

การพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ โดยใช้เทคนิควิธีเหมืองข้อมูล Commercial Bank Credit Amount Forecasting using Data Mining Technique

อรรณพ กางกั้น
Unnop Kangkan

คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยสยาม
Faculty of Information Technology, Siam University

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนา และเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย โดยใช้เทคนิควิธีเหมืองข้อมูล 3 เทคนิค ได้แก่ 1) การถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย ข้อมูลที่นำมาใช้ศึกษาเป็นข้อมูลที่เกี่ยวข้องกับปัจจัยการกำหนดปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย ได้แก่ ปริมาณเงินฝากของธนาคารพาณิชย์ อัตราดอกเบี้ยเงินกู้ ปริมาณเงินกู้ที่ไม่ก่อให้เกิดรายได้ และปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย โดยใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม 2554 ถึงวันที่ 31 ธันวาคม 2564 ซึ่งผู้วิจัยได้ทำการแบ่งชุดข้อมูล 10 ชุดสำหรับสร้างแบบจำลอง และชุดข้อมูล 1 ชุด สำหรับทดสอบแบบจำลอง จากการเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ ที่มีความเหมาะสมกับชุดข้อมูลในปี 2564 พบว่า แบบจำลองการถดถอยเชิงเส้น มีประสิทธิภาพโดยรวมสูงที่สุด มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 2.54% รองลงมาคือ แบบจำลองซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 3.62%

คำสำคัญ: การพยากรณ์ ปริมาณสินเชื่อของธนาคารพาณิชย์ เทคนิควิธีเหมืองข้อมูล

Abstract

The purposes of this research were to develop and compare the effectiveness of the model of forecasting the amount of credit of the commercial banks in Thailand using the data mining techniques by 3 methods, as follows; 1) Linear Regression, 2) Multi-Layer Perceptron and 3) Support Vector Machine for Regression. The data used for this research is related to all factors used for identifying the amount of credit of the commercial banks in Thailand i.e. the amount of deposit of the commercial banks, minimum loan rate, non-performing loan and the amount of credit of the commercial banks in Thailand. The data is from 1 January 2011 to 31 December 2021 and divided into 10 data sets for model creation and 1 data set for model testing. The comparison of the effectiveness for the model of forecasting the amount of credit of the commercial banks that suitable for testing data in 2021 found that the Linear Regression was the highest performance model for forecasting at 2.54% of MMRE and the Support Vector Machine for Regression was at 3.62% of MMRE subsequently.

Keywords: Forecasting, Commercial Bank Credit Amount, Data Mining Technique

1. บทนำ

สถาบันการเงินมีบทบาทสำคัญในการขับเคลื่อนเศรษฐกิจการเงินของประเทศ ธนาคารพาณิชย์จัดเป็นสถาบันทางการเงินที่ทำหน้าที่ในการระดมเงินฝากและให้กู้ยืมเงินแก่ภาคครัวเรือน และภาคธุรกิจในรูปแบบของสินเชื่อ ซึ่งรายได้หลักของธนาคารพาณิชย์ มาจากสามช่องทาง ประกอบด้วย ดอกเบี้ยเงินกู้ ค่าธรรมเนียมต่างๆ และจากการลงทุนของธนาคาร ดังนั้นปริมาณสินเชื่อของธนาคารพาณิชย์จึงเป็นสิ่งสำคัญที่ผู้บริหารจะต้องให้ความสนใจ

การพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์จึงมีความจำเป็นอย่างยิ่ง เพื่อประเมินความเหมาะสมเกี่ยวกับสภาพคล่องของธนาคารพาณิชย์ งานวิจัยนี้จึงนำเสนอการเปรียบเทียบประสิทธิภาพ เลือกเทคนิคที่เหมาะสม และพัฒนาแบบจำลองในการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย เพื่อช่วยให้ธนาคารพาณิชย์ทั้งภาครัฐ และภาคเอกชน สามารถนำไปใช้คาดการณ์ปริมาณสินเชื่อได้ล่วงหน้า เพื่อการวางแผนตัดสินใจ กำหนดนโยบายบริหารจัดการสินเชื่อได้อย่างมีประสิทธิภาพ ทั้งนี้ผู้วิจัยได้นำเทคนิควิธีเหมืองข้อมูล 3 เทคนิค ได้แก่ การวิเคราะห์การถดถอย (Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) มาประยุกต์ใช้ในการพัฒนาแบบจำลอง และเปรียบเทียบประสิทธิภาพของแบบจำลอง โดยศึกษาจากข้อมูลของปัจจัยกำหนดปริมาณสินเชื่อ และปริมาณสินเชื่อที่มีการรวบรวมไว้ตั้งแต่วันที่ 1 มกราคม 2554 ถึงวันที่ 31 ธันวาคม 2564

