

บทที่ 2

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

ในบทนี้จะกล่าวถึงความน่าเชื่อถือ องค์ประกอบที่มีผลกระทบต่อความน่าเชื่อถือ ความน่าเชื่อถือของเว็บไซต์ในระบบพาณิชย์อิเล็กทรอนิกส์ รวมทั้งเทคนิคทางด้านคอมพิวเตอร์ที่จะนำมาใช้เพื่อประเมินระดับความน่าเชื่อถือของเว็บไซต์ ทั้งในเชิงทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ความน่าเชื่อถือ (Trust)

McKnight และ Chervany (1996) ได้ศึกษาและรวบรวมข้อมูลเกี่ยวกับความน่าเชื่อถือที่ได้จากมุมมองที่แตกต่างกันของงานวิจัยและหนังสือจำนวน 60 เรื่อง และได้สรุปว่าความน่าเชื่อถือเกิดขึ้นจากโครงสร้าง 6 ประการ ที่มีความสัมพันธ์กัน คือ

1. **ความตั้งใจที่จะเชื่อถือ (Trusting Intention)** หมายถึง การที่บุคคลหนึ่งตั้งใจที่จะเชื่อถือบุคคลอื่นด้วยความรู้สึกที่ปลอดภัยเมื่อมีสถานการณ์เฉพาะเกิดขึ้น ถึงแม้ว่าผลในทางที่ไม่ดีอาจจะเกิดขึ้นได้หลังจากนั้น โดยความตั้งใจที่จะเชื่อถือเกิดขึ้นจากหน่วยพื้นฐาน 5 ประการ ได้แก่

-โอกาสที่จะเกิดผลในทางที่ไม่ดีตามมา (Potential negative consequences) หรืออีกนัยหนึ่งคือ การเกิดความเสียหาย ถ้าหากไม่มีความเสี่ยง สถานการณ์จะไม่สร้างความเสียหายให้เกิดความตั้งใจที่จะเชื่อถือ

-การขึ้นกับบุคคลอื่น (Dependence) สามารถแปลความหมายได้ว่าเป็นการวางใจให้บุคคลอื่นอยู่ในตำแหน่งที่มีอำนาจเหนือกว่าตน ดังนั้นบุคคลที่มีความตั้งใจที่จะเชื่อถือ คือ บุคคลที่ปรารถนาให้บุคคลอื่นมีอำนาจเหนือกว่าตนในสถานการณ์ที่กำหนด

-ความรู้สึกปลอดภัย (Feeling of security) บุคคลที่อยู่ในความเสียเปรียบปรารถนาที่จะขึ้นอยู่กับผู้อื่นด้วยความรู้สึกปลอดภัย

-ความเฉพาะของสถานการณ์ (Situation-specific) บุคคลจะมีการสร้างโมเดลขึ้นในใจว่าควรจะขึ้นกับบุคคลใดเมื่ออยู่ในสถานการณ์ใด ยกตัวอย่างเช่น บุคคลจะเชื่อถือแพทย์ในการวินิจฉัยโรค แต่จะไม่เชื่อถือแพทย์ในการขับเครื่องบิน

-ไม่ขึ้นอยู่กับการควบคุม (Lack of reliance on control) บุคคลที่มีความตั้งใจที่จะเชื่อถือจะวางใจในความเชื่อถือ (rely on trust) แต่จะไม่ให้ความสำคัญกับกลไกในการควบคุม ยกตัวอย่างเช่น หัวหน้าที่ใช้กล้องวงจรปิดในการเฝ้าดูการทำงานของลูกน้อง จะขึ้นกับกล้องวงจร

ปิดในการทำให้มั่นใจว่าลูกน้องประพฤติตนอย่างเหมาะสม แต่ในกรณีนี้หัวหน้าไม่ได้มีความตั้งใจที่จะเชื่อถือในพฤติกรรมของลูกน้องอย่างแท้จริง

2.ความเชื่อถือในพฤติกรรม (Trusting Behavior) ความตั้งใจที่จะเชื่อถือเป็นสิ่งนำไปสู่การเชื่อถือในบุคคลอื่นอย่างแท้จริง ดังนั้น ความเชื่อถือในพฤติกรรม หมายถึง การที่บุคคลอาสาสมัครที่จะขึ้นกับบุคคลอื่นในสถานการณ์เฉพาะด้วยความรู้สึที่ปลอดภัย ถึงแม้ว่าผลในทางที่ไม่ดีอาจจะเกิดขึ้นได้หลังจากนั้น เนื่องจากการ"ขึ้นกับ" เป็นพฤติกรรม จึงทำให้ความเชื่อถือในพฤติกรรมแตกต่างจากความตั้งใจที่จะเชื่อถือ

3.ความเชื่อในความน่าเชื่อถือ (Trusting Beliefs) หมายถึง การที่บุคคลเชื่อว่าบุคคลอื่นเป็นผู้ที่ควรค่าแก่การเชื่อถือ (trustworthiness) เมื่ออยู่ในสถานการณ์ โดยความเชื่อในความน่าเชื่อถือเกิดจากการรวมกันของความเชื่อ 4 กลุ่ม คือ

- ความมีเมตตา (Benevolence) หมายถึงความห่วงใยในผู้อื่นและตั้งใจช่วยเหลือในสิ่งที่ผู้อื่นต้องการ

- ความซื่อสัตย์ (Honesty) หมายถึงการปฏิบัติตามข้อตกลง

- ศักยภาพ (Competence) หมายถึงความสามารถในการทำเพื่อผู้อื่นในสิ่งที่ผู้อื่นต้องการได้

- สามารถคาดการณ์ได้ (Predictability) หมายถึงการกระทำที่มีความสม่ำเสมอที่ผู้อื่นสามารถคาดการณ์ได้ว่าผู้นั้นจะกระทำเช่นไรเมื่อมีสถานการณ์เกิดขึ้น

การรวมกันของความเชื่อทั้ง 4 กลุ่ม ก่อให้เกิดพื้นฐานที่มั่นคงของความตั้งใจที่จะเชื่อและความเชื่อถือในพฤติกรรม นั่นคือ ถ้าบุคคลเป็นผู้ที่มีความสม่ำเสมอหรือสามารถคาดการณ์ได้สามารถพิสูจน์ได้ว่าเป็นผู้ที่มีเมตตาและสามารถที่จะทำในสิ่งที่ผู้อื่นต้องการได้ด้วยความซื่อสัตย์ ผู้นั้นจะเป็นผู้ที่ควรค่าแก่การเชื่อถือ

4.ระบบของความเชื่อถือ (System Trust) หมายถึง การที่บุคคลเชื่อว่ามีโครงสร้างที่เหมาะสมเพื่อทำให้บุคคลสามารถคาดการณ์ความพยายามที่จะสำเร็จในอนาคตได้ โดยโครงสร้างที่เหมาะสมแบ่งเป็น 2 ประเภท

- โครงสร้างของการรับประกัน (Structural assurance) ได้แก่ กฎเกณฑ์ การรับประกัน หรือ ข้อตกลง ยกตัวอย่างเช่น การที่บริษัทมั่นใจว่าโครงการใหม่จะประสบผลสำเร็จเนื่องจากเคยประสบผลสำเร็จมาแล้วหลายครั้ง

