

## บทที่ 2

### แนวคิด ทฤษฎี และผลงานวิจัยที่เกี่ยวข้อง

#### 2.1 ประวัติธนาคารกรุงศรีอยุธยา จำกัด (มหาชน)

ธนาคารกรุงศรีอยุธยา จำกัด (มหาชน) (2556) ก่อตั้งขึ้นเมื่อวันที่ 27 มกราคม 2488 และได้นำหลักทรัพย์เข้าจดทะเบียนในตลาดหลักทรัพย์แห่งประเทศไทยในปี 2520 ปัจจุบันเป็นธนาคารพาณิชย์ที่มีเงินให้สินเชื่อและเงินฝากที่ใหญ่เป็นอันดับ 5 ของประเทศ และเป็นธนาคารที่ให้บริการทางการเงินอย่างครบวงจรแก่ลูกค้าธุรกิจและลูกค้าบุคคล

การเปลี่ยนแปลงครั้งสำคัญของธนาคารกรุงศรีอยุธยา จำกัด (มหาชน) มีดังนี้

1. 3 มกราคม 2550 จีอี มั่นนี้ ซึ่งเป็นสถาบันการเงินเพื่อรายย่อยชั้นนำของโลกได้บรรลุข้อตกลงการเป็นพันธมิตรทางธุรกิจกับธนาคารกรุงศรีอยุธยา

2. 14 กุมภาพันธ์ 2551 ธนาคารฯ เข้าซื้อกิจการบริษัท จีอี แคปิตอล ออโต้ ลีส จำกัด (มหาชน) (GECAL) ทำให้พอร์ตสินเชื่อของธนาคารฯ เติบโตขึ้นประมาณร้อยละ 17 และต่อมา GECAL เปลี่ยนชื่อเป็น บริษัท อยุธยา แคปิตอล ออโต้ ลีส จำกัด (มหาชน) (AYCAL)

3. 8 เมษายน 2552 ธนาคารฯ เข้าซื้อหุ้นธนาคาร เอไอจี เพื่อรายย่อย จำกัด (มหาชน) (AIGRB) และ บริษัท เอไอจี คาร์ด (ประเทศไทย) จำกัด (AIGCC) ทำให้ธนาคารฯ มีสินทรัพย์เพิ่มขึ้นประมาณ 32,800 ล้านบาท มีฐานสินเชื่อเพิ่มขึ้น 21,900 ล้านบาท เงินฝากเพิ่มขึ้น 18,600 ล้านบาท และมีจำนวนบัตรเครดิตเพิ่มขึ้นประมาณ 222,000 บัตร

4. 9 กันยายน 2552 ธนาคารฯ เข้าซื้อกิจการบริษัท ซีเอฟจี เซอร์วิสเซส จำกัด (CFGS) ซึ่งเป็นบริษัทในเครือของบริษัทอเมริกัน อินเตอร์เนชั่นแนล กรุ๊ป ینگค์ จำกัด ทั้งนี้ CFGS เป็นหนึ่งในผู้ให้บริการไมโครไฟแนนซ์โดยเฉพาะสินเชื่อที่ใช้ทะเบียนรถเป็นหลักประกัน มีสาขากว่า 163 แห่งทั่วประเทศ และเป็นที่ยู่อักในนาม "ศรีสวัสดิ์ เงินดีลื้อ"

5. 5 พฤศจิกายน 2552 ธนาคารฯ เข้าซื้อธุรกิจการเงินเพื่อผู้บริโภคของจีอี แคปิตอล ในประเทศไทย ทำให้ธุรกิจการเงินเพื่อลูกค้ารายย่อยของธนาคารฯ เติบโตเร็วขึ้น และทำให้สัดส่วนสินเชื่อรายย่อยของธนาคารเพิ่มขึ้นจาก 36% เป็น 42% ของสินเชื่อรวม การผนวกธุรกิจของจีอี มั่นนี้ ประเทศไทย เข้ากับธนาคารทำให้ธนาคารฯ กลายเป็นผู้ให้บริการบัตรเครดิตรายใหญ่ที่สุดใน

ประเทศไทย ด้วยจำนวนบัตรหมุนเวียนในระบบมากกว่า 3 ล้านใบ และให้บริการลูกค้ากว่า 8 ล้านราย

6. 25 มกราคม 2555 คณะกรรมการธนาคารมีมติอนุมัติเห็นชอบให้ธนาคารฯ ดำเนินการรับ โอนธุรกิจลูกค้ารายย่อยในประเทศไทยของธนาคารฮ่องกงและเซี่ยงไฮ้เบงกิงคอร์ปอเรชั่น จำกัด (HSBC) ซึ่งประกอบธุรกิจบัตรเครดิต ธุรกิจสินเชื่อส่วนบุคคล ธุรกิจสินเชื่อสหกรณ์ทรัพย์ ธุรกิจเงินฝากส่วนบุคคลและตัวแลกเงิน

7. 18 ธันวาคม 2556 มิตซูบิชิ ยูเอฟเจ ไฟแนนเชียล กรุ๊ป (MUFG) เข้ามาเป็นพันธมิตรทางธุรกิจกับธนาคารฯ แทนจีอี มั่นนี้

## 2.2 ความหมายสินเชื่อในทางเศรษฐศาสตร์

สำนักงานเศรษฐกิจการคลัง (2549) สินเชื่อ หมายถึง อำนาจในการเข้าถึงการใช้สินค้าและบริการ โดยสัญญาว่าจะชดใช้คืนในอนาคต โดยสินเชื่ออาจอยู่ในรูปของสินค้าและบริการ หรืออยู่ในรูปของเงินก็ได้

