ระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติ ด้วยเทคนิควิธีการเรียนรู้ของเครื่อง

AUTOMATED SYSTEM FOR EMPLOYEE WELFARE LOAN APPROVAL THROUGH MACHINE LEARNING TECHNIQUES

สุดากาญจน์ ติดตะ[1]

ดร.ธนภัทร ฆังคะจิตร[2]

บทคัดย่อ

เนื่องจากสภาวะเศรษฐกิจถดถอยจากสถานการณ์โควิด 19 บุคลากรมหาวิทยาลัยจำนวนมากได้รับผลกระทบ ส่งผลให้จำนวนผู้ขอกู้กองทุนสวัสดิการมีจำนวนเพิ่มมากขึ้นอย่างมาก เป็นการเพิ่มภาระให้เจ้าหน้าที่ต้องตรวจสอบเอกสารและรายละเอียดของผู้ขอกู้จำนวนมากขึ้น จากงานวิจัยก่อนหน้าที่พบว่าปัจจัยที่มีผลต่อการชำระหนี้เงินกู้ได้แก่ ประวัติสินเชื่อของผู้ขอสินเชื่อและความรับผิดของผู้ค้ำประกันร่วมกับเทคนิคการเรียนรู้ของเครื่อง ซึ่งให้ความแม่นยำค่อนข้างสูง ดังนั้นงานนี้จึงได้นำเสนอแบบจำลองสำหรับพิจารณาอนุมัติเงินกู้สวัสดิการโดยใช้เทคนิคการเรียนรู้ของเครื่อง เพื่อลดภาระการทำงานของเจ้าหน้าที่ในการอนุมัติเอกสารขอกู้เงิน โดยใช้ปัจจัยที่เกี่ยวข้องกับข้อมูลผู้กู้ ผู้ค้ำประกัน และความเสี่ยงของผู้กู้และผู้ค้ำประกัน ผลการศึกษาพบว่าเทคนิคแรนดอมฟอร์เรสให้ความแม่นยำสูงสุดที่ 99.56% โดยปัจจัยที่มีความสำคัญในการทำนายการอนุมัติเงินกู้สวัสดิการ ได้แก่ ความเสี่ยงในการค้ำประกันและเงินเดือนพนักงานของผู้สมัครและผู้ค้ำประกันตามลำดับ ซึ่งสามารถนำปัจจัยสำคัญเหล่านี้ไปใช้ในการกำหนดนโยบายขอกู้เงินในอนาคตได้อย่างเหมาะสม

คำสำคัญ: การอนุมัติการขอกู้เงินสวัสดิการ การจำแนกประเภท และทำนายการเรียนรู้ด้วยของเครื่อง

ABSTRACT

According to the COVID-19 pandemic, Thailand's economy was severely affected, especially on the issue of labor market. The cash shortage situation has widely expanded due to the declines in income. Consequently, the significantly increasing number of university employees have applied for welfare loans. The loan officers have to put more effort to examine more documents. The former research has shown that the important factors for loan approval are the credit history of loan applicants and surety's liability. This work attempts to propose a loan welfare approval model by using machine learning techniques. The input features include several properties of loan applicants and surety, especially their jobs status, financial data, and risk scores. Experimental results showed that the highest accuracy of 97.50% can be obtained by random forests technique. The most important factors are surety risk and employee salary of applicant, and surety, respectively. Therefore, these factors are capable of formulating suitable future loan policies.

Keywords: Approval of loan welfare, Classification, Prediction Machine learning

บทนำ

การจัดสวัสดิการและสิทธิประโยชน์เป็นกระบวนการหนึ่งของการบริหารงานบุคคล ซึ่งมีความมุ่งหมายที่จะให้ความสะดวกสบายความพึงพอใจของผู้ปฏิบัติงานตลอดจนความอบอุ่นใจ หมดกังวล ในเรื่องต่าง ๆ และให้ผู้ปฏิบัติงานมีกำลังใจในการทำงานเกิดความเชื่อมั่นในองคก์ร ซึ่งจะเป็นปัจจัยเสริมส่งการบริหารงานขององค์การดำเนินไปอย่างมีประสิทธิภาพ และบรรลุวัตถุประสงค์ (อุทัย หิรัญโต, 2531)

ทางมหาวิทยาลัยได้นำหลักการมาใช้เป็นแนวทางในการบริหารโดยจัดให้มีสวัสดิการมหาวิทยาลัยเพื่อให้สมาชิกได้รับสิทธิประโยชน์ โดยกำหนดให้มีสวัสดิการต่าง ๆ ดังนี้ 1. กองทุนสำรองเลี้ยงชีพ 2. เงินกู้กองทุนสวัสดิการ 3. เงินกู้เพื่อที่อยู่อาศัย/ เอนกประสงค์ 4. เงินช่วยเหลือกรณีประสบภัยพิบัติ 5. เงินช่วยเหลือกรณีเสียชีวิต 6. เงินช่วยเหลือกรณีประสบอุบัติเหตุ 7. เงินช่วยเหลือค่ารักษาพยาบาล