-ความเป็นปกติของสถานการณ์ (Situational normality) ซึ่งอยู่บนพื้นฐานของความเข้าใจว่าเหตุการณ์โดยทั่วไปมีความปกติ ทั้งนี้เพื่อช่วยลดความไม่แน่นอน และทำให้บุคคลมีความรู้สึกปลอดภัยมากขึ้นเมื่อต้องรับความเสี่ยงกับบุคคลอื่น

5.การถ่ายทอดของความน่าเชื่อถือ (Dispositional Trust) ประกอบด้วย 2 เหตุผล ได้แก่

-ความเชื่อในบุคคล (Belief-in-people) บุคคลเชื่อว่าโดยทั่วไปแล้วผู้คนมีความน่าเชื่อถือ ซึ่งเป็นสิ่งที่สนับสนุนความเชื่อในความน่าเชื่อถือ

-ท่าทีที่นำเชื่อถือ (Trusting stance) บุคคลเชื่อว่าไม่ว่าคนจะดีหรือไม่ดี บุคคลสามารถได้รับผลที่ดีจากการเชื่อถือคนเหล่านั้น ดังนั้นบุคคลควรเชื่อถือผู้อื่น ซึ่งเป็นสิ่งที่สนับสนุนความตั้งใจที่จะเชื่อถือ

6.การตัดสินใจที่จะเชื่อถือขณะที่มีสถานการณ์เกิดขึ้น (Situation decision to trust) หมายถึง การที่บุคคลสร้างความตั้งใจที่จะเชื่อถือผู้อื่นในทุกๆครั้งที่มีสถานการณ์เกิดขึ้น ทั้งนี้เนื่องจากผลประโยชน์ที่จะได้รับมีมากกว่าผลลัพธ์ในทางไม่ดีที่อาจจะเกิดขึ้นจากการเชื่อถือในครั้งนี้

2.2 ความน่าเชื่อถือ และความไม่น่าเชื่อถือ (Trust and Distrust)

Lewicki และ Bies (1998) ได้นำเสนอโครงสร้างเชิงทฤษฎีเพื่อให้เข้าใจความสัมพันธ์ของความน่าเชื่อถือและความไม่น่าเชื่อถือ โดยให้คำจำกัดความของความน่าเชื่อถือว่าเป็นความคาดหวังการกระทำของผู้อื่นในทางที่ดี และให้คำจำกัดความของความไม่น่าเชื่อถือว่าเป็นความคาดหวังการกระทำของผู้อื่นในทางที่ไม่ดี อย่างไรก็ตาม ความน่าเชื่อถือและความไม่น่าเชื่อถือเป็นโครงสร้างที่แยกออกจากกัน ความสัมพันธ์ระหว่างความน่าเชื่อถือและความไม่น่าเชื่อถือ มี 4 รูปแบบ ดังนี้

1.มีความน่าเชื่อถือน้อยและมีความไม่น่าเชื่อถือน้อย (Low trust / Low distrust)

ภายใต้ความสัมพันธ์นี้บุคคลไม่จำเป็นต้องมีความเชื่อมั่นหรือไม่จำเป็นต้องคอยเฝ้าระวังในการกระทำของผู้อื่น เมื่อเริ่มมีความสัมพันธ์ระหว่างกันมากขึ้น ความตระหนักถึงบุคคลอื่นจะพัฒนาขึ้นและก่อตัวเป็นความเชื่อว่าบุคคลอื่นควรค่าแก่ความเชื่อถือหรือความไม่น่าเชื่อถือ

2. มีความน่าเชื่อถือสูงและมีความไม่น่าเชื่อถือน้อย (High trust / Low distrust)

ภายใต้ความสัมพันธ์นี้บุคคลมีเหตุผลที่จะเชื่อมั่นในบุคคลอื่นและไม่มีเหตุผลที่จะสงสัยในบุคคลอื่น พื้นฐานของความสัมพันธ์จะเกิดจากประสบการณ์ในด้านบวกที่เกิดขึ้นหลาย

ครั้ง โดยประสบการณ์นี้จะเป็นสิ่งที่สร้างสังคมซึ่งทำให้ผู้ที่มีความเชื่อถือซึ่งกันและกันให้การสนับสนุนผู้ที่มีความน่าเชื่อถือ แต่ฝ่ายจะหาทางที่จะพัฒนาความสัมพันธ์อย่างต่อเนื่องและซับซ้อนขึ้น และรับรู้ถึงผลกระทบในทางบวกจากความเชื่อถือ

3.มีความน่าเชื่อถือน้อยและมีความไม่น่าเชื่อถือสูง (Low trust / High distrust)

ภายใต้ความสัมพันธ์นี้บุคคลไม่มีเหตุผลที่จะเชื่อมั่นในบุคคลอื่น แต่มีเหตุผลที่จะคอยเฝ้าระวังบุคคลอื่น สถานการณ์เช่นนี้เป็นการยากต่อการรักษาความสัมพันธ์ระหว่างกันให้คงอยู่ในระยะยาว พื้นฐานของความสัมพันธ์เกิดจากประสบการณ์ในด้านลบที่เกิดขึ้นหลายครั้ง ซึ่งทำให้เกิดเป็นความไม่น่าเชื่อถือ หากต้องปฏิสัมพันธ์กับผู้ที่มีความไม่น่าเชื่อถือ อีกฝ่ายหนึ่งจะสละทรัพยากรที่สำคัญไปเพื่อการเฝ้าระวังการกระทำของผู้ที่ไม่น่าเชื่อถือ และเตรียมตัวพร้อมสำหรับการกระทำที่ไม่น่าเชื่อถือ

4.มีความน่าเชื่อถือสูงและมีความไม่น่าเชื่อถือสูง (High trust / High distrust)

ภายใต้ความสัมพันธ์นี้บุคคลมีเหตุผลที่จะเชื่อมั่นในบุคคลอื่นในบางเรื่อง และมีเหตุผลที่จะคอยเฝ้าระวังบุคคลอื่นในบางเรื่อง พื้นฐานของความสัมพันธ์เกิดจากประสบการณ์ทางด้านบวกที่เกิดขึ้นหลายครั้งซึ่งก่อให้เกิดความน่าเชื่อถือ และประสบการณ์ทางด้านลบที่เกิดขึ้นหลายครั้งซึ่งก่อให้เกิดความไม่น่าเชื่อถือ

2.3 ความเชื่อถือได้ (Credibility)

McKnight (2007) ได้ทำงานวิจัยเกี่ยวกับปัจจัยและผลกระทบของความเชื่อถือได้ของข้อมูล และให้ความหมายของความเชื่อถือได้ของข้อมูลว่า เกิดจากการสุ่มเลือกและทดลองปฏิบัติตามคำแนะนำและพิจารณาจากผลลัพธ์ของการทดลองปฏิบัติเป็นระยะเวลายาวนาน ยกตัวอย่างเช่น เว็บไซต์เกี่ยวกับการลงทุนทางการเงินที่ให้คำแนะนำซึ่งอาจจะส่งผลให้ได้ผลตอบแทนในทางที่เพิ่มขึ้นหรือลดลง และมีผู้ปฏิบัติตามคำแนะนำในเว็บไซต์แล้วสังเกตผลที่ได้รับ

