

## บทที่ 4

### ผลการดำเนินการวิจัย

#### 4.1 การวิเคราะห์และสรุปผลการดำเนินการวิจัย

จากการประมวลผลข้อมูลด้วยโปรแกรม WEKA เวอร์ชัน 3.6.2 สามารถสรุปผลการประเมินประสิทธิภาพโมเดลการตัดสินใจของทั้ง 2 เทคนิค ได้แก่ เทคนิค J48 และเทคนิค ID3 ได้ดังรายละเอียดในตารางต่อไปนี้

ตารางที่ 4.1 เปรียบเทียบความสามารถในการจำแนกข้อมูล หรือการพยากรณ์ของโมเดล

Technique	J48	ID3
Measurement		
Correctly Classified (No. Instances, (%))	12,495 (96.9131%)	12,489 (96.8665%)
Incorrectly Classified (No. Instances, (%))	398 (3.0869%)	400 (3.1025 )
Unclassified Instances	-	4
<b>Total Instances</b>	<b>12,893</b>	<b>12,893</b>
Root Mean Squared Error	0.1671	0.1659

หมายเหตุ. คำว่า “อินสแตนซ์ (Instance)” ในโปรแกรม WEKA หมายถึง เร็คคอร์ด (Record) หรือแถว (Row) และคำว่า “แอตทริบิวต์ (Attribute)” ในโปรแกรม WEKA หมายถึง คอลัมน์ (Column) หรือฟิลด์ (Field)

จากตารางที่ 4.1 ผลการเปรียบเทียบความสามารถในการจำแนกข้อมูลของโมเดลที่ได้จากเทคนิค J48 และ ID3 สรุปผลได้ดังนี้

1. ค่าความถูกต้องแม่นยำในการจำแนกข้อมูล (Correctly Classified) ของเทคนิค J48 และ ID3 มีค่าเท่ากับ 96.91% และ 96.86% ตามลำดับ
2. ค่าความผิดพลาดในการจำแนกข้อมูล (Incorrectly Classified) ของเทคนิค J48 และ ID3 มีค่าเท่ากับ 3.09% และ 3.10% ตามลำดับ
3. จำนวนเรีคคอร์ด ที่ไม่สามารถจำแนกข้อมูลได้ (Unclassified Instances) ของเทคนิค J48 สามารถจำแนกข้อมูลได้ทุกเรีคคอร์ด และ ID3 มีจำนวน 4 เรีคคอร์ด
4. จำนวนเรีคคอร์ด ที่ใช้ห้รับการเรียนรู้และสร้างรูปแบบ โมเดลนี้มีจำนวน 12,893 เรีคคอร์ด
5. ค่าดัชนีความคลาดเคลื่อนในการจำแนกข้อมูลระหว่างค่าจริงและค่าพยากรณ์ (Root Mean Squared Error) ของเทคนิค J48 และ ID3 มีค่าเท่ากับ 0.1671 และ 0.1659 ตามลำดับ

ตารางที่ 4.2 สูตรคำนวณค่า TP, FP และ Precision

True Positive Rate (TP) or Recall Rate		False Positive Rate (FP)		Precision	
Pass	No	Pass	No	Pass	No
$a/(a+b)$	$d/(c+d)$	$c/(c+d)$	$b/(a+d)$	$a/(a+c)$	$d/(b+d)$

ตารางที่ 4.3 การเปรียบเทียบจำนวนข้อมูลจำแนกตามคลาสจริง กับผลการพยากรณ์ของโมเดล

Technique & Classified as Fact Value	J48		ID3	
	Pass	No	Pass	No
Pass (11,868)	11,774 (a)	94 (b)	11,769 (a)	95 (b)
No (1,025)	304 (c)	721 (d)	305 (c)	720 (b)