ในทางเศรษฐศาสตร์มองว่า หน่วยเศรษฐกิจในระบบเศรษฐกิจนั้นอาจมีลักษณะการใช้ทรัพยากรที่ไม่สมดุลกัน กล่าวคือ บางหน่วยเศรษฐกิจอาจมีความต้องการใช้สินค้าและบริการที่มากกว่าทรัพยากรที่ตนมีอยู่ในปัจจุบัน ในขณะที่บางหน่วยเศรษฐกิจอาจมีทรัพยากรเหลือใช้เกินความต้องการ ดังนั้น หากมีการโอนทรัพยากรส่วนที่เหลือใช้นั้นไปให้ผู้ที่มีความต้องการใช้ทรัพยากรและก่อให้เกิดประโยชน์ได้ในอนาคต โดยมีข้อตกลงว่าจะมีการชำระคืนแก่เจ้าของเดิมในอนาคต การจัดสรรการใช้ทรัพยากรต่างๆ ในระบบเศรษฐกิจก็จะมีประสิทธิภาพมากขึ้น

## 2.3 ความสัมพันธ์ระหว่างสินเชื่อกับระบบเศรษฐกิจ

สำนักงานเศรษฐกิจการคลัง (2549) ความสัมพันธ์ระหว่างสินเชื่อกับระบบเศรษฐกิจ โดยมีภาคการเงินเป็นส่วนส่งเสริมให้เกิดการใช้ทรัพยากรในภาคเศรษฐกิจได้อย่างมีประสิทธิภาพ เริ่มจากการที่หน่วยเศรษฐกิจที่มีรายได้มากกว่าความต้องการใช้จ่าย ทำให้เกิดส่วนรั่วไหลจากระบบเศรษฐกิจกลายเป็นการออมเงินในรูปแบบต่างๆ เช่น การออมเงินผ่านระบบธนาคารพาณิชย์ เป็นต้น จากนั้นหน่วยเศรษฐกิจที่มีความต้องการใช้จ่ายมากกว่าทรัพยากรที่ตนมีอยู่ ก็จะไปขอกู้เงินหรือขอสินเชื่อจากหน่วยเศรษฐกิจที่รับออมเงิน โดยมีอัตราดอกเบี้ยเป็นราคาของทุนและใช้เป็นเครื่องมือช่วยให้การจัดสรรทรัพยากรจากผู้ออมไปสู่ผู้กู้ ให้เป็นไปอย่างมีประสิทธิภาพ โดยผู้กู้ไม่ว่าจะเป็นกรณีการใช้นโยบายขาดดุลงบประมาณของภาครัฐ หรือภาคเอกชน เม็ดเงินจากการกู้ยืมเหล่านี้จะถูกนำไปใช้จ่าย ซึ่งเป็นการอัดฉีดเงินที่รั่วไหลกลับเข้าสู่ภาคเศรษฐกิจอีกครั้ง

## 2.4 ความรู้ทั่วไปด้านสินเชื่อ (ภิเชก ชัยนิรันดร์, 2553, น. 9 - 22)

1. ความหมายของสินเชื่อ Joseph French Johnson นักเศรษฐศาสตร์ชาวอเมริกัน ให้นิยามของสินเชื่อไว้ว่า “Credit is the power to obtain goods or service by given a promise to pay at a specified date in the future”

สินเชื่อ คือ ความสามารถที่จะได้มาซึ่งสินค้าหรือบริการ โดยการให้คำมั่นสัญญาว่าจะชำระคืนตามวันที่กำหนดไว้ในอนาคต

จากนิยามข้างต้นแสดงถึงองค์ประกอบของสินเชื่อ 3 ส่วน ได้แก่

1.1 ความสามารถ หมายถึง ความสามารถของผู้กู้ที่จะสร้างความน่าเชื่อถือให้ประจักษ์ต่อผู้ให้กู้เกิดความมั่นใจว่า หากให้สินเชื่อแก่ผู้กู้แล้ว ผู้กู้มีศักยภาพเพียงพอที่จะบริหารจัดการองค์การได้อย่างมีประสิทธิภาพ นำพาธุรกิจขององค์การเจริญก้าวหน้าได้อย่างยั่งยืน และมีรายได้เพียงพอกับการจ่ายชำระหนี้ในอนาคต

1.2 การได้มาซึ่งสินค้าและบริการ หมายถึง ผู้กู้ต้องนำเงินจากการกู้ยืมนั้นไปใช้ให้ตรงตามวัตถุประสงค์ของการอนุมัติสินเชื่อ

1.3 คำมั่นสัญญาว่าจะชำระคืนตามวันที่กำหนด หมายถึง พันธสัญญาร่วมกันระหว่างผู้กู้มีต่อผู้ให้กู้ที่จะหาเงินมาจ่ายชำระหนี้ ภายในวันหรือระยะเวลาที่กำหนดไว้

## 2. ความหมายของสินเชื่อ ในกรณีกู้ยืมเงินตามประมวลกฎหมายแพ่งและพาณิชย์

การกู้ยืมเงิน ตามประมวลกฎหมายแพ่งและพาณิชย์ถือเป็นสัญญายืมใช้สິณเปลือง ที่มีวัตถุประสงค์สัญญาเป็นเงินตรา ตามมาตรา 650 บัญญัติไว้ ดังนี้

“อันว่ายืมใช้สິณเปลือง คือ สัญญาซึ่งผู้ให้ยืมโอนกรรมสิทธิ์ในทรัพย์สินชนิดใช้ไปสິณไปนั้น เป็นปริมาณที่กำหนดให้แก่ผู้ยืมและผู้ยืมตกลงว่าจะคืนทรัพย์สินประเภท ชนิด และปริมาณเช่นเดียวกันให้แก่แทนทรัพย์สินซึ่งให้ยืมนั้น สัญญาย่อมบริบูรณ์ต่อเมื่อส่งมอบทรัพย์สินที่ยืม”