Fogg (2003) ได้สรุปความหมายของความเชื่อถือได้จากงานวิจัยที่ย้อนหลังไปกว่า 5 ทศวรรษว่า ความเชื่อถือได้คือการรับรู้ถึงคุณภาพ (perceived quality) ซึ่งเป็นสิ่งที่ไม่ได้อยู่ในวัตถุบุคคล หรือ ข้อมูล ความเชื่อถือได้ไม่สามารถสัมผัส ได้ยิน หรือเห็นได้ด้วยตา แต่ความเชื่อถือได้คือการรับรู้ที่เกิดขึ้นเมื่อมีการประเมินค่าของบุคคล วัตถุ หรือข้อมูล ความเชื่อถือได้เกิดจากมิติที่สำคัญ 2 มิติ คือ ความควรค่าแก่การเชื่อถือ (trustworthiness) และ ความเชี่ยวชาญ (expertise)

-ความควรค่าแก่การเชื่อถือ เป็นการรับรู้ถึงคุณค่าหรือคุณธรรมของแหล่งกำเนิดในแง่ของอาชีพ ผู้พิพากษา แพทย์ นักบุญ และกรรมการ โดยทั่วไปจะได้รับการรับรู้ว่าเป็นผู้ที่เชื่อถือได้ (trustworthy) สิ่งสำคัญที่นำไปสู่การรับรู้ถึงความควรค่าแก่การเชื่อถือ ได้แก่ การรับรู้ว่าแหล่งกำเนิดมีความยุติธรรมและไม่มีอคติ การรับรู้ว่าแหล่งกำเนิดมีความซื่อสัตย์อย่างแท้จริง และการรับรู้ถึงความเหมือนกัน เช่น การมีความเหมือนกันทางด้านการคิดเห็น ภาษา หรือปุมหลัง เป็นต้น

-ความเชี่ยวชาญ คือการรับรู้ถึงการมีความรู้ ความชำนาญการ และประสบการณ์ของแหล่งกำเนิด สิ่งนำไปสู่การรับรู้ถึงความเชี่ยวชาญ ได้แก่ สัญลักษณ์ที่ประกาศว่าบุคคลมีความเชี่ยวชาญ เช่น คำนำหน้าว่า ศาสตราจารย์ หรือนายแพทย์ เอกสารที่แสดงถึงความสำเร็จ เช่น ใบรับรองความสามารถ หรือ สัญลักษณ์ที่ปรากฏ เช่น แบบฟอร์มของทหาร

การรวมกันของการรับรู้ถึงความควรค่าแก่การเชื่อถือและการรับรู้ถึงความเชี่ยวชาญทำให้เกิดการรับรู้ถึงความเชื่อถือได้

2.4 ความน่าเชื่อถือในระบบพาณิชย์อิเล็กทรอนิกส์

Mahmood (2006) ได้นำเสนอโมเดลทางความคิดซึ่งแสดงถึงปัจจัยสำหรับพิจารณาความน่าเชื่อถือของผู้ใช้ระบบพาณิชย์อิเล็กทรอนิกส์ โดยโมเดลนี้เกิดขึ้นจากการวิเคราะห์ความน่าเชื่อถือในเชิงลึกทั้งทางด้านสังคมวิทยา จิตวิทยา และทางด้านอิเล็กทรอนิกส์ รวมทั้งได้รวบรวมการศึกษาทฤษฎีและการทดลอง (Empirical Studies) ที่เกี่ยวข้องกับความน่าเชื่อถือทางจิตวิทยาและทางด้านอิเล็กทรอนิกส์ในมุมมองของผู้ใช้อินเทอร์เน็ตและธุรกิจ งานวิจัยนี้ได้สรุปองค์ประกอบทางด้านพาณิชย์อิเล็กทรอนิกส์ที่มีผลกระทบต่อความน่าเชื่อถือ ดังต่อไปนี้

1. เทคโนโลยีทางด้านความปลอดภัย โดยทั่วไปเชื่อว่า ความปลอดภัยเป็นองค์ประกอบที่มีความสำคัญสูงสุดในการพิจารณาระดับความเชื่อของผู้ใช้บริการระบบพาณิชย์อิเล็กทรอนิกส์
2. ความน่าเชื่อถือในบุคคลที่สาม ปัจจุบันมีการใช้สัญลักษณ์ความน่าเชื่อถือจากบุคคลที่สาม (Third parties seals) เช่น TRUSTe หรือ Webtrust เพื่อรับรองความน่าเชื่อถือของเว็บไซต์
3. ความน่าเชื่อถือในชื่อเสียง ผู้ใช้บริการจะมีการตอบสนองในทางบวกที่จะซื้อสินค้า ถ้าได้รับทราบผลตอบรับในเชิงบวกจากผู้ใช้บริการก่อนหน้านี้

4. ความน่าเชื่อถือจากการอ้างอิง ข้อมูลเกี่ยวกับสินค้า หรือข้อมูลเกี่ยวกับบริษัท ที่ได้รับจากผู้ให้บริการรายอื่น มีผลกระทบต่อความน่าเชื่อถือในธุรกิจ
5. รูปแบบของเว็บไซต์ เว็บไซต์ที่มีการติดต่อกับผู้ใช้บริการอย่างมีประสิทธิภาพ และใช้ภาพที่สวยงามมีความละเอียดสูง มีผลให้มีความน่าเชื่อถือสูง
6. แหล่งข้อมูล ความถูกต้อง และความชัดเจนของข้อมูล มีผลต่อความน่าเชื่อถือของเว็บไซต์
7. ความคิดเห็นจากผู้มีชื่อเสียง
8. การรักษาข้อมูลส่วนบุคคล ผู้ใช้บริการจะยอมให้ข้อมูลส่วนตัว ก็ต่อเมื่อมั่นใจว่าระบบมีการรักษาข้อมูลที่ได้ให้ไว้

Fogg (2003) ได้ทำงานวิจัยเชิงสำรวจโดยการออกแบบสอบถามเพื่อประเมินองค์ประกอบที่มีผลกระทบต่อความน่าเชื่อถือของเว็บไซต์กับหน่วยตัวอย่างมากกว่า 2500 คน เป็นผู้หญิง 58.1% เป็นผู้ชาย 41.9% มีอายุเฉลี่ย 39.9 ปี มีจำนวนชั่วโมงการใช้อินเทอร์เน็ตต่อสัปดาห์ 19.6 ชั่วโมง โดยในการทำการทดลองจะให้ผู้เข้าร่วมทำการทดลองเข้าชมเว็บไซต์ที่เตรียมไว้ 2 เว็บไซต์ ประเมินความน่าเชื่อถือของเว็บไซต์ เรียงลำดับเว็บไซต์ตามความน่าเชื่อถือ และให้เขียนความคิดเห็นเกี่ยวกับความน่าเชื่อถือของแต่ละเว็บไซต์ จากนั้นนำความคิดเห็นของผู้ร่วมทดลองที่รวบรวมมาได้ทั้งหมด มาจำแนกด้วยเทคนิคการเรียนรู้ของเครื่องว่าความคิดเห็นนั้นเกี่ยวข้องกับองค์ประกอบใด ทั้งนี้ไม่ได้แยกว่าเป็นความคิดเห็นในเชิงบวกหรือเชิงลบ จากนั้นเรียงลำดับองค์ประกอบที่ได้รับการแสดงความคิดเห็น ซึ่งผลของการทดลอง ดังแสดงในตารางที่ 2.1 แสดงให้เห็นว่าองค์ประกอบที่ผู้ร่วมทดลองให้ความสนใจในการแสดงความคิดเห็นมากที่สุด คือ รูปแบบของเว็บไซต์ (Design Look) ซึ่งมีจำนวนความคิดเห็น 46.1% อันดับที่สองคือ การออกแบบการวางข้อความ 28.5%