8. เงินชดเชย 9. เงินบำเหน็จ/ บำนาญ เนื่องด้วยสถานการณ์โควิด-19 ทำให้บุคลากรมหาวิทยาลัยได้รับผลกระทบจำนวนมาก ส่งผลให้จำนวนผู้ขอกู้เงินกองทุนสวัสดิการมีจำนวนที่เพิ่มมากขึ้น

ตั้งแต่เดือนมกราคม-พฤษภาคม 2564 เทียบกับ เดือนมกราคม-พฤษภาคม 2563 เฉลี่ยเพิ่มขึ้น 10 % ส่งผลให้เจ้าหน้าที่ที่ต้องตรวจสอบเอกสารและรายละเอียดของผู้ขอกู้ต้องจำนวนมากขึ้น

จากการศึกษางานวิจัยโมเดลเพื่อทำนายความสามารถในการชำระหนี้เงินกู้ กรณีศึกษาสำนักงานส่งเสริมสวัสดิการและสวัสดิภาพครูและบุคลากรทางการศึกษา จังหวัดอุดรธานี (ญาณินี สิทธิสารและ พุธษดี ศิริแสงตระกูล, 2561) ได้หาปัจจัยที่มีผลต่อการชำระหนี้เงินกู้ โดยการศึกษาตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลเพื่อพิจารณาปัจจัยที่ผลต่อการเกิดหนี้ค้างชำระเงินกู้ได้แก่ จำนวนบุตร จำนวนเงินงวด อายุ จำนวนเงินกู้ รายได้ต่อเดือน ภาระหนี้ การทำประกัน เพศ ซึ่งแนวทางการวิจัยนี้ผู้วิจัยได้นำมาใช้ในการหาปัจจัยที่มีความสัมพันธ์กับการอนุมัติเอกสารการขอกู้เงินสวัสดิการช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย สอดคล้องกับงานวิจัยของ Wei Li และคณะ (Wei Li, 2010) ได้ศึกษาการจำแนกประเภทของลูกค้าที่มีบัตรเครดิต โดยใช้ตัวแปรผลที่ได้จากการจำแนกประเภทลูกค้าบัตรเครดิตโดยใช้เทคนิคต้นไม้การตัดสินใจ (Decision Tree) โดยมีปัจจัยที่สำคัญคือ รายได้ส่วนตัวต่อเดือน รายได้ครอบครัวต่อเดือน ค่าเฉลี่ยของจำนวนครั้งที่รูดบัตรต่อเดือนและสถานะเครดิตส่วนตัวส่งผลให้ความถูกต้องสูงกว่าวิธีอื่น และงานวิจัย Eiman Kambal และคณะ (Eiman, 2013) ปัจจัยสถานะเครดิตมีผลต่อการศึกษาปัจจัยการให้สินเชื่อของธนาคารในการให้กู้ยืม

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาโมเดลในการช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย ผู้วิจัยได้ศึกษาและรวบรวมปัจจัยที่มีผลต่อการทำนายการอนุมัติการขอกู้เงินสวัสดิการจำนวน 14 ปัจจัย ปัจจัยที่เป็นข้อมูลคุณภาพได้แก่ สถานะภาพ ประเภทพนักงาน ระดับการศึกษา เพศ สถานะการอนุมัติเงินกู้ ประเภทสายงาน และปัจจัยที่เป็นข้อมูลเชิงปริมาณ ได้แก่ อายุผู้กู้ อายุการทำงาน เงินเดือนผู้กู้ เงินเดือนผู้ค้ำประกัน อายุงานผู้ค้ำประกัน จำนวนเดือนที่กู้เงิน จำนวนเงินกู้ จำนวนบุตร โดยการทำนายการอนุมัติการขอกู้เงินสวัสดิการด้วยวิธีการจำแนกประเภทและทำนาย (Classification and Prediction) ด้วยการนำเทคนิควิธีนาอีฟเบย์ (Naïve Bayes) เทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor) เทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) ในการพัฒนาระบบซึ่งผลการวิเคราะห์จะแสดงการแบ่งกลุ่มลูกค้าออกเป็น 2 ประเภทคือกลุ่มผู้กู้ที่ได้รับการอนุมัติและกลุ่มผู้กู้ที่ไม่ได้รับการอนุมัติ หลังจากนั้นจึงหาความสัมพันธ์ของข้อมูลเพื่อเป็นแนวทางในการออกหลักเกณฑ์ในการกู้สวัสดิการ

วัตถุประสงค์

1. เพื่อสร้างแบบจำลองที่มีประสิทธิภาพสำหรับช่วยเจ้าหน้าที่พิจารณาการอนุมัติเอกสารการขอกู้เงินสวัสดิการมหาวิทยาลัย
2. เพื่อหาปัจจัยที่มีความสัมพันธ์กับการอนุมัติเอกสารการขอกู้เงินสวัสดิการช่วยในการออกหลักเกณฑ์ในการกู้สวัสดิการ