## 3. หลักการพิจารณาในการเลือกใช้สินเชื่อระยะสั้นและสินเชื่อระยะยาว

โดยทั่วไปแหล่งที่มาของเงินทุนที่ใช้เพื่อการดำเนินธุรกิจมาจาก 2 แหล่ง คือ จากส่วนของเจ้าของ หรือเรียกว่า “ทุน” และจากการกู้ยืม สำหรับหลักการในการพิจารณาว่าการที่กิจการจะลงทุนในสินทรัพย์ถาวรและสินทรัพย์หมุนเวียนนั้น ควรจะเลือกใช้เงินทุนจากแหล่งใด และถ้าหากเลือกใช้วิธีการกู้ยืมจะใช้สินเชื่อประเภทใดและมีระยะเวลาแบบใด มีหลักการพิจารณาอยู่ 3 รูปแบบ ดังนี้

3.1 พิจารณาตามหลักการจับคู่ตามระยะเวลา (Maturity Matching Approach) หลักการ คือ กรณีที่กิจการต้องการลงทุนในสินทรัพย์ถาวร เช่น เพื่อซื้อที่ดิน อาคาร และอุปกรณ์ ควรใช้สินเชื่อระยะยาว เนื่องจากสินทรัพย์ถาวรเหล่านี้ ต้องใช้ระยะเวลานานพอสมควร ในการที่จะสร้างรายได้

และสามารถนำเงินมาชำระหนี้ได้ ในกรณีกิจการที่ต้องการลงทุนในสินทรัพย์หมุนเวียน เช่น เพื่อซื้อสินค้าเข้าร้าน ควรใช้สินเชื่อระยะสั้น เช่น สินเชื่อเงินกู้เบิกเกินบัญชี (Bank Overdraft) หรือเรียกว่าวงเงินโอดี ทั้งนี้ เพื่อเป็นการเสริมสภาพคล่องของกิจการให้มีเงินทุนเพียงพอต่อการซื้อสินค้าหรือบริการ ในสถานการณ์ที่กิจการค้าขายดี กิจการก็มีจำเป็นต้องสำรองสินค้าไว้เพื่อขายในปริมาณมากขึ้นจากเดิม และเมื่อมีรายได้ก็นำไปชำระหนี้เงินต้นพร้อมดอกเบี้ย จากนั้นวงเงินของบัญชีเงินกู้เบิกเกินบัญชีที่เคยอนุมัติไว้ ก็จะกลับมาคงดังเดิมและสามารถใช้เป็นเงินทุนหมุนเวียนในกิจการต่อไปได้อีกเรื่อยๆ จนกว่าจะสิ้นสุดระยะเวลาของสัญญากู้ยืม

3.2 พิจารณาตามหลักการที่ค่อนข้างเสี่ยง (Aggressive Approach) หลักการคือ ในสถานการณ์ที่อัตราดอกเบี้ยของสินเชื่อระยะสั้นต่ำกว่าอัตราดอกเบี้ยสินเชื่อระยะยาว กิจการอาจพิจารณาเลือกใช้สินเชื่อระยะสั้น เพื่อนำมาลงทุนในสินทรัพย์ถาวรของกิจการ หลักการนี้ข้อเสียคือสินทรัพย์ถาวรมันอาจสร้างรายได้ไม่ทันกับรอบการจ่ายชำระหนี้ อาจทำให้โอกาสที่จะไม่สามารถนำเงินมาจ่ายชำระหนี้ภายในระยะเวลาที่กำหนดมีค่อนข้างสูง

3.3 พิจารณาตามหลักความปลอดภัยไว้ก่อน (Conservative Approach) หลักการคือ กิจการอาจใช้จุดแข็งสินเชื่อระยะยาวที่มีกำหนดระยะเวลาการชำระคืนค่อนข้างนาน แต่อาจมีความเสี่ยงเรื่องความผันผวนของอัตราดอกเบี้ยลอยตัวในอนาคต ดังนั้น บางกิจการอาจนำเงินทุนที่เหลือใช้จากสินเชื่อระยะยาวไปลงทุนในสินทรัพย์หมุนเวียนที่มีสภาพคล่องสูงและสามารถขายคืนหรือแปลงเป็นเงินสดได้เร็ว

#### 4. ประเภทของสินเชื่อ แบ่งตามลักษณะต่างๆ ได้ดังนี้

4.1 ประเภทสินเชื่อแบ่งตามลักษณะของการชำระคืน ได้แก่ สินเชื่อชำระครั้งเดียว (Single Payment Credit) และสินเชื่อผ่อนส่ง (Installment Credit)

4.2 ประเภทสินเชื่อแบ่งตามระยะเวลา ได้แก่ สินเชื่อระยะสั้น (Short Term Credit) เป็นสินเชื่อที่มีกำหนดระยะเวลาชำระคืนภายใน 1 ปี สินเชื่อระยะปานกลาง (Intermediate Term Credit) เป็นสินเชื่อที่มีกำหนดระยะเวลาชำระคืนตั้งแต่ 1 - 5 ปี และสินเชื่อระยะยาว (Long Term Credit) เป็นสินเชื่อที่มีกำหนดระยะเวลาชำระคืนตั้งแต่ 5 ปี ขึ้นไป ในทางปฏิบัติโดยทั่วไปนิยมเรียกสินเชื่อที่มีกำหนดระยะเวลาชำระคืนตั้งแต่ 1 ปีขึ้นไปว่า “สินเชื่อระยะยาว”

4.3 ประเภทสินเชื่อแบ่งตามหลักประกัน ได้แก่ สินเชื่อที่มีหลักประกัน (Secured Credit) สินเชื่อที่ไม่มีหลักประกัน (Unsecured Credit)

