DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติ ด้วยเทคนิควิธีการเรียนรู้ด้วยของเครื่อง

สุดากาญจน์ ติดตะ

สารนิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
วิศวกรรมศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมข้อมูลขนาดใหญ่
วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
ปีการศึกษา 2564

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

AUTOMATED SYSTEM FOR EMPLOYEE WELFARE LOAN APPROVAL USING MACHINE LEARNING TECHNIQUES

SUDAKAN TIDTA

An Independent Study Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Big Data Engineering,

College of Innovative Technology and Engineering,

Dhurakij Pundit University

Academic Year 2021

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ใบรับรองงานสารนิพนธ์

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิตย์

ปริญญา วิศวกรรมศาสตรมหาบัณฑิต

หัวข้อสารนิพนธ์ ระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติด้วยเทคนิควิธีการเรียนรู้ของเครื่อง

เสนอโดย สุดากาญจน์ ติดตะ

สาขาวิชา วิศวกรรมข้อมูลขนาดใหญ่

อาจารย์ที่ปรึกษาสารนิพนธ์ ดร.ธนภัทร ฆังคะจิตร

ได้พิจารณาเห็นชอบโดยคณะกรรมการสอบสารนิพนธ์แล้ว

.....ประธานกรรมการ

(ดร.สรรพฤทธิ์ มฤคทัต)

..... กรรมการและอาจารย์ที่ปรึกษา

(ดร.ธนภัทร ฆังคะจิตร)

..... กรรมการ

(ผู้ช่วยศาสตราจารย์ ดร.ดวงใจ จิตคงชื่น)

วิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์รับรองแล้ว

.....

(ดร.ชัยพร เขมะภาตะพันธ์)

คณบดีวิทยาลัยนวัตกรรมด้านเทคโนโลยีและวิศวกรรมศาสตร์

วันที่ 14 เดือน ธันวาคม พ.ศ. 2564

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

หัวข้อสารนิพนธ์	ระบบการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติด้วยเทคนิควิธีการเรียนรู้ด้วยของเครื่อง
ชื่อผู้เขียน	สุดากาญจน์ ติดตะ
อาจารย์ที่ปรึกษา	ดร. ธนภัทร ฆังคะจิตร
หลักสูตร	วิศวกรรมศาสตรมหาบัณฑิต
ปีการศึกษา	2564

บทคัดย่อ

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

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

Independent Study Title	AUTOMATED SYSTEM FOR EMPLOYEE WELFARE LOAN APPROVAL USING MACHINE LEARNING TECHNIQUES
Author	Sudakan Tidta
Independent Study Advisor	Dr. Thanapat Kangkachit
Department	Big Data Engineering
Academic Year	2021

ABSTRACT

According to the COVID-19 pandemic, Thailand's economy was severely impacted especially on the labor market. The cash shortage situation spread widely due to the declines in income. Consequently, the significantly increasing numbers of university employees have applied for welfare loans. The loan officers have to spend more effort to examine extra 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 proposes 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 show that the highest accuracy of 97.50% can be obtained by using the 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.

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

กิตติกรรมประกาศ

สารนิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้โดยการให้ความช่วยเหลือของ ดร.ธนภัทร ฆังคะจิตร ซึ่งเป็นอาจารย์ที่ปรึกษาสารนิพนธ์ ที่ได้กรุณาให้คำแนะนำตรวจสอบ และแก้ไขข้อบกพร่องต่างๆ มาโดยตลอดเพื่อให้สารนิพนธ์ฉบับนี้สมบูรณ์ผู้เขียนจึงขอกราบขอบพระคุณไว้ณโอกาสนี้

ผู้เขียนขอกราบขอบพระคุณ ดร.สรรพฤทธิ์ มฤคทัต ที่กรุณาให้เกียรติ เป็นประธานโดยมี ผศ.ดร.ดวงใจ จิตคงชื่น เป็นกรรมการในการสอบสารนิพนธ์ ซึ่งได้กรุณาตรวจแก้ไขสารนิพนธ์ฉบับนี้ให้ถูกต้องสมบูรณ์ยิ่งขึ้นและนางสาวกุลธิดา รอดบุญ รวมถึงเจ้าหน้าที่บัณฑิตมหาวิทยาลัยธุรกิจบัณฑิตย์ ทุกท่านที่ให้ความสะดวกด้านอำนวยการและประสานงาน และขอขอบพระคุณหัวหน้าฝ่ายบริหารข้อมูลทรัพยากร นางสาวสุทิสา ลี้อนันต์ศักดิ์ศิริ ในการสนับสนุนการทำสารนิพนธ์ให้กับผู้เขียน ทำให้การจัดทำสารนิพนธ์ของผู้เขียนในครั้งนี้สำเร็จลุล่วงไปด้วยดี

สุดากาญจน์ ติดตะ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ

	หน้า
บทคัดย่อภาษาไทย ……..	ฆ
บทคัดย่อภาษาอังกฤษ……..…..……	ง
กิตติกรรมประกาศ……	จ
สารบัญตาราง..……....	ซ
สารบัญภาพ ……	ฌ
บทที่	
1. บทนำ……..……..	1
1.1 ที่มาและความสำคัญของปัญหา……..……..……..	1
1.2 วัตถุประสงค์ของงานวิจัย……...…….	2
1.3 ขอบเขตงานวิจัย……..……...…….	2
1.4 ประโยชน์ที่คาดว่าจะได้รับ…….…….….....	3
1.5 นิยามศัพท์……....……...	3
2. ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง……..…….…….……..	4
2.1 เทคนิคการวิเคราะห์ข้อมูลด้วยการจำแนกประเภทและการทำนายข้อมูล (Classification and Prediction)……..……..……..…….….	4
2.2 การจัดระดับความเสี่ยงของผู้กู้ เมื่อพิจารณาจากความเพียงพอในการชำระหนี้...	8
2.3 การวัดประสิทธิภาพโมเดล.....	8
2.4 งานวิจัยที่เกี่ยวข้อง……...	10
3. วิธีวิจัย ……..……...	12
3.1 ขอบเขตของการวิจัยและรูปแบบการวิจัย……	12
3.2 ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย..…….....	13
3.3 เครื่องมือที่ใช้ในการวิจัย...……...	13
3.3 กระบวนการรวบรวมข้อมูลและวิเคราะห์ข้อมูล…….....	14
4. ผลการศึกษา ……...	24

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญ (ต่อ)

บทที่		หน้า
	4.1 ความเข้าใจธุรกิจประกันชีวิต (Business Understanding)……	24
	4.2 การเตรียมข้อมูล (Data Preparation)...……...……...	24
	4.3 การพัฒนาแบบจำลอง (Modeling)……...	27
	4.4 การตรวจสอบและประเมินผล (Evaluation)..……..……	34
	4.5 การประยุกต์ใช้งาน (Deployment)……...	35
5.	บทสรุปและข้อเสนอแนะ……..……...	38
	5.1 สรุปผลการศึกษา.……..……	38
	5.2 ข้อเสนอแนะและการพัฒนางานในอนาคต...…….....	41
บรรณานุกรม……		43
ประวัติผู้เขียน ……..		46

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญตาราง

ตารางที่ **หน้า**

3.1 รายละเอียดคุณสมบัติที่นำมาวิเคราะห์ข้อมูล..... 21
4.1 ข้อมูลที่นำมาใช้ในการทำนายข้อมูล..... 26
4.2 เทคนิควิธีแรนดอมฟอร์เรสด้วยคุณสมบัติใหม่ (Random Forest)..... 32
5.1 ปัจจัย (Feature) ที่ใช้ในการทำนายของโมเดล..... 40

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ

ภาพที่ **หน้า**

2.1 ตัวอย่างสมการทฤษฎีของเบย์……..……..…… 5

2.2 ตัวอย่างของความใกล้เคียงกันมากที่สุด……..… 6

2.3 ตัวอย่าง SVM ใน 2 มิติ…….....……..... 7

2.4 เทคนิควิธีแรนดอมฟอร์เรส (Random Forest)…… 7

2.5 ตัวอย่างเมตริกซ์การวัดประสิทธิภาพสาหรับการจาแนกประเภทข้อมูล 2 กลุ่ม... 9

3.1 ขั้นตอนการพัฒนาแบบจำลองเพื่อทำนายการอนุมัติเงินกู้สวัสดิการพนักงาน ปรับจากขั้นตอน CRISP-DM……..…….. 15

3.2 Personal Loan Approval Percentage…….…… 18

3.3 เพิ่ม 4 คุณสมบัติและสูตรคำนวณ……..……..……..... 20

3.4 สถาปัตยกรรมของโปรแกรมคอมพิวเตอร์ (Application Architectures)..……... 23

4.1 ตารางการวิเคราะห์ข้อมูลเบื้องต้น……..……..……..…… 25

4.2 Confusion Matrix for Naïve Bayes Model ……..…….. 28

4.3 Naïve Bayes Model Classification Report...……..…….……... 28

4.4 Confusion Matrix for K-Nearest neighbor Model...…...…..…….…… 29

4.5 K-Nearest neighbor Model Classification Report ……..…….. 29

4.6 Confusion Matrix for Support Vector Model ……. 30

4.7 Support Vector Model Classification Report ……..…….. 30

4.8 Confusion Matrix for Random Forest Model ……....…… 31

4.9 Random Forest Model Classification Report ……..... 31

4.10 Confusion Matrix for Random Forest New Model...……...……. 33

4.11 Random Forest New Model Classification Report ……..... 33

4.12 ตารางแนวทางการสกัด Rule จากเทคนิควิธีแรนดอมฟอร์เรส (Random Forest)...…….. 34

4.13 ภาพตัวอย่างหน้าจอข้อมูลพนักงานผู้กู้...……...… 35

4.14 ภาพตัวอย่างหน้าจอข้อมูลพนักงานผู้ค้ำประกัน..…...……... 36

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สารบัญภาพ (ต่อ)

ภาพที่ **หน้า**

4.16 ตัวอย่างหน้าจอคำนวณการผ่อนชำระเงินกู้..... 36
4.16 ตัวอย่างหน้าจอกรณีผ่านการพิจารณาอนุมัติ..... 37
4.17 ตัวอย่างหน้าจอกรณีไม่ผ่านการพิจารณาอนุมัติ..... 37
5.1 ผลประสิทธิภาพการทำนายของโมเดล..... 38
5.2 ปัจจัยที่มีผลต่อการทำนายของโมเดล..... 39
5.3 สถาปัตยกรรมการพัฒนางานในอนาคตของโปรแกรมคอมพิวเตอร์ (Application Architectures)..... 42

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 1
บทนำ

1.1 ที่มาและความสำคัญของปัญหา

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

จากการศึกษางานวิจัยโมเดลเพื่อทำนายความสามารถในการชำระหนี้เงินกู้ กรณีศึกษาสำนักงานส่งเสริมสวัสดิการและสวัสดิภาพครูและบุคลากรทางการศึกษา จังหวัดอุดรธานี (ญาณินี สิทธิสารและพุธษดี ศิริแสงตระกูล, 2561) ได้หาปัจจัยที่มีผลต่อการชำระหนี้เงินกู้ โดยการศึกษาตัวแปรที่ใช้ในการวิเคราะห์ข้อมูลเพื่อพิจารณาปัจจัยที่ผลต่อการเกิดหนี้ค้างชำระเงินกู้ได้แก่ จำนวนบุตร จำนวนเงินงวด อายุ จำนวนเงินกู้ รายได้ต่อเดือน ภาระหนี้ การทำประกัน เพศ ซึ่งแนวทางการวิจัยนี้ผู้วิจัยได้นำมาใช้ในการหาปัจจัยที่มีความสัมพันธ์กับการอนุมัติเอกสารการขอกู้เงินสวัสดิการช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย สอดคล้องกับงานวิจัยของ Wei Li และคณะ (Wei Li, 2010) ได้ศึกษาการจำแนกประเภทของลูกค้าที่มีบัตรเครดิตโดยใช้ตัวแปรผลที่ได้จากการ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จำแนกประเภทลูกค้าบัตรเครดิตโดยเทคนิค Decision tree model C5. โดยมีปัจจัยที่สำคัญคือ รายได้ส่วนตัวต่อเดือน รายได้ครอบครัวต่อเดือน ค่าเฉลี่ยของจำนวนครั้งที่รูดบัตรต่อเดือนและสถานะเครดิตส่วนตัวส่งผลให้ความถูกต้องสูงกว่าวิธีอื่น และงานวิจัย Eiman kambal และคณะ (Eiman kambal, 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 ประเภทคือกลุ่มผู้กู้ที่ได้รับการอนุมัติกลุ่มผู้กู้ที่ไม่ได้รับการอนุมัติ และหาความสัมพันธ์ของข้อมูลเพื่อเป็นแนวทางในการออกหลักเกณฑ์ในการกู้สวัสดิการ

1.2 วัตถุประสงค์ของงานวิจัย

1.2.1 เพื่อสร้างแบบจำลองที่มีประสิทธิภาพสำหรับช่วยเจ้าหน้าที่พิจารณาการอนุมัติเอกสารการขอกู้เงินสวัสดิการมหาวิทยาลัย

1.2.2 เพื่อหาปัจจัยที่มีความสัมพันธ์กับการอนุมัติเอกสารการขอกู้เงินสวัสดิการช่วยในการออกหลักเกณฑ์ในการกู้สวัสดิการนโยบาย

1.3 ขอบเขตของงานวิจัย

1.3.1 ข้อมูลที่นำมาทำนายใช้ข้อมูลการขอกู้เงินสวัสดิการมหาวิทยาลัยตั้งแต่เดือนตุลาคม 2562 ถึงพฤษภาคม 2564 จำนวน 5,870 รายการ

1.3.2 ปัจจัยที่นำมาทำนายผลการอนุมัติขอกู้เงินสวัสดิการมหาวิทยาลัยมีทั้งหมดจำนวน 14 คุณสมบัติ ได้แก่ สถานะภาพ ประเภทพนักงาน ระดับการศึกษา เพศ สถานะการอนุมัติกู้เงิน ประเภท

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

สายงาน อายุ อายุการทำงาน เงินเดือนผู้กู้ เงินเดือนผู้ค้ำประกัน อายุงานผู้ค้ำประกัน จำนวนเดือนที่กู้เงิน จำนวนเงินกู้ จำนวนบุตร

1.4 ประโยชน์ที่คาดว่าจะได้รับ

1.4.1 เพื่อช่วยลดขั้นตอนการขออนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย

1.4.2 เพื่อเป็นเครื่องมือในการช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิให้เจ้าหน้าที่

1.5 นิยามศัพท์

1.5.1 ผู้กู้ หมายถึง บุคคลใดบุคคลหนึ่งที่ต้องการกู้ยืมเงินเป็นสัญญาอย่างหนึ่ง

1.5.2 ผู้ค้ำประกัน หมายถึง บุคคลหนึ่งที่ยอมรับที่จะชำระหนี้แทนอีกบุคคล หากบุคคลนั้นไม่ชำระหนี้ให้แก่เจ้าหนี้

1.5.3 ผู้ให้กู้ บุคคลใดบุคคลหนึ่งที่ยินยอมให้กู้ยืมเงินโดยคิดดอกเบี้ยจากผู้กู้ตามอัตราที่ตกลงกันไว้เป็นการตอบแทนตามเอกสารสัญญา

1.5.4 การกู้ยืมเงิน หมายถึง การที่บุคคลหนึ่ง(ผู้กู้) ได้ยืมเงินจากอีกบุคคลหนึ่งผู้(ให้กู้) เพื่อนำไปใช้ประโยชน์ตามวัตถุประสงค์ที่ต้องการ

1.5.5 จำนวนเงินกู้ คือ จำนวนเงินที่ผู้กู้ต้องการกู้ยืม

1.5.6 จำนวนเงินผ่อนชำระรายเดือน คือ จำนวนเงินที่ต้องส่งคืนต่อเดือนบวกจำนวนดอกเบี้ย

1.5.7 อัตราดอกเบี้ย หมายถึง ผลตอบแทนจากการให้กู้ยืมเงิน โดยคิดจากผู้กู้ไม่เกินร้อยละ 15 ต่อปี หรือในอัตราร้อยละ1.25 ต่อเดือน

1.5.8 จำนวนงวด คือ จำนวนเดือนที่ต้องส่งชำระหนี้

1.5.9 หลักฐานการยืม หมายถึง เอกสารประกอบการการพิจารณาการขอยืนกู้เงิน

1.5.10 Accuracy หมายถึง ค่าความแม่นยำของโมเดลที่ใช้ในการพยากรณ์

1.5.11 Precision หมายถึง คือค่าที่บอกว่าโมเดลพยากรณ์ได้ว่าจริงถูกต้องเท่าไหร่

1.5.12 Recall หมายถึง ค่าที่โมเดลพยากรณ์ได้ว่าจริง เป็นอัตราส่วนเท่าไหร่เทียบกับของจริงทั้งหมด

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 2
ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้มีวัตถุประสงค์เพื่อทำนายการอนุมัติเอกสารการขอกู้เงินสวัสดิการมหาวิทยาลัย โดยมีรายละเอียดของทฤษฎีและงานวิจัยที่เกี่ยวข้องดังนี้

2.1 เทคนิคการวิเคราะห์ข้อมูลด้วยการจำแนกประเภทและการทำนายข้อมูล (Classification and Prediction)

เป็นวิธีในการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่าง ๆ ที่ได้มีการกำหนดไว้แล้ววิธีนี้เหมาะกับการสร้างตัวแบบเพื่อการพยากรณ์ค่าข้อมูล (Predictive Modeling) ในอนาคตเป็นการเรียนรู้แบบมีผู้สอน (Supervised Learning) โดยเป็นการสร้างโมเดลจำแนกประเภทเพื่อทำนายกลุ่มของข้อมูลใหม่ การสร้างโมเดลจำแนกประเภทข้อมูลเกิดจากการหาความสัมพันธ์ของข้อมูลโดยข้อมูลทั้งหมดจะแบ่งเป็น 2 ส่วน คือส่วนแรกข้อมูลเรียนรู้ (Training Data) เป็นชุดข้อมูลที่ใช้ในการสร้างโมเดลจำแนกประเภทข้อมูลขึ้นมาใหม่ เพื่อให้โมเดลที่สร้างได้เรียนรู้ข้อมูลและส่วนที่สองของข้อมูลที่ใช้ในการทดสอบโมเดลที่สร้างขึ้นมา (Testing Data) เป็นชุดข้อมูลประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูล (Hengphraphrom K, Hengphraphrom S and Makwibunchai S, 2014, p. 1) โดยงานวิจัยนี้ได้ทดสอบข้อมูลใช้เทคนิคการจำแนกประเภท 4 วิธี

2.1.1 เทคนิควิธีนาอีฟเบย์ (Naïve Bayes)

คือ ตัวจำแนกประเภทอาศัยหลักการความน่าจะเป็นตามทฤษฎีของเบย์ (Bayes Theorem) ซึ่งมีอลักอริทึมที่ไม่ซับซ้อนเป็นขั้นตอนวิธีในการจำแนกข้อมูล โดยการเรียนรู้ปัญหาที่เกิดขึ้นเพื่อนำมาสร้างเงื่อนไขการจำแนกข้อมูลใหม่ หลักการของนาอีฟเบย์ใช้การคำนวณหาความน่าจะเป็นในการทำนายผลเป็นเทคนิคในการแก้ปัญหาแบบจำแนกประเภทที่สามารถคาดการณ์ผลลัพธ์ได้จะทำการวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรเพื่อใช้ในการสร้างเงื่อนไขความน่าจะเป็นสำหรับแต่ละความสัมพันธ์หมาะกับกรณีของเซตตัวอย่างงที่มีจำนวนมากและคุณสมบัติ (Attribute) ของตัวอย่างไม่

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ขึ้นต่อกันโดยกำหนดให้ความน่าจะเป็นของข้อมูลเท่ากับสมการ (Vilailuck S, Jaroenpuntaruk V, Wichadakul D, 2015, p. 2) ดังแผนภาพที่ 2

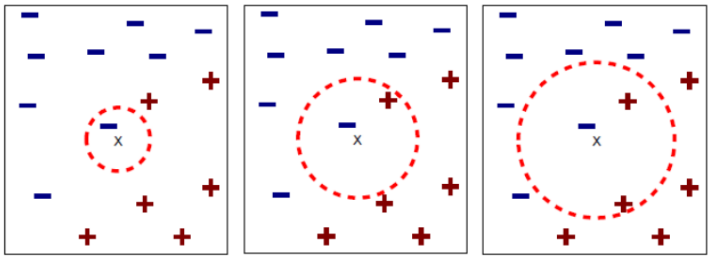
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

ภาพที่ 2.1 ตัวอย่างสมการทฤษฎีของเบย์

2.1.2 เทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor)

วิธีที่ใช้ในการจัดแบ่งคลาส โดยเทคนิคนี้จะตัดสินใจว่าคลาสใดที่จะแทนเงื่อนไขหรือกรณีใหม่ ๆ ไดบ้าง โดยการตรวจสอบจำนวนบางจำนวน K ในขั้นตอน วิธีการหาเพื่อนบ้านใกลที่สุดของกรณีหรือเงื่อนไขที่เหมือนกันหรือใกล้เคียงกันมากที่สุด โดยจะหาผลรวม (Count Up) ของจำนวนเงื่อนไขหรือกรณีต่าง ๆ สำหรับแต่ละคลาสและกำหนดเงื่อนไขใหม่ ๆ ให้คลาสที่เหมือนกับคลาสที่ใกล้เคียงกันมากที่สุด การนำเทคนิคของขั้นตอนวิธีการเพื่อนบ้านใกลที่สุดไปใช้นั้นเป็นการหาระยะห่างระหว่างแต่ละตัวแปร (Attribute) ในข้อมูลจากนั้นก็คำนวณค่าออกมา ซึ่งวิธีนี้จะเหมาะสำหรับข้อมูลแบบตัวเลขแต่ตัวแปรที่เป็นค่าแบบไม่ต่อเนื่องนั้นก็สามารถทำได้เพียงแต่ต้องการการจัดการแบบพิเศษเพิ่มขึ้น อย่างเช่น ถ้าเป็นเรื่องของสีจะใช้อะไร วัดความแตกต่างระหว่างสีน้ำเงินกับสีเขียวต่อจากนั้นต้องมีวิธีในการรวมค่าระยะห่างของ Attribute ทุกค่าที่วัดมาได้ เมื่อสามารถคำนวณระยะห่างระหว่างเงื่อนไขหรือกรณีต่าง ๆ ได้จากนั้นก็เลือกชุดของเงื่อนไขที่ใช้จัดคลาสมาเป็นฐานสำหรับการจัดคลาสในเงื่อนไขใหม่ ๆ ได้แล้วจะตัดสินได้ว่าขอบเขตของจุดข้างเคียงที่ควรเป็นนั้น ควรมีขนาดใหญ่เท่าไรและอาจมีการตดัสินใจได้ด้วยว่าจะนับจำนวนจุดข้างเคียงตัวมัน (T.L. Saaty, 2006, p. 4-5)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

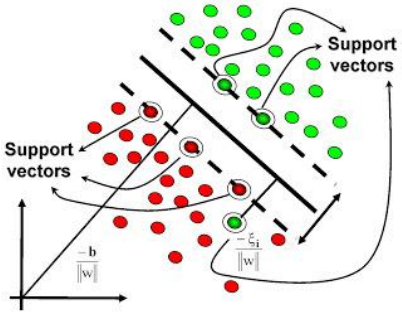


ภาพที่ 2.2 ตัวอย่างของความใกล้เคียงกันมากที่สุด: (a) ความใกล้เคียงกันมากที่สุดโดยพิจารณาจากข้อมูล 1 ตัว, (b) ความใกล้เคียงกันมากที่สุดโดยพิจารณาจากข้อมูล 2 ตัว และ (c) ความใกล้เคียงกันมากที่สุดโดย พิจารณาจากข้อมูล 3 ตัว

2.1.3 เทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)

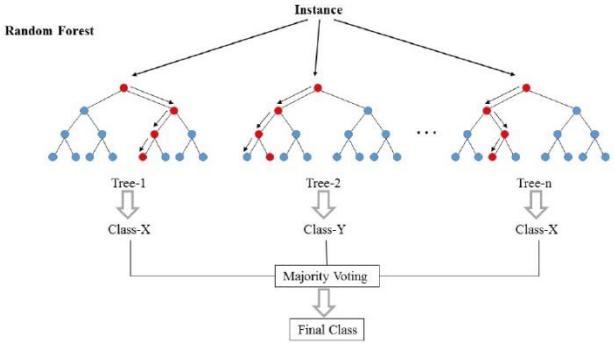
เป็นสมการที่ใช้จำแนกค่าคุณลักษณะของ 2 กลุ่ม ที่วางตัวอยู่ในพื้นที่คุณลักษณะ (Feature Space) ออกจากกันโดยจะสร้างเส้นแบ่ง (Plane) ที่เป็นเส้นตรงขึ้นมาและเพื่อให้ทราบว่าเส้นตรงที่แบ่ง 2 กลุ่ม ออกจากกันนั้น เส้นตรงใดที่เป็นเส้นที่ดีที่สุด โดยเส้นตรงนั้นจะเพิ่มเส้นขอบ (Margin) ออกไปทั้งสองข้าง โดยเส้นขอบที่เพิ่มนั้นจะขนานกับเส้นเดิมเสมอ เส้นขอบที่เพิ่มขึ้นมานี้จะขยายออกไปจนกว่าจะสัมผัสกับค่าของกลุ่มตัวอย่างที่ใกล้ที่สุด เคอร์เนล (Kernel) ในโลกความเป็นจริงนั้นข้อมูล 2 กลุ่ม ไม่ได้วางตัวในพื้นที่คุณลักษณะและไม่สามารถแบ่งได้โดยเส้นตรงแต่ข้อมูลอาจจะจับกลุ่มกันในตำแหน่งต่าง ๆ ดังนั้นจึงเป็นปัญหาทำให้ไม่สามารถที่จะใช้สมการซัพพอร์ตเวกเตอร์แมชชีนแบบเชิงเส้นได้ ดังนั้นจะต้องมีเครื่องมือมาช่วยให้ข้อมูลเหล่านั้นเรียงตัวใหม่ในพื้นที่เรียกว่าพื้นที่หลายมิติ (Higher Dimensional Space) (อานนท์ นามสนิท, 2549, หน้า 53)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 2.3 ตัวอย่าง SVM ใน 2 มิติ

2.1.4 เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) เกิดจากการรวมกลุ่มกันของโครงสร้างต้นไม้ ซึ่งค่าความความเคลื่อนโดยรวมของป่าไม้จะถูกเปลี่ยนให้เป็นค่าลิมิต ทำให้จำนวนต้นไม้ในป่าเพิ่มขึ้น ค่าความคลาดเคลื่อนโดยรวมจะขึ้นกับความมั่นคง (Strength) ของต้นไม้แต่ละต้น รวมถึงความสัมพันธ์กันระหว่างต้นไม้เหล่านั้น โดยจะใช้วิธีการสุ่มเลือกคุณสมบัติเพื่อการแบ่งแยกโหนดทำให้ ค่าความผิดพลาดลดลง ขั้นตอนวิธีนี้จะมีประสิทธิภาพมากเมื่อนำไปใช้วิเคราะห์เกี่ยวกับการประมาณการขนาดใหญ่ เราสามารถสร้างแบบจำลองที่ใช้ต้นไม้ต้นหลายๆต้นในการตัดสินใจเพื่อนำมาประมวลผล ซึ่งมีความแม่นยำสูงสามารถจัดการข้อมูลได้มากและเหมาะสมสำหรับข้อมูลที่มีความสำคัญ (L. Breiman, 2001, p. 532) ดังภาพที่ 2.4



ภาพที่ 2.4 เทคนิควิธีแรนดอมฟอร์เรส (Random Forest)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2.2 การจัดระดับความเสี่ยงของผู้กู้ เมื่อพิจารณาจากความเพียงพอในการชำระหนี้

1. การคำนวณใช้สูตร DSCR (Debt-Service Coverage Ratio) หรืออัตราหนี้สินต่อรายได้ เพื่อคำนวณภาระหนี้สิน โดยมีสูตรการคำนวณดังนี้

$$\text{DSCR} = \frac{\text{ภาระหนี้ต่อเดือน}}{\text{รายได้ต่อเดือน}} \; x \; 100$$

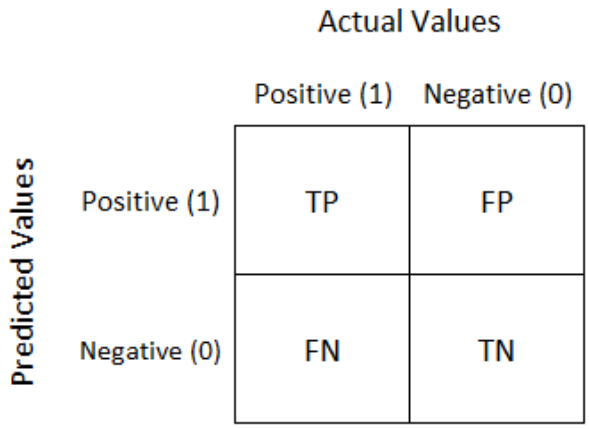
2. ระดับความเสี่ยงของผู้กู้

ระดับความเสี่ยง/Ration	DSCR
ต่ำ	น้อยกว่า 15 %
ปานกลาง	น้อยกว่า 40 %
สูง	ประมาณ 50 %
สูงมาก	มากกว่า 70 %

2.3 การวัดประสิทธิภาพโมเดล

การวัดประสิทธิภาพของการจำแนกข้อมูลแต่ละเทคนิคด้วยการวัดประสิทธิภาพของผลลัพธ์การทำนายเปรียบเทียบกับผลลัพธ์จริง (Mohammed J. Zaki and Wagner Meira, Jr, 2014) โดยมีวิธีการคำนวณดังนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 2.5 ตัวอย่างเมตริกซ์การวัดประสิทธิภาพสำหรับการจำแนกประเภทข้อมูล 2 กลุ่ม

ที่มา: Sarang Narkhede-2018)

1) ค่าความถูกต้อง (Accuracy) คือค่าความแม่นยำของโมเดลที่ใช้ในการพยากรณ์ มีสูตรคำนวณ คือ

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

2) ค่าความแม่นยำ (Precision) คือค่าที่บอกว่าโมเดลพยากรณ์ได้ว่าจริงถูกต้องเท่าไร มีสูตรคำนวณ คือ

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

3) ค่าความระลึก (Recall) ค่าที่บอกว่าโมเดลพยากรณ์ได้ว่าจริงเป็นอัตราส่วนเท่าไรของจริงทั้งหมด มีสูตรคำนวณ คือ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

โดยที่

True Positive (TP) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าจริงและในความเป็นจริงก็เป็นเรื่องจริง

True Negative (TN) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าไม่จริงและในความเป็นจริงก็เป็นเรื่องไม่จริง

False Positive (FP) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าจริงแต่ในความเป็นจริงเป็นเรื่องไม่จริงFalse Negative (FN) คือ จำนวนข้อมูลของสิ่งที่ทำนายว่าไม่จริงแต่ในความเป็นจริงเป็นเรื่องจริง

2.4 งานวิจัยที่เกี่ยวข้อง

การใช้เหมืองข้อมูลทำนายการจำแนกประเภทผู้สมัครบัตรเครดิต Yap Bee Wah และ Irma Rohaiza Ibrahim (Bee Wah, Y, 2010) โดยใช้แบบจำลองการทำนาย สามารถช่วยในกระบวนการประเมินการคัดเลือกลูกค้าที่จะสมัครบัตรเครดิต โดยใช้เทคนิคเหมืองข้อมูล (Data Mining) 3 วิธี ได้แก่ Logistic regression (LR) model, Classification and regression tree (CART) model และ Neural Network (NN) model ผู้สมัครบัตรเครดิตซึ่งข้อมูลที่นำมาเป็นข้อมูลในอดีตและปัจจุบัน ตัวอย่างเช่น เพศ อายุ สภานภาพการสมรส เงินเดือน รายได้เสริม ประเภทที่อยู่อาศัย และระดับการศึกษา ผลที่ได้จากการเปรียบเทียบ Classification and regression tree (CART) model สามารถทำนายถูกต้องแม่นยำสูงกว่าวิธีอื่น

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 ค่าเฉลี่ยของจำนวนครั้งที่รูดบัตรต่อเดือนและสถานะเครดิตส่วนตัวมาใช้เป็นตัวแปรผลที่ได้จากการจำแนกประเภทลูกค้าบัตรเครดิตโดยเทคนิค Decision tree model C5.ความถูกต้องสูงกว่าวิธีอื่น

จากข้อมูลของธนาคารซูดาน Eiman kambal และคณะ (Eiman kambal, 2013) ได้พบว่าหนึ่งในปัจจัยความสำเร็จที่สำคัญขององค์กรสินเชื่อและธนาคารในการให้กู้ยืมโดยเฉพาะการประเมินความ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

คุ้มค่าของผู้กู้รูปแบบการจัดลำดับเครดิตได้ถูกนำมาใช้โดยนักวิจัยจำนวนมากที่จะปรับปรุงกระบวนการของการประเมินเครดิต โดยให้ผู้กู้ยืมเงินในอนาคตอยู่บนพื้นฐานของความน่าจะเป็นของการชำระหนี้ 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

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 3
วิธีการวิจัย

การศึกษาวิจัยเรื่อง การสร้างโมเดลการทำนายข้อมูลการพิจารณาอนุมัติสินเชื่อสวัสดิการพนักงานด้วยเทควิธีการเรียนรู้ด้วยของเครื่องเพื่อประยุกต์ใช้ในการพัฒนาเครื่องมือสนับสนุนการปฏิบัติงานของกองทุนสวัสดิการมหาวิทยาลัย ผู้วิจัยได้นำแนวคิด ทฤษฎีและผลงานวิจัยที่เกี่ยวข้องมากำหนดขั้นตอนในการศึกษา และเลือกใช้ขั้นตอนตามแบบ CRISP-DM (Chapman, et al., 2000) ซึ่งมีกระบวนการประมวลผลดังต่อไปนี้

3.1 ขอบเขตของการวิจัยและรูปแบบการวิจัย

ในการศึกษาวิจัยครั้งนี้ได้นำข้อมูลการขอกู้เงินสวัสดิมหาวิทยาลัยที่ได้ทำการจัดเก็บไว้แล้วของมหาวิทยาลัยแห่งหนึ่งจากเอกสารทำการนำเข้าฐานข้อมูลที่มีอยู่มาทำการศึกษาค้นหาความรู้ความสัมพันธ์ที่แฝงอยู่ในข้อมูลเงื่อนไขในการอนุมัติเอกสารด้วยกระบวนการ CRISP-DM ร่วมกับวิธีการจำแนกประเภทและทำนาย (Classification and Prediction) เพื่อนำเสนอวิธีการในการอนุมัติเอกสารการขอกู้เงินสวัสดิการมหาวิทยาลัย

3.1.1 การวิจัยเชิงพรรณนา (Descriptive Research)

การวิจัยเชิงพรรณนาเพื่อหาหลักการแนวคิด และทฤษฎีที่เกี่ยวข้องจากแหล่งทุติยภูมิ โดยค้นคว้าจากหนังสือ สิ่งพิมพ์ อินเทอร์เน็ต บทความและงานวิจัยที่เกี่ยวข้องกับทฤษฎี หลักการ รวมทั้งแนวคิดและหลักการในการทำนายข้อมูลการพิจารณาอนุมัติสินเชื่อสวัสดิการ

3.1.2 การวิจัยเชิงสำรวจ (Exploratory Research)

การวิจัยเชิงสำรวจเป็นการนำข้อมูลการขอกู้เงินสวัสดิการมหาวิทยาลัยได้ทำการจัดเก็บไว้แล้วในฐานข้อมูล ซึ่งเป็นข้อมูลทุติยภูมิในฐานข้อมูล โดยข้อมูลที่เลือกมาใช้จะเลือกเฉพาะข้อมูลที่ผู้กู้ส่งเอกสารยื่นขอกู้แล้วเท่านั้น ตัวอย่างข้อมูล ข้อมูลเชิงคุณภาพ เช่น รหัสการกู้ ระดับการศึกษา ประเภทพนักงาน ประเภทสายงาน เพศ สถานะการอนุมัติกู้เงิน ประวัติการชำระล่าช้า ข้อมูลเชิงปริมาณ เช่น อายุ อายุการทำงานผู้กู้ เงินเดือนผู้กู้ เงินเดือนผู้ค้ำประกัน อายุการทำงานผู้ค้ำประกัน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

จำนวนเดือนที่กู้เงิน จำนวนเงินกู้ จำนวนบุตร มาทำการศึกษาหาความสัมพันธ์ของข้อมูลและจัดทำโมเดลทำนาย ซึ่งดำเนินการตามขั้นตอนกระบวนการของ CRISP-DM

3.2 ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ข้อมูลที่นำมาศึกษาวิจัยเป็นข้อมูลจากฐานข้อมูลมหาวิทยาลัย ซึ่งเป็นข้อมูลประวัติการขอยื่นกู้กองทุนสวัสดิการมหาวิทยาลัย (กรณีศึกษา) โดยการคัดเลือกข้อมูลตั้งแต่ตุลาคม พ.ศ.2562 จน ถึง พฤษภาคม พ.ศ. 2564 มีจำนวนทั้งสิ้น 5,870 รายการ 14 ปัจจัย จากนั้นดำเนินการโดยการศึกษาค้นคว้าข้อมูลจากเอกสารและงานวิจัยต่างๆ ที่เกี่ยวข้องนำมาเป็นแนวคิด ทฤษฎี สำหรับการเนินงานวิจัย

3.3 เครื่องมือที่ใช้ในการวิจัย

3.3.1 ภาษาที่ใช้ในการพัฒนา

1) ภาษาไพธอน (Python language)

ไพธอนคือชื่อภาษาที่ใช้ในการเขียนโปรแกรมภาษาหนึ่งซึ่งถูกพัฒนาขึ้นมาโดยไม่ยึดติดกับแพลตฟอร์ม กล่าวคือสามารถรันภาษา Python ได้ทั้งบนระบบ Unix, Linux , Windows NT, Windows 2000, Windows XP หรื/อแม้แต่ระบบ FreeBSD อีกอย่างหนึ่งภาษาตัว นี้เป็น Open Source เหมือนอย่าง PHP ทำให้ทุกคนสามารถที่จะนำ Python มาพัฒนาโปรแกรมของเราได้ฟรีๆ โดยไม่ต้องเสียค่าใช้จ่าย และความเป็น Open Source ทำให้มีคนเข้ามาช่วยกันพัฒนาให้ Python มีความสามารถสูงขึ้น และใช้งานได้ครบคุมกับทุกลักษณะงาน (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2561)

2) ภาษาเอสคิวแอล (SQL)

ภาษาเอสคิวแอล เป็นภาษาที่ใช้งานร่วมกับฐานข้อมูลเชิงสัมพันธ์ เมื่อผ่านกระบวนการออกแบบอีอาร์ไดอะแกรมเสร็จเรียบร้อยแล้ว และได้นำไดอะแกรมไปสร้างฐานข้อมูลผ่านภาษาเอสคิวแอล ซึ่งภาษาเอสคิวแอลเป็นสิ่งจำเป็นมากเพราะเป็นชุดคำสั่งเพื่อนำไปใช้ในการสืบค้นข้อมูล จัดการโครงสร้างฐานข้อมูลและจัดการข้อมูลที่อยู่ในฐานข้อมูลเชิงสัมพันธ์ โดยสามารถแยกประเภทของชุดคำสั่งออกได้เป็น 3 ประเภทหลัก ได้แก่ 1) ชุดคำสั่งการค้นคืนข้อมูล 2) ชุดคำสั่งจัดการกับข้อมูลและ 3) ชุดคำสั่งที่จัดการโครงสร้างฐานข้อมูล (Gavin Powell, 2005)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.3.2. เครื่องมือในการพัฒนาระบบ

1) Google Colab

Google Colab เป็นโครงการจาก Google Research ซึ่งเป็นสภาพแวดล้อมที่ใช้ Jupyter ฟรี ซึ่งช่วยให้เราสร้างสมุดบันทึก Jupyter [การเขียนโปรแกรม] เพื่อเขียนและเรียกใช้งาน Python [16] และเครื่องมือของบุคคลที่สามที่ใช้ Python และเฟรมเวิร์กการเรียนรู้ของเครื่องเช่น Pandas, PyTorch, Tensorflow, Keras, Monk, OpenCV และอื่น ๆ ในเว็บเบราว์เซอร์ สมุดบันทึกการเขียนโปรแกรมคือเชลล์หรือเคอร์เนลประเภทหนึ่งในรูปแบบของโปรแกรมประมวลผลคำซึ่งเราสามารถเขียนและรันโค้ดได้ ข้อมูลที่จำเป็นสำหรับการประมวลผลใน Google Colab สามารถติดตั้งลงใน Google ไดรฟ์หรือนำเข้าจากแหล่งใดก็ได้บนอินเทอร์เน็ต Project Jupyter เป็นองค์กรซอฟต์แวร์โอเพ่นซอร์สที่พัฒนาและสนับสนุนโน้ตบุ๊ก Jupyter สำหรับการประมวลผลแบบโต้ตอบ (Jupyter Available, 2021)

2) Microsoft SQL Server Management Studio

เครื่องมือใช้ในการบริหารจัดการฐานข้อมูลเชิงสัมพันธ์ (RDBMS) ทำหน้าที่ในการจัดการงานต่างๆ มากมาย เช่น การจัดการฐานข้อมูล การสร้างตารางใหม่ การเพิ่มข้อมูลต่างๆ ในตาราง การ Backup และ Restore (Recovery) ข้อมูล การสร้าง Package สาหรับถ่ายโอนข้อมูล การกำหนดเกี่ยวกับระบบรักษาความปลอดภัย การตรวจสอบข้อผิดพลาดระหว่างระบบ เป็นต้น (กิตติ ภักดีวัฒนะกุล และจำลอง ครูอุตสาหะ, 2546)

3.4 กระบวนการรวบรวมข้อมูลและวิเคราะห์ข้อมูล

การวิจัยครั้งนี้ใช้ขั้นตอนการเก็บข้อมูลตามกระบวนการและวิเคราะห์ข้อมูลมีกระบวนการมาตรฐานที่เรียกว่า Cross-Industry Standard Process for Data Mining (CRISP-DM) ซึ่งแบ่งเป็นขั้นตอน ดังนี้

3.4.1 ความเข้าใจทางธุรกิจ (Business Understanding)

3.4.2 ความเข้าใจเกี่ยวกับข้อมูล (Data Understanding)

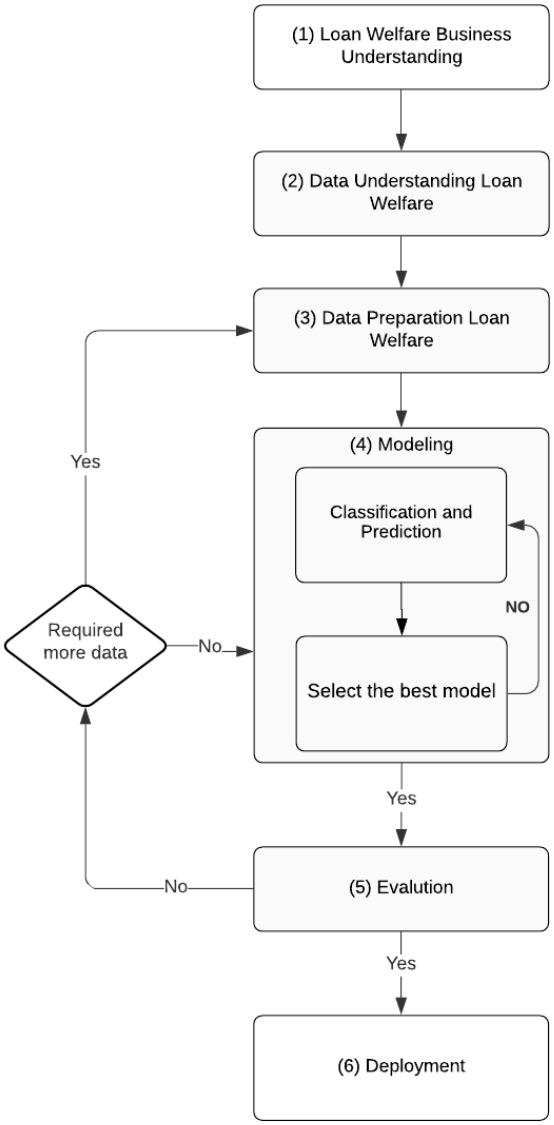
3.4.3 การเตรียมข้อมูล (Data Preparation)

3.4.4 การพัฒนาแบบจำลอง (Modeling)

3.4.5 การทดสอบแบบจำลอง (Evaluation)

3.4.6 การนำแบบจำลองไปใช้ (Deployment)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.1 ขั้นตอนการพัฒนาแบบจำลองเพื่อทำนายการอนุมัติเงินกู้สวัสดิการพนักงานปรับจากขั้นตอน CRISP-DM **ที่มา:** (Chapman et al, 2000)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.4.1 ความเข้าใจทางธุรกิจ (Business Understanding)

1) การกำหนดวัตถุประสงค์ทางธุรกิจ (Determine Business Objective)

เป็นขั้นตอนที่ต้องทำเป็นสิ่งแรก เนื่องจากความเข้าใจในวัตถุประสงค์ของธุรกิจอย่างชัดเจน หากเข้าใจวัตถุประสงค์ของธุรกิจผิดอาจส่งผลต่อผลลัพธ์ที่ไม่ถูกต้องหลักจากเสร็จกระบวนการ ลักษณะของหน่วยงานที่นำมาศึกษาครั้งนี้ คือ มหาวิทยาลัยรัฐบาลแห่งหนึ่ง โดยนำข้อมูลการยื่นขอเงินกู้กองทุนสวัสดิการที่มีอยู่จริงรวมไปถึงปัญหาที่พบในการขอยื่นกู้นำมาพัฒนาพัฒนาระบบให้ลดระยะเวลาในการยื่นเอกสารและเพิ่มประสิทธิภาพการทำงานดียิ่งขึ้น

2) การประเมินสถานการณ์ปัจจุบัน (Assess Situation) ในขั้นตอนนี้จะทำการประเมินเกี่ยวกับรายละเอียดที่เกี่ยวข้องต่างๆ เช่น ข้อมูลมีอยู่จริง ข้อจำกัดกระบวนการทางธุรกิจ สมมติฐานและปัจจัยอื่นๆ ที่มีผลต่อการกำหนดวัตถุประสงค์และวางแผนโครงการ รวมถึงปัญหาในอดีตที่เกิดขึ้น ควรนำมาพิจารณาเพื่อช่วยในการตัดสินใจด้วยเช่นกัน เพื่อที่จะวางแผนเกี่ยวกับเป้าหมายในการวิเคราะห์ข้อมูล

3) การกำหนดเป้าหมายของการทำนายข้อมูล (Determine Data Predict Goals) เป็นการกำหนดวัตถุประสงค์และเป้าหมายของโครงการจากการทำนายข้อมูลให้อยู่ในทิศทางเดียวกันกับเป้าหมายการดำเนินธุรกิจ ซึ่งการวิจัยนี้เลือกใช้วิธีการจำแนกประเภทและการวัดระดับความเสี่ยงของผู้กู้จากการคำนวณอัตราส่วนหนี้สินต่อรายได้

3.4.2 ความเข้าใจข้อมูล (Data Understanding)

1) การเก็บรวบรวมข้อมูลขั้นต้น (Collect Initial Data) ข้อมูลที่นำมาใช้วิเคราะห์หาความสัมพันธ์เป็นข้อมูลการขอกู้เงินสวัสดิการที่ถูกจัดเก็บไว้ในฐานข้อมูลของมหาวิทยาลัยแห่งหนึ่ง ซึ่งการวิจัยครั้งนี้เป็นการคัดเลือกข้อมูลจากฐานข้อมูลจริงเก็บไว้ยังฐานข้อมูลที่จำลองขึ้นเพื่อทำการศึกษาและที่สำคัญต้องระบุถึงปัญหาที่เจอ บันทึกปัญหาที่พบและวิธีการ ตลอดจนแนวทางในการแก้ไขปัญหา

2) การอธิบายข้อมูล (Describe Data) อธิบายถึงข้อมูลที่ต้องการ เช่น อธิบายว่าข้อมูลแต่ละฟิวส์คือข้อมูลอะไร รวมถึงรูปแบบของข้อมูล ปริมาณข้อมูลทั้งในเชิงเทคนิคและเชิงธุรกิจและนำข้อมูลที่ได้มาทำความเข้าใจโครงสร้างของข้อมูลทั้งหมด ซึ่งจากงานวิจัยนี้ได้เก็บรวบรวมข้อมูล เพื่อใช้ในการสร้างโมเดลโดยใช้ข้อมูลการขอยื่นกู้กองทุนสวัสดิการมหาวิทยาลัยรัฐบาลแห่งหนึ่ง ดังนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1. Qualitative features: ข้อมูลเชิงคุณภาพ

1.1 Nominal Variable

1.1.1 MaritalStatus (สถานะภาพ) ประกอบด้วย โสด, สมรส,หย่า

1.2 Ordinal Variable

1.2.1 PersonType (ประเภทพนักงาน) ประกอบด้วย พนักงานมหาวิทยาลัย, ข้าราชการ, ลูกจ้างประจำ

1.3 Binary Variable

1.3.1 Degree (ระดับการศึกษา) ประกอบด้วย ระดับปริญญาตรีขึ้นไปและต่ำกว่าปริญญาตรี

1.3.2 Gender (เพศ) ประกอบด้วย ผู้หญิง, ผู้ชาย

1.3.3 LoanApprove (สถานะการอนุมัติกู้เงิน) ประกอบด้วย ไม่อนุมัติ, อนุมัติ

1.3.4. JobLevel (ประเภทสายงาน) ประกอบด้วย สายวิชาการ, สายสนันสนุน

2. Quantitative features: ข้อมูลเชิงปริมาณ

2.1 Age (อายุ)

2.2 WorkAge (อายุการทำงาน)

2.3 EmployeeSalary (เงินเดือนผู้กู้)

2.4 SuretySalary (เงินเดือนผู้ค้ำประกัน)

2.5 SuretyWorkAge (อายุงานผู้ค้ำประกัน)

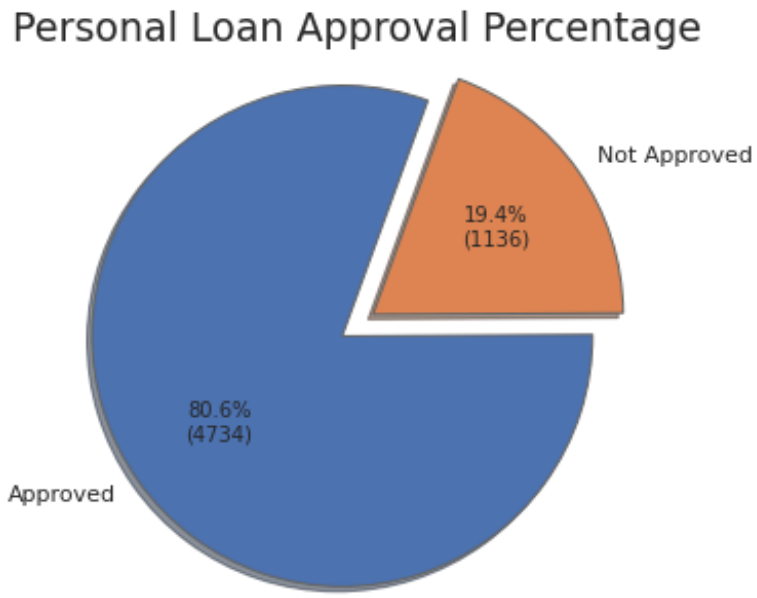
2.6 Period (จำนวนเดือนที่กู้เงิน)

2.7 LoanAmount (จำนวนเงินกู้)

2.8 Children (จำนวนบุตร)

3) การสำรวจข้อมูล (Explore Data) ทำการสืบค้นข้อมูลจากแหล่งข้อมูลที่มีหรือจินตนาการ (Visualization) และเทคนิคการรายงาน (Reporting Techniques) เพื่อให้เกิดความเข้าใจรวมไปถึงการสำรวจข้อมูล โดยมีรายละเอียดสัดส่วนข้อมูลดังต่อไปนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.2 Personal Loan Approval Percentage

Number of samples: 5870

Number of features: 15

Number of Target features: 1

Number of Approved Loan: 4734 (80.65%)

Number of Not Approved Loan: 1136 (19.35%)

Imbalance Ratio ≈ 4:1

4 การตรวจสอบคุณภาพของข้อมูล (Verify data quality)

การตรวจสอบคุณภาพข้อมูลที่นำมาวิเคราะห์ เช่น ข้อมูลที่ได้มาสมบูรณ์ มีความซ้ำซ้อนหรือข้อมูลนั้นมีความถูกต้องหรือไม่รวมไปถึงข้อมูลมีความครอบคลุมเพียงพอกับความต้องการหรือไม่หรือหากข้อมูลนั้นมีความผิดพลาดไม่ครบถ้วน จะมีวิธีจัดการอย่างไร โดยมีการจัดทำรายการคุณภาพข้อมูลที่ได้ทาการตรวจสอบ โดยทั่วไปแล้ววิธีการแก้ไขปัญหาของคุณภาพข้อมูลขึ้นอยู่กับ 2 ปัจจัยด้วยกัน คือ ข้อมูลและกระบวนการทางธุรกิจระบวนการทางธุรกิจ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.4.3 การเตรียมข้อมูล (Data Preparation)

1) การคัดเลือกข้อมูล (Select Data)

ในขั้นตอนการเตรียมข้อมูลที่จะนำไปใช้ในการวิเคราะห์ เป็นการใช้ข้อมูลจากฐาน ข้อมูลที่มีอยู่ทำการคัดเลือกข้อมูลที่เกี่ยวข้องกับเป้าหมายของงานวิจัย เช่น เพศ อายุ สถานภาพ อายุงาน รายได้ จำนวนเงินกู้ จำนวนงวด หรือเลือกตัดข้อมูลที่ไม่จาเป็นต่อการวิเคราะห์ เพื่อลดขนาดของข้อมูล เช่น เลขประจาตัวประชาชน ชื่อ-สกุล ตำแหน่งงาน หน่วยงานที่สังกัด เป็นต้น

2) ทำความสะอาดข้อมูล (Clean Data)

การหลังการคัดกรองข้อมูลเป็นที่เรียบร้อยแล้ว ผู้วิจัย ทำความสะอาดข้อมูลด้วยSQL SRVER ที่มีความซ้ำซ้อนหรือผิดปกติของข้อมูลด้วยการ Remove missing values เติมข้อมูลเงินเดือนที่เป็นค่าว่างตามนโยบายมหาลัยด้วยการ Replace Missing Values เพื่อสร้างคุณภาพให้แก่ข้อมูลที่นำมาใช้อย่างแม่นยำ โดยมีการจัดทำเป็นรายงานของข้อมูลที่ถูกนำมาทำความสะอาดที่อธิบายถึงการตัดสินใจและการกระทำที่ได้ทำไปเพื่อระบุถึงปัญหาคุณภาพของข้อมูลในระหว่างการทาการตรวจสอบและผลกระทบที่อาจเกิดขึ้นระหว่างการวิเคราะห์ก็เป็นเรื่องที่จะต้องนำมาพิจารณาด้วย

3) โครงสร้างข้อมูลใหม่ (Construct Data)

กระบวนการนี้รวมไปถึงวิธีการดำเนินการในการจัดเตรียมข้อมูล เช่น การสร้างลักษณะประจำใหม่ การทำข้อมูลชุดใหม่ให้สมบูรณ์ ในกระบวนการเตรียมข้อมูลนั้นบางครั้งอาจต้องการข้อมูลเพิ่มเติมเพื่อทำให้ข้อมูลสมบูรณ์ หรือสร้างข้อมูลใหม่เพื่อให้โครงสร้างข้อมูลเหมาะกับขั้นตอนวิธีที่จะนำมาใช้ทำวิเคราะห์ข้อมูล การปรับบรรทัดฐานข้อมูล (Max-Min Normalization) เงินเดือนผู้กู้ เงินเดือนผู้ค้ำประกัน จำนวนเงินกู้ เนื่องด้วยข้อมูลมีความแตกต่างกันในเรื่องของช่วงของข้อมูล ซึ่งประกอบด้วยค่าที่มี 1 หลัก (หลักหน่วย) และค่าที่มี 6 เพื่อไม่ให้เกิดความแตกต่างก่อนการนำเข้าข้อมูลจำเป็นต้องมีการปรับช่วงของข้อมูลให้อยู่ในบรรทัดฐานเดียวกัน โดยใช้หลักการ Max-Min Normalization (เอกสิทธิ์ พัชรวงศ์ศักดา , 2557) จะทำให้ได้ข้อมูลอยู่ในช่วงระหว่าง 0 ถึง 1 ซึ่งสามารถปรับช่วงค่าข้อมูลได้ดังสมการที่ (1)

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

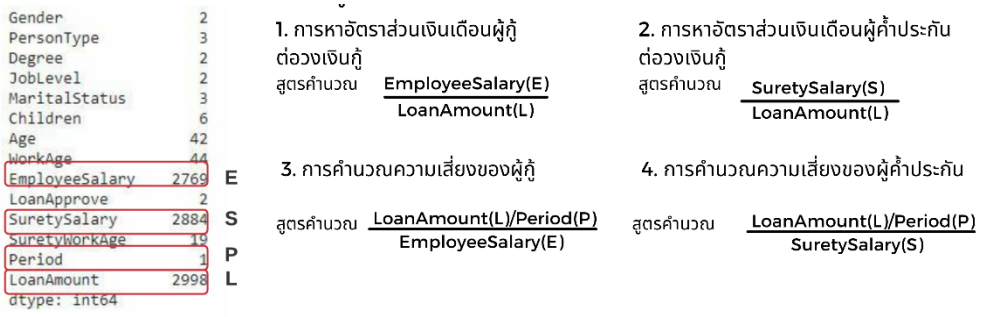
DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4) การรวมข้อมูล (Integrate Data)

เป็นขั้นตอนในการรวบรวมข้อมูลจากแฟ้มข้อมูลที่มีจำนวนตั้งแต่ 2 คอลัมน์ขึ้นไป หรือจากหลาย ๆ แหล่งที่มาที่มีความสัมพันธ์กัน กำหนดตัวแปรที่ต้องการทำนายคือ การอนุมัติเงินกู้(Loan Approve)

5) รูปแบบของข้อมูล (Format Data)

เพื่อให้ข้อมูลเหมาะสมกับระบบที่นำไปใช้และง่ายต่อการวิเคราะห์ให้ตรงเป้าหมายของธุรกิจ จากงานวิจัยนี้ผู้วิจัยได้นำข้อมูลผู้ขอกู้กองทุนสวัสดินำมาแปลงเป็นข้อมูลที่เหมือนกันและแทนด้วยสัญลักษณ์หรือตัวอักษรเอาไว้ เช่นการนำข้อมูลเข้ารหัส ระดับการศึกษา (Degree) ประเภทพนักงาน (Person Type) ประเภทสายงาน (Job Level) เพศ (Gender) สถานะการอนมัติกู้เงิน (LoanApprove) และในขั้นตอนนี้พบว่าจำนวน 14 ปัจจัย ทำให้ไม่เพียงพอต่อการวิเคราะห์ข้อมูล ผู้วิจัยได้ทำการเพิ่ม 4 ปัจจัย มีรายละเอียดสูตรคำนวณดังภาพที่ 1 และรายละเอียดคุณสมบัติดังตารางที่ 2



ภาพที่ 3.3 เพิ่ม 4 คุณสมบัติและสูตรคำนวณ

รายละเอียดคุณสมบัติที่ใช้ในการวิเคราะห์ข้อมูลพิจารณาอนุมัติเงินกู้สวัสดิการพนักงาน 18 ปัจจัย โดยมีรายละเอียดดังนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ตารางที่ 3.1 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

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3.4.4 การพัฒนาแบบจำลอง (Modeling)

1) การเลือกเทคนิคสร้างแบบจำลอง (Select Modeling Technique)

ขั้นตอนในการเลือกเทคนิคที่เหมาะสมกับประเภทของงานข้อจำกัด การแก้ไขปัญหาและสามารถนำมาใช้งานได้จริง ซึ่งในงานวิจัยนี้ผู้วิจัยได้มีการพัฒนาโมเดลเปรียบเทียบความแม่นยำทั้งหมด 4 แบบ ด้วยวิธีเทคนิควิธีนาอีฟเบย์ (Naïve Bayes) เทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor) เทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) เทคนิควิธีแรนดอมฟอร์เรส (Random Forest)

2) การสร้างแบบการทดสอบ (Generate Test Design) ก่อนการสร้างแบบจำลองจะต้องทำการทดสอบความถูกต้องของโมเดล เช่น การนำระดับความผิดพลาดมาใช้ในการวัดคุณภาพของข้อมูลที่ใช้ในการทำเหมืองข้อมูล ดังนั้น การออกแบบการทดสอบควรจะมีการแยกข้อมูลออกเป็น 2 กลุ่ม คือ ชุดข้อมูลที่ใช้ในการทดสอบการสร้างแบบจำลองเพื่อประเมินคุณภาพ (Training set) และชุดข้อมูลที่ใช้งานจริง (Test Set)

3) การสร้างแบบจำลอง (Build Model) ในขั้นตอนนี้ผู้วิจัยได้สร้างโมเดลแบ่งเป็น 2 ขั้นตอน ขั้นตอนแรกเป็นการทำนายการอนุมัติการขอกู้เงินสวัสดิการด้วยวิธีการจำแนกประเภทและทำนาย (Classification and Prediction) ในการพัฒนาระบบซึ่งผลการวิเคราะห์จะแสดงการแบ่งกลุ่มลูกค้าออกเป็น 2 ประเภทคือกลุ่มผู้กู้ที่ได้รับการอนุมัติกลุ่มผู้กู้ที่ไม่ได้รับการอนุมัติ

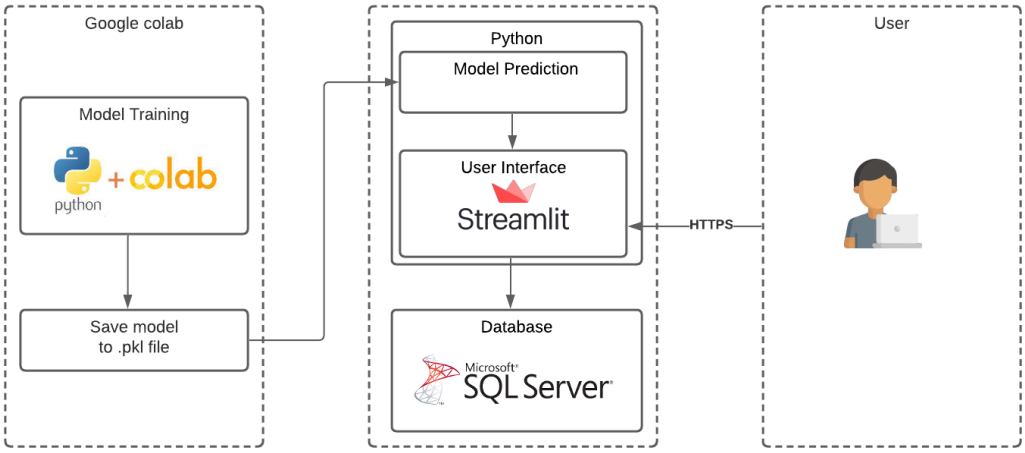
4) การประเมินแบบจำลอง (Assess Model) บันทึกผลการวิเคราะห์จากโมเดล โดนการประเมินแบบจาลองเพื่อให้มั่นใจว่ามีความน่าเชื่อถือ และผ่านมาตรฐานการตัดสินใจ และทำการปรับปรุงการตั้งค่าพารามิเตอร์และประเมินผลจนกว่าจะเชื่อมั่นได้ว่าเป็นรูปแบบของโมเดลดีที่ดีสุด

5) การประเมินแบบจำลอง (Evaluation) เป็นการประเมินโมเดลที่ได้ เพื่อพิจารณาความเหมาะสมในการนำแบบจำลองไปประยุกต์ใช้ว่ามีความแม่นยำกับการทำนายมากน้อยเพียงใด

3.4.6 การนำแบบจำลองไปใช้ (Deployment)

1) การวางแผนการปรับใช้แบบจำลอง (Plan Deployment) ขั้นตอนนี้เป็นการนำผลการประเมินและกลยุทธ์ในการตัดสินใจไปใช้งานจริง ผู้วิจัยนำโมเดลมาประยุกต์ในการพัฒนาระบบ โดยมีรายละเอียดสถาปัตยกรรมของโปรแกรมคอมพิวเตอร์ (Application Architectures) ดังภาพที่ 3.2

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 3.4 สถาปัตยกรรมของโปรแกรมคอมพิวเตอร์ (Application Architectures)

2) แผนการติดตามและการบำรุงรักษา (Plan Monitoring and Maintenance) ขั้นตอนนี้ถือเป็นขั้นตอนที่สำคัญเพราะด้วยพฤติกรรมของความต้องการทางธุรกิจขององค์กรที่เปลี่ยนแปลงอยู่เสมอส่งผลโดยตรงต่อโมเดล ทำให้ผู้พัฒนาจะต้องมีการเตรียมแผนการซ่อมบารุง ซึ่งจะช่วยหลีกเลี่ยงการใช้งานข้อมูลจากข้อมูลเหมืองที่ไม่ถูกต้องได้ และโครงการควรจัดทำแผนการติดตามการดาเนินงานเสมอ

3) จัดทำรายงานขั้นสุดท้าย (Produce Final Report) เมื่อสิ้นสุดโครงการ จะต้องมีการจัดทำรายงาน ซึ่งขึ้นอยู่กับการปรับใช้แผน รายงานนี้อาจนำเสนอผลสรุปโครงการ ตลอดจนการนำเสนอผลลัพธ์ของการทำนายข้อมูล

4) ทบทวนโครงการ (Review Project) การประเมินครั้งสุดท้ายของการทำโครงการว่าสิ่งที่ได้จัดทำนั้นถูกต้องหรือไม่และมีสิ่งที่ต้องดำเนินการปรับปรุงปรับเปลี่ยนหรือไม่ โดยการจัดทำเอกสารที่เกิดจากประสบการณ์การทำงานที่สำคัญที่ได้รับจากการดำเนินงาน เช่น การบันทึกข้อผิดพลาด โดยมุ่งเน้นเพื่อมองหาลู่ทางในการแก้ไข และพัฒนาโครงการครั้งต่อไปให้ดียิ่งขึ้น รวมถึงใช้ส่วนหนึ่งประกอบการตัดสินใจในโครงการที่คล้ายคลึงกัน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 4
ผลการศึกษา

จากการวิจัยการพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานอัตโนมัติด้วยเทควิธีการเรียนรู้ด้วยของเครื่อง ผู้วิจัยประมวลผลข้อมูลตามขั้นตอนของ CRISP-DM ดังที่ได้กล่าวไว้ในบทที่ 3 ซึ่งได้มีการปรับให้เหมาะสมกับกระบวนการ โดยผลการวิจัยและการวิเคราะห์ข้อมูล ดังนี้

4.1 ความเข้าใจธุรกิจประกันชีวิต (Business Understanding)

ลักษณะขององค์กรที่นำมาศึกษาครั้งนี้ คือ มหาวิทยาลัยรัฐบาลแห่งหนึ่ง โดยมีสวัสดิการมหาวิทยาลัย เพื่อให้สมาชิกได้รับสิทธิประโยชน์ โดยกำหนดให้มีสวัสดิการต่าง ๆ ดังนี้ 1. กองทุนสำรองเลี้ยงชีพ 2. เงินกู้กองทุนสวัสดิการ 3. เงินกู้เพื่อที่อยู่อาศัย/ เอนกประสงค์ 4. เงินช่วยเหลือกรณีประสบภัยพิบัติ 5. เงินช่วยเหลือกรณีเสียชีวิต 6. เงินช่วยเหลือกรณีประสบอุบัติเหตุ 7. เงินช่วยเหลือค่ารักษาพยาบาล 8. เงินชดเชย 9. เงินบาเหน็จ/ บำนาญ จากสถานะการโควิด-19 มีผลกระทบต่อการพิจารณาอนุมัติการขอกู้เงินสวัสดิการ เจ้าหน้าการคลังต้องตรวจสอบเอกสารและตรวจสอบวงเงินในการกู้เป็นจำนวนมากและใช้เวลานานเนื่องจากมีผู้ขอกู้จำนวนที่เพิ่มมากขึ้น ดังนั้นผู้วิจัยจึงใช้เทคนิคการทำนายการอนุมัติการขอกู้เงินสวัสดิการด้วยวิธีการจำแนกประเภทและทำนาย (Classification and Prediction) ในการพัฒนาระบบซึ่งผลการวิเคราะห์จะแสดงการแบ่งกลุ่มลูกค้าออกเป็น 2 ประเภทคือ กลุ่มผู้กู้ที่ได้รับการอนุมัติกลุ่มผู้กู้ที่ไม่ได้รับการอนุมัติ ซึ่งผลที่วิเคราะห์ที่ได้จะนำมาออกหลักเกณฑ์การให้กู้เงินสวัสดิการ

4.2 การเตรียมข้อมูล (Data Preparation)

ในขั้นตอนการเตรียมข้อมูลสำหรับนำไปใช้วิเคราะห์ จากที่ผู้วิจัยได้ฐานข้อมูลผู้ขอกู้เงินสวัสดิการ ซึ่งในบางส่วนของข้อมูลไม่มีความจำเป็นในการนำมาใช้วิเคราะห์ ผู้วิจัยจึงได้ทำการคัดกรองข้อมูลให้เหลือเฉพาะส่วนที่สอดคล้องกับวัตถุประสงค์ในการวิจัยครั้งนี้ โดยมีการวิเคราะห์ข้อมูลเบื้องต้นดังนี้ ค่าเฉลี่ยอายุของพนักงานอยู่ที่ 40 ปี และ ต่ำสุด 20 ปี และสูงสุด 61 ปี 75% ของพนักงานที่

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ขอยืนกู้เป็นเพศชายสถานภาพโสด อายุงานประมาณ 15 ปี การศึกษาอยู่ในระดับปริญญาตรีขึ้นไป สายงานสนับสนุน 75 % ยอดวงเงินกู้อยู่ที่ 20,000 บาท อัตราส่วนเงินเดือนผู้กู้ต่อวงเงินกู้เป็น 2 เท่า อัตราส่วนเงินเดือนผู้ค้ำประกันต่อวงเงินกู้เป็น 3 เท่า 75 % ของระดับความเสี่ยงของผู้กู้อยู่ที่ 5 % 75 % ของระดับความเสี่ยงของผู้ค้ำประกันอยู่ที่ 3 %

	count	mean	std	min	25%	50%	75%	max
Id	5870.0	2935.500000	1694.667371	1.00	1468.25	2935.50	4402.75	5870.00
Gender	5870.0	0.298126	0.457474	0.00	0.00	0.00	1.00	1.00
PersonType	5870.0	1.115843	0.439804	1.00	1.00	1.00	1.00	3.00
Degree	5870.0	1.017376	0.130681	1.00	1.00	1.00	1.00	2.00
JobLevel	5870.0	1.726576	0.445755	1.00	1.00	2.00	2.00	2.00
MaritalStatus	5870.0	0.174446	0.380422	0.00	0.00	0.00	0.00	2.00
Children	5870.0	0.532709	0.831647	0.00	0.00	0.00	1.00	8.00
Age	5870.0	40.024532	9.891113	20.00	32.00	40.00	47.00	61.00
WorkAge	5870.0	10.814140	7.877537	0.00	4.00	10.00	15.00	44.00
EmployeeSalary	5870.0	32216.553152	14946.329363	10170.00	21980.00	27070.00	39360.00	122810.00
SuretySalary	5870.0	47570.453748	22895.981338	9140.00	32475.00	40057.50	59040.00	184215.00
SuretyWorkAge	5870.0	15.458433	2.937393	2.00	13.00	15.00	18.00	20.00
Period	5870.0	10.000000	0.000000	10.00	10.00	10.00	10.00	10.00
LoanAmount	5870.0	17878.094463	10157.100570	5040.00	11125.00	14420.00	21243.75	101200.00
EmployeeSalary/ LoanAmount	5870.0	1.922595	0.380851	0.21	2.00	2.00	2.00	5.31
SuretySalary/ LoanAmount	5870.0	2.827537	0.620370	0.31	3.00	3.00	3.00	7.97
DSCR Employee	5870.0	0.057402	0.032522	0.02	0.05	0.05	0.05	0.48
DSCR Surety	5870.0	0.036639	0.023331	0.01	0.03	0.03	0.03	0.32
LoanApprove	5870.0	0.806474	0.395096	0.00	1.00	1.00	1.00	1.00

ภาพที่ 4.1 ตารางการวิเคราะห์ข้อมูลเบื้องต้น

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4.2.1 การคัดเลือกข้อมูล (Data Selection)

งานวิจัยที่ต้องการหาความสัมพันธ์ของการขอกู้เงินสวัสดิการ ผู้วิจัยจึงคัดเลือกเฉพาะข้อมูลผู้ขอยื่นกู้ที่ได้รับการยินยอมค้ำประกันแล้วจำนวน 5,870 รายการ ซึ่งประกอบด้วย เพศ, อายุ, สถานภาพ , อายุงาน, รายได้, จำนวนเงินกู้, จำนวนงวด หรือเลือกตัดข้อมูลที่ไม่จาเป็นต่อการวิเคราะห์ เพื่อลดขนาดของข้อมูล เช่น เลขประจาตัวประชาชน ชื่อ-สกุล ตำแหน่งงาน หน่วยงานที่สังกัด เป็นต้น ดังตารางที่ 4.1

ตารางที่ 4.1 ข้อมูลที่นำมาใช้ในการทำนายข้อมูล

ข้อมูล	คำอธิบายข้อมูล
Degree	ระดับการศึกษา
PersonType	ประเภทพนักงาน
JobLevel	ประเภทสายงาน
Gender	เพศ
LoanApprove	สถานะการอนุมัติกู้เงิน
LoanStatus	ประวัติการชำระล่าช้า
Age	อายุ
WorkAge	อายุการทำงาน
Children	จำนวนบุตร
EmployeeSalary	เงินเดือนผู้กู้
SuretySalary	เงินเดือนผู้ค้ำประกัน
SuretyWorkAge	อายุผู้ค้ำประกัน
Period	จำนวนเดือนที่กู้เงิน
LoanAmount	จำนวนเงินกู้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4.2.2 ทำความสะอาดข้อมูล (Clean Data)

จากการคัดกรองข้อมูลเป็นที่เรียบร้อยแล้ว ผู้วิจัยทำความสะอาดข้อมูลด้วย SQL SRVER ที่มีความซ้ำซ้อนหรือผิดปกติของข้อมูลด้วยการ Remove missing values เติมข้อมูลเงินเดือนที่เป็นค่าว่างตามนโยบายมหาลัยด้วยการ Replace Missing Values เพื่อสร้างคุณภาพให้แก่ข้อมูลที่นำมาใช้อย่างแม่นยำ

4.2.3 การทำข้อมูลให้อยู่ในรูปแบบมาตรฐาน (Data Transformation)

เมื่อรวบรวมข้อมูลทั้งหมดแล้ว พบว่าข้อมูลบางส่วนยังไม่สามารถนำไปสร้าแบบจำลองได้ 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 (1)

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

4.3 การพัฒนาแบบจำลอง (Modeling)

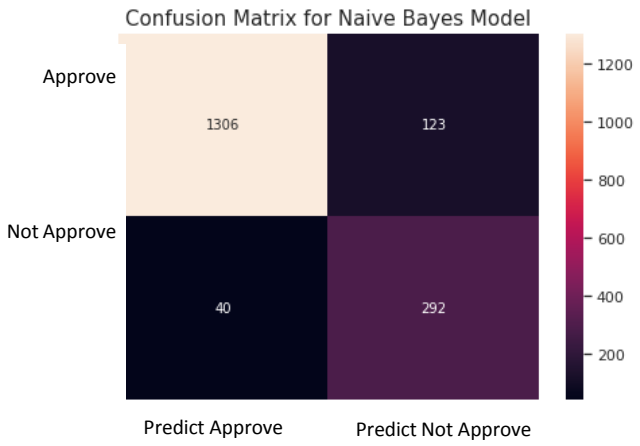
ในการพัฒนาแบบจำลอง ผู้วิจัยได้มีการพัฒนาแบบจำลองเปรียบเทียบความแม่นยำทั้งหมด 4 แบบด้วยวิธีเทคนิควิธีนาอีฟเบย์ (Naïve Bayes) เทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor) เทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) และเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) สำหรับจำแนกประเภทและทำนาย (Classification and Prediction)

4.3.1 ผลการวัดประสิทธิภาพความถูกต้องของโมเดล

ข้อมูลที่ใช้ในการทดสอบจำนวน 5,870 รายการ โดยแบ่งชุดข้อมูลที่ใช้ในการทดสอบการสร้างแบบจำลองเพื่อประเมินคุณภาพ (Training set) 70% จำนวน 4,109 รายการประกอบด้วย ไม่อนุมัติ 804 กรณีและอนุมัติ 3,305 กรณี สำหรับชุดข้อมูลที่ใช้งานจริง (Test Set) 30% จำนวน 1,761 รายการ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1) เทคนิควิธีนาอีฟเบย์ (Naïve Bayes)



ภาพที่ 4.2 Confusion Matrix for Naïve Bayes Model

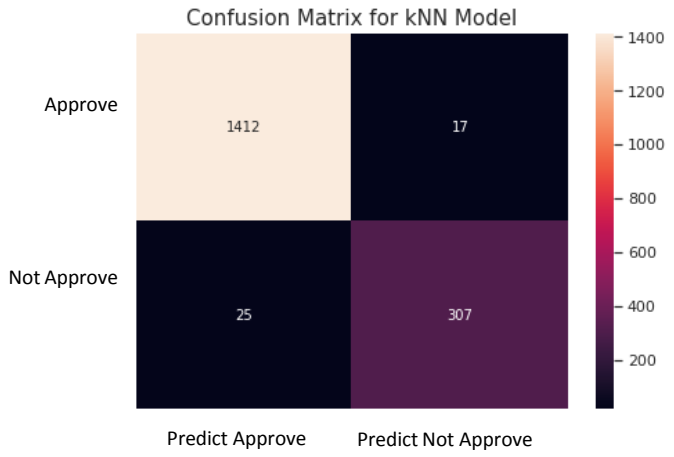
Naive Bayes Model - Classification Report

	precision	recall	f1-score	support
1	0.97	0.91	0.94	1429
0	0.70	0.88	0.78	332
accuracy			0.91	1761
macro avg	0.84	0.90	0.86	1761
weighted avg	0.92	0.91	0.91	1761

ภาพที่ 4.3 Naïve Bayes Model Classification Report

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

2) เทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor)



ภาพที่ 4.4 Confusion Matrix for K-Nearest neighbor Model

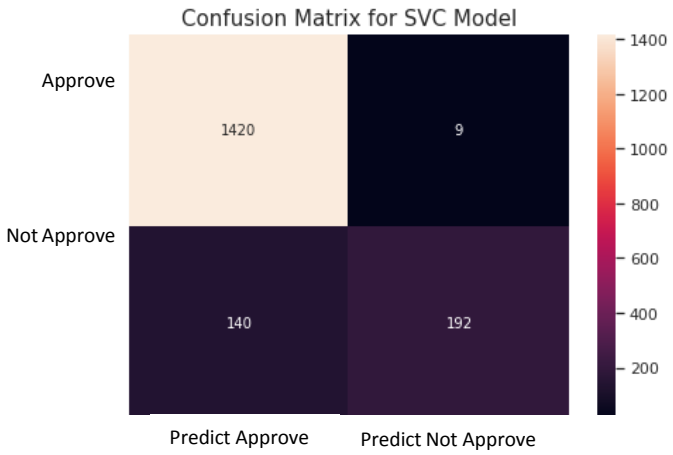
kNN Model - Classification Report

	precision	recall	f1-score	support
1	0.98	0.99	0.99	1429
0	0.95	0.92	0.94	332
accuracy			0.98	1761
macro avg	0.97	0.96	0.96	1761
weighted avg	0.98	0.98	0.98	1761

ภาพที่ 4.5 K-Nearest neighbor Model Classification Report

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3) เทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine)



ภาพที่ 4.6 Confusion Matrix for Support Vector Model

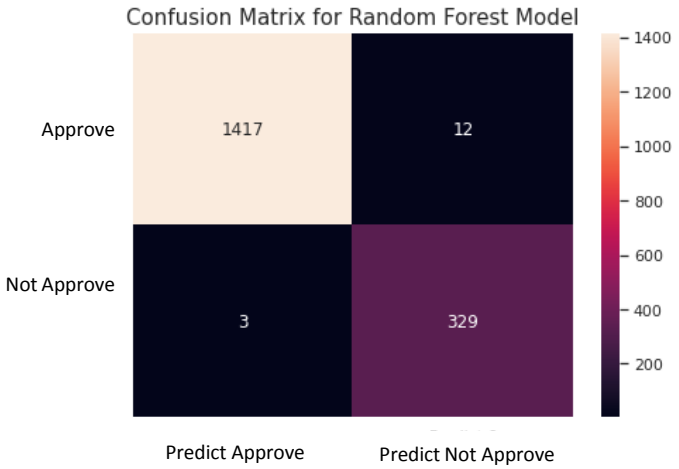
SVC Model - Classification Report

	precision	recall	f1-score	support
1	0.91	0.99	0.95	1429
0	0.96	0.58	0.72	332
accuracy			0.92	1761
macro avg	0.93	0.79	0.84	1761
weighted avg	0.92	0.92	0.91	1761

ภาพที่ 4.7 Support Vector Model Classification Report

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4) เทคนิควิธีแรนดอมฟอร์เรส (Random Forest)



ภาพที่ 4.8 Confusion Matrix for Random Forest Model

Random Forest Model - Classification Report

	precision	recall	f1-score	support
1	1.00	0.99	0.99	1429
0	0.96	0.99	0.98	332
accuracy			0.99	1761
macro avg	0.98	0.99	0.99	1761
weighted avg	0.99	0.99	0.99	1761

ภาพที่ 4.9 Random Forest Model Classification Report

4.3.2 เลือกโมเดลที่มีประสิทธิภาพมากที่สุด

จากผลการวัดประสิทธิภาพความถูกต้องของโมเดลพบว่า เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) มีความแม่นจำสูงที่สุด (Accuracy) 0.9915 สาเหตุที่เกิดข้อผิดพลาดในการทำนายมีรายละเอียดดังนี้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

1) กรณีเกิด False Positives (FP), Type I Error = 3 เนื่องจากอายุงานพนักงาน (WorkAge) ต่ำกว่า 3 ปี แต่อายุงานผู้ค้ำประกัน (SuretyWorkAge) มากกว่า 3 ปี Class Probabilities False Positives (FP) มีค่าเท่ากับ 0.946

2) กรณีเกิด False Negatives (FN), Type II Error = 12 เนื่องจากอายุงานพนักงาน (WorkAge) ต่ำกว่า 3 ปี และอายุงานผู้ค้ำประกัน (SuretyWorkAge) น้อยกว่า 3 ปี Class Probabilities False Positives (FP) มีค่าเท่ากับ 0.967

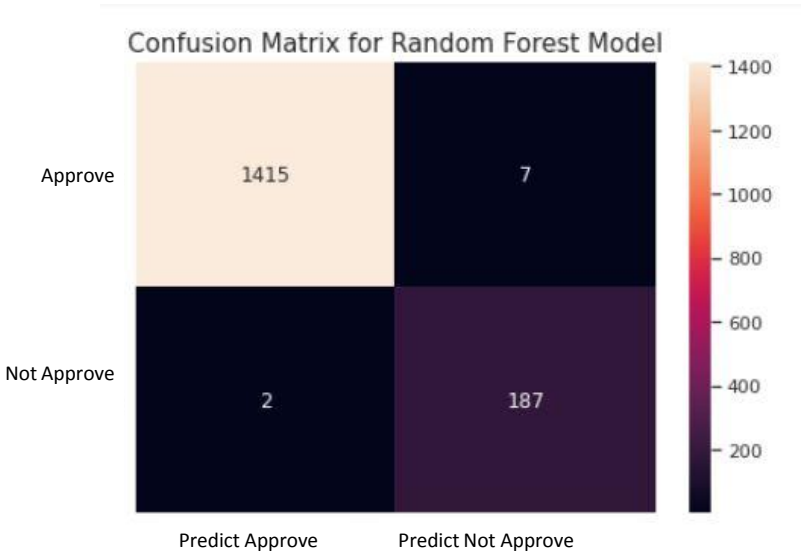
เทคนิควิธีแรนดอมฟอร์เรส (Random Forest) ทำนายการอนุมัติเอกสารการขอกู้เงินสวัสดิการมหาวิทยาลัยอีกครั้งด้วยปัจจัย (Feature) ดังตารางที่ 4.2

ตารางที่ 4.2 เทคนิควิธีแรนดอมฟอร์เรสด้วยคุณสมบัติใหม่ (Random Forest)

ข้อมูล	คำอธิบายข้อมูล
Gender	เพศ
PersonType	ประเภทพนักงาน
Degree	ระดับการศึกษา
JobLevel	ประเภทสายงาน
MaritalStatus	สถานะภาพสมรส
Children	จำนวนบุตร
Age	อายุผู้กู้
WorkAge	อายุงานผู้กู้
SuretyAge	อายุงานผู้ค้ำประกัน
(EmployeeSalary/LoanAmount)	สัดส่วนเงินเดือนผู้กู้ต่อวงเงินกู้
(SuretySalary/LoanAmount)	สัดส่วนเงินเดือนผู้ค้ำประกันต่อวงเงินกู้
LoanAmount/Period) /EmployeeSalary	ความเสี่ยงของผู้กู้
(LoanAmount/Period) /SuretySalary	ความเสี่ยงของผู้ค้ำประกัน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

3) เทคนิควิธีแรนดอมฟอร์เรสด้วยคุณสมบัติใหม่ (Random Forest)



ภาพที่ 4.10 Confusion Matrix for Random Forest New Model

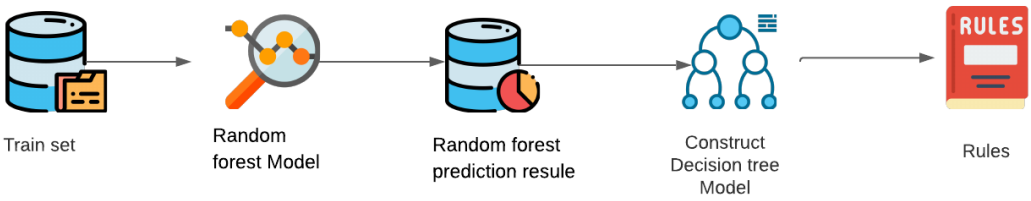
Random Forest Model - Classification Report

	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 4.11 Random Forest New Model Classification Report

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4) แนวทางการสกัด Rule จากเทคนิควิธีแรนดอมฟอร์เรสด้วยคุณสมบัติใหม่ (Random Forest)



ภาพที่ 4.12 ตารางแนวทางการสกัด Rule จากเทคนิควิธีแรนดอมฟอร์เรส (Random Forest)

4.1 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%]

4.4 การตรวจสอบและประเมินผล (Evaluation)

จากแบบจำลองพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานที่ได้จาก google colab สามารถตรวจสอบความแม่นยำและความเหมาะสมในการนำแบบจำลองไปประยุกต์ใช้นั้นจะแม่นยำมากน้อยเพียงใด โดยการประเมินจะแบ่งตามลักษณะของการทำเหมืองข้อมูล ดังนี้

4.4.1 ตรวจสอบและประเมินผล

ผู้วิจัยได้วิธีการประเมินผลโดยการนำข้อมูลชุดทดสอบ (Test Set) มาทำการประมวลผลเพื่อหาค่าความแม่นยำ (Accuracy) จากการจำแนกประเภท ซึ่งค่าแตกต่างระหว่างข้อมูลชุดทดสอบ (Test Set) และข้อมูลที่นำมาใช้ในการวิจัย (Training Set) พบว่าค่าความแม่นยำ (Accuracy) ของข้อมูล

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

มีค่ามากที่สุด ส่วนการวิเคราะห์ข้อมูลความสัมพันธ์นั้น สามารถประเมินผลที่ได้จากการพิจารณาค่าความเชื่อมั่น (Confidence) ของความสัมพันธ์

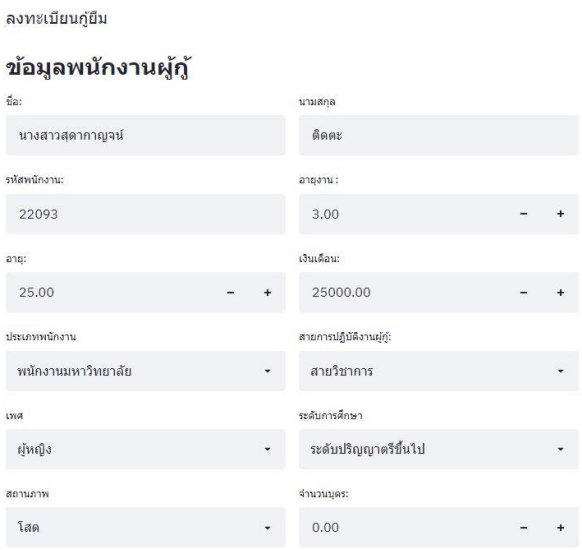
4.4.2 ตรวจสอบและประเมินผลจากการปรึกษาผู้เชี่ยวชาญ

เมื่อตรวจสอบและประเมินผลการวิจัยจากโปรแกรม google colab แล้ว ผู้วิจัยได้นำผลการจำแนกประเภท (Classification) และการวิเคราะห์ความสัมพันธ์ของข้อมูลไปปรึกษาผู้เชี่ยวชาญฝ่ายการคลัง เพื่อให้ผู้เชี่ยวชาญได้ช่วยดูความเหมาะสมของการจำแนกประเภท (Classification) การอนุมัติและไม่อนุมัติเงินกู้สวัสดิการพนักงาน ช่วยประเมินความสัมพันธ์ของข้อมูลแต่ละกลุ่ม ถึงความเป็นไปได้และแนวทางในการออกนโยบายการกู้เงินสวัสดิการพนักงาน เพื่อให้งานวิจัยชิ้นนี้มีประสิทธิภาพมากยิ่งขึ้น

4.5 การประยุกต์ใช้งาน (Deployment)

เมื่อได้แบบจำลองพิจารณาอนุมัติเงินกู้สวัสดิการพนักงานนำไปประยุกต์ใช้ในระบบสวัสดิการ กองทุนกู้ยืมมหาวิทยาลัย โดยให้ผู้กู้บันทึกข้อมูลการขอกู้ผ่านระบบ เพื่อทดสอบความแม่นยำของแบบจำลองมีรายละเอียดดังนี้

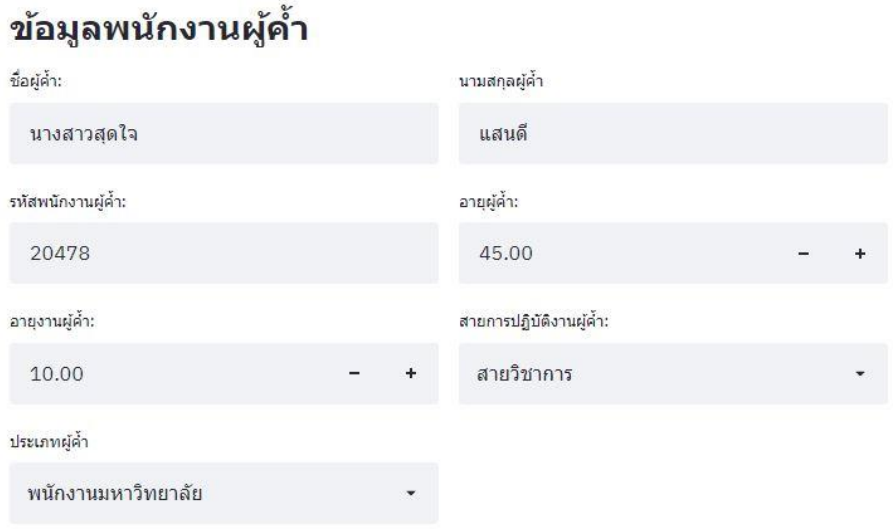
4.5.1 ตัวอย่างหน้าจอข้อมูลพนักงานผู้กู้



ภาพที่ 4.13 ภาพตัวอย่างหน้าจอข้อมูลพนักงานผู้กู้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4.5.2 ตัวอย่างหน้าจอข้อมูลพนักงานผู้ค้ำประกัน



ภาพที่ 4.14 ภาพตัวอย่างหน้าจอข้อมูลพนักงานผู้ค้ำประกัน

4.5.3 ตัวอย่างหน้าจอคำนวณการผ่อนชำระเงินกู้

ภาพที่ 4.15 ตัวอย่างหน้าจอคำนวณการผ่อนชำระเงินกู้

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

4.5.4 ตัวอย่างหน้าจอการแสดงผลเมื่อส่งพิจารณาการขอกู้เงินสวัสดิการ

1) กรณีผ่านการพิจารณาอนุมัติ

ภาพที่ 4.16 ตัวอย่างหน้าจอกรณีผ่านการพิจารณาอนุมัติ

2) กรณีผ่านการพิจารณาอนุมัติ

ภาพที่ 4.17 ตัวอย่างหน้าจอกรณีไม่ผ่านการพิจารณาอนุมัติ

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บทที่ 5
บทสรุปและข้อเสนอแนะ

งานวิจัยนี้ได้นำเสนอเกี่ยวกับการพัฒนาโมเดลเพื่อช่วยพิจารณาอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย โดยสามารถสรุปผลการวิจัยได้ดังนี้

5.1 สรุปผลการศึกษา

5.1.1 โมเดลที่มีประสิทธิภาพดีที่สุดในการทำนายการอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัย

ผลการศึกษาพบว่า ประสิทธิภาพการทำนายของโมเดลเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) พบค่าความแม่นยำสูงสุดที่ 99.56โมเดลเทคนิควิธีความใกล้เคียงกันมากที่สุด (K-Nearest neighbor) ที่ค่าความแม่นยา 97.50 โมเดลเทคนิควิธีซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine) ที่ค่าความแม่นยำ 91.53 และโมเดลเทคนิควิธีนาอีฟเบย์ (Naïve Bayes) ที่ค่าความแม่นยำที่ 90.74 ตามลำดับ

	Naive Bayes	kNN	SVM	Random Forest
Accuracy	0.907439	0.975014	0.915389	0.9956

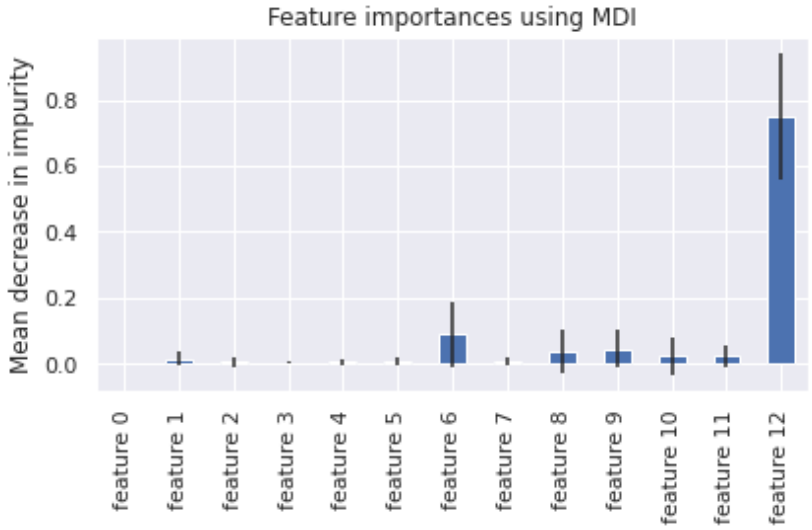
ภาพที่ 5.1 ผลประสิทธิภาพการทำนายของโมเดล

5.1.2 ปัจจัยที่มีความสำคัญ (Feature importance)

จากการสร้างแบบจำลองในการทำนายการกู้เงินสวัสดิการมหาวิทยาลัย มีการคัดเลือกปัจจัยทั้งหมด 13 ปัจจัย จากการทำนายปัจจัยที่มีผลมากที่สุดคือ Feature 12 ความเสี่ยงของผู้ค้ำประกัน (LoanAmount/Period)/SuretySalary รองลงมาคือ Feature 6 อายุผู้กู้ (Age) Feature 9 อัตราส่วน

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

เงินเดือนผู้กู้ต่อวงเงิน (LoanAmount/Period)/EmployeeSalary Feature 8 อายุงานผู้ค้ำประกัน (Surety Age) Feature 10 อัตราส่วนเงินเดือนผู้ค้ำประกันต่อวงเงินกู้ (SuretySalary/LoanAmount) Feature 11 ความเสี่ยงของผู้กู้ (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 5.1



ภาพที่ 5.2 ปัจจัยที่มีผลต่อการทำนายของโมเดล

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ตารางที่ 5.1 ปัจจัย (Feature) ที่ใช้ในการทำนายของโมเดล

ปัจจัย (Feature)	คำอธิบายข้อมูล
Feature 0	เพศ (Gender)
Feature 1	ประเภทพนักงาน (PersonType)
Feature 2	ระดับการศึกษา (Degree)
Feature 3	ประเภทสายงาน (JobLevel)
Feature 4	สถานะภาพสมรส (MaritalStatus)
Feature 5	จำนวนบุตร (Children)
Feature 6	อายุผู้กู้ Age
Feature 7	อายุงานผู้กู้ WorkAge
Feature 8	อายุงานผู้ค้ำประกัน SuretyAge
Feature 9	สัดส่วนเงินเดือนผู้กู้ต่อวงเงินกู้ (EmployeeSalary/LoanAmount)
Feature 10	สัดส่วนเงินเดือนผู้ค้ำประกันต่อวงเงินกู้ (SuretySalary/LoanAmount)
Feature 11	ความเสี่ยงของผู้กู้ (LoanAmount/Period) /EmployeeSalary
Feature 12	ความเสี่ยงของผู้ค้ำประกัน (LoanAmount/Period) /SuretySalary

5.1.3 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%]

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

5.1.4 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

5.1.5 ข้อผิดพลาดในการทำนาย

จากการวัดประสิทธิภาพดีที่สุดในการทำนายการอนุมัติเอกสารขอกู้เงินสวัสดิการมหาวิทยาลัยพบว่าเทคนิควิธีแรนดอมฟอร์เรส (Random Forest) พบค่าความแม่นยำสูงสุดที่ 99.56 ซึ่งเกิดข้อผิดพลาดในการทำนายทั้งหมด 15 กรณีโดยมีสาเหตุดังนี้

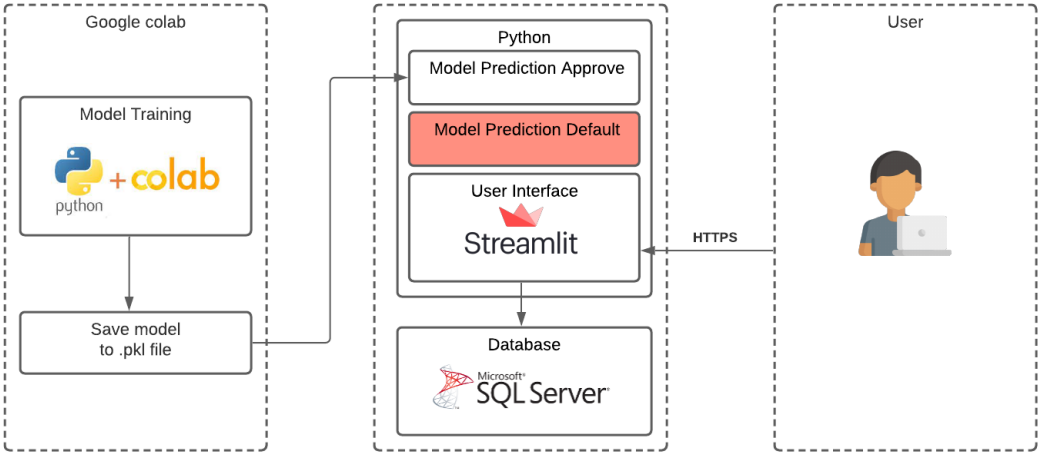
1. กรณีเกิด False Positives (FP), Type I Error = 2 เนื่องจากอายุงานพนักงาน (WorkAge) ต่ำกว่า 3 ปี แต่อายุงานผู้ค้ำประกัน (SuretyWorkAge) มากกว่า 3 ปี

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

5.2 ข้อเสนอแนะและการพัฒนางานในอนาคต

เนื่องจากการทำนายการขอกู้เงินสวัสดิการมหาวิทยาลัยไม่สามารถอธิบายการเกิดหนี้เสียได้เนื่องจากผู้ขอกู้ไม่สามารถชำระหนี้ให้กับมหาวิทยาลัยตามที่กำหนด ส่งผลให้ผู้ค้ำประกันต้องการชำระหนี้แทนผู้ขอกู้ โดยผู้วิจัยวางแผนการพัฒนางานในอนาคต (Future Work) หลังจากทำนายการอนุมัติขอกู้เงิน ผลการทำนายในส่วนที่ผ่านการอนุมัติกู้มาวิเคราะห์โอกาสการเกิดหนี้เสีย โดยมีรายละเอียดสถาปัตยกรรมของโปรแกรมคอมพิวเตอร์ (Application Architectures) ดังภาพที่ 5.3

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์



ภาพที่ 5.3 สถาปัตยกรรมการพัฒนางานในอนาคตของโปรแกรมคอมพิวเตอร์ (Application Architectures)

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

บรรณานุกรม

ภาษาไทย

กิตติ ภักดีวัฒนะกุล และจำลอง ครูอุตสาหะ (2546): การออกแบบฐานข้อมูล Database Design (พิมพ์ครั้งที่ 4). กรุงเทพฯ: เคทีพี คอมพ์ แอนด์ คอนซัลท์

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

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

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

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

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (2561): “เทคโนโลยี (วิทยาการคำนวณ)”. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

ภาษาต่างประเทศ

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); 2013, 26-28 August; Khartoum, sudan; 2013. pp.378-383;

Gavin Powell (2005): Beginning Database Design. Indiana: Wiley Publishing, Inc.

Hengphraphrom K, Hengphraphrom S, MakwibunchaiS (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 Interdiscip Res. 2014; 9(1):71–80

L. Breiman (2001): Random Forests. Machine Learning Journal, Vol. 45, No. 1, 2001, p. 532.

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

VilailuckS, Jaroenpuntaruk V, 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. 2015;2(2):1–17

T.L. Saaty (2006): Fundamentals of Decision Making and Priority Theory with the Analytic Hierarchy

Process. (2nd edition), Pittsburgh, PA: RWS, 2006, pp. 4-5.

Mohammed J. Zaki and Wagner Meira, Jr.(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.

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); 2010, 11-14 December; Nanning, pp.378-383.

DPU DHURAKIJ PUNDIT UNIVERSITY มหาวิทยาลัยธุรกิจบัณฑิตย์

ประวัติผู้เขียน

ชื่อ-นามสกุล	นางสาวสุดากาญจน์ ติดตะ
ประวัติการศึกษา	วิทยาศาสตรบัณฑิต สาขาวิศวกรรมซอฟต์แวร์ มหาวิทยาลัยนอร์ทกรุงเทพ ปีการศึกษา 2560