก้ไขหนี้ค้าง ชำระให้เหมาะสมมีประสิทธิภาพ

จากที่กล่าวมาข้างต้น ผู้วิจัยได้ตระหนักถึงปัจจัยต่าง ๆ ที่มีผลต่อการพิจารณาอนุมัติสินเชื่อเพื่อ ลดขั้นตอนและลดระยะเวลาในการปฏิบัติงาน จึงได้นำกระบวนการเหมืองข้อมูล (Data Mining) มา ช่วยในการพิจารณาการให้สินเชื่อของลูกค้าธนาคาร

2. วัตถุประสงค์การวิจัย

2.1 เพื่อศึกษาผลกระทบจากการแปลงข้อมูลให้เป็นมาตรฐานกับการจำแนกข้อมูลอนุมัติสินเชื่อของลูกค้า

2.2 เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลของการอนุมัติสินเชื่อของลูกค้าด้วยเทคนิคเหมืองข้อมูล ได้แก่ เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbor) ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม(Neural Network)

3. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

3.1 ทฤษฎีที่เกี่ยวข้อง

3.1.1 การทำเหมืองข้อมูล (Data mining) เป็นกระบวนการในการสกัดเพื่อค้นหารูปแบบหรือความรู้ จากฐานข้อมูลขนาดใหญ่ เพื่อให้ได้สารสนเทศที่น่าสนใจที่ไม่คาดคิดมาก่อนและเป็นประโยชน์มาใช้ สนับสนุนการตัดสินใจในองค์กร

3.1.2 เพื่อนบ้านใกล้ที่สุดเค (K – nearest neighbor) หรือ เคเอ็นเอ็น (KNN) เป็นกระบวนการเรียนรู้แบบมีผู้สอนสำหรับปัญหาการจำแนกประเภทข้อมูลที่ถือว่าเป็นวิธีการพื้นฐานและง่ายที่สุดวิธีการหนึ่งในศาสตร์การเรียนรู้ของเครื่อง โดยจัดเป็นการเรียนรู้แบบขี้เกียจ (lazy learning) กล่าวคือ ไม่มีการสร้างตัวแบบจากชุดข้อมูลสอน แต่จะเริ่มคำนวณเมื่อมีการสอบถามเข้ามา

3.1.3 ต้นไม้ตัดสินใจ (Decision Tree) เป็นกระบวนการเรียนรู้แบบมีผู้สอนสำหรับปัญหาการจำแนกประเภทข้อมูล โดยใช้ทฤษฎีสารสนเทศ (information theory) และโครงสร้างต้นไม้ (tree structure) ในการสร้างตัวแบบ

3.1.4 เครือข่ายประสาทเทียม (Artificial Neuron Network: ANN) เป็นกระบวนการเรียนรู้แบบมีผู้สอนโดยจำลองกระบวนการทำงานมาจากระบบประสาทของมนุษย์

3.1.5 การแปลงข้อมูลให้เป็นมาตรฐานซี (Z-score) เป็นการปรับค่าของคุณลักษณะโดยการแปลงให้เป็นคะแนนมาตรฐานซีเพื่อปรับให้ คาเฉลี่ยของข้อมูลมีค่าเป็น 0 และ ค่าส่วนเบี่ยงเบนมาตรฐานมีค่าเป็น 1 คำนวณได้ตามสมการ (1)

$$V_i^{'} = \frac{v_i - Mean_v}{SD_v} \quad (1)$$

โดย $v_i^{'}$ แทน ค่าใหม่ของคุณลักษณะตัวที่ i ของข้อมูล v

v_i แทน คาของคุณลักษณะตัวที่ i ของข้อมูล v เดิม

$Mean_v$ แทน คาเฉลี่ยของคุณลักษณะนั้น

SD_v แทน ส่วนเบี่ยงเบนมาตรฐานของคุณลักษณะนั้น

3.1.6 การแปลงข้อมูลให้เป็นมาตรฐาน Mean คล้ายกับ Rescaling ด้านบน แตกต่างกันที่ใช้ Mean แทน Min ทำให้ช่วงของ Output [-0.5, 0.5] มีทั้งบวกและลบ Balance กัน ตรงเลข 0 (ขยับ Mean มาตรง 0) คำนวณได้ตามสมการ (2)