ขอบเขตของการวิจัย

1. ข้อมูลที่นำมาทำนายใช้ข้อมูลการขอกู้เงินสวัสดิการมหาวิทยาลัยตั้งแต่เดือนตุลาคม 2562 ถึงพฤษภาคม 2564 จำนวน 5,870 รายการ
2. ปัจจัยที่นำมาทำนายผลการอนุมัติขอกู้เงินสวัสดิการมหาวิทยาลัยมีทั้งหมดจำนวน 14 คุณสมบัติ ได้แก่ สถานะภาพ ประเภทพนักงาน ระดับการศึกษา เพศ สถานะการอนุมัติกู้เงิน ประเภทสายงาน อายุ อายุการทำงาน เงินเดือนผู้กู้ เงินเดือนผู้ค้ำประกัน อายุงานผู้ค้ำประกัน จำนวนเดือนที่กู้เงิน จำนวนเงินกู้ จำนวนบุตร

ประโยชน์ที่คาดว่าจะได้รับ

1. เพื่อช่วยลดขั้นตอนการขออนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย
2. เพื่อเป็นเครื่องมือในการช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิให้เจ้าหน้าที่

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

การศึกษาวิจัยนี้ เป็นการเทคนิคการวิเคราะห์ข้อมูลด้วยการจำแนกประเภทและการทำนายข้อมูล (Classification and Prediction) การอนุมัติเอกสารการขอกู้เงินสวัสดิการมหาวิทยาลัย โดยจำเป็นต้องศึกษาเอกสารและงานวิจัยเกี่ยวข้อง ดังรายการต่อไปนี้

1. เทคนิควิธีนาอีฟเบย์ (Naïve Bayes) มีพื้นฐานมาจากทฤษฎีของเบย์ (Bayes Theorem) เป็นเทคนิคที่ไม่ซับซ้อน เป็นขั้นตอนในการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้น เพื่อนำมาสร้างเงื่อนไขการจำแนกใหม่ โดยการใช้การคำนวณหาความน่าจะเป็นในการทำนายผล เป็นเทคนิคในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้ จะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์เหมาะกับกรณีของเซตตัวอย่างที่มีจำนวนมากและคุณสมบัติของตัวอย่างไม่ขึ้นต่อกัน (Vilailuck S, Jaroenpuntaruk V, Wichadakul D, 2015, p. 1-17)

2. เทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor) วิธีที่ใช้ในการจัดแบ่งคลาส โดยการตรวจสอบจำนวนบางจำนวน K ในขั้นตอน วิธีการหาเพื่อนบ้านใกลที่สุดของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไขหรือกรณีต่าง ๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกับคลาสที่ ใกล้เคียงกันมากที่สุด (T.L. Saaty, 2006, p. 4-5)

3. เทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เป็นสมการที่ใช้จำแนกค่าคุณลักษณะของ 2 กลุ่ม ที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกันโดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมาและเพื่อให้ทราบว่าเส้นตรงที่แบ่ง 2 กลุ่ม ออกจากกันนั้น เส้นตรงใดที่เป็นเส้นที่ดีที่สุด โดยเส้นตรงนั้นจะเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้นจะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มขึ้นมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด เคอร์เนล (Kernel) (อานนท์ นามสนิท, 2549, หน้า 53)

4. เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) เกิดจากการรวมกลุ่มกันของโครงสร้างต้นไม้ ซึ่งค่าความความเคลื่อนโดยรวมของป่าไม้จะถูกเปลี่ยนให้เป็นค่าลิมิต ทำให้จำนวนต้นไม้ในป่าเพิ่มขึ้น ค่าความคลาดเคลื่อนโดยรวมจะขึ้นกับความมั่นคง (Strength) ของต้นไม้แต่ละต้น รวมถึงความสัมพันธ์กันระหว่างต้นไม้เหล่านั้น โดยจะใช้วิธีการสุ่มเลือกคุณสมบัติเพื่อการแบ่งแยกโหนดทำให้ค่าความผิดพลาดลดลง (L. Breiman, 2001, p. 532)

งานวิจัยที่เกี่ยวข้อง

Wei Li และคณะ (Wei Li, 2010) ได้ศึกษาการจำแนกประเภทของลูกค้าที่มีบัตรเครดิตเป็น 4 รูปแบบ เพื่อเป็นข้อมูลพื้นฐานในการสร้างระบบทำนายความสามารถในการชำระหนี้ของลูกค้าบัตรเครดิตให้แก่ธนาคาร โดยใช้รูปแบบ C5.0, Neural network, Chi-squared automatic in traction detector และ Classification and regression tree ข้อมูลประกอบด้วยรายได้ส่วนตัวต่อเดือนและรายได้ครอบครัวต่อเดือน ค่าเฉลี่ยของจำนวนครั้งที่รูดบัตรต่อเดือนและสถานะเครดิตส่วนตัวมาใช้เป็นตัวแปรผลที่ได้จากการจำแนกประเภทลูกค้าบัตรเครดิต โดยใช้เทคนิค Decision Tree Model C5.0 ความถูกต้องสูงกว่าวิธีอื่น