ตารางที่ 2.1

องค์ประกอบของเว็บไซต์ที่มีผลกระทบต่อระดับความน่าเชื่อถือของเว็บไซต์

องค์ประกอบที่มีผลต่อความน่าเชื่อถือของเว็บไซต์	จำนวนความคิดเห็น (%)
รูปลักษณ์ของเว็บไซต์ (Design Look)	46.1
โครงสร้างการวางข้อความ (Information Design/Structure)	28.5
เน้นการให้ข้อมูล (Information Focus)	25.1
เกี่ยวกับบริษัท (Company Motive)	15.5
ข้อมูลมีประโยชน์ (Usefulness of Information)	14.8
ความถูกต้องของข้อมูล (Accuracy of Information)	14.3
ความมีชื่อเสียง (Name Recognition & Reputation)	14.1
โฆษณา (Advertising)	13.8
อคติ (Bias)	11.6
แนวการเขียน (Tone of the Writing)	9.0
แจ้งเว็บไซต์ที่สนับสนุน (Identify of Site Sponsor)	8.8
เว็บไซต์ทำหน้าที่ได้ดี (Functionality of Site)	8.6
ลูกค้าสัมพันธ์ (Customer Service)	6.4
ประสบการณ์เกี่ยวกับเว็บไซต์ (Past Experience with Site)	4.6
ความชัดเจนของข้อมูล (Information Clarity)	3.7
ประสิทธิภาพจากการทดลอง (Performance on a Test)	3.6
อ่านง่าย (Readability)	3.6
ความเชื่อมโยง (Affiliation)	3.4

Lumsden (2009) ได้ศึกษาความสำคัญขององค์ประกอบที่มีผลกระทบต่อความน่าเชื่อถือของเว็บไซต์ โดยให้ผู้เข้าร่วมทำการทดลองเข้าไปในเว็บไซต์ที่เตรียมไว้ แล้วตอบแบบสอบถามเกี่ยวกับระดับความสำคัญของแต่ละองค์ประกอบที่มีผลกระทบต่อความน่าเชื่อถือของเว็บไซต์ เกณฑ์การให้คะแนนเป็นการเรียงลำดับความเห็นด้วย (Likert scale) จากนั้นน้อยไปมาก โดยมีระดับคะแนน 1 – 5 โดย ระดับ 1 หมายถึง ผู้ตอบแบบสอบถามเห็นด้วยว่าองค์ประกอบนั้นมีผลกระทบต่อระดับความน่าเชื่อถือของเว็บไซต์น้อยที่สุด และระดับ 5 หมายถึง ผู้ตอบแบบสอบถามเห็นด้วยว่าองค์ประกอบนั้นมีผลกระทบต่อระดับความน่าเชื่อถือของเว็บไซต์มากที่สุด องค์ประกอบที่นำมาใช้ในแบบสอบถาม มีดังนี้คือ

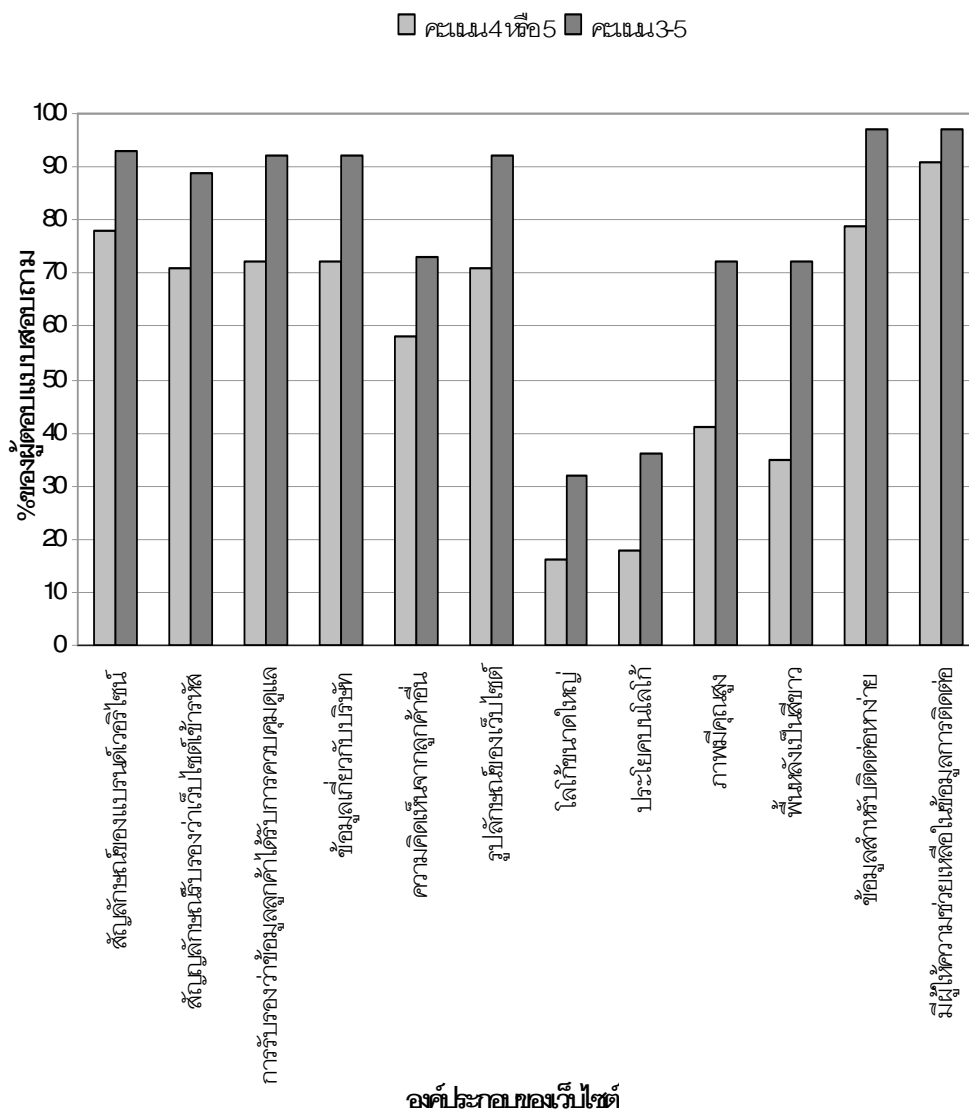
1. สัญลักษณ์ของแบรนด์เวอร์ซิไซน์ ที่รับรองว่าเว็บไซต์มีการใช้ระบบเข้ารหัส (Verisign Security Certificate)
2. สัญลักษณ์ที่รับรองว่าเว็บไซต์มีการใช้ระบบเข้ารหัส (Security Sockets Layer)
3. ข้อความที่รับรองข้อมูลของลูกค้าที่ส่งผ่านเว็บไซต์ได้รับการควบคุมดูแลอย่างดี (Privacy and Term information)
4. ข้อมูลเกี่ยวกับบริษัท (Company Profile information)
5. ความคิดเห็นจากลูกค้าอื่น (Testimonials from other customers)
6. รูปลักษณ์เว็บไซต์ที่ดูเป็นมืออาชีพ (Professional looking website design)
7. โลโก้ขนาดใหญ่ (Large Logo)
8. ประโยคบนโลโก้ (Statement on logo)
9. ภาพมีคุณภาพสูง (High quality graphics)
10. พื้นหลังเป็นสีขาว (Ample white space)
11. ข้อมูลสำหรับติดต่อหาง่าย (Easy to find contact information)
12. มีผู้ให้ความช่วยเหลือในข้อมูลการติดต่อ (Contact information includes live person support)