$$V_i^{'} = \frac{v_i - Mean_v}{Max_v - Min_v} \quad (2)$$

โดย $v_i^{'}$ แทน ค่าใหม่ของคุณลักษณะตัวที่ i ของข้อมูล v

v_i แทน คาของคุณลักษณะตัวที่ i ของข้อมูล v เดิม

$Mean_v$ แทน คาเฉลี่ยของคุณลักษณะนั้น

Max_v แทน คาที่มากที่สุดของคุณลักษณะนั้น

Min_v แทน ค่าที่น้อยที่สุดของคุณลักษณะนั้น

3.1.7 การแปลงข้อมูลให้เป็นมาตรฐาน Min-Max

การปรับช่วงข้อมูล (min-max) เป็นการปรับค่าคุณลักษณะของข้อมูลให้อยู่ในช่วงค่าน้อยสุด และค่ามากสุดตามที่กำหนด ซึ่งนิยมใช้ค่าน้อยสุดเป็น 0 และ คามากสุดเป็น 1 บางครั้งนิยมเรียกวิธีการนี้ว่า การทำข้อมูลให้เป็นปกติแบบ 0-1 (0-1 Normalization) วิธีการแปลงค่าคำนวณได้จากสมการ (3)

$$V_i^{'} = \frac{V_i - Min_v}{Max_v - Min_v} \quad (3)$$

โดยที่ v_i' แทน ค่าใหม่ของคุณลักษณะตัวที่ i ของข้อมูล v

v_i แทน คาของคุณลักษณะตัวที่ i ของข้อมูล v เดิม

Max_v แทน คาที่มากที่สุดของคุณลักษณะนั้น

Min_v แทน ค่าที่น้อยที่สุดของคุณลักษณะนั้น

3.2 งานวิจัยที่เกี่ยวข้อง

สัณหณัฐ [1] การวิเคราะห์ข้อมูลการเกิดอุบัติเหตุทางถนนโดยใช้เทคนิคเหมืองข้อมูล ทำการศึกษาเพื่อพัฒนาโมเดลการเกิดอุบัติเหตุบนถนน ด้วยเทคนิคการทำเหมืองข้อมูลโดยรวบรวมข้อมูลจากกรมป้องกันและบรรเทาสาธารณภัย ปี 2548-2562 จำนวนมากกว่า 177,816 ชุด ซึ่งทำการแบ่งกลุ่มข้อมูลออกเป็น 2 กลุ่ม แบ่งตามเกณฑ์การเสียชีวิต ในการสร้างโมเดลเปรียบเทียบระหว่างเทคนิคการจำแนกประเภทข้อมูล ควบคู่กับขั้นตอนวิธีเชิงพันธุกรรม (Genetic Algorithm : GA) ในการคัดเลือกลักษณะและวัดประสิทธิภาพ ด้วยวิธี Cross Validation จากผลกาดำเนินงานวิจัยพบว่า Neural Net มีประสิทธิภาพดีกว่าทุกโมเดลที่ทำการทดลองในการวัดประสิทธิภาพของทุกกลุ่มข้อมูล ดังนั้นจากการวิจัยจึงได้โมเดล Neural Net สำหรับพยากรณ์ผลการเกิดอุบัติเหตุบนถนนและหาปัจจัยที่ที่ผลต่อการเกิดอุบัติเหตุบนถนนให้มีประสิทธิภาพ