จากข้อมูลของธนาคารซูดาน Eiman kambal และคณะ (Eiman, 2013) ได้พบว่าหนึ่งในปัจจัยความสำเร็จที่สำคัญขององค์กรสินเชื่อและธนาคารในการให้กู้ยืมโดยเฉพาะการประเมินความคุ้มค่าของผู้กู้ รูปแบบการจัดลำดับเครดิตได้ถูกนำมาใช้โดยนักวิจัยจำนวนมากที่จะปรับปรุงกระบวนการของการประเมินเครดิต โดยให้ผู้กู้ยืมเงินในอนาคตอยู่บนพื้นฐานของความน่าจะเป็น

ของการชำระหนี้ การทดลองสร้างแบบจำลองการจัดลำดับเครดิตที่เหมาะสมสำหรับธนาคารซูดาน คือ Decision Tree (DT) และเครือข่ายประสาทเทียม (ANN)

โมเดลเพื่อทำนายความสามารถในการชำระหนี้เงินกู้ กรณีศึกษาสำนักงานส่งเสริมสวัสดิการและสวัสดิภาพครูและบุคลากรทางการศึกษา จังหวัดอุดรธานี (ญาณินี สิทธิสารและพุธษดี ศิริแสงตระกูล, 2561) สร้างโมเดลเพื่อทำนายความสามารถในการชำระหนี้เงินกู้ โดยใช้ต้นไม้การตัดสินใจ เครือข่ายความเชื่อเบย์ และโครงข่ายประสาทเทียม ปัจจัยที่มีผลต่อการชำระหนี้เงินกู้ โดยผลการศึกษา 8 ตัวแปรใช้ในการวิเคราะห์ข้อมูล จำนวนบุตร จำนวนเงินงวด อายุ จำนวนเงินกู้ รายได้ต่อเดือน ภาระหนี้ การทำประกัน เพศ ผลประสิทธิภาพความแม่นยำการทำนายของโมเดลโครงข่ายประสาทเทียม 98.77 โมเดลเครือข่ายความเชื่อเบย์ 98.7 และโมเดลต้นไม้การตัดสินใจ 98.74

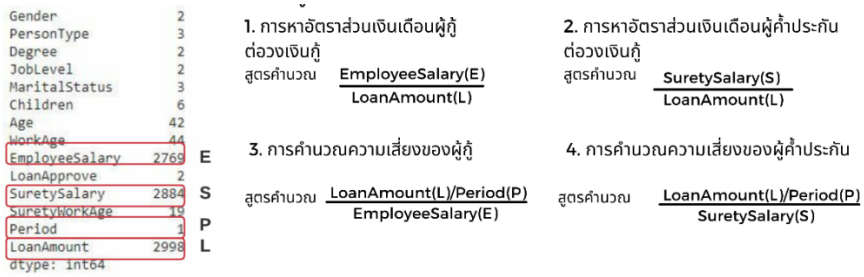
ระเบียบวิธีวิจัย

การศึกษาวิจัยเรื่องระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติด้วยเทคนิควิธีการเรียนรู้ด้วยของเครื่อง โดยมีแนวทางการวิจัยต่อไปนี้

1 การสำรวจข้อมูลและการเตรียมข้อมูล (Explore Data and Data Preparation)

1.1 การคัดเลือกข้อมูลและการอธิบายข้อมูล (Data Selection and Describe Data)

งานวิจัยนี้ได้เก็บรวบรวมข้อมูล เพื่อใช้ในการสร้างโมเดลโดยใช้ข้อมูลการขอยื่นกู้กองทุนสวัสดิการมหาวิทยาลัยรัฐบาลแห่งหนึ่ง โดยมีการคัดเลือกเฉพาะข้อมูลผู้ขอยื่นกู้ที่ได้รับการยินยอมค้ำประกันแล้วจำนวน 5,870 รายการ จากจำนวน 14 คุณสมบัติ ในขั้นตอนตรวจสอบคุณภาพของข้อมูล (Verify data quality) ทำให้ทราบว่าจำนวน 14 คุณสมบัติทำให้ไม่เพียงพอต่อการวิเคราะห์ข้อมูล ผู้วิจัยได้ทำการเพิ่ม 4 คุณสมบัติ มีรายละเอียดสูตรคำนวณดังภาพที่ 1 และรายละเอียดคุณสมบัติดังตารางที่ 2



ภาพที่ 1 เพิ่ม 4 คุณสมบัติและสูตรคำนวณ

รายละเอียดคุณสมบัติที่ใช้ในการวิเคราะห์ข้อมูลพิจารณาอนุมัติเงินกู้สวัสดิการพนักงาน 18 คุณสมบัติ โดยมีรายละเอียดดังนี้