ผลการทดลองดังแสดงในภาพที่ 2.1 องค์ประกอบที่ผู้เข้าร่วมทำการทดลองเห็นด้วยว่ามีผลกระทบต่อระดับความน่าเชื่อถือของเว็บไซต์ที่มีระดับคะแนน 3-5 เกินกว่า 80% มีอยู่ 7 องค์ประกอบ ได้แก่ องค์ประกอบที่ (1) สัญลักษณ์ของแบรนด์เวอร์ซิไซน์ (2) สัญลักษณ์ที่รับรองว่าเว็บไซต์มีการใช้ระบบเข้ารหัส (3) ข้อความที่รับรองข้อมูลของลูกค้าที่ส่งผ่านเว็บไซต์ได้รับการ

ควบคุมดูแลอย่างดี (4)ข้อมูลเกี่ยวกับบริษัท (5)รูปลักษณ์เว็บไซต์ที่ดูเป็นมืออาชีพ (11)ข้อมูล
สำหรับติดต่อหาง่าย และ (12)มีผู้ให้ความช่วยเหลือในข้อมูลการติดต่อ

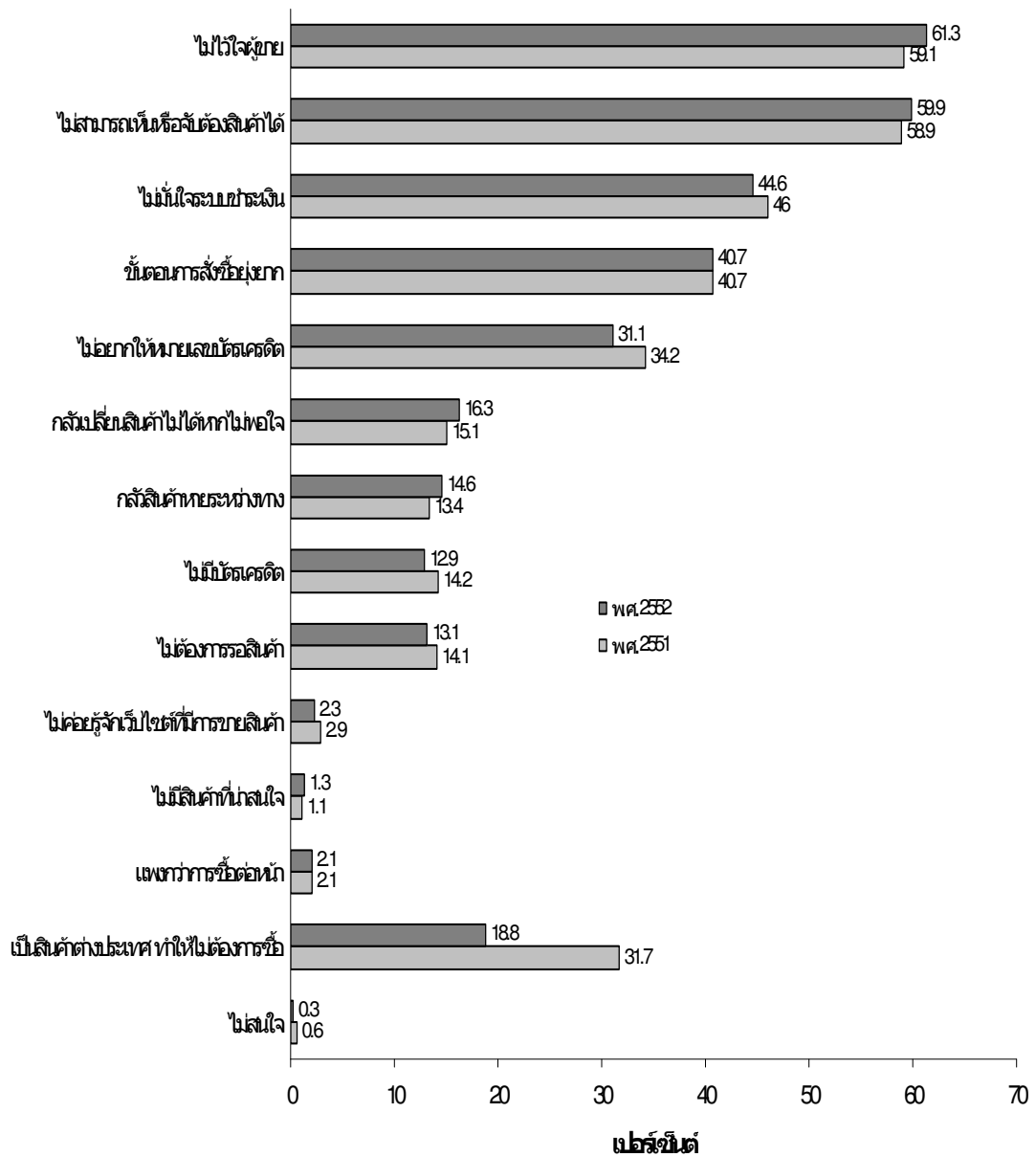
ภาพที่ 2.1

องค์ประกอบของเว็บไซต์ที่มีผลต่อระดับความน่าเชื่อถือของเว็บไซต์จากงานวิจัยของ



จากรายงานผลการสำรวจกลุ่มผู้ใช้อินเทอร์เน็ตในประเทศไทย ปี 2552 ที่จัดทำขึ้น โดยศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ โดยให้ผู้ใช้อินเทอร์เน็ตตอบแบบสอบถามออนไลน์ สำหรับหัวข้อการซื้อสินค้าหรือบริการทางอินเทอร์เน็ต ที่มีผู้ตอบทั้งหมด 10,352 คน พบว่ามีผู้เคยซื้อสินค้าและบริการทางอินเทอร์เน็ต 47.8% และผู้ไม่เคยซื้อสินค้าและบริการทางอินเทอร์เน็ต 52.2% ในจำนวนผู้ไม่เคยซื้อสินค้าและบริการทางอินเทอร์เน็ตให้เหตุผลสำคัญในการไม่ซื้อสินค้าและบริการผ่านทางอินเทอร์เน็ต 5 อันดับแรกว่า ไม่ไว้วางใจผู้ขายว่ามีสินค้านั้นจริงหรือจะส่งสินค้าให้จริงร้อยละ 61.3 ไม่สามารถเห็นสินค้าหรือจับต้องได้ร้อยละ 59.9 ไม่มั่นใจในระบบการชำระเงินร้อยละ 44.6 ขั้นตอนการสั่งซื้อยุ่งยากร้อยละ 40.7 และไม่ต้องการส่งข้อมูลบัตรเครดิตผ่านทางอินเทอร์เน็ตร้อยละ 31.1 ซึ่งเหตุผลของการไม่ซื้อสินค้าทางอินเทอร์เน็ตดังแสดงในภาพที่ 2.2