ประเทือง และคณะ [2] การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัย ของผู้ประกอบอาชีพด้านคอมพิวเตอร์ ได้ทำการศึกษาเพื่อสร้างและเปรียบเทียบประสิทธิภาพตัวแบบความสัมพันธ์ของข้อมูล จากเทคนิค Neural Networks, Naive Bayes และ Decision Tree เพื่อค้นหาลักษณะนิสัยของ ผู้ประกอบอาชีพด้านคอมพิวเตอร์ตามทฤษฎีการเลือกอาชีพของฮอลแลนด์โดยพิจารณาจากตัวแบบ จากค่าความถูกต้อง (Accuracy) และความแม่นยำ (Precision) ดำเนินการเก็บรวบรวมข้อมูลจากผู้ประกอบอาชีพด้านคอมพิวเตอร์ ใช้การสุ่มตัวอย่างจาก จำนวนผู้ที่สำเร็จการศึกษาระดับปริญญาตรี หลักสูตรวิศวกรรมคอมพิวเตอร์ วิทยาการคอมพิวเตอร์ และวิทยาศาสตร์คอมพิวเตอร์ กลุ่มตัวอย่าง 395 คน หลังจากรวบรวมข้อมูล แล้วจัดเตรียมข้อมูลให้ อยู่ในรูปแบบไฟล์ CSV จากนั้นวิเคราะห์ด้วยวิธี5 Fold-Cross-Validation แล้วสร้างตัวแบบทั้ง 3 เทคนิค จากนั้นวัดค่าประสิทธิภาพและค่าความแม่นยำของตัวแบบ

ขวัญนภา [3] การใช้เหมืองข้อมูลสำหรับพิจารณาการให้สินเชื่อสำหรับธนาคาร ได้ศึกษาการนำเทคนิคเหมืองข้อมูลมาใช้เพื่อระบุปัจจัยสำคัญในการพิจารณาสินเชื่อ โดยพิจารณาจากข้อมูลผู้ที่ยื่น สินเชื่อโครงการธนาคารประชาชนจำนวน 14 ปัจจัย จากข้อมูลผู้ยื่นสินเชื่อทั้งหมด 1,000 ตัวอย่าง โดยมีวัตถุประสงค์เพื่อเปรียบเทียบกระบวนการอนุมัติสินเชื่อของธนาคารโดยไม่ใช้วิธีการคัดเลือก คุณลักษณะ และใช้วิธีการคัดเลือกคุณลักษณะ ซึ่งประกอบด้วยวิธี Chi-Square และวิธี Information Gain จากนั้นนำข้อมูลที่ได้ไปสร้างโมเดลด้วยวิธีซัพพอร์ทเว็กเตอร์แมชชีน (Support Vector Machine: SVM) โครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron: MLP) และต้นไม้ตัดสินใจ (Decision Tree) เพื่อใช้สำหรับพิจารณาการได้สินเชื่อของธนาคาร ในการทดสอบได้เลือกใช้วิธีK-Fold Cross-Validation โดยกำหนดให้ K มีค่าเท่ากับ 10

จิราภรณ์ [4] การพยากรณผลสัมฤทธิ์ทางการเรียน ด้วยเทคนิคเหมืองข้อมูลโดยใช้Rapid Mineได้ทำการศึกษาและนำเสนอการทำเหมืองข้อมูลทางการศึกษา โดยการจำแนกประเภทข้อมูลการวิเคราะห์องค์ประกอบหลักของข้อมูล เพื่อหาความสัมพันธ์ของตัวแปร และเปรียบเทียบประสิทธิภาพของอัลกอริทึม ซึ่งเป็นการศึกษาด้วยเทคนิค ต้นไม้ตัดสินใจ เทคนิคป่าแห่งการทำนาย การเรียนรู้เบย์ และ K-NN โดยใช้ข้อมูลของนักเรียนระดับมัธยมศึกษาในโรงเรียนโปรตุเกสประกอบด้วยข้อมูลด้านผลการเรียน ด้านความเป็นอยู่ และความเชื่อมโยงทางสังคมและโรงเรียนจาก UCI Machine Learning Repository มีข้อมูล 649 รายการ 31 แอตทริบิวต์

ซอและ และคณะ [5] การวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล เพื่อวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษา หลักสูตรวิทยาการ คอมพิวเตอร์และหลักสูตรเทคโนโลยีสารสนเทศ มหาวิทยาลัยราชภัฏยะลา ซึ่งใช้ข้อมูลของนักศึกษาที่กำลังศึกษาและ สำเร็จการศึกษาแล้วระหว่างปีการศึกษา 2555 – 2558 เป็นข้อมูลจากหลักสูตรวิทยาการคอมพิวเตอร์ จำนวน 97 ระเบียน และหลักสูตรเทคโนโลยีสารสนเทศ จำนวน 202 ระเบียน ประกอบด้วย 26 คุณลักษณะ โดยใช้เทคนิค เหมืองข้อมูล ประกอบด้วยเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เทคนิคโครงข่ายประสาทเทียมแบบย้อนกลับ (Back Propagation Neural Network : BP-NN) และเทคนิคซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine : SVM) มาสร้างตัวแบบพยากรณ์และเปรียบเทียบตัวแบบด้วยการทดสอบประสิทธิภาพแบบ 10-fold Cross Validation