4.4 ประเภทสินเชื่อแบ่งตามลักษณะของผู้ให้สินเชื่อ ได้แก่ สินเชื่อรัฐบาลหรือสินเชื่อสาธารณะ (Public Credit) เช่น ตัวเงินคลัง พันธบัตรรัฐบาล การกู้ยืมทางตรงของรัฐบาลทั้งในและต่างประเทศ และสินเชื่อในภาคเอกชน (Private Credit) ได้แก่ สินเชื่อทุกประเภทของสถาบัน

การเงิน (Banking Credit) และสินเชื่อการค้า (Trade Credit) หรือการให้เครดิตวันชำระเงินค่าสินค้าระหว่างกิจการคู่ค้าด้วยกันเอง รวมถึงการออกตราสารหนี้เพื่อระดมทุนของกิจการ เป็นต้น

## 2.5 แนวทางการแสวงหาความรู้

ณรงค์ โภธิพิทักษ์นันท์ (2556, น. 16 - 29) ในอดีตมนุษย์ค้นหาความรู้ต่างๆ จากสภาพแวดล้อมทางธรรมชาติรอบตัวและวิธีการแสวงหาความรู้ดังกล่าวยังไม่เป็นระบบระเบียบเท่าที่ควร แนวทางการแสวงหาความรู้ มีดังนี้

1. การลองผิดลองถูก เป็นความรู้ที่ได้จากการปฏิบัติแล้วจดจำไปใช้ในคราวต่อไป
2. ความบังเอิญ เป็นความรู้ที่ได้มาโดยไม่คาดคิดมาก่อน
3. ความเชื่อหรือขนบธรรมเนียมประเพณีที่สืบทอดมาแต่โบราณ
4. ความรู้จากผู้รู้ หรือนักปราชญ์ หรือผู้เชี่ยวชาญ
5. ความรู้จากประสบการณ์จริง หรือประสาทสัมผัส
6. ความรู้จากการหยั่งรู้ หรือการรู้แจ้ง
7. ความรู้จากการหาเหตุผล มี 2 วิธี คือ

7.1 การหาเหตุผลแบบนิรนัย (Deductive Method) หรือวิธีอนุมาน เป็นการค้นหาความรู้จากการเก็บรวบรวมข้อเท็จจริงใหญ่ไปหาข้อเท็จจริงย่อย แล้วสรุปผล การหาความรู้รูปแบบนี้ อริสโตเติล (Aristotle) เป็นผู้ใช้เป็นคนแรก เรียกว่า “Syllogistic Method” หรือ “Aristotle Deduction”

7.2 การหาเหตุผลแบบอุปนัย (Inductive Method) หรือวิธีอุปมาน เป็นการค้นหาความรู้จากการเก็บรวบรวมข้อเท็จจริงย่อยไปหาข้อเท็จจริงใหญ่ แล้วสรุปผล เป็นทฤษฎีการหาเหตุผลในแบบของ ฟรานซิส เบคอน (Francis Bacon) ซึ่งมีความคิดเห็นขัดแย้งและต้องการแก้ไขจุดอ่อนของทฤษฎีการค้นหาคำความรู้แบบนิรนัยของอริสโตเติล (Aristotle) รูปแบบการหาความรู้แบบอุปนัย แบ่งออกเป็น 2 แบบ คือ

7.2.1 อุปนัยแบบสมบูรณ์ (Perfect Induction) เป็นการหาความรู้โดยการเก็บรวบรวมข้อมูลทุกหน่วยของประชากร (Population) ที่จะศึกษา

7.2.2 อุปนัยแบบไม่สมบูรณ์ (Imperfect Induction) เป็นการหาความรู้โดยการเก็บรวบรวมข้อมูลจากบางหน่วยของประชากร (Population) ที่จะศึกษา แล้วนำผลที่ได้ไปอ้างอิงว่าเป็นการศึกษาของประชากรทั้งหมด

8. วิธีทางวิทยาศาสตร์ จากแนวคิดของ ชาลส์ ดาร์วิน (Charles Darwin) ที่ได้นำรูปแบบการค้นหาคำรู้ทั้งแบบนิรนัยและอุปนัยมาใช้ในกระบวนการหาความรู้ทางวิทยาศาสตร์ธรรมชาติ

โดยวิธีการค้นหาความรู้แบบนี้ เรียกว่า วิธีการอนุมาน-อุปมาน (Deductive-Inductive Method) คือ การใช้วิธีการอนุมานคำตอบหรือตั้งสมมติฐานก่อน แล้วใช้วิธีอุปมานเก็บรวบรวมข้อมูลเพื่อทดสอบสมมติฐาน ต่อมา จอห์น ดิวอี้ (John Dewey) ได้ปรับปรุงแนวคิดดังกล่าวที่เรียกว่า “Reflective Thinking” โดยแบ่งขั้นตอนการแก้ปัญหาออกเป็น 5 ขั้นตอน และในเวลาต่อมาเรียกว่า “วิธีการทางวิทยาศาสตร์” ประกอบด้วย

- 8.1 กำหนดปัญหา (Problem)
- 8.2 ตั้งสมมติฐาน (Hypothesis)
- 8.3 เก็บรวบรวมข้อมูล (Collection of Data)
- 8.4 วิเคราะห์ข้อมูล (Analysis of Data)
- 8.5 สรุปผล (Conclusion)

## 2.6 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูลเป็นเทคนิคที่ใช้สำหรับการวิเคราะห์ข้อมูล เพื่อค้นหาความรู้ที่มีประโยชน์จากฐานข้อมูลขนาดใหญ่ (Knowledge Discovery in Databases - KDD) โดยอาศัยหลักสถิติ หลักคณิตศาสตร์ และการเรียนรู้ของเครื่องจักร โดยเครื่องจักรในที่นี้หมายถึง โปรแกรมวิเคราะห์ข้อมูลที่ติดตั้งไว้ในเครื่องคอมพิวเตอร์ มากระทำกับข้อมูลจำนวนมากโดยอัตโนมัติ ความรู้ที่ได้จะแสดงออกมาในลักษณะที่เป็นรูปแบบ (Pattern) กฎ (Rule) และความสัมพันธ์ที่ซ่อนอยู่ในข้อมูลเหล่านั้น