เหมืองข้อมูล (Data Mining) เป็นการกระทำกับข้อมูล (Data) โดยการขุดค้น (Mining) เป็นกระบวนการจำแนกสิ่งที่ต้องการสิ่งต่างจากเดิม สิ่งที่มีประโยชน์ เป็นการค้นหาสารสนเทศที่มีค่าจากข้อมูลขนาดใหญ่ โดยวิธีการอัตโนมัติหรือกึ่งอัตโนมัติ ทำให้เข้าใจในรูปแบบ (Pattern) และกฎเกณฑ์ (Rule) ของข้อมูล (Pyle, 2003)

การวิเคราะห์การถดถอย (Linear Regression) เป็นการประมาณการตัวแปรอิสระที่มีความสัมพันธ์เชิงเส้นตรงในการกำหนดค่าตัวแปรตามที่เป็นไปได้มากที่สุด ลักษณะของสมการ Linear Regression จะแสดงค่าความสัมพันธ์ของตัวแปรอิสระที่มีผลต่อตัวแปรตามในรูปแบบของข้อมูลเชิงปริมาณ รูปแบบทั่วไปของสมการมีดังนี้ (Tryfos, 1998)

$$Y = a + bx_1 + cx_2 + dx_3 + \dots \quad (1)$$

โดยที่ Y คือ ตัวแปรตาม และ x_1 , x_2 , x_3 คือ ตัวแปรอิสระ

โครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) เป็นการจำลองจากระบบโครงข่ายประสาทของสมองมนุษย์ที่มีจุดเชื่อมต่อเซลล์ประสาทในการแลกเปลี่ยนข้อมูลเป็นพันล้านจุด โดยจะมีโครงสร้างหลายชั้น แบ่งเป็นชั้นข้อมูลเข้า (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นผลลัพธ์ (Output Layer) แต่ละชั้นประกอบด้วยโหนด (Node) และเส้นเชื่อมโยงแต่ละโหนด (Olson & Shi, 2007) โดยชั้นข้อมูลเข้าจะมี 1 ชั้น และสร้างโหนดตามจำนวนแอตทริบิวต์ของข้อมูล และส่งค่าของโหนดกับค่าน้ำหนักไปตามเส้นเชื่อมโยงเข้าสู่ชั้นซ่อน (อาจมีได้หลายชั้น) ภายในชั้นซ่อนจะมีการคำนวณค่าผลรวมจากค่าโหนดคูณน้ำหนักในแต่ละเส้นเชื่อม สุดท้ายจะคำนวณค่าส่งออกโดยใช้ฟังก์ชันกระตุ้น (Activation Function) แล้วส่งไปยังชั้นผลลัพธ์ที่มี 1 ชั้น (Michalski, Bratko, & Kubat, 1999)

ซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นเทคนิคในการแบ่งกลุ่มข้อมูลเพื่อการจำแนกข้อมูล โดยหาขอบเขตที่เหมาะสมที่สุดในการแบ่งข้อมูล (Pyle, 2003) ระบายจะอยู่กึ่งกลางระหว่างแนวขอบของข้อมูลทั้ง 2 ฝ่าย ข้อมูลที่อยู่ใกล้ขอบมากที่สุดจะเป็นซัพพอร์ตเวกเตอร์ที่ใช้เป็นตัวแทนของข้อมูลแต่ละส่วน โดยตำแหน่งของซัพพอร์ตเวกเตอร์ในข้อมูลแต่ละส่วนจะเป็นแนวขอบของระบาย ซึ่งจะต้องไม่มีข้อมูลเกินออกมาจากแนวขอบของระบายทั้ง 2 ฝ่าย

จากวิทยานิพนธ์ของ พิมพ์ภักษ์ พัฒนชัย (2559) เรื่องปัจจัยที่มีผลต่อปริมาณเงินกู้ยืมของธนาคารพาณิชย์ไทย พบว่าปัจจัยที่มีผลต่อปริมาณเงินกู้ยืมของธนาคารพาณิชย์มากที่สุดคือ อัตราส่วนของปริมาณสินเชื่อต่อปริมาณเงินฝาก โดยมีความสัมพันธ์ในทิศทางเดียวกัน และวิทยานิพนธ์ของสุนันทา พรหมมาศ (2559) เรื่อง ปัจจัยที่ส่งผลกระทบต่อปริมาณการปล่อยสินเชื่อของธนาคารไทย