ภาพที่ 2.2
เหตุผลของผู้ที่ไม่ซื้อสินค้าหรือบริการทางอินเทอร์เน็ต



วงศ์วี วิบูลย์สุข (2545) ทำงานวิจัยเกี่ยวกับการออกแบบหน้าจอบล็อกเว็บไซต์และการจัดการข้อผิดพลาดที่มีผลกับความไว้วางใจในพาณิชย์อิเล็กทรอนิกส์ โดยปัจจัยที่เป็นองค์ประกอบในการออกแบบหน้าจอบล็อกเว็บเพจประกอบด้วย ตำแหน่งการจัดวางองค์ประกอบบนอินเทอร์เน็ตเพสคุณสมบัติของสี ชนิดของตัวอักษร และขนาดของตัวอักษร โดยแบ่งเป็น 2 ประเภท ได้แก่

- องค์ประกอบในการออกแบบหน้าจอบล็อกเว็บเพจประเภทที่ 1 มีลักษณะตำแหน่งการจัดวางขององค์ประกอบสอดคล้องกันทุกหน้า ขนาดขององค์ประกอบแต่ละประเภทจะมีขนาดเท่ากันทุกหน้า เช่นปุ่มมีขนาดเท่ากัน สีของพื้นหลังเป็นสีอ่อน ไม่ฉูดฉาด ใช้ตัวอักษร Tahoma ขนาด 13 พิกเซล
- องค์ประกอบในการออกแบบหน้าจอบล็อกเว็บเพจประเภทที่ 2 มีลักษณะตำแหน่งการจัดวางขององค์ประกอบไม่สอดคล้องกันทุกหน้า ขนาดขององค์ประกอบแต่ละประเภทจะมีขนาดไม่เท่ากัน สีของพื้นหลังเป็นสีฉูดฉาด ใช้ตัวอักษร Comic Sans MS ขนาด 18 พิกเซล

สำหรับปัจจัยปริมาณข้อผิดพลาดและการจัดการเมื่อเกิดข้อผิดพลาดขึ้นประกอบด้วย ปริมาณลิงค์ที่ใช้งานไม่ได้ จำนวนคำที่สะกดผิด ชนิดของข้อความที่ใช้อธิบายข้อผิดพลาดที่เกิดขึ้น และการมีข้อความแนะนำการใช้งานเมื่อมีข้อผิดพลาดเกิดขึ้น โดยแบ่งเป็น 2 ประเภท ได้แก่

- ปริมาณข้อผิดพลาดและการจัดการข้อผิดพลาดประเภทที่ 1 ในเว็บเพจแต่ละหน้าจะไม่มีลิงค์ที่ใช้งานไม่ได้ ไม่มีคำที่สะกดผิด ระบบจะแสดงข้อความอธิบายข้อผิดพลาดให้ผู้เข้าใจ แนะนำการใช้งานเบื้องต้นเมื่อเกิดข้อผิดพลาดขึ้น เช่นการกรอกข้อมูลไม่ถูกต้องตามรูปแบบที่ระบบกำหนด ผู้ใช้ใส่รหัสผ่านไม่ตรงกับฐานข้อมูล เป็นต้น
- ปริมาณข้อผิดพลาดและการจัดการข้อผิดพลาดประเภทที่ 2 ในเว็บเพจจะมีลิงค์ที่ใช้งานไม่ได้ มีคำที่สะกดผิด ระบบจะแสดงข้อความของข้อผิดพลาดจากระบบโดยตรง (System error message) และไม่มีข้อความแนะนำการใช้งานเมื่อเกิดข้อผิดพลาดขึ้น

ผู้วิจัยได้ทำการทดลองโดยให้หน่วยทดลองจำนวน 48 คน ใช้เว็บไซต์ที่มีลักษณะแตกต่างกัน 4 เว็บไซต์ ดังนี้

เว็บไซต์ที่ 1: มีองค์ประกอบในการออกแบบหน้าจอบล็อกเว็บเพจประเภทที่ 1 และมีปริมาณข้อผิดพลาดและการจัดการข้อผิดพลาดประเภทที่ 1

เว็บไซต์ที่2: มีองค์ประกอบในการออกแบบหน้าจอเว็บเพจประเภทที่ 2 และมีปริมาณ
ข้อผิดพลาดและการจัดการข้อผิดพลาดประเภทที่ 1

เว็บไซต์ที่3: มีองค์ประกอบในการออกแบบหน้าจอเว็บเพจประเภทที่ 1 และมีปริมาณ
ข้อผิดพลาดและการจัดการข้อผิดพลาดประเภทที่ 2

เว็บไซต์ที่4: มีองค์ประกอบในการออกแบบหน้าจอเว็บเพจประเภทที่ 2 และมีปริมาณ
ข้อผิดพลาดและการจัดการข้อผิดพลาดประเภทที่ 2

ผลการทดลองพบว่าเว็บไซต์ที่มีลักษณะตำแหน่งการจัดวางขององค์ประกอบ
สอดคล้องกันทุกหน้า เช่น มีปุ่มขนาดเท่ากัน สีของพื้นหลังเป็นสีอ่อน ใช้ตัวอักษร Tahoma ขนาด
13 พิกเซล และ ในเว็บไซต์ไม่มีลิงค์ที่ใช้งานไม่ได้ ไม่มีคำที่สะกดผิด เป็นเว็บไซต์ที่ได้รับความ
น่าเชื่อถือมากที่สุด แต่ถ้าเว็บไซต์มีลิงค์ที่ใช้งานไม่ได้ มีคำที่สะกดผิด และไม่มีคำแนะนำการ
จัดการเมื่อเกิดข้อผิดพลาด เว็บไซต์นั้นจะไม่ได้ได้รับความน่าเชื่อถือแม้จะมีการออกแบบหน้าจอที่ดี

ซุซanne ดั้งมณี และ วรวิทย์ อชินานนท์ (2551) ได้ทำงานวิจัยเกี่ยวกับการ
เปรียบเทียบความน่าเชื่อถือต่อเว็บไซต์ขายหนังสือเรียน เมื่อใช้ตรารับรองของกระทรวงพาณิชย์
และเมื่อใช้แบบอักษรแตกต่างกัน โดยใช้แบบอักษร Microsoft sans serif สำหรับแบบที่เป็น
ทางการ และ แบบอักษร Luxamon สำหรับแบบที่ไม่เป็นทางการ โดยทำการทดลองแบบ 2 x 2
(แสดงหรือไม่แสดงตรารับรอง x อักษรแบบที่เป็นทางการหรือไม่เป็นทางการ) ทำการทดสอบกับ
นิสิตคณะพาณิชยศาสตร์และการบัญชี จุฬาลงกรณ์มหาวิทยาลัย จำนวนรวม 168 คน หรือ
จำนวน 42 คนต่อลักษณะของการทดลอง ผลการทดลองพบว่าความเชื่อถือต่อเว็บไซต์ขายหนังสือ
เรียนเมื่อแสดงตรารับรองและไม่แสดงตรารับรองมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ
และความเชื่อถือต่อเว็บไซต์ขายหนังสือเรียนเมื่อใช้แบบอักษรที่เป็นทางการและไม่เป็นทางการมี
ความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ โดยเว็บไซต์ที่แสดงตรารับรองได้รับความเชื่อถือโดย
เฉลี่ยสูงกว่าเว็บไซต์ที่ไม่แสดงตรารับรอง และเว็บไซต์ที่ใช้แบบอักษรที่เป็นทางการได้รับความ
เชื่อถือโดยเฉลี่ยสูงกว่าเว็บไซต์ที่ใช้แบบอักษรที่ไม่เป็นทางการ