วิธีการที่จะทำให้โปรแกรมวิเคราะห์ข้อมูลที่ติดตั้งในคอมพิวเตอร์เรียนรู้ข้อมูล (เอกสิทธิ์ พัชรวงศ์ศักดิ์, 2556, น. 42) มี 2 แนวทาง คือ

1. การวิเคราะห์ข้อมูลโดยการเรียนรู้ข้อมูลในอดีตที่มีชุดคำตอบเตรียมไว้ล่วงหน้า (Supervised Learning) เพื่อใช้เป็นแบบแผนในการวิเคราะห์ข้อมูล หรือเรียกว่าเป็นการสอนคอมพิวเตอร์จากข้อมูลที่ถูกจัดเตรียมแนวคำตอบไว้เรียบร้อยแล้ว
2. การวิเคราะห์ข้อมูลโดยไม่มีการเตรียมชุดคำตอบไว้ล่วงหน้า (Unsupervised Learning) เป็นการให้โปรแกรมคอมพิวเตอร์ทำการวิเคราะห์ข้อมูลโดยการหาความเหมือนหรือความคล้ายคลึงกัน หรือที่เรียกว่า รูปแบบความสัมพันธ์กันของข้อมูลโดยอัตโนมัติ

### 2.6.1 รูปแบบวิธีการวิเคราะห์ข้อมูลของการทำเหมืองข้อมูล มีดังนี้

1. วิธีจำแนกข้อมูล (Classification) เพื่อใช้ในการพยากรณ์ (Prediction) หรือเพื่อการทำนายผล หรือเพื่อการคาดคะเนคำตอบ วิธีนี้จะพยากรณ์ข้อมูลตามโมเดลต้นแบบที่ได้จากการเรียนรู้ข้อมูลในอดีต

2. วิธีหาความสัมพันธ์ของข้อมูล (Association) วิธีนี้ใช้สำหรับค้นหาความสัมพันธ์ของข้อมูลรายการที่เกิดขึ้นซ้ำกันบ่อยๆ

3. วิธีจัดกลุ่มข้อมูล (Clustering) วิธีนี้ใช้เพื่อจัดกลุ่มข้อมูลที่มีลักษณะคล้ายกันไว้ด้วยกัน

## 2.6.2 มาตรฐานในการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล

ในปี 2539 บริษัท Daimler Chrysler บริษัท SPSS และบริษัท NCR (เอกสิทธิ์ พัทธวงค์ ศักดา, 2556, น. 10-12) ได้ร่วมกันกำหนดมาตรฐานของการวิเคราะห์ข้อมูลด้วยการทำเหมืองข้อมูล เรียกว่า “Cross-Industry Standard Process for Data Mining: CRISP-DM” ประกอบด้วย 6 ขั้นตอน ดังนี้

1. ขั้นตอนการทำความเข้าใจปัญหาทางธุรกิจ และวางแผนแนวทางการดำเนินการเบื้องต้น (Business Understanding)

2. ขั้นตอนการเก็บรวบรวม การตรวจสอบ และวิเคราะห์ข้อมูลที่เป็นต้องใช้ในการวิเคราะห์ข้อมูล (Data Understanding)

3. ขั้นตอนการเตรียมข้อมูล (Data Preparation) คือ การปรับปรุงคุณภาพข้อมูลให้อยู่ในสภาพที่สามารถนำไปใช้ในการทำเหมืองข้อมูลได้ วิธีการที่ใช้สำหรับการเตรียมข้อมูล (ฉวีวรรณ เพ็ชรศิริ, 2556; Jiawei Han, Jian Pei & Micheline Kamber, 2011) มี 4 วิธี ดังนี้

3.1 การทำความสะอาดข้อมูล (Data Cleansing) คือ การจัดการข้อมูลในกรณีข้อมูลจริงไม่สมบูรณ์หรือมีค่าที่ขาดหายไป (Missing Value)

3.2 การผสานข้อมูล (Data Integration) คือ การผสานข้อมูลที่รวบรวมมาจากแหล่งข้อมูลที่ต่างกัน เพื่อลดความซ้ำซ้อนของข้อมูล (Data Redundancies) และช่วยลดปัญหาความไม่สอดคล้องกันของข้อมูล (Data Inconsistencies)

3.3 การแปลงข้อมูล (Data Transformation) คือ การทำนอร์มอลไลซ์ (Normalization) ด้วยการแปลงค่าข้อมูลให้อยู่ในช่วงสั้นๆ เช่น (0,1)

3.4 การลดรูปข้อมูล (Data Reduction) คือ การจัดการข้อมูลให้อยู่ในรูปแบบกะทัดรัด ประหยัดพื้นที่ในการเก็บข้อมูล

4. ขั้นตอนการวิเคราะห์และออกแบบ (Modeling) เพื่อให้ได้รูปแบบหรือโมเดล

5. ขั้นตอนการวัดประสิทธิภาพของรูปแบบหรือโมเดล (Evaluation) ว่ามีความถูกต้องตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่

6. ขั้นตอนการนำผลลัพธ์ที่ได้จากการวิเคราะห์ข้อมูลไปใช้ให้เกิดประโยชน์ต่อองค์กร (Deployment)