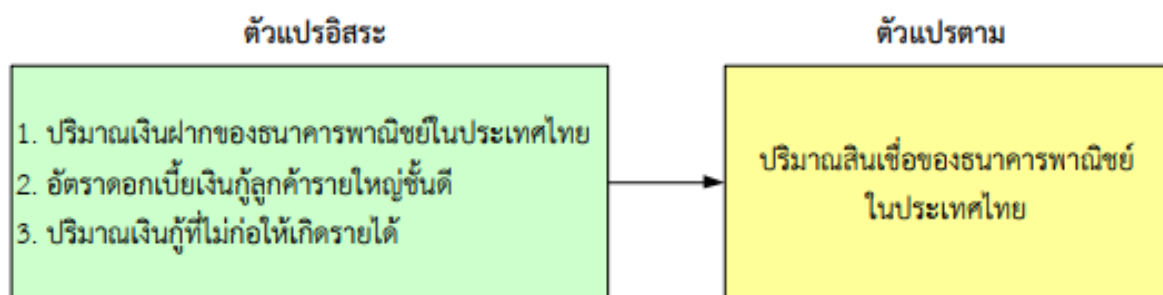
พาณิชย์ จำกัด (มหาชน) พบว่า ปัจจัยที่มีผลต่อปริมาณสินเชื่อรวมในทิศทางเดียวกันที่ระดับความเชื่อมั่นร้อยละ 99 คือ ปริมาณเงินฝากรวม ซึ่งมีระดับนัยสำคัญทางสถิติ และอัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี ซึ่งไม่มีระดับนัยสำคัญทางสถิติ ส่วนปัจจัยที่มีผลในทางตรงข้ามที่ระดับความเชื่อมั่นร้อยละ 99 คือ ปริมาณสินเชื่อที่ไม่ก่อให้เกิดรายได้ ซึ่งมีระดับนัยสำคัญทางสถิติ จากรายงานวิจัยของ นครินทร์ ภูมิตะ และ ศิริขวัญ เจริญวิริยะกุล (2563) เรื่อง ปัจจัยที่มีอิทธิพลต่อปริมาณเงินให้สินเชื่อของธนาคารกรุงเทพ ธนาคารกรุงไทย และธนาคารไทยพาณิชย์ จำกัด (มหาชน) พบว่า ปริมาณเงินฝาก และอัตราดอกเบี้ยเงินให้สินเชื่อมีผลกระทบเชิงลบต่อปริมาณเงินให้สินเชื่อทั้งสามธนาคาร และรายงานวิจัยของ เจนจิรา ช่วยคงมา และอภิชาติ พงศ์สุพัฒน์ (2564) เรื่อง ปัจจัยที่มีผลกระทบต่อปริมาณเงินให้สินเชื่อภาคธุรกิจของธนาคารพาณิชย์ในประเทศไทย พบว่า ปัจจัยที่มีผลกระทบต่อปริมาณเงินให้สินเชื่อภาคธุรกิจของธนาคารพาณิชย์ในประเทศไทย อย่างมีนัยสำคัญทางสถิติที่ระดับ 0.05 คือ ดัชนีราคาผู้บริโภค และอัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี โดยมีความสัมพันธ์ในทิศทางเดียวกัน แต่สำหรับอัตราส่วนสินเชื่อจัดชั้นด้อยคุณภาพ และอัตราส่วนเงินให้สินเชื่อต่อเงินรับฝากมีความสัมพันธ์ในทิศทางตรงข้าม

2. วัตถุประสงค์การวิจัย

1. เพื่อพัฒนาแบบจำลองในการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย ด้วยเทคนิควิธีเหมืองข้อมูล
2. เพื่อเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย

3. วิธีดำเนินการวิจัย

จากงานวิจัยที่เกี่ยวข้อง ปัจจัยที่ส่งผลต่อปริมาณสินเชื่อของธนาคารพาณิชย์ ประกอบด้วย ปริมาณเงินฝาก อัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี และ ปริมาณเงินกู้ที่ไม่ก่อให้เกิดรายได้ ผู้วิจัยจึงนำมาพัฒนาเป็นกรอบแนวคิดในการวิจัย เรื่องการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ โดยเทคนิควิธีเหมืองข้อมูล มีตัวแปรอิสระ คือ ปริมาณเงินฝากของธนาคารพาณิชย์ในประเทศไทย อัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี ปริมาณเงินกู้ที่ไม่ก่อให้เกิดรายได้ และตัวแปรตาม คือ ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดในการวิจัยเรื่องการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์โดยเทคนิควิธีเหมืองข้อมูล