ตารางที่ 1 รายละเอียดคุณสมบัติที่นำมาวิเคราะห์ข้อมูล

ข้อมูล	ประเภทข้อมูล	คำอธิบายข้อมูล
Degree	Qualitative data (Binary Variable)	ระดับการศึกษา
PersonType	Quantitative data (Ordinal Variable)	ประเภทพนักงาน
JobLevel	Quantitative data (Binary Variable)	ประเภทสายงาน
Gender	Quantitative data (Binary Variable)	เพศ
MaritalStatus	Quantitative data (Nominal Variable)	สถานะภาพ
LoanApprove	Quantitative data (Binary Variable)	สถานะการอนุมัติกู้เงิน
Age	Quantitative data	อายุ
WorkAge	Quantitative data	อายุการทำงาน
Children	Quantitative data	จำนวนบุตร
EmployeeSalary	Quantitative data	เงินเดือนผู้กู้
SuretySalary	Quantitative data	เงินเดือนผู้ค้ำประกัน
SuretyWorkAge	Quantitative data	อายุผู้ค้ำประกัน
Period	Quantitative data	จำนวนเดือนที่กู้เงิน
LoanAmount	Quantitative data	จำนวนเงินกู้
(EmployeeSalary/LoanAmount)	Quantitative data	สัดส่วนเงินเดือนผู้กู้ต่อวงเงินกู้
(SuretySalary/LoanAmount)	Quantitative data	สัดส่วนเงินเดือนผู้ค้ำประกันต่อวงเงินกู้
(LoanAmount/Period) /EmployeeSalary	Quantitative data	ความเสี่ยงของผู้กู้
(LoanAmount/Period) /SuretySalary	Quantitative data	ความเสี่ยงของผู้ค้ำประกัน

1.2 ทำความสะอาดข้อมูล (Clean Data)

จากการคัดกรองข้อมูลเป็นที่เรียบร้อยแล้ว ผู้วิจัยทำความสะอาดข้อมูลด้วย SQL SERVERที่มีความซ้ำซ้อนหรือผิดปกติของข้อมูลด้วยการ Replace Missing Values เติมข้อมูลเงินเดือนที่เป็นค่าว่างตามนโยบายมหาลัย เพื่อสร้างคุณภาพให้แก่ข้อมูลที่นำมาใช้อย่างแม่นยำ

1.3 การทำข้อมูลให้อยู่ในรูปแบบมาตรฐาน (Data Transformation)

เมื่อรวบรวมข้อมูลทั้งหมดแล้ว พบว่าข้อมูลบางส่วนยังไม่สามารถนำไปสร้างแบบจำลองได้ เนื่องจากรูปแบบของข้อมูลยังไม่เหมาะสมกับการสร้างแบบจำลอง ผู้วิจัยต้องแปลงข้อมูลจากข้อความเป็นตัวเลข เช่น ระดับการศึกษาประกอบด้วย 1 ปริญญาตรีขึ้นไป 2 ต่ำกว่าปริญญาตรี, ประเภทพนักงาน ประกอบด้วย 0 พนักงานมหาวิทยาลัย 1 ข้าราชการ 2 ลูกจ้างประจำ, เพศ ประกอบด้วย 0 ผู้หญิง 1 ผู้ชาย,สถานะการอนุมัติกู้เงิน ประกอบด้วย 0 ไม่อนุมัติ 1 อนุมัติ ประเภทสายงานประกอบด้วย 1 สายวิชาการ 2 สายสนันสนุน และได้มีการปรับบรรทัดฐานข้อมูลโดยใช้หลักการ Max-Min Normalization (เอกสิทธิ์ พัชรวงศ์ศักดา, 2557) จะทำ ให้ได้ข้อมูลอยู่ในช่วงระหว่าง 0 ถึง 1 ซึ่งสามารถปรับช่วงค่าข้อมูล เงินเดือนผู้กู้ เงินเดือนผู้ค้ำประกัน จำนวนเงินกู้ โดยใช้สมการที่ (2)

$$v' = \frac{v - min_A}{max_A - min_A}(new_max_A - new_min_A) + new_min_A \qquad (2)$$

2 การพัฒนาแบบจำลอง (Modeling)

2.1 การเลือกเทคนิคสร้างแบบจำลอง (Select Modeling Technique)

ขั้นตอนในการเลือกเทคนิคที่เหมาะสมกับประเภทของงานข้อจำกัด การแก้ไขปัญหาและสามารถนำมาใช้งานได้จริง ซึ่งในงานวิจัยนี้ผู้วิจัยได้มีการพัฒนาโมเดลเปรียบเทียบความแม่นยำทั้งหมด 4 แบบ ด้วยวิธี 1) Naive Bayes algorithm 2) K-Nearest neighbor algorithm 3) Support vector machines algorithm และ 4) Random Forest algorithm สำหรับจำแนกประเภทการทำนายผลการอนุมัติหรือไม่อนุมัติการขอกู้เงินสวัสดิการ

2.2 การสร้างแบบการทดสอบ (Generate Test Design)

ก่อนการสร้างแบบจำลองจะต้องทำการทดสอบความถูกต้องของโมเดล เช่น การนำระดับความผิดพลาดมาใช้ในการวัดคุณภาพของข้อมูลที่ใช้ในการทำเหมืองข้อมูล ดังนั้น การออกแบบการทดสอบควรจะมีการแยกข้อมูลออกเป็น 2 กลุ่ม คือ ชุดข้อมูลที่ใช้ในการทดสอบการ