## 2.7 ตัวจำแนกข้อมูลแบบ C4.5 หรือ J48

ตัวจำแนกข้อมูลแบบ C4.5 หรือในโปรแกรม WEKA เรียกว่าเทคนิค “J48” (เสถียร วิชาเรื่อง, 2553, น. 52) เป็นอัลกอริทึมสำหรับการจำแนกข้อมูลในกลุ่มเทคนิคแผนภาพต้นไม้ตัดสินใจ (Decision Tree) เป็นแบบจำลองหรือโมเดลการตัดสินใจทางคณิตศาสตร์สถิติ ใช้สำหรับหาทางเลือกที่ดีที่สุด เหมาะสำหรับการใช้เพื่อแก้ปัญหาที่ไม่ซับซ้อน และมีทางเลือกเพื่อการตัดสินใจน้อย จุดเด่นของแผนภาพต้นไม้ตัดสินใจ คือ ช่วยให้ผู้ใช้งานมองเห็นภาพรวมของรูปแบบการตัดสินใจชัดเจนขึ้น

$$\text{อัตราส่วนความน่าจะเป็น} = \frac{\text{ความน่าจะเป็นที่จะเกิดเหตุการณ์}}{\text{ความน่าจะเป็นที่จะไม่เกิดเหตุการณ์}}$$

แผนภาพต้นไม้ตัดสินใจเป็นแบบจำลองที่ได้จากการเรียนรู้รูปแบบขององค์ประกอบที่ใช้ในการตัดสินใจโดยโปรแกรมคอมพิวเตอร์ เพื่อพยากรณ์หรือคาดคะเนคำตอบที่ควรจะเป็นตามหลักการความน่าจะเป็น จากข้อมูลชุดที่ใช้เพื่อการเรียนรู้ การเรียนรู้แบบนี้เรียกว่า “การเรียนรู้แบบมีครู”(Supervised Learning) (สิทธิโชค มุกดาสกุลภินาด, 2551; Jiawei Han, Jian Pei & Micheline Kamber, 2011; N. Vandana Sawant, Ketan Shah & Vinayak Ashok Bharadi, 2011) องค์ประกอบของแผนภาพต้นไม้ตัดสินใจ มี 4 ส่วน ดังนี้

1. Root node คือ node แรกตัวบนสุด
2. Branch คือ กิ่งก้านสาขาของ node มีทั้งกิ่งด้านซ้าย (Left Branch) และกิ่งด้านขวา (Right Branch)
3. Child คือ ลูกๆ ของ Branch
4. Leaf node คือ node ที่อยู่ลำดับสุดท้าย

Berry, Michael J.A. & Linoff, Gordon S. (2004, pp. 166 - 167) ได้เปรียบเทียบลักษณะของการจำแนกข้อมูลไว้ว่า การจำแนกข้อมูลมีลักษณะคล้ายกับการเล่นเกมสควอช 20 คำถาม โดยที่ผู้เล่นคนหนึ่งๆ พยายามค้นหาคำตอบโดยจะถามคำถามเชิงกว้างแล้วค่อยๆ แคบลงไปเรื่อยๆ จนเกิดความมั่นใจและสามารถระบุคำตอบได้ว่าสิ่งนั้นคืออะไร

Ohlhorst, Frank. (2013, p.4) ให้นิยามของการทำเหมืองข้อมูล (Data Mining) ไว้ว่า “Data Mining is process in which data are analyzed from different perspectives and then turned into summary data that are deemed useful. Data Mining is normally used with data at rest or with archival data. Data Mining techniques focus on modeling and knowledge discovery for predictive,

rather than purely descriptive, purposed—an ideal process for uncovering new patterns from data sets.”

กระบวนการสร้างแบบจำลอง (Model) หรือการสร้างโมเดลตามทฤษฎีของการทำเหมืองข้อมูล ต้องมีการแบ่งข้อมูลออกเป็น 2 ชุด ข้อมูลชุดแรก เรียกว่า Training Data เป็นข้อมูลที่ใช้สำหรับสอน โปรแกรมคอมพิวเตอร์ให้ทำการเรียนรู้รูปแบบปัญหาและหาข้อสรุป แล้วนำไปสร้างเป็นแบบจำลองหรือที่เรียกว่าโมเดล และข้อมูลชุดที่สอง เรียกว่า Testing Data เป็นข้อมูลที่ใช้สำหรับทดสอบประสิทธิภาพของโมเดลที่โปรแกรมวิเคราะห์ข้อมูลนั้นค้นพบจากข้อมูลชุด Training Data ว่ามีความถูกต้องน่าเชื่อถือมากน้อยเพียงใด ทั้งนี้ ขนาดของข้อมูลและรูปแบบการเลือกใช้ข้อมูลเพื่อการ Training และ Testing นั้นขึ้นอยู่กับความต้องการและความเชี่ยวชาญของผู้วิเคราะห์ข้อมูล

งานวิจัยนี้ได้ทำการประมวลผลและวิเคราะห์ข้อมูลโดยใช้โปรแกรม WEKA เวอร์ชัน 3.6.2 ซึ่งมีความสามารถพิเศษที่ผู้ใช้สามารถเลือกได้ว่าจะทำการแบ่งข้อมูลเพื่อการเรียนรู้ (Training) และเพื่อการทดสอบ (Testing) โมเดลแบบใดได้ในคราวเดียวกัน

## 2.8 งานวิจัยที่เกี่ยวข้อง

ปิยะมาศ กรณียกคุณ (2556) ได้ทำการวิจัยเกี่ยวกับปัจจัยที่มีอิทธิพลต่อค่าความแม่นยำของการจำแนกข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล เปรียบเทียบระหว่างอัลกอริทึม Naïve Bayes Simple และ ID3 แนวทางการวิจัยได้เน้นให้เห็นถึงความสำคัญของการเตรียมข้อมูลก่อนการสร้างโมเดล โดยทำการแบ่งข้อมูลชุดเรียนรู้รูปแบบโมเดลออกเป็น 2 ชุด เรียกว่า ชุด A และชุด B โดยมีความแตกต่างของข้อมูลตรงที่การจัดกลุ่มค่าข้อมูลในแต่ละแอตทริบิวต์ ซึ่งค่าของข้อมูลชุด B มีความละเอียดและมีความเป็นเอกลักษณ์ที่ชัดเจนกว่าชุด A จากนั้นทำการวิเคราะห์เปรียบเทียบผลการทดลองจากทั้ง 2 อัลกอริทึม พบว่า การจัดกลุ่มค่าข้อมูลที่มีความละเอียดและมีความเป็นเอกลักษณ์ที่ชัดเจนกว่าจะทำให้การจำแนกข้อมูลมีความถูกต้องยิ่งขึ้น และพบว่าอัลกอริทึม Naïve Bayes Simple มีความสามารถในการจำแนกข้อมูลที่ดีกว่า ID3