4. วิธีดำเนินการวิจัย

4.1 ข้อมูลสำหรับการวิจัย

ข้อมูลสำหรับการทำวิจัยเป็นข้อมูลที่นำมาจาก Gaggle โดยมีชื่อไฟล์ว่า Customer data เป็นข้อมูลเกี่ยวกับชุดข้อมูลจำแนกความเสี่ยงด้านเครดิต โดยมีข้อมูลภายใน 12 คอลัมน์ 977 แถว คอลัมน์ที่ 1 ชื่อว่า fea_1 เป็นคุณลักษณะที่ 1 คอลัมน์ที่ 2 ชื่อว่า fea_2 เป็นคุณลักษณะที่ 2 คอลัมน์ที่ 3 ชื่อว่า fea_3 เป็นคุณลักษณะที่ 3 คอลัมน์ที่ 4 ชื่อว่า fea_4 เป็นคุณลักษณะที่ 4 คอลัมน์ที่ 5 ชื่อว่า fea_5 เป็นคุณลักษณะที่ 5 คอลัมน์ที่ 6 ชื่อว่า fea_6 เป็นคุณลักษณะที่ 6 คอลัมน์ที่ 7 ชื่อว่า fea_7 เป็นคุณลักษณะที่ 7 คอลัมน์ที่ 8 ชื่อว่า fea_8 เป็นคุณลักษณะที่ 8 คอลัมน์ที่ 9 ชื่อว่า fea_9 เป็นคุณลักษณะที่ 9 คอลัมน์ที่ 10 ชื่อว่า fea_10 เป็นคุณลักษณะที่ 10 คอลัมน์ที่ 11 ชื่อว่า fea_11 เป็นคุณลักษณะที่ 11 คอลัมน์ที่ 12 ชื่อว่า Class เป็นการจำแนกเครดิต