จากตารางที่ 4.3 แสดงการเปรียบเทียบจำนวนข้อมูลจำแนกตามคลาสจริง กับผลการ

พยากรณ์ของโมเดลที่ได้จากการใช้เทคนิค J48 และ ID3 สรุปผลได้ดังนี้

1. จำนวนเรีคคอร์ดที่มีคลาสจริงมีค่าเป็น Pass จำนวน 11,868 นั้น พบว่าผลการพยากรณ์ของโมเดลที่ได้จากการใช้เทคนิค J48 ที่พยากรณ์เป็น Pass จำนวน 11,774 เรีคคอร์ด และพยากรณ์เป็น No จำนวน 94 เรีคคอร์ด และพบว่าผลการพยากรณ์ของเทคนิค ID3 ที่พยากรณ์เป็น Pass จำนวน 11,769 เรีคคอร์ด และพยากรณ์เป็น No จำนวน 95 เรีคคอร์ด แสดงว่าโมเดลที่ได้จากเทคนิค ID3 ไม่สามารถพยากรณ์ข้อมูลได้ จำนวน 4 เรีคคอร์ด

2. จำนวนเรีคคอร์ดที่มีคลาสจริงมีค่าเป็น No จำนวน 1,025 เรีคคอร์ด พบว่ามีผลการพยากรณ์ของโมเดลที่ได้จากการใช้เทคนิค J48 ที่พยากรณ์เป็น Pass มีจำนวน 304 เรีคคอร์ด และพยากรณ์เป็น No มีจำนวน 721 เรีคคอร์ด และพบว่าผลการพยากรณ์ของเทคนิค ID3 ที่พยากรณ์เป็น Pass จำนวน 305 เรีคคอร์ด และพยากรณ์เป็น No จำนวน 720 เรีคคอร์ด

ตารางที่ 4.4 เปรียบเทียบค่าดัชนีชี้วัดประสิทธิภาพในการจำแนกข้อมูล หรือการพยากรณ์ของโมเดล

Class \ Measure	J48			ID3		
	Pass	No	Average	Pass	No	Average
True Positive Rate	0.992	0.703	0.969	0.992	0.702	0.969
False Positive Rate	0.297	0.008	0.274	0.298	0.008	0.275
Precision	0.975	0.885	0.968	0.975	0.883	0.967

จากตารางที่ 4.4 แสดงการเปรียบเทียบค่าดัชนีชี้วัดประสิทธิภาพในการจำแนกข้อมูล หรือการพยากรณ์ของโมเดล ซึ่งสามารถคำนวณได้จากการนำข้อมูลในตารางที่ 4.3 ซึ่งผู้วิจัยนำข้อมูลมาจากค่าที่แสดงอยู่ในส่วนของ Confusion Matrix ของโปรแกรม WEKA จากนั้นนำมาเข้าสู่สูตรคำนวณตามสูตรในตารางที่ 4.2 และเมื่อพิจารณาค่าดัชนีชี้วัดประสิทธิภาพโมเดลจากค่าเฉลี่ย (Average) สามารถสรุปผลได้ดังนี้

1. ค่า True Positive Rate (Average) หรือค่าที่ใช้วัดผลการพยากรณ์ตามรูปแบบโมเดลที่ถูกค้นพบ แล้วให้คำตอบตรงกับข้อมูลจริง หรือเรียกว่าค่า “Recall” พบว่าทั้งเทคนิค J48 และ ID3 มีค่าเท่ากันคือ 0.969

2. ค่า False Positive Rate (Average) หรือค่าที่ใช้วัดผลการพยากรณ์ตามรูปแบบโมเดลที่ถูกค้นพบ แล้วให้คำตอบเป็นเท็จ หมายถึง การที่ข้อมูลจริงไม่ได้อยู่ในคลาสนั้นแต่คำพยากรณ์จากโมเดลบอกว่าอยู่ในคลาสนั้น พบว่าของเทคนิค J48 และ ID3 มีค่าเท่ากับ 0.274 และ 0.275 ตามลำดับ

3. ค่า Precision (Average) หรือค่าที่ใช้วัดความเชื่อมั่นของผลการพยากรณ์ พบว่าของเทคนิค J48 และ ID3 มีค่าเท่ากับ 0.968 และ 0.967 ตามลำดับ