อย่างไรก็ตาม ผู้วิจัยมีประเด็นที่พิจารณาแล้วเห็นว่า แม้ว่าอัลกอริทึม Naïve Bayes Simple จะมีประสิทธิภาพในการจำแนกหรือพยากรณ์ข้อมูลได้ดีกว่า ID3 จริง ในแง่ของการนำไปอธิบายต่อผู้ใช้งานคนอื่นๆ มีความยุ่งยากที่จะอธิบายให้บุคคลเหล่านั้นมองเห็นภาพรวมของรูปแบบการตัดสินใจที่ชัดเจนได้ ในขณะที่อัลกอริทึม ID3 เป็นอัลกอริทึมในกลุ่มแผนภาพต้นไม้ตัดสินใจ (Decision Tree) ซึ่งมีจุดเด่นด้านการอธิบายรูปแบบการตัดสินใจเป็นแบบแผนภาพที่เข้าใจง่าย แต่มีข้อเสีย คือ ในกรณีที่ข้อมูลชุดเรียนรู้รูปแบบการตัดสินใจแบบเดียวกัน และมี

คำตอบทั้งเห็นด้วยและปฏิเสธ อัลกอริทึม ID3 จะไม่พยากรณ์ความน่าจะเป็นของคำตอบของรูปแบบการตัดสินใจดังกล่าว

การไม่แสดงผลการพยากรณ์ของอัลกอริทึม ID3 นี้ ผู้วิจัยมีความเห็นว่าเป็นทั้งจุดเด่นและจุดด้อย ขึ้นอยู่กับการไปใช้กับงานแต่ละประเภท งานมีความละเอียดมากน้อยเพียงใด เช่น ถ้านำไปใช้กับงานวินิจฉัยโรค ก็จะถือว่าเป็นจุดเด่น เพราะบางครั้งต้องใช้ความรู้ความเชี่ยวชาญเฉพาะของแพทย์นั้นๆ มาประกอบการตัดสินใจ แต่สำหรับบางงานอาจไม่จำเป็นต้องใช้ความละเอียดในการพิจารณาที่ยู่ยากซับซ้อนมาก หรือเพียงแค่ต้องการคำตอบในภาพรวม หรือเพื่อวัตถุประสงค์ที่จะนำไปเป็นแนวทางในการทำงานที่เกี่ยวข้องต่อไปเท่านั้น ซึ่งกรณีนี้อาจกล่าวได้ว่าเป็นจุดด้อยของอัลกอริทึม ID3

จากเหตุผลข้างต้น ผู้วิจัยจึงได้นำข้อดีและข้อเสียในแต่ละประเด็นที่จากการวิจัยครั้งก่อนมาพิจารณาปรับปรุงให้ดีขึ้น พร้อมกับทดลองหารูปแบบการตัดสินใจที่เหมาะสมที่สุดจากอัลกอริทึมอื่นๆ ดังที่จะแสดงไว้ในงานวิจัยครั้งนี้

ณัฐพร อาทรธรรมคุณ (2552, น. 43 - 62) ได้ทำการประยุกต์ใช้วิธีการจำแนกข้อมูล (Classification) ของเทคนิคการทำเหมืองข้อมูลโดยใช้เทคนิค Classifiers.Trees.J48 ของโปรแกรม WEKA ทำการประมวลผลวิเคราะห์หารูปแบบโมเดลจากข้อมูลผลการเรียนของนักเรียนเตรียมทหาร โรงเรียนเตรียมทหาร สถาบันวิชาการป้องกันประเทศ กองบัญชาการกองทัพไทย ที่จบการศึกษาตั้งแต่ปี พ.ศ. 2547-2549 จำนวน 4 รายวิชา ได้แก่ คณิตศาสตร์ ภาษาอังกฤษ ฟิสิกส์ และเคมี เพื่อนำผลลัพธ์จากโมเดลมาเป็นแนวทางสำหรับพยากรณ์ผลการเรียนของนักเรียนในชั้นปีปัจจุบัน

จากการศึกษาวิจัยนี้ ทำให้ผู้วิจัยมีความเชื่อมั่นว่าเทคนิค Classifiers.Trees.J48 น่าจะสามารถนำมาประยุกต์ใช้กับงานทบทวนวงเงินได้เช่นเดียวกัน

เสถียร วิชาเรือง (2553, น. 38 - 73) ได้ทำการประยุกต์ใช้วิธีการจำแนกข้อมูล (Classification) ของเทคนิคการทำเหมืองข้อมูลในกระบวนการอนุมัตินิเชื่อ โดย ได้ทำการวิจัยเชิงเปรียบเทียบประสิทธิภาพของอัลกอริทึมที่มีในโปรแกรม WEKA จำนวน 3 อัลกอริทึม ได้แก่ 1) เทคนิคการจำแนกกลุ่มแบบ Decision Rules : Part หรือในโปรแกรม WEKA เรียกว่า Classifiers.rules.PART 2) เทคนิคการจำแนกกลุ่มแบบ Decision Tree : C4.5 หรือในโปรแกรม WEKA เรียกว่า เทคนิค Classifiers.Trees.J48 และ 3) เทคนิคการจำแนกกลุ่มแบบ Decision Table หรือในโปรแกรม WEKA เรียกว่าเทคนิค Classifiers.Decision Table โดยทำการประมวลผลข้อมูลเพื่อวิเคราะห์หารูปแบบโมเดลจากข้อมูลที่ใช้ประกอบการพิจารณาอนุมัตินิเชื่อของบริษัทเอกชนหลวงลิสซิ่ง จำกัด ตั้งแต่ปี พ.ศ. 2545 - 2550 จำนวน 17,964 สัญญา จากนั้นวิเคราะห์เปรียบเทียบ