สร้างแบบจำลองเพื่อประเมินคุณภาพ (Training Set) และชุดข้อมูลที่ใช้งานจริง (Test Set) ตั้งแต่เดือนตุลาคม 2562 ถึงพฤษภาคม 2564 จำนวน 5,870 รายการ

2.3 การสร้างแบบจำลอง (Build Model)

ในขั้นตอนนี้ผู้วิจัยได้สร้างโมเดลแบ่งเป็น 2 ขั้นตอน ขั้นตอนแรกเป็นการทำนายการอนุมัติการขอกู้เงินสวัสดิการด้วยวิธีการจำแนกประเภทการทำนายผลการอนุมัติหรือไม่อนุมัติการขอกู้เงินสวัสดิการด้วยเทคนิคในข้อ 2.1 ในการพัฒนาระบบซึ่งผลการวิเคราะห์จะแสดงการแบ่งกลุ่มลูกค้าออกเป็น 2 ประเภทคือกลุ่มผู้กู้ที่ได้รับการอนุมัติ (Training Set) สำหรับการ Train จำนวน 3,305 รายการ และ (Test Set) สำหรับทดสอบจำนวน 804 รายการ กลุ่มผู้กู้ที่ไม่ได้รับการอนุมัติ (Training Set) สำหรับการ Train จำนวน 1,429 รายการ และ (Test Set) สำหรับทดสอบจำนวน 332 รายการ

2.4 การประเมินแบบจำลอง (Assess Model)

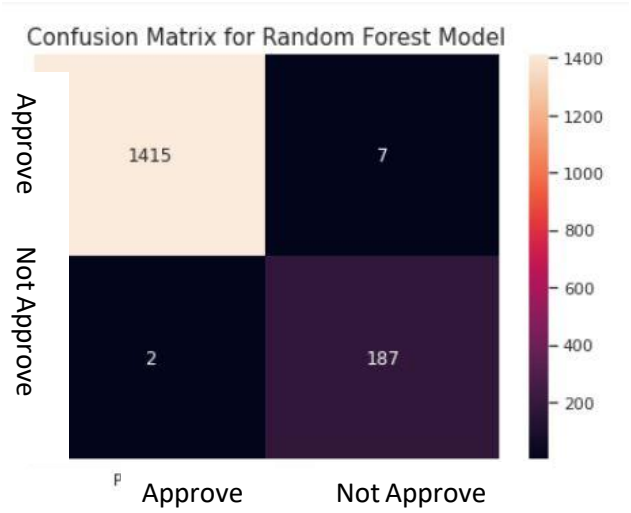
บันทึกผลการวิเคราะห์จากโมเดล โดนการประเมินแบบจำลองเพื่อให้มั่นใจว่ามีความน่าเชื่อถือ และผ่านมาตรฐานการตัดสินใจ และทำการปรับปรุงการตั้งค่าพารามิเตอร์และประเมินผลจนกว่าจะเชื่อมั่นได้ว่าเป็นรูปแบบของโมเดลดีที่ดีสุด โดยใช้วิธีการวิเคราะห์ข้อมูลและวัดประสิทธิภาพด้วย 4 ค่า คือ ค่าความระลึก (Recall) ค่าความแม่นยำ (Precision) ค่าเฉลี่ยของ Precision และ Recall (F1 Score) และค่าความถูกต้อง (Accuracy)

ผลการวิจัย

1. ผลการวัดประสิทธิภาพความถูกต้องของโมเดล

จากผลการวัดประสิทธิภาพความถูกต้องของโมเดลพบว่า เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) มีความแม่นจำสูงที่สุด (Accuracy) 0.9956 โดยมีผลการทำนายดังนี้

1) ผลการทำนายโดยใช้เทคนิควิธีแรนดอมฟอร์เรส (Random Forest)



ภาพที่ 2 ตาราง Random Forest Confusion matrix

Random Forest Model - Classification Report

	precision	recall	f1-score	support
1	1.00	1.00	1.00	1422
0	0.96	0.99	0.98	189
accuracy			0.99	1611
macro avg	0.98	0.99	0.99	1611
weighted avg	0.99	0.99	0.99	1611

ภาพที่ 3 ตาราง Random Forest Classification matrix

จากการวิเคราะห์ข้อมูล ผลการทำนายโดยใช้เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) ข้อมูลและวัดประสิทธิภาพด้วย 4 ค่า คือ ค่าความระลึก (Recall) ค่าความแม่นยำ (Precision) ค่าเฉลี่ยของ Precision และ Recall (F1 Score) และค่าความถูกต้อง (Accuracy) โดยมีรายเอียดดังตารางที่ 3

ตารางที่ 3 ประสิทธิภาพของโมเดลเทคนิควิธีแรนดอมฟอร์เรส (Random Forest)