วรรณิดา กลิ่นดี (2551) ได้ทำงานวิจัยเกี่ยวกับปัจจัยทางอินเทอร์เน็ตฟลที่มีผลกระทบ
ต่อความน่าเชื่อถือของเว็บไซต์พาณิชย์อิเล็กทรอนิกส์ โดยได้คัดเลือกปัจจัยทางอินเทอร์เน็ตฟลที่มี
ความสำคัญต่อความน่าเชื่อถือของเว็บไซต์จำนวน 30 ปัจจัยมาใช้ในการวิจัย ได้แก่ (1)การแสดง
ราคาสินค้าและค่าขนส่ง (2)การแสดงรูปภาพของพนักงานและผู้บริหาร (3)การแสดงความน่าเชื่อถือ
อยู่ เบอร์โทรศัพท์ติดต่อ (4) การแสดงวิธีการชำระเงิน (5)การอัปเดตข้อมูลของเว็บไซต์ (6)การ
แสดงประวัติของบริษัท (7)การแสดงนโยบายการซื้อและคืนสินค้า (8)ปัจจัยด้านการแสดงข้อมูลผู้

ถือหุ้น (9) คำสะกดบนเว็บไซต์ (10) ชื่อโดเมนลงท้าย.com (11) การแสดงรูปภาพตัวอย่างสินค้า (12) การแสดงระบบช่วยเหลือทางอีเมล และระบบสนทนาออนไลน์ (13) การแสดงระบบตรวจสอบการสั่งซื้อสินค้า (14) การแสดงระบบตอบข้อซักถาม (15) การแสดงระบบนำทาง (16) การแสดงข้อความเรียกร้องให้ผู้ใช้งานลงทะเบียนกับเว็บไซต์ (17) การแสดงแผนผังเว็บไซต์ (18) เวลาในการแสดงผลของอินเทอร์เน็ตเฟส (19) การแสดงระบบยกเลิกการสั่งซื้อ (20) การจัดการข้อผิดพลาดที่เกิดขึ้นบนอินเทอร์เน็ตเฟส (21) การจัดวางเมนูบนอินเทอร์เน็ตเฟส (22) การใช้รูปแบบตัวอักษร (23) โทนสีพื้นของเว็บไซต์ (24) จำนวนสื่อโฆษณาบนเว็บไซต์ (25) จำนวนภาพเคลื่อนไหว (26) ขนาดตัวอักษรของข้อความบนเว็บไซต์ (27) การแสดงเครื่องหมายรับประกันการรักษาข้อมูลส่วนบุคคลของผู้ใช้งาน (28) การแสดงเครื่องหมายรับประกันความปลอดภัยในการใช้งาน (29) การแสดงตราสัญลักษณ์สินค้า และ (30) การแสดงรางวัลและหนังสือรับรองของบริษัท

งานวิจัยนี้ใช้การออกแบบสอบถามออนไลน์เพื่อสำรวจความคิดเห็นของผู้บริโภคว่าปัจจัยใดที่ผู้บริโภคเห็นว่ามีความกระทบต่อความน่าเชื่อถือของเว็บไซต์ขายหนังสือออนไลน์ โดยดูจากปัจจัยที่ผู้บริโภคเลือกที่มีความกระทบต่อความน่าเชื่อถือของเว็บไซต์และระดับคะแนนที่ผู้บริโภคให้กับแต่ละปัจจัยที่เลือก ผลการทดลองพบว่า ผู้ที่มีประสบการณ์และผู้ที่ไม่มีความประสบการณ์ในการซื้อสินค้าออนไลน์ให้ความสำคัญต่อยังปัจจัยทางอินเทอร์เน็ตเฟสที่มีความกระทบต่อความน่าเชื่อถือของเว็บไซต์แตกต่างกันถึง 15 ปัจจัย แต่ปัจจัยที่ผู้ที่มีประสบการณ์และผู้ที่ไม่มีความประสบการณ์ในการซื้อสินค้าออนไลน์ให้ความสำคัญ 9 อันดับแรกเป็นปัจจัยที่เหมือนกัน ดังแสดงในตารางที่ 2.2 โดยทั้ง 9 ปัจจัยเป็นปัจจัยที่เหมือนกันเพียงแต่ผู้ที่มีประสบการณ์และผู้ที่ไม่มีความประสบการณ์ให้ลำดับความสำคัญแตกต่างกันเท่านั้น สำหรับลำดับคะแนนความน่าเชื่อถือเฉลี่ยของผู้ที่มีประสบการณ์และไม่มีความประสบการณ์ของทุกปัจจัยแสดงในภาพที่ 2.3

ตารางที่ 2.2

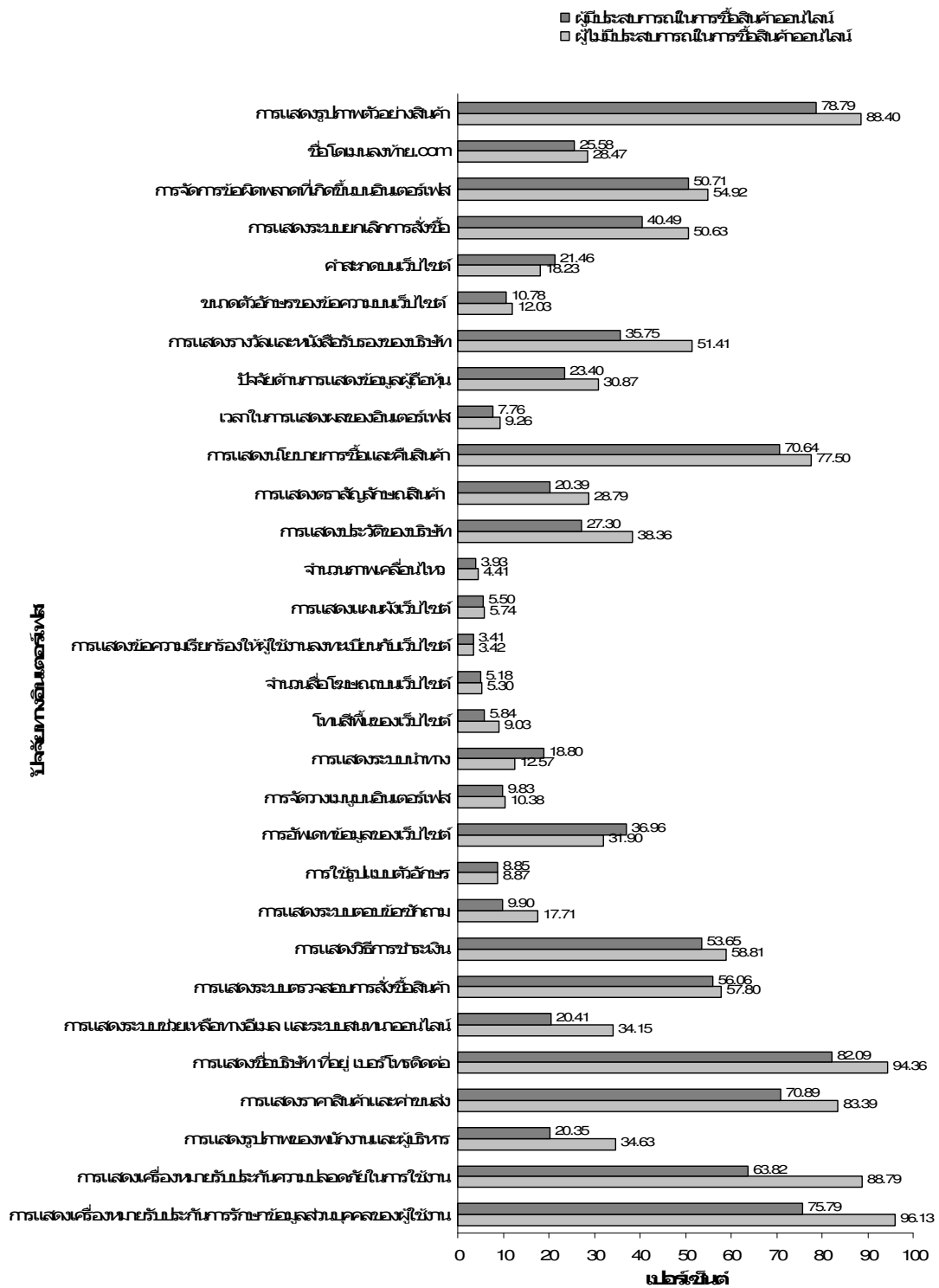
ปัจจัยที่ผู้มีประสบการณ์และผู้ที่ไม่ประสบการณ์ในการซื้อสินค้าออนไลน์