ความถูกต้องของแต่ละโมเดลและเลือกรูปแบบที่เหมาะสมตรงกับความเป็นจริงมากที่สุด โดยผลวิจัยได้คัดเลือกโมเดลที่ได้จากเทคนิค Classifiers.rules.PART เนื่องจากมีความถูกต้องและมีเหตุผลสอดคล้องในทางธุรกิจมากที่สุด และนำรูปแบบเงื่อนไขดังกล่าวไปสร้างเป็นโปรแกรมจำแนกลูกค้าว่าเป็นลูกค้ากลุ่มดีหรือไม่ดี

จากการศึกษาวิจัยข้างต้น ผู้วิจัยมีความเห็นว่า รูปแบบการตัดสินใจที่ได้จากเทคนิค Classifiers.rules.PART เป็นลักษณะของการใช้คำเชื่อม ซึ่งเหมาะกับการนำไปเขียนเป็นโปรแกรม แต่ไม่เหมาะกับการที่จะนำมาใช้อธิบายแก่ผู้ใช้งาน เพราะอาจก่อให้เกิดความสับสนและเข้าใจผิดพลาดได้

สิทธิโชค มุกดาสกุลภิบาล (2551, น. 48 - 92) ได้ทำการวัดประสิทธิภาพของขั้นตอนวิธีตัวจำแนก C4.5, ADTree และ Naïve Bayes ในการจำแนกข้อมูลการชุกช่อนสิ่งเสพติดสำหรับไปรษณีย์ระหว่างประเทศ ได้ผลสรุปว่า ตัวจำแนกแต่ละตัวมีข้อดีข้อเสียแตกต่างกัน ในแง่ความถูกต้องตัวจำแนก ADTree ให้ค่าความถูกต้องได้ดี อยู่ที่ช่วงร้อยละ 94 - 97 แต่ก่อนข้างใช้เวลาในการประมวลผลนานกว่าตัวจำแนกอื่นๆ ขณะที่ตัวจำแนก C4.5 และ Naïve Bayes ใช้เวลาในการประมวลผลใกล้เคียงกัน ตัวจำแนก C4.5 สามารถทำนายข้อมูลที่ไม่มีสิ่งเสพติดชุกช่อนได้ดี อยู่ที่ช่วงร้อยละ 99.0 - 99.9 และตัวจำแนก Naïve Bayes สามารถทำนายข้อมูลที่มีสิ่งเสพติดชุกช่อนได้ดี อยู่ที่ช่วงร้อยละ 21.0 - 84.4 และจะมากกว่าตัวจำแนกอื่นเมื่อเพิ่มขนาดของข้อมูลการเรียนรู้ที่สูงขึ้น

จากการศึกษาวิจัยนี้ ทำให้ผู้วิจัยมีความเชื่อมั่นว่าเทคนิคตัวจำแนก C4.5 หรือที่โปรแกรม WEKA เรียกว่า “Classifiers.Trees.J48” น่าจะมีประสิทธิภาพในการจำแนกข้อมูลที่ดี และน่าจะสามารถนำมาประยุกต์ใช้กับงานทบทวนวงเงินได้

ชุตินา อุดมะมุณี และ ประสงค์ ปราณีตพลกรัง (2553) ได้ทำการวิจัยเชิงเปรียบเทียบประสิทธิภาพของอัลกอริทึม C4.5 และ Bayesian Networks พบว่า Bayesian Networks มีประสิทธิภาพในการจำแนกข้อมูลได้ดีกว่าอัลกอริทึม C4.5 จึงได้นำโมเดลการจำแนกข้อมูลของ Bayesian Networks ไปพัฒนาเป็นระบบสนับสนุนการตัดสินใจแบบอัตโนมัติออนไลน์สำหรับการเลือกสาขาวิชาเรียนของนักศึกษาระดับอุดมศึกษา

จากการศึกษาวิจัยนี้ ทำให้ผู้วิจัยได้ทราบถึงจุดด้อยของการจำแนกข้อมูลด้วยอัลกอริทึม C4.5 ที่ไม่เหมาะกับงานบางประเภท

นโรดม อารีกุล และ ชุติรัตน์ จรัสกุลชัย (2554) ได้ทำการวิจัยเชิงเปรียบเทียบประสิทธิภาพของอัลกอริทึม Decision trees (J48) และอัลกอริทึม Bayes (BayesNet) จากการประมวลผลข้อมูลเพื่อวิเคราะห์หารูปแบบโมเดลของโปรแกรม WEKA พบว่า อัลกอริทึม Decision

trees (J48) มีความสามารถในการจำแนกข้อมูลค้นหาผู้หลบหลีกภาษีได้ดีกว่า อัลกอริทึม Bayes และมีประสิทธิภาพในการพยากรณ์ถูกต้องเฉลี่ยร้อยละ 71.97

จากการศึกษาวิจัยนี้ ทำให้ผู้วิจัยมีความเชื่อมั่นว่าการวิเคราะห์รูปแบบโมเดลด้วยโปรแกรม WEKA โดยเลือกใช้เทคนิค Classifiers.Trees.J48 น่าจะเป็นเทคนิคที่เหมาะสมและควรนำมาประยุกต์ใช้กับงานทบทวนวงเงิน