จากการเปรียบเทียบข้อดีข้อเสียของทั้ง 2 เทคนิค พบว่ารูปแบบโมเดลการตัดสินใจที่ได้จากเทคนิค J48 มีข้อดีมากกว่าเทคนิค ID3 ทั้งในแง่ของลำดับขั้นการตัดสินใจที่ไม่เยอะจนเกินไป ดังที่แสดงไว้ในบทที่ 3 และมีค่าดัชนีที่ใช้วัดประสิทธิภาพของรูปแบบโมเดลการตัดสินใจ ที่แสดงถึงความมีประสิทธิภาพในการจำแนกข้อมูลหรือสามารถพยากรณ์ข้อมูลของเทคนิค J48 นั้น ดีกว่าเทคนิค ID3

#### 4.2 การประยุกต์ใช้โมเดลการตัดสินใจ

จากการวิเคราะห์และสรุปผลการประเมินประสิทธิภาพโมเดลการตัดสินใจข้างต้น ผู้วิจัยจึงได้พัฒนาโปรแกรมพยากรณ์โอกาสที่บัญชีเงินกู้เบิกเกินบัญชีจะได้รับการทบทวนวงเงิน ที่มีลักษณะการใช้งานแบบ Web Application โดยใช้ภาษา PHP และกำหนดให้การพยากรณ์ข้อมูลนั้นเรียกใช้รูปแบบโมเดลการตัดสินใจที่ได้จากเทคนิค J48 จากโปรแกรม WEKA โดยตรง

ส่วนประกอบของโปรแกรมพยากรณ์โอกาสที่บัญชีเงินกู้เบิกเกินบัญชีจะได้รับการทบทวนวงเงิน ประกอบด้วย 3 ส่วน ดังนี้

1. หน้าจอหลัก ใช้สำหรับการพยากรณ์
2. หน้าจอแสดง Decision Tree ใช้สำหรับผู้จัดทำข้อมูล เพื่อทำความเข้าใจรูปแบบการตัดสินใจของตัวโปรแกรมพยากรณ์ โดยตัวเลขที่อยู่ในวงเล็บใช้สำหรับอธิบายข้อมูล หรือเป็นการแจกแจงให้ทราบเกี่ยวกับข้อมูลทางเทคนิคสำหรับผู้วิเคราะห์ข้อมูล ตัวเลขด้านหน้า คือ จำนวนเรีคคอร์ดของข้อมูลชุดเรียนรู้ที่มีเงื่อนไขรูปแบบการตัดสินใจแบบนั้นๆ ว่ามีกี่เรีคคอร์ด และตัวเลขด้านหลัง คือ จำนวนเรีคคอร์ดที่ข้อมูล Class ของผลการพิจารณาทบทวนวงเงินจริงขัดแย้งกับผลการพยากรณ์ของโมเดล

3. หน้าจอแสดง Decision Tree ใช้สำหรับผู้ใช้ทั่วไป เพื่อทำความเข้าใจรูปแบบการตัดสินใจของตัวโปรแกรมพยากรณ์ ทั้งนี้ เพื่อความเข้าใจง่าย ผู้วิจัยจึงได้แปลงข้อมูลทางเทคนิคในข้อ 2. ให้แสดงผลเป็นเปอร์เซ็นต์ (%) ของความน่าจะเป็นตามผลการพยากรณ์

**Auto Review**

Prediction Decision Model Decision Model(User)

**OD A/C Info.**

Account Number: xxx-0-xxxx-x  
 Utilization(%): <=80  
 LTV(%): <=75  
 Movement(%): 10-14.99  
 Payment History(%): Bad