ให้ความสำคัญ 9 อันดับแรก

	ผู้มีประสบการณ์		ผู้ไม่มีประสบการณ์	
	อันดับ	คะแนน	อันดับ	คะแนน
การแสดงชื่อบริษัท ที่อยู่ เบอร์โทรศัพท์ติดต่อ	1	82.09	2	94.36
การแสดงรูปภาพตัวอย่างสินค้า	2	78.79	4	88.40
การแสดงเครื่องหมายรับประกันการรักษาข้อมูลส่วนบุคคลของผู้ใช้งาน	3	75.79	1	96.13
การแสดงราคาสินค้าและค่าขนส่ง	4	70.89	5	83.39
การแสดงนโยบายการซื้อและคืนสินค้า	5	70.64	6	77.50
การแสดงเครื่องหมายรับประกันความปลอดภัยในการใช้งาน	6	63.82	3	88.79
การแสดงระบบตรวจสอบการสั่งซื้อสินค้า	7	56.06	8	57.80
การแสดงวิธีการชำระเงิน	8	53.65	7	58.81
การจัดการข้อผิดพลาดที่เกิดขึ้นบนอินเทอร์เน็ต	9	50.71	9	54.92

ภาพที่ 2.3

คะแนนความน่าเชื่อถือเว็บไซต์ของผู้ที่มีประสบการณ์และผู้ที่ไม่ประสบการณ์
ในการซื้อสินค้าออนไลน์



2.5 เทคนิคการเรียนรู้ของเครื่อง

เนื่องจากในฐานข้อมูลมีข้อมูลอยู่เป็นจำนวนมากที่เป็นข้อมูลแฝงและสามารถนำมาใช้ประโยชน์เพื่อการตัดสินใจที่ถูกต้องได้ การวิเคราะห์ข้อมูลเพื่อสกัดโมเดลที่สามารถอธิบายถึงกลุ่มของข้อมูล และนำไปใช้เพื่อทำนายแนวโน้มของข้อมูลในอนาคตจึงเป็นเรื่องที่สำคัญ โดยรูปแบบหนึ่งของการวิเคราะห์ข้อมูลที่สามารถนำมาใช้เพื่ออธิบายกลุ่มของข้อมูลและทำนายแนวโน้มของข้อมูลในอนาคต คือ การจัดกลุ่มข้อมูล ทั้งนี้การจัดกลุ่มข้อมูลมีอยู่หลายเทคนิคและมีประสิทธิภาพแตกต่างกัน โดยในหัวข้อนี้จะกล่าวถึง การจัดกลุ่มข้อมูล และเทคนิคการจัดกลุ่มข้อมูล 4 เทคนิค ได้แก่ เทคนิคต้นไม้ตัดสินใจ เทคนิคโครงข่ายประสาทเทียม เทคนิคนาอ็อฟเบย์ และเทคนิคตรรกศาสตร์คลุมเครือ

2.5.1 การจัดกลุ่มข้อมูล (Classification)

Han และ Kamber (2000) ได้อธิบายถึงการจัดกลุ่มข้อมูลซึ่งเป็นหนึ่งในเทคนิคการเรียนรู้ของเครื่องว่า เป็นวิธีการที่เกี่ยวข้องกับการวิเคราะห์ข้อมูลและการดึงรูปแบบของข้อมูลเป็นกลุ่มที่สามารถอธิบายลักษณะเฉพาะของกลุ่มได้ การจัดกลุ่มข้อมูลมี 2 ขั้นตอน คือ

1.การสร้างโมเดลเพื่ออธิบายชุดของกลุ่มข้อมูล โดยโมเดลจะถูกสร้างขึ้นมาจากการวิเคราะห์ข้อมูลในแต่ละแถวของฐานข้อมูล (tuple) ที่สามารถอธิบายได้ด้วยคุณลักษณะ (attribute) ข้อมูลในแต่ละแถวของฐานข้อมูลจะถูกสมมติว่าเป็นสมาชิกของกลุ่มใดกลุ่มหนึ่งที่ได้มีการกำหนดไว้ล่วงหน้าแล้ว และจะเรียกข้อมูลในแต่ละแถวว่าตัวอย่าง (sample) โดยตัวอย่างที่ถูกนำมาวิเคราะห์อย่างต่อเนื่องเพื่อสร้างโมเดลจะถูกเรียกว่าชุดของข้อมูลที่นำมาสอน (training data set) และข้อมูลแต่ละแถวที่ประกอบกันเป็นชุดของข้อมูลที่นำมาสอนจะเรียกว่าตัวอย่างที่นำมาสอน (training sample) ซึ่งจะถูกเลือกอย่างสุ่มมาจากตัวอย่างประชากร (sample population) เนื่องจากได้มีการกำหนดกลุ่มของตัวอย่างที่นำมาสอนไว้ล่วงหน้าแล้ว จึงเรียกขั้นตอนนี้อีกอย่างหนึ่งว่าขั้นตอนการเรียนรู้แบบมีผู้สอน (supervised learning) เพราะการเรียนรู้ของโมเดลเกิดจากการบอกให้โมเดลรู้ว่าแต่ละแถวของข้อมูลอยู่ในกลุ่มใด โดยทั่วไปโมเดลของการเรียนรู้จะแสดงในรูปของกฎการจัดกลุ่ม ต้นไม้ตัดสินใจ หรือสูตรทางคณิตศาสตร์ โดยที่กฎสามารถนำมาใช้เพื่อจำแนกตัวอย่างข้อมูล (data sample) ในอนาคตได้ อีกทั้งทำให้เข้าใจเนื้อหาในฐานข้อมูลได้

2. การนำโมเดลมาจัดกลุ