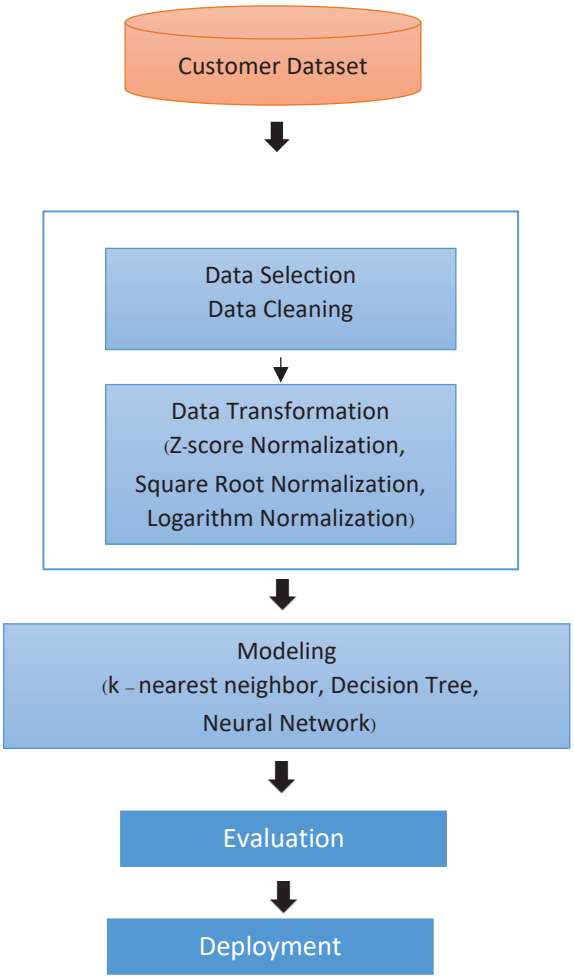
	A	B	C	D	E	F	G	H	I	J	K	L
1	class	fea_1	fea_2	fea_3	fea_4	fea_5	fea_6	fea_7	fea_8	fea_9	fea_10	fea_11
2	1	5	1245.5	3	77000	2	15	5	109	5	151300	244.949
3	0	4	1277	1	113000	2	8	-1	100	3	341759	207.1738
4	0	7	1298	1	110000	2	11	-1	101	5	72001	1
5	1	7	1335.5	1	151000	2	11	5	110	3	60084	1
6	0	6	1217	3	56000	2	6	-1	100	3	60091	1
7	1	4	1304	3	35000	2	8	9	85	5	60069	1
8	1	5	1256	3	78000	2	15	-1	111	3	60030	1
9	0	5	1323.5	3	218000	2	15	5	112	4	151300	282.8427
10	0	7	1314.5	1	483000	2	11	9	101	3	151300	282.8427
11	0	4	1250	3	95000	2	8	9	111	4	60047	1
12	0	7	1223	3	81000	2	11	5	114	4	151300	194.9359
13	0	7	1365.5	1	96000	2	11	-1	78	3	60025	178.8854
14	0	7	1257.5	3	126000	2	11	5	105	3	151300	1
15	0	4	1214	3	81000	2	8	5	111	1	60043	202.4846
16	1	7	1241	3	78000	1	11	5	105	5	60074	1
17	1	7	1241	1	111000	2	11	5	90	5	350068	223.6068
18	0	4	1305.5	1	129000	2	8	5	81	5	60014	1
19	0	7	1277	1	150000	2	11	5	95	5	60015	1
20	0	4	1304	1	35000	2	8	5	105	3	340019	181.659
21	0	4	1320.5	3	164000	1	8	5	105	5	151300	187.0829
22	0	7	1290.5	3	119000	2	11	9	85	3	151300	200
23	0	5	1455.5	3	35000	2	15	5	92	5	60058	1
24	1	7	1242.5	3	71000	1	11	5	108	5	60005	173.2051
25	1	7	1239.5	3	79000	2	11	8	111	5	71003	1
26	0	7	1335.5	3	150000	2	11	9	105	5	350050	412.3106
27	0	4	1355	3	177000	2	8	5	87	5	151101	1
28	0	5	1338.5	3	30000	2	15	5	110	4	71000	223.6068
29	0	7	1259	3	74000	2	11	5	95	5	60036	207.3186

data | minmax | mean | z-score | root | log

ภาพที่ 1 ข้อมูลเกี่ยวกับความเสี่ยงด้านเครดิต

4.2 ขั้นตอนในการดำเนินการวิจัย

ในการวิจัยนี้ได้แบ่งขั้นตอนการวิจัยออกเป็นขั้นตอน ดังได้นำเสนอในกรอบแนวคิดในการวิจัยดังได้นำเสนอในภาพที่ 2 ดังต่อไปนี้



ภาพที่ 2 กรอบแนวคิดในการวิจัย

สำหรับในการวิจัยครั้งนี้ได้แบ่งกรอบแนวคิดในการวิจัย ประกอบด้วย 7 ขั้นตอนดังนี้

4.2.1 สำรวจหาข้อมูลที่ต้องการ และเหมาะสมในการทำวิจัย

4.2.2 ทำการสำรวจและเก็บรวบรวมข้อมูลวิจัยที่เกี่ยวข้องกับการใช้เทคนิคการทำเหมืองข้อมูล เพื่อศึกษาแนวทางในการทำวิจัย

4.2.3 ศึกษาข้อมูลที่เลือกมา และทำความสะอาดข้อมูล (Clean Data) ข้อมูลที่ไม่เกี่ยวข้องออกไป

4.2.4 การแปลงข้อมูลให้เป็นมาตรฐาน การแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน ในการวิจัยนี้จะเลือกการแปลงข้อมูลให้เป็นมาตรฐานด้วย Min-Max, Mean และRescaling หรือ Z-score เพื่อทำการเปรียบเทียบว่าข้อมูลที่แปลงมาแล้วนั้นข้อมูลไหนเหมาะสมกับเทคนิคเหมืองข้อมูลที่เลือกมาข้างต้น