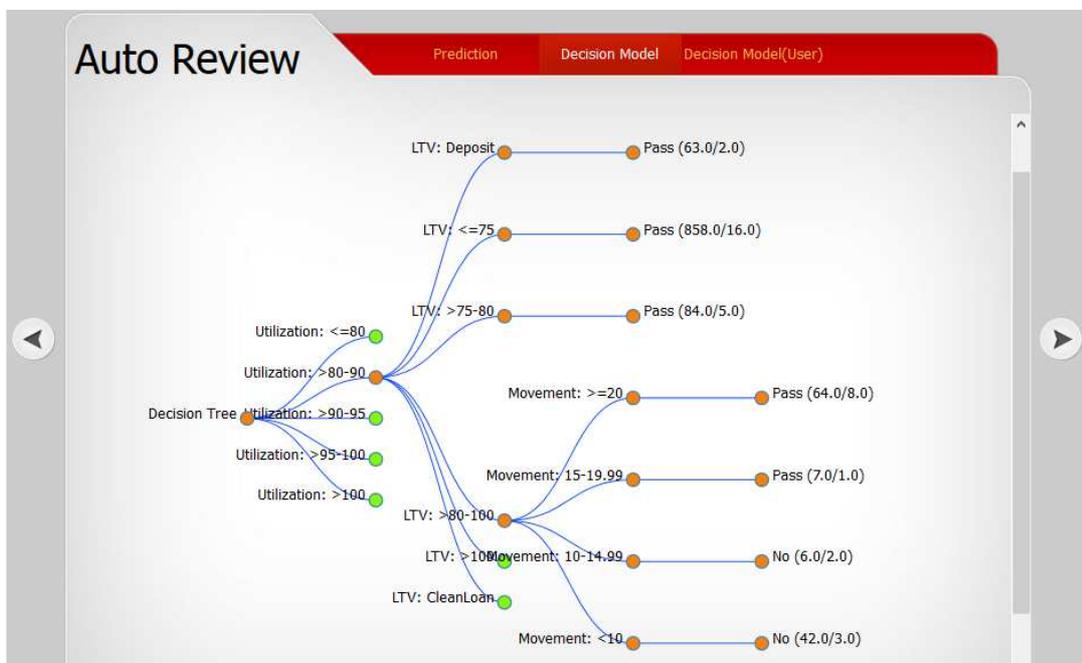
Clear Predict

---

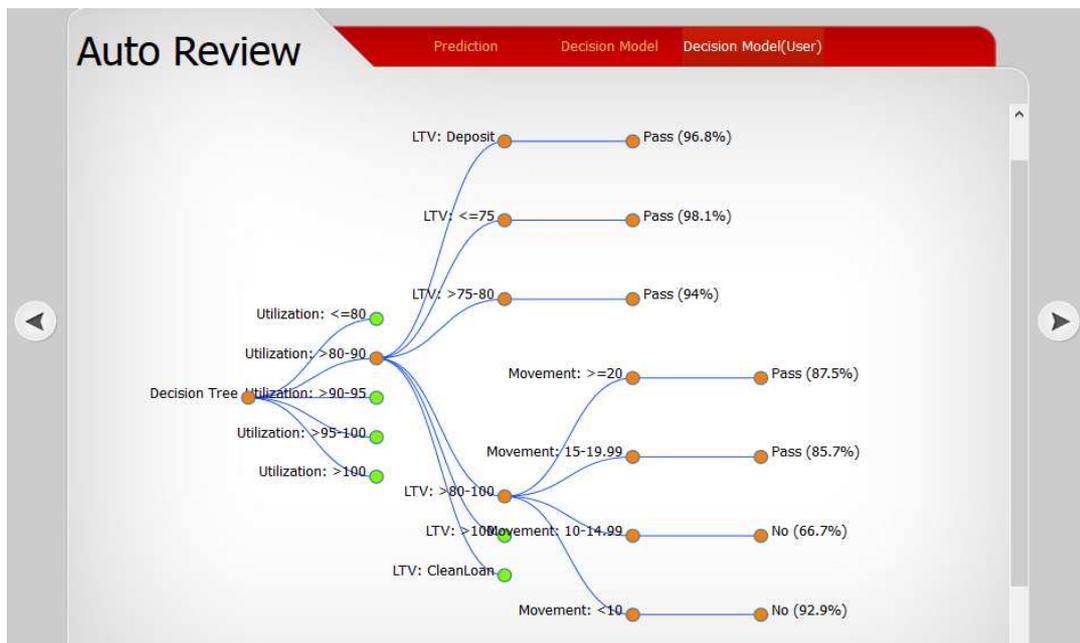
**Prediction Result**

Account No: xxx-0-xxxx-x  
 Prediction: Pass  
 Probability: 0.993  
 Sign:

ภาพที่ 4.1 หน้าจอแสดงผลที่ใช้สำหรับพยากรณ์



ภาพที่ 4.2 หน้าจอแสดงผล Decision Tree สำหรับผู้จัดทำข้อมูล



ภาพที่ 4.3 หน้าจอแสดงผล Decision Tree สำหรับผู้ใช้ทั่วไป

#### 4.3 ผลการเปรียบเทียบผลลัพธ์จากโปรแกรมพยากรณ์ กับผลการพิจารณาทบทวนวงเงินจากข้อมูลจริง

จากการนำข้อมูลทดสอบชุดที่ 2 จำนวน 872 เร็คคอร์ด มาทำการทดสอบเพื่อประเมินประสิทธิภาพของโปรแกรมพยากรณ์ โอกาสที่บัญชีเงินกู้เบิกเกินบัญชีจะได้รับการทบทวนวงเงินแล้วเปรียบเทียบผลลัพธ์จากโปรแกรมพยากรณ์กับผลการพิจารณาทบทวนวงเงินจากข้อมูลจริง โดยสามารถแสดงผลการเปรียบเทียบ จำแนกตามรูปแบบ โมเดลการตัดสินใจของเทคนิค J48 ทั้ง 55 รูปแบบ ดังนี้

ตารางที่ 4.5 เปรียบเทียบผลลัพธ์จากโปรแกรมพยากรณ์กับผลการพิจารณาทบทวนวงเงินจากข้อมูลจริง

Seq.	Utilization	LTV	Movement	Payment	Predict. Result	Prob.	% Prob.	No. Test Case	TRUE	FALSE	% False/Total Test Case
1	<=80				Pass	0.993	99.3	520	517	3	1%
2	>80-90	Deposit			Pass	0.968	96.8	6	6	0	0%
3	>80-90	<=75			Pass	0.981	98.1	49	48	1	2%
4	>80-90	>75-80			Pass	0.94	94	2	2	0	0%
5	>80-90	>80-100	>=20		Pass	0.875	87.5	6	6	0	0%
6	>80-90	>80-100	15-19.99		Pass	0.857	85.7	1	1	0	0%
7	>80-90	>80-100	10-14.99		No	0.667	66.7	1	0	1	100%
8	>80-90	>80-100	<10		No	0.929	92.9	1	1	0	0%
9	>80-90	>100	>=20		Pass	0.871	87.1	6	6	0	0%
10	>80-90	>100	15-19.99		Pass	1	100				
11	>80-90	>100	10-14.99		Pass	0.889	88.9				
12	>80-90	>100	<10		No	0.682	68.2	1	0	1	100%
13	>80-90	CleanLoan			Pass	0.895	89.5	10	8	2	20%
14	>90-95	Deposit			Pass	0.96	96	2	2	0	0%
15	>90-95	<=75			Pass	0.95	95	43	43	0	0%
16	>90-95	>75-80	>=20		Pass	0.946	94.6	1	1	0	0%
17	>90-95	>75-80	15-19.99		Pass	1	100	1	1	0	0%
18	>90-95	>75-80	10-14.99		Pass	1	100	1	1	0	0%
19	>90-95	>75-80	<10		No	0.963	96.3	1	1	0	0%
20	>90-95	>80-100	>=20		Pass	0.8	80	2	1	1	50%
21	>90-95	>80-100	15-19.99		Pass	0.8	80				
22	>90-95	>80-100	10-14.99		No	1	100	1	1	0	0%
23	>90-95	>80-100	<10		No	0.84	84	2	2	0	0%
24	>90-95	>100	>=20		Pass	0.905	90.5	3	3	0	0%
25	>90-95	>100	15-19.99		No	0.531	53.1				
26	>90-95	>100	10-14.99		No	1	100				
27	>90-95	>100	<10		No	0.833	83.3	3	3	0	0%
28	>90-95	CleanLoan			Pass	0.938	93.8	7	7	0	0%
29	>95-100	Deposit			Pass	0.984	98.4	11	11	0	0%
30	>95-100	<=75			Pass	0.926	92.6	77	71	6	8%
31	>95-100	>75-80	>=20	Good	Pass	0.84	84				
32	>95-100	>75-80	>=20	Bad	No	1	100				
33	>95-100	>75-80	15-19.99		Pass	1	100				
34	>95-100	>75-80	10-14.99		Pass	0.778	77.8	1	0	1	100%
35	>95-100	>75-80	<10		No	0.904	90.4	7	7	0	0%
36	>95-100	>80-100	>=20		Pass	0.789	78.9				
37	>95-100	>80-100	15-19.99		No	1	100				
38	>95-100	>80-100	10-14.99		No	0.833	83.3	3	3	0	0%
39	>95-100	>80-100	<10		No	0.875	87.5	10	8	2	20%