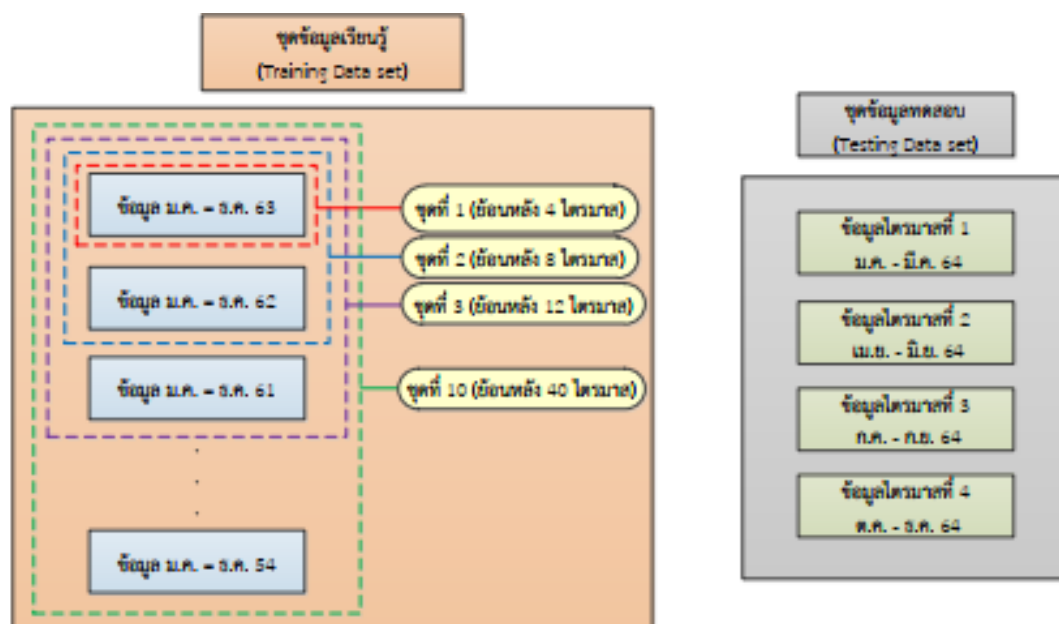
ในการดำเนินการวิจัยมีขั้นตอนที่เกี่ยวข้องแบ่งได้เป็น 2 ส่วน คือ

1. การเก็บรวบรวมข้อมูล ผู้วิจัยได้รวบรวมข้อมูลทุติยภูมิเกี่ยวกับปัจจัยดังกล่าวซึ่งเป็นข้อมูลรายไตรมาส โดยใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม 2554 ถึงวันที่ 31 ธันวาคม 2564 จากเว็บไซต์ที่เผยแพร่ของธนาคารแห่งประเทศไทย (ธนาคารแห่งประเทศไทย, 2565) ประกอบด้วย ปริมาณเงินฝากของธนาคารพาณิชย์ในประเทศไทย อัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี ปริมาณเงินกู้ที่ไม่ก่อให้เกิดรายได้ และปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย สำหรับรายละเอียดของปัจจัยแสดงได้ดังตารางที่ 1

ตารางที่ 1 ปัจจัยที่มีผลต่อปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย

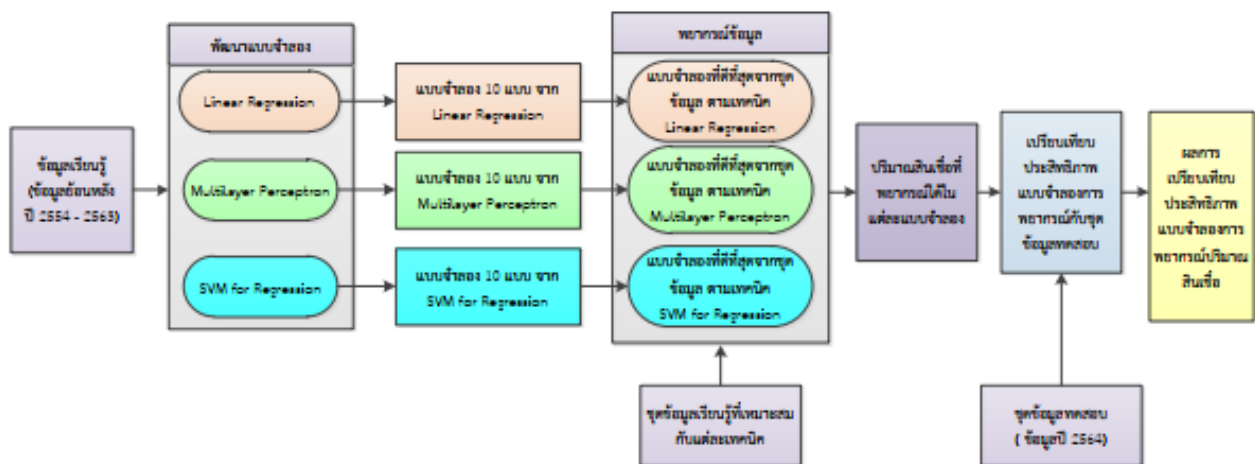
ลำดับ	ปัจจัย	ชื่อย่อ	ค่าตัวแปร	หน่วย	รายละเอียด
1	Deposit	DEP	ตัวเลข	ล้านบาท	ปริมาณเงินฝากของธนาคารพาณิชย์ในประเทศไทย
2	Minimum Loan Rate	MLR	ตัวเลข	ร้อยละ	อัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี
3	Non-Performing Loan	NPL	ตัวเลข	ล้านบาท	ปริมาณเงินกู้ที่ไม่ก่อให้เกิดรายได้
4	Credit	CDT	ตัวเลข	ล้านบาท	ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย

ทั้งนี้ผู้วิจัยได้จัดแบ่งข้อมูลออกเป็น 2 ส่วน (ดังภาพที่ 2) ได้แก่ ชุดข้อมูลเรียนรู้ (Training Data set) เป็นชุดข้อมูลสำหรับสร้างแบบจำลองการพยากรณ์ โดยใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม 2554 ถึงวันที่ 31 ธันวาคม 2563 และชุดข้อมูลทดสอบ (Testing Data set) ใช้ข้อมูลตั้งแต่วันที่ 1 มกราคม ถึง 31 ธันวาคม 2564 เพื่อเปรียบเทียบประสิทธิภาพของเทคนิควิธีเหมือนข้อมูลที่ใช้ในการสร้างแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ การเตรียมข้อมูลเพื่อนำเข้าสู่อัลกอริธึมในการสร้างแบบจำลองนั้น ผู้วิจัยใช้โปรแกรมสเปรดชีตในการจัดการข้อมูลให้อยู่ในรูปแบบของแฟ้มข้อมูลชนิด CSV โดยได้จัดกลุ่มชุดข้อมูลเรียนรู้ (Training Data set) แบบข้อมูลย้อนหลัง (Lagged Data) ซึ่งมีชุดข้อมูลทั้งหมด 10 กลุ่ม ได้แก่ ชุดย้อนหลัง 4 ไตรมาส (ม.ค. - ธ.ค. 63) ชุดย้อนหลัง 8 ไตรมาส (ม.ค. 62 - ธ.ค. 63) ชุดย้อนหลัง 12 ไตรมาส (ม.ค. 61 - ธ.ค. 63) ชุดย้อนหลัง 16 ไตรมาส (ม.ค. 60 - ธ.ค. 63) ชุดย้อนหลัง 20 ไตรมาส (ม.ค. 59 - ธ.ค. 63) ชุดย้อนหลัง 24 ไตรมาส (ม.ค. 58 - ธ.ค. 63) ชุดย้อนหลัง 28 ไตรมาส (ม.ค. 57 - ธ.ค. 63) ชุดย้อนหลัง 32 ไตรมาส (ม.ค. 56 - ธ.ค. 63) ชุดย้อนหลัง 36 ไตรมาส (ม.ค. 55 - ธ.ค. 63) และชุดย้อนหลัง 40 ไตรมาส (ม.ค. 54 - ธ.ค. 63) ในการป้อนเข้าสู่กระบวนการวิเคราะห์ข้อมูล ด้วยเทคนิคเหมือนข้อมูลทั้ง 3 แบบ เพื่อใช้ในการวิเคราะห์หาชุดข้อมูลย้อนหลังที่มีความเหมาะสมในการสร้างแบบจำลองการพยากรณ์ปริมาณสินเชื่อ โดยแบบจำลองที่ได้จะถูกนำไปทดสอบประสิทธิภาพกับชุดข้อมูลทดสอบ (Testing Data set) คือชุดข้อมูล 4 ไตรมาส (ม.ค. - ธ.ค. 64) และนำเสนอค่าความคลาดเคลื่อนในรูปแบบของข้อมูลรายไตรมาส



ภาพที่ 2 ชุดข้อมูลเรียนรู้ (Training Data set) และชุดข้อมูลทดสอบ (Testing Data set)

2. การวิเคราะห์ข้อมูล การสร้างแบบจำลองการพยากรณ์ปริมาณสินเชื่อใช้เทคนิคเหมือนข้อมูล 3 แบบ ได้แก่ การวิเคราะห์การถดถอย (Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Vector Machine for Regression) โดยนำแบบจำลองที่ได้จากเทคนิคแต่ละแบบมาทำการเปรียบเทียบประสิทธิภาพ ซึ่งในการวิจัยนี้แบ่งการวิเคราะห์ออกเป็น 2 ส่วน คือ 1) การวิเคราะห์เพื่อหาชุดข้อมูลเรียนรู้ (Training Data set) ที่มีความสามารถในการสร้างแบบจำลองที่มีประสิทธิภาพและเหมาะสมกับเทคนิควิธีเหมือนข้อมูลทั้ง 3 เทคนิค โดยใช้การวัดค่าเฉลี่ยความคลาดเคลื่อนสัมบูรณ์ (Mean Absolute Error : MAE) และค่ารากของความคลาดเคลื่อนกำลังสอง (Root Mean Square Error : RMSE) 2) การทดสอบประสิทธิภาพของแบบจำลองกับชุดข้อมูลทดสอบ (Testing Data set) โดยประมาณการความแม่นยำในการพยากรณ์จากค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error : MRE) และใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error : MMRE) ในการทดสอบประสิทธิภาพโดยรวมของแบบจำลอง ดังกระบวนการตามภาพที่ 3