4.2.5 จากนั้นนำข้อมูลที่ผ่านการแปลงข้อมูลให้เป็นมาตรฐานแล้ว เข้า Weka และใช้เทคนิคเหมืองข้อมูลที่เลือกมา คือ เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbor) ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม (Neural Network) มาสร้างโมเดลสำหรับการจำแนกข้อมูลเพื่อเปรียบเทียบประสิทธิภาพของโมเดล

4.2.6 ทำการประเมินผล (Evaluation) เพื่อนำผลที่ได้จากการนำเข้า Weka มาพิจารณาว่าเทคนิคเหมืองข้อมูลชนิดไหนมีประสิทธิภาพที่เหมาะสมที่สุดด้วยวิธีหาค่าความแม่นยำในการจำแนกข้อมูล (Accuracy)

4.2.7 ทำการปรับใช้ (Deployment) เพื่อให้งานวิจัยเกิดประสิทธิภาพมากขึ้น และอาจจะนำผลที่ได้มาใช้ให้เกิดประโยชน์เกี่ยวกับการให้เครดิตลูกค้าธนาคาร

4.3 การประเมิลประสิทธิภาพของการวิจัย

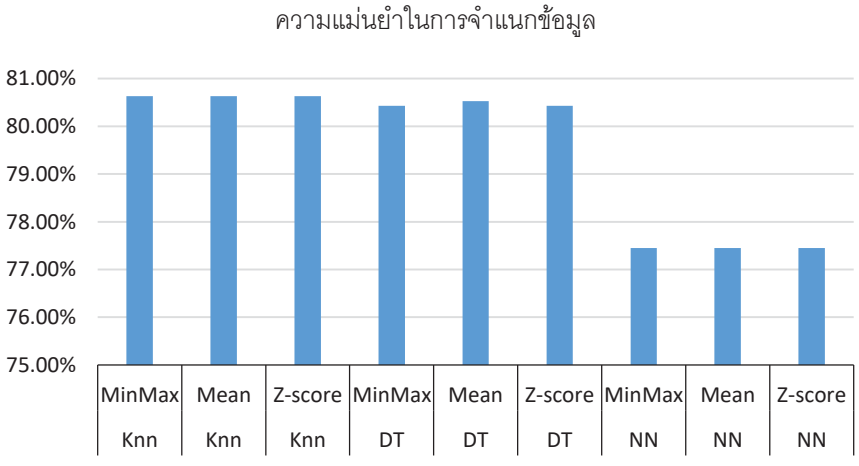
การประเมินประสิทธิภาพของงานวิจัยนี้ได้ใช้วิธีวัดค่าความแม่นยำในการจำแนกข้อมูล (Accuracy) ซึ่งเป็นมาตรวัดค่าความแม่นตรง คือ ค่าที่บอกว่าการทำนาย ทำนายได้แม่นยำแค่ไหน สามารถคำนวณได้ดังนี้

$$Accuracy = \frac{TP+FP}{TP+FP+TN+FN}$$

โดยที่ TP คือค่า True Positive, TN คือค่า True Negative,
FP คือค่า False Positive, FN คือค่า False Negative

5. ผลการวิจัย

ผลการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลการอนุมัติสินเชื่อธนาคาร ด้วยการแปลงข้อมูลให้เป็นมาตรฐาน 3 วิธี ได้แก่ Min-Max, Z-score และ Mean ด้วยเทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbor) ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม(Neural Network) ผลปรากฏว่า เทคนิค Knn-MinMax, Knn-Mean และ Knn-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.63 % ส่วนเทคนิคที่ให้ประสิทธิภาพรองลงมา คือ DT-Mean ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.53 % รองลงมาคือ DT-MinMax และ DT-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.43 % ถัดมาเทคนิค NN-MinMax, NN-Mean, NN-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 77.45 % ดังแสดงในตารางที่ 1 และในภาพที่ 3 ตามลำดับ



ภาพที่ 3 ผลการเปรียบเทียบความแม่นยำในการจำแนกข้อมูล

ตารางที่ 1 เปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลด้วยแปลงข้อมูลแต่ละแบบ

เทคนิค	tranfrom	ค่า accuracy
KNN	MinMax	*80.63%
KNN	Mean	*80.63%
KNN	Z-score	*80.63%
DT	MinMax	80.43%
DT	Mean	80.53%
DT	Z-score	80.43%
NN	MinMax	77.45%
NN	Mean	77.45%
NN	Z-score	77.45%