ตัววัด	Precision	Recall	F1-score	Accuracy
Random Forest	0.98	0.99	0.99	0.99

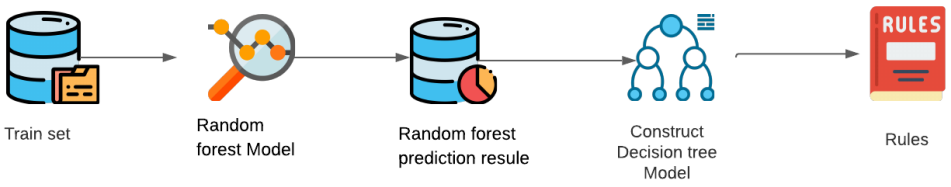
2. ค่าคลาดเคลื่อนในการประมาณการน้ำหนัก

จากการวัดประสิทธิภาพดีที่สุดในการทำนายการอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัยพบว่าเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) พบค่าความแม่นยำสูงสุดที่ 99.56 ซึ่งเกิดข้อผิดพลาดในการทำนายทั้งหมด 15 กรณีโดยมีสาเหตุดังนี้

1. กรณีเกิด False Positives (FP), Type I Error = 2 เนื่องจากอายุงานพนักงาน (WorkAge) ต่ำกว่า 3 ปี แต่อายุงานผู้ค้ำประกัน (SuretyWorkAge) มากกว่า 3 ปี

2. กรณีเกิด False Negatives (FN), Type II Error = 7 เนื่องจากอายุงานพนักงาน (WorkAge) ต่ำกว่า 3 ปี และอายุงานผู้ค้ำประกัน (SuretyWorkAge) น้อยกว่า 3 ปี

3. แนวทางการสกัด Rule จากเทคนิควิธีแรนดอมฟอร์เรส (Random Forest)



ภาพที่ 4 แนวทางการสกัด Rule จากเทคนิควิธีแรนดอมฟอร์เรส (Random Forest)

4. วิเคราะห์ข้อมูลจากโมเดลเงื่อนไขในการอนุมัติกู้เงิน

1. อายุงานผู้กู้ >= 4 ปี [3246/3757*100=86.39%]
2. อายุงานผู้กู้ <= 4 ปี
   2.1 อายุงานผู้ค้ำประกัน <= 4 และ > 2.63 และ เงินเดือนผู้กู้ต่อวงเงินกู้ <=3.63 [31/34*100=91.17%]

2.2 อายุงานผู้ค้ำประกัน <= 4 และ > 1.23 เงินเดือนผู้ค้ำประกันต่อวงเงินกู้ > 2.63 [6/26*100=23.07]

2.3 อายุงานผู้ค้ำประกัน <= 4 และ เงินเดือนผู้ค้ำประกันต่อวงเงินกู้ <= 1.23 และ เงินเดือนผู้กู้ต่อวงเงินกู้ <= 0.65 [1/3*100=33.33%]

5. หลักเกณฑ์การอนุมัติเงินกู้สวัสดิการจากการวิเคราะห์ข้อมูลจากโมเดลเงื่อนไขในการอนุมัติกู้เงิน

1. อนุมัติกู้เงินอายุงาน 3 ปี ขึ้นไป

1.1 ผู้กู้ต้องมีสัญญาจ้างตั้งแต่ 3 ปี ขึ้นไป

1.2 วงเงินกู้ที่กำหนดให้กู้ยืมได้ไม่เกินร้อยละ 70 ของเงินเดือน กรณีเกินให้พิจารณาเงินเดือน

1.3 ผู้ค้ำประกันต้องมากกว่าจำนวนวงเงินกู้ 1.5 เท่า

2. อนุมัติกู้เงินอายุงานต่ำกว่า 3 ปี

2.1 สัญญาจ้างต่ำกว่า 3 ปี ให้พิจารณาเงินเดือนผู้ค้ำประกันต้องมากกว่าจำนวนวงเงินกู้ 1.5 เท่า

2.2 วงเงินกู้ที่กำหนดให้กู้ยืมได้ไม่เกินร้อยละ 70 ของเงินเดือน

สรุปและอภิปรายผลการวิจัย

งานวิจัยนี้ได้นำเสนอเกี่ยวกับการพัฒนาโมเดลเพื่อช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย โดยสามารถสรุปผลการวิจัยได้ดังนี้

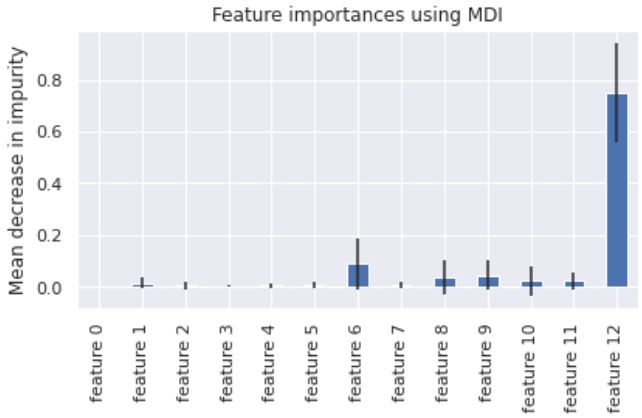
1. โมเดลที่มีประสิทธิภาพดีที่สุดในการทำนายการอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย ผลการศึกษาพบว่า ประสิทธิภาพการทำนายของโมเดลเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) พบค่าความแม่นยำสูงสุดที่ 99.56

	Naive Bayes	kNN	SVM	Random Forest
Accuracy	0.907439	0.975014	0.915389	0.9956

ภาพที่ 5 ผลประสิทธิภาพการทำนายของโมเดล

2. ปัจจัยที่มีความสำคัญ (Feature importance)

จากการสร้างแบบจำลองในการทำนายการกู้เงินสวัสดิการมหาวิทยาลัย มีการคัดเลือกปัจจัยทั้งหมด 13 ปัจจัย จากการทำนายปัจจัยที่มีผลมากที่สุดคือ Feature 12 ความเสี่ยงของผู้ค้ำประกัน (LoanAmount/Period)/SuretySalary รองลงมาคือ Feature 6 อายุผู้กู้ (Age) Feature 9 อัตราส่วนเงินเดือนผู้กู้ต่อวงเงิน (LoanAmount/Period)/EmployeeSalary Feature 8 อายุงานผู้ค้ำประกัน (Surety Age) Feature 10 อัตราส่วนเงินเดือนผู้ค้ำประกันต่อวงเงินกู้ (SuretySalary/LoanAmount) Feature 11ความเสี่ยงของผู้กู้ (LoanAmount/Period)/EmployeeSalary Feature 7 อายุการทำงาน (WorkAge) Feature 1 ประเภทพนักงาน(PersonType) Feature 2 ระดับการศึกษา (Degree) Feature 4 สถานะภาพ (MaritalStatus) Feature 5 จำนวนบุตร (Children) Feature 3 ประเภทสายงาน (JobLevel) Feature 0 เพศ (Gender) ผลการวิเคราะห์ดังภาพที่ 6



ภาพที่ 6 ผลประสิทธิภาพการทำนายของโมเดล

ข้อเสนอแนะ

ควรเพิ่มการทำนายโอกาสการเกิดหนี้เสีย หลังจากทำนายการอนุมัติขอกู้เงินนำผลจากการทำนายในส่วนที่ผ่านการอนุมัติกู้มาแล้วนำมาวิเคราะห์โอกาสการเกิดหนี้เสีย

บรรณานุกรม

ภาษาไทย

ญาณินี สิทธิสารและพุธษดี ศิริแสงตระกูล(2561): โมเดลเพื่อทำนายความสามารถในการชำระหนี้เงินกู้ กรณีศึกษาสำนักงานส่งเสริมสวัสดิการและสวัสดิภาพครูและบุคลากรทางการศึกษา จังหวัดอุดรธานี. วารสารการประชุมวิชาการ เสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 19, มหาวิทยาลัยขอนแก่น, ขอนแก่น, 1 น.

อานนท์ นามสนิท (2549): การจำแนกกลุ่มเพลงไทยโดยใช้ซัพพอร์ตเวกเตอร์แมชชีน, วิทยานิพนธ์ปริญญาโท, สถาบันเทคโนโลยีพระจอมเกล้าพระนครเหนือ, กรุงเทพฯ, 53

เอกสิทธิ์ พัชรวงศ์ศักดา (2557): การวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์เบื้องต้น. พิมพ์ครั้งที่ 2 . กรุงเทพฯ: เอเชียดิจิตอลการพิมพ์

อุทัย หิรัญโต. (2531): หลักการบริหารงานบุคคล. พิมพ์ครั้งที่ 2. กรุงเทพฯ: โอ.เอสพริ้นติ้งเฮ้า.

ภาษาต่างประเทศ

Breiman, L. (2001). Random Forests. Machine Learning Journal, 45(1), 2001, p. 532.

Eiman Kambal, Izzeldin Oaman, Methag Taha, Noon Mohammed, & Sara Mohammed. Credit Scoring Using Data Mining Techniques with Particular Reference to Sudanese Banks. (2013). *International Conference on Computing, Electrical and Electronic Engineering(ICCEEE)*, 26-28, pp.378-383.

Hengphraphrom K, Hengphraphrom S, & Makwibunchai. (2014). The Knowledge Discovery of Students' Critical Characteristics towards Learning Achievement of Nakhon Pathom Rajabhat University Students' Program in Computer via Data Mining Technique. *J Thai Interdiscipe Res*, 9(1):71–80.

Mohammed, J., Meira Wagner, J.R. (2014). *Data Mining and Analysis Fundamental Concepts and Algorithms.* Cambridge University Press. New York. USA.

Pete Chapman, Julian Clinton, et al. (2000). *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS inc 78.

T.L. Saaty. (2006). *Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy Process* (2nd ed.), Pittsburgh, PA: RWS, 2006, pp. 4-5.

VilailuckS, Jaroenpuntaruk V, and Wichadakul D. (2015). Utilizing Data Mining Techniques to Forecast Student Academic Achievement of Kasetsart University Laboratory School Kamphaeng Saen Campus Educational Research and Development Center. *Veridian E-J Sci Technol Silpakorn Univ*, 2(2), 1–17.

Wei Li, Xuemei Wu, Yayun Sun, Quanju Zhang. (2010). Credit Card Customer Segmentation and Target Marketing Based on Data Mining. *2010 International Conference on Computational Intelligence and Security (CIS)*, 11-14, pp.378-383.