ภาพที่ 3 กระบวนการพัฒนาแบบจำลองและการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์

4. ผลการวิจัย

ผู้วิจัยได้ทำการสร้างและเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ ด้วยโปรแกรม WEKA Version 3.9.6 โดยใช้เทคนิควิธีเหมือนข้อมูลทั้ง 3 เทคนิค กับชุดข้อมูลเรียนรู้ (Training Data set) และชุดข้อมูลทดสอบ (Testing Data set) ตามขั้นตอนการวิจัย ผลการทดสอบอธิบายได้เป็น 2 ส่วนดังนี้

1. การวิเคราะห์เพื่อหาชุดข้อมูลเรียนรู้ (Training Data set) ที่มีความสามารถในการสร้างแบบจำลองที่มีประสิทธิภาพ จากเทคนิควิธีเหมือนข้อมูลทั้ง 3 เทคนิค และเปรียบเทียบความคลาดเคลื่อนด้วยการวัดค่า MAE และ RMSE ดังตารางที่ 2

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์โดยชุดข้อมูลเรียนรู้

ชุดข้อมูลย้อนหลัง	Linear Regression		Multilayer Perceptron		SMOreg	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
4 ไตรมาส	1,127,367.83	1,337,213.63	1,266,851.73	1,564,992.72	2,406,653.03	4,358,101.11
8 ไตรมาส	252,755.26	348,729.61	324,531.22	385,871.30	188,882.16	312,018.43
12 ไตรมาส	307,343.61	368,213.50	307,114.66	384,507.15	277,102.27	350,881.87
16 ไตรมาส	423,180.51	563,590.18	361,344.65	439,179.89	406,037.00	571,509.26
20 ไตรมาส	462,552.59	539,225.35	505,940.58	591,336.13	542,310.50	610,162.52

ตารางที่ 2 การเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ปริมาณสินค้าของธนาคารพาณิชย์โดยชุดข้อมูลเรียนรู้ (ต่อ)

ชุดข้อมูลย้อนหลัง	Linear Regression		Multilayer Perceptron		SMOreg	
	MAE	RMSE	MAE	RMSE	MAE	RMSE
24 ไตรมาส	427,234.20	514,659.63	324,100.33	404,587.37	421,186.63	531,859.92
28 ไตรมาส	440,588.57	525,030.08	327,151.11	408,399.48	459,795.82	570,390.80
32 ไตรมาส	377,007.03	470,339.68	393,169.40	455,189.99	453,433.84	534,960.75
36 ไตรมาส	421,648.54	526,601.31	489,290.39	593,476.97	511,853.19	613,694.13
40 ไตรมาส	459,722.83	561,316.36	526,060.01	624,389.61	482,247.47	601,454.63

จากตารางที่ 2 แสดงการเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ปริมาณสินค้าโดยใช้ชุดข้อมูลเรียนรู้จากเทคนิคการวิเคราะห์การถดถอย (Linear Regression) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น (Multilayer Perceptron) และซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย (Support Vector Machine for Regression) เมื่อพิจารณาค่า MAE และค่า RMSE เป็นดัชนีชี้วัดประสิทธิภาพในการสร้างแบบจำลองที่ดีที่สุด พบว่าการใช้ชุดข้อมูลย้อนหลัง 8 ไตรมาสมีความเหมาะสมกับเทคนิคการวิเคราะห์การถดถอย โดยมีค่า MAE ต่ำสุดเท่ากับ 252,755.26 และค่า RMSE ต่ำสุดเท่ากับ 348,729.61 ชุดข้อมูลย้อนหลัง 12 ไตรมาสมีความเหมาะสมกับเทคนิคแบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีค่า MAE ต่ำสุดเท่ากับ 307,114.66 และค่า RMSE ต่ำสุดเท่ากับ 384,507.15 ชุดข้อมูลย้อนหลัง 8 ไตรมาสมีความเหมาะสมกับเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีค่า MAE ต่ำสุดเท่ากับ 188,882.16 และค่า RMSE ต่ำสุดเท่ากับ 312,018.43 จากผลการทดลองผู้วิจัยจึงเลือกชุดข้อมูล และเทคนิควิธีเหมืองข้อมูลที่มีประสิทธิภาพดีที่สุดในการสร้างแบบจำลองเพื่อการพยากรณ์ปริมาณสินค้ากับชุดข้อมูลทดสอบ

2. การทดสอบประสิทธิภาพของแบบจำลองกับชุดข้อมูลทดสอบ จากการนำแบบจำลองการพยากรณ์ที่มีประสิทธิภาพมากที่สุดมาทดสอบกับชุดข้อมูลทดสอบ แล้วใช้ค่าความคลาดเคลื่อนสัมพัทธ์ (Magnitude of Relative Error : MRE) ชี้วัดประสิทธิภาพในการพยากรณ์รายไตรมาส และใช้ค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์ (Mean Magnitude of Relative Error : MMRE) ในการทดสอบประสิทธิภาพโดยรวมของแบบจำลอง ดังตารางที่ 3