6. สรุปผล

จากผลการทดลองเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลการอนุมัติสินเชื่อธนาคาร จากการแปลงข้อมูลให้เป็นมาตรฐาน 3 วิธี ได้แก่ Min-Max, Z-score และ Mean ด้วยเทคนิคการจำแนกข้อมูล 3 เทคนิค ได้แก่ เทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbor) ต้นไม้ตัดสินใจ (Decision Tree) และ โครงข่ายประสาทเทียม(Neural Network) จะเห็นได้ว่า เทคนิค Knn-MinMax , Knn-Mean และ Knn-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.63 % ส่วนเทคนิคที่ให้ประสิทธิภาพรองลงมา คือ DT-Mean ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.53 % รองลงมาคือ DT-MinMax และ DT-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 80.43 % ถัดมาเทคนิค NN-MinMax, NN-Mean, NN-Z-score ให้ประสิทธิภาพในการจำแนกข้อมูลที่ 77.45 %

ดังนั้น เทคนิคการแปลงข้อมูลให้เป็นมาตรฐานในรูปแบบ Min-Max, Z-score และ Mean มีผลต่อการจำแนกข้อมูลด้วยเทคนิคเพื่อนบ้านที่ใกล้ที่สุด (K-Nearest Neighbor) สูงสุดเนื่องจากให้ประสิทธิภาพในการจำแนกข้อมูลสูงสุด เทคนิคที่ให้ประสิทธิภาพในการจำแนกข้อมูลรองลงมา คือ เทคนิคต้นไม้ตัดสินใจ (Decision Tree) และสุดท้าย คือ โครงข่ายประสาทเทียม (Neural Network)

7. เอกสารอ้างอิง

[1] สัณหณัฐ รักราช. (2562). **การวิเคราะห์ข้อมูลการเกิดอุบัติเหตุทางถนนโดยใช้เทคนิคเหมืองข้อมูล.**[ออนไลน์] สืบค้นเมื่อวันที่ 27 มีนาคม 2565 จาก: http://coms.kru.ac.th/tee/Projects/ShowPdf?name=637225442801621215.pdf&chk=False

[2] ประเทือง วงษ์ทอง, วาฤทธิ์ กันแก้ว ,อภิชชญา ขวัญแก้ว และยุวดี โฉมแดง. (2563). **การประยุกต์เทคนิคเหมืองข้อมูลค้นหาลักษณะนิสัย.** [ออนไลน์] สืบค้นเมื่อวันที่ 27 มีนาคม 2565 จาก: https://research.rmutsb.ac.th/fullpaper/2562/research.rmutsb-2562-20201006100035839.pdf

[3] ขวัญนภา พิมพ์ชารีย์. (2564). **การใช้เหมืองข้อมูลสำหรับพิจารณาการให้สินเชื่อสำหรับธนาคาร.**[ออนไลน์] สืบค้นเมื่อวันที่ 27 มีนาคม 2565 จาก :http://olarik.it.msu.ac.th/wpcontent/uploads/2021/07/61011284502.pdf

[4] จิราภรณ์ เจริญยิ่ง. (2563). **การพยากรณผลสัมฤทธิ์ทางการเรียน ด้วยเทคนิคเหมืองข้อมูลโดยใช้Rapid Mine**. [ออนไลน์] สืบค้นเมื่อวันที่ 27 มีนาคม 2565 จาก :http://irithesis.swu.ac.th/dspace/bitstream/123456789/1231/1/gs591130025.pdf

[5] ซอและ เกปัน, พิมลพรรณ ลีลาภัทรพันธุ์ และ อัจฉราพร ยกขุน. (2561**). การวิเคราะห์ปัจจัยที่มีผลต่อการพ้นสภาพของนักศึกษาโดยใช้เทคนิคเหมืองข้อมูล.** [ออนไลน์] สืบค้นเมื่อวันที่ 27 มีนาคม 2565 จาก: https://profile.yru.ac.th/storage/academic-articles/June2020/rCVJ2PUndQdukFwGT1tj.pdf