ตารางที่ 4.5 (ต่อ)

Seq.	Utilization	LTV	Movement	Payment	Predict. Result	Prob.	% Prob.	No. Test Case	TRUE	FALSE	% False/Total Test Case
40	>95-100	>100	>=20		Pass	0.867	86.7	1	0	1	100%
41	>95-100	>100	15-19.99		No	0.857	85.7				
42	>95-100	>100	10-14.99		No	0.8	80				
43	>95-100	>100	<10		No	0.663	66.3	9	7	2	22%
44	>95-100	CleanLoan			Pass	0.91	91	38	37	1	3%
45	>100	Deposit			Pass	0.955	95.5	2	2	0	0%
46	>100	<=75	>=20	Good	Pass	0.75	75	1	0	1	100%
47	>100	<=75	>=20	Bad	No	1	100	1	1	0	0%
48	>100	<=75	15-19.99		Pass	0.5	50				
49	>100	<=75	10-14.99	Good	Pass	0.667	66.7	1	1	0	0%
50	>100	<=75	10-14.99	Bad	No	1	100				
51	>100	<=75	<10		No	0.986	98.6	23	22	1	4%
52	>100	>75-80			No	0.917	91.7	6	5	1	17%
53	>100	>80-100			No	0.9	90	5	3	2	40%
54	>100	>100			No	0.682	68.2	2	1	1	50%
55	>100	CleanLoan			Pass	0.971	97.1	4	4	0	0%
<b>Total</b>								<b>872</b>	<b>844</b>	<b>28</b>	<b>3.21%</b>

จากตารางที่ 4.5 ผลการทดสอบโปรแกรมพยากรณ์กับข้อมูลจริง จำนวน 872 เร็คคอร์ด สามารถสรุปผลได้ดังนี้

1. ผลลัพธ์จากโปรแกรมพยากรณ์ กับผลการพิจารณาทบทวนวงเงินจากข้อมูลจริงมีความถูกต้องตรงกัน (Correctly Classified) จำนวน 844 เร็คคอร์ด คิดเป็นร้อยละ 96.79 หรือเรียกว่าค่า True Positive Rate หรือค่า Recall มีค่าเท่ากับ 0.9679

2. ผลลัพธ์จากโปรแกรมพยากรณ์ไม่ตรง กับผลการพิจารณาทบทวนวงเงินจากข้อมูลจริง (Incorrectly Classified) จำนวน 28 เร็คคอร์ด มีค่าเท่ากับ 0.0321 หรือคิดเป็นร้อยละ 3.21

3. รูปแบบโมเดลการตัดสินใจที่ไม่ได้ทดสอบ เนื่องจากข้อมูลในชุดสอบไม่มีรูปแบบการตัดสินใจที่ตรงกับรูปแบบการตัดสินใจของโมเดล จำนวน 13 รูปแบบ ได้แก่ รูปแบบลำดับที่ 10, 11, 21, 25, 26, 31, 32, 33, 36, 37, 41, 42, 48 และ 50

4. รูปแบบโมเดลการตัดสินใจที่ผลการพยากรณ์ผิดพลาด จากผลการพิจารณาทบทวนวงเงินจากข้อมูลจริง ตั้งแต่ร้อยละ 20 ขึ้นไป มีจำนวน 11 รูปแบบ ได้แก่ รูปแบบลำดับที่ 7, 12, 13, 20, 34, 39, 40, 43, 46, 53 และ 54