ตารางที่ 3 การเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ปริมาณสินค้าของธนาคารพาณิชย์ในแต่ละไตรมาส

ชุดข้อมูล ทดสอบ ปี 2564	Linear Regression			Multilayer Perceptron		SMOreg	
	ข้อมูลย้อนหลัง			ข้อมูลย้อนหลัง		ข้อมูลย้อนหลัง	
	8 ไตรมาส			12 ไตรมาส		8 ไตรมาส	
ไตรมาส	ค่าจริง (ล้านบาท)	ค่าพยากรณ์ (ล้านบาท)	MRE	ค่าพยากรณ์ (ล้านบาท)	MRE	ค่าพยากรณ์ (ล้านบาท)	MRE
ม.ค. - มี.ค.	51,554,705	50,440,012	0.0216	50,048,820	0.0292	50,240,156	0.0255
เม.ย. - มิ.ย.	52,847,100	51,190,846	0.0313	50,098,850	0.0520	50,723,645	0.0402
ก.ค. - ก.ย.	52,577,464	51,288,144	0.0245	50,103,863	0.0470	50,782,845	0.0341
ต.ค. - ธ.ค.	53,298,400	52,003,247	0.0243	50,117,677	0.0597	50,904,445	0.0449
MMRE			2.54%		4.70%		3.62%

การเปรียบเทียบประสิทธิภาพแบบจำลองการพยากรณ์ปริมาณสินเชื่อในแต่ละไตรมาส โดยเปรียบเทียบกับข้อมูลทดสอบ 4 ไตรมาส พบว่าแบบจำลองการพยากรณ์ด้วยเทคนิคการวิเคราะห์การถดถอย มีประสิทธิภาพโดยรวมสูงสุดพิจารณาจากค่า MMRE เท่ากับ 2.54% รองลงมาคือ การพยากรณ์ด้วยแบบจำลองจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีค่า MMRE เท่ากับ 3.62% และเมื่อพิจารณาการพยากรณ์ในแต่ละไตรมาสพบว่าเทคนิคการวิเคราะห์การถดถอย มีประสิทธิภาพสูงสุดในไตรมาสที่ 1 (ม.ค. – มี.ค. 64) ค่า MRE เท่ากับ 0.0216 รองลงมาคือในไตรมาสที่ 4 (ต.ค. – ธ.ค. 64) มีค่า MRE เท่ากับ 0.0243

5. อภิปรายผลการวิจัยและข้อเสนอแนะ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนา และเปรียบเทียบประสิทธิภาพของแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทยด้วยเทคนิควิธีเหมืองข้อมูล 3 เทคนิค ได้แก่ 1) การถดถอยเชิงเส้น 2) แบบจำลองโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น และ 3) ซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย โดยชุดข้อมูลที่นำมาใช้ศึกษาเป็นข้อมูลรายไตรมาส ประกอบด้วยปริมาณเงินฝากของธนาคารพาณิชย์ อัตราดอกเบี้ยเงินกู้ลูกค้ารายใหญ่ชั้นดี ปริมาณเงินกู้ที่ไม่ก่อให้เกิดรายได้ และปริมาณสินเชื่อของธนาคารพาณิชย์ในประเทศไทย ตั้งแต่วันที่ 1 มกราคม 2554 ถึงวันที่ 31 ธันวาคม 2564 โดยแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ชุดข้อมูลเรียนรู้ (Training Data set) เป็นชุดข้อมูลสำหรับสร้างแบบจำลองการพยากรณ์ปริมาณสินเชื่อ และชุดข้อมูลทดสอบ (Testing Data set) เพื่อเปรียบเทียบประสิทธิภาพของเทคนิควิธีเหมืองข้อมูล

จากผลการทดลองพบว่าชุดข้อมูลย้อนหลัง 8 และ 12 ไตรมาสมีความเหมาะสมในการใช้เป็นชุดข้อมูลเพื่อสร้างแบบจำลองในการพยากรณ์ และเมื่อพิจารณาประสิทธิภาพของแบบจำลองกับชุดข้อมูลทดสอบในปี 2564 พบว่าแบบจำลองการพยากรณ์ด้วยเทคนิคการวิเคราะห์การถดถอยมีประสิทธิภาพโดยรวมสูงสุด มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 2.54% รองลงมาคือแบบจำลองจากเทคนิคซัพพอร์ตเวกเตอร์แมชชีนสำหรับการถดถอย มีค่าเฉลี่ยความคลาดเคลื่อนสัมพัทธ์เท่ากับ 3.62%

จากผลการวิจัยแสดงให้เห็นประเด็นสำคัญเกี่ยวกับการใช้ข้อมูลย้อนหลังในการพยากรณ์ปริมาณสินเชื่อมีความเหมาะสมอยู่ในช่วง 8 และ 12 ไตรมาส ซึ่งอาจจะสะท้อนให้เห็นว่าการพัฒนาแบบจำลองไม่ได้ขึ้นกับปริมาณข้อมูล แต่อาจจะขึ้นกับความสอดคล้องของข้อมูลกับเทคนิควิธีที่เลือกใช้ในการพยากรณ์ และจากการเปรียบเทียบเทคนิคการเลือกใช้แบบจำลองในการพยากรณ์ แสดงให้เห็นว่าการใช้แบบจำลองด้วยเทคนิคโครงข่ายประสาทเทียมแบบเปอร์เซ็ปตรอนหลายชั้น มีความผิดพลาดสูงที่สุดโดยค่า MMRE เท่ากับ 4.70% แต่ก็ยังอยู่ในเกณฑ์ต่ำ ดังนั้นจึงขึ้นกับผู้ใช้งานที่จะกำหนดเกณฑ์ความผิดพลาดที่สามารถยอมรับได้

จากการศึกษาและพัฒนาแบบจำลองการพยากรณ์ปริมาณสินเชื่อของธนาคารพาณิชย์ โดยเทคนิควิธีเหมืองข้อมูล สามารถสรุปข้อเสนอแนะได้ดังต่อไปนี้

1. เนื่องจากอาจมีปัจจัยอื่นๆ ที่ส่งผลกระทบต่อปริมาณสินเชื่อของธนาคารพาณิชย์ จึงสามารถปรับเปลี่ยนปัจจัยเพื่อเพิ่มประสิทธิภาพในการพัฒนาแบบจำลองในการพยากรณ์ได้
 1. การพยากรณ์อาจใช้เทคนิคอื่นเพิ่มเติม ที่มีการพิจารณาถึงปัจจัยที่เกี่ยวข้องเหล่านี้เข้ามาร่วมในการวิเคราะห์ เพื่อให้ได้แบบจำลองที่มีประสิทธิภาพในการพยากรณ์มากขึ้น
 2. การนำแบบจำลองการพยากรณ์ที่ได้ไปใช้พัฒนาร่วมกับโปรแกรมประยุกต์ เพื่อให้สามารถรับข้อมูลที่เป็นปัจจัยในการพยากรณ์ได้ จะช่วยให้เกิดประสิทธิภาพ และความคล่องตัวในการใช้ผลของการพยากรณ์ได้มากขึ้น

6. เอกสารอ้างอิง

- เจนจิรา ช่วยคงมา และอภิชาติ พงศ์สุพัฒน์. (2564). *ปัจจัยที่มีผลกระทบต่อปริมาณเงินให้สินเชื่อภาคธุรกิจของธนาคารพาณิชย์ในประเทศไทย* (รายงานผลการวิจัย). กรุงเทพฯ: คณะบริหารธุรกิจ มหาวิทยาลัยเกษตรศาสตร์.
- ธนาคารแห่งประเทศไทย. (2565). *สถิติเงินรับฝาก สินเชื่อ และเงินกู้ยืม*. สืบค้น 3 พฤษภาคม 2565, จาก <https://www.bot.or.th/Thai/Statistics/FinancialInstitutions/Pages/StatDepositsAndLoans.aspx>

- นครินทร์ ภูมิตะ และ ศิริขวัญ เจริญวิริยะกุล. (2563). *ปัจจัยที่มีอิทธิพลต่อปริมาณเงินให้สินเชื่อของธนาคารกรุงเทพ ธนาคารกรุงไทย และธนาคารไทยพาณิชย์ จำกัด (มหาชน)* (รายงานผลการวิจัย). ชลบุรี: คณะเศรษฐศาสตร์ ศรีราชา มหาวิทยาลัยเกษตรศาสตร์.
- พิมพ์ลักษณ์ พัฒนชัย. (2559). *ปัจจัยที่มีผลต่อปริมาณเงินกู้ยืมของธนาคารพาณิชย์ไทย* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยกรุงเทพ.
- สุนันทา พรหมมาศ. (2559). *ปัจจัยที่ส่งผลกระทบต่อปริมาณการปล่อยสินเชื่อของธนาคารไทยพาณิชย์ จำกัด (มหาชน)* (วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยกรุงเทพ.
- Michalski, R. S., Bratko, I., & Kubat, M. (1999). *Machine learning and data mining methods and applications*. USA: John Wiley & Sons, Inc.
- Olson, D., & Shi, Y. (2007). *Introduction to business data mining*. Singapore: McGraw-Hill.
- Pyle, D. (2003). *Business modeling and data mining*. USA: Morgan Kaufmann.
- Tryfos, P. (1998). *Methods for business analysis and forecasting : text and cases*. USA: John Wiley & Sons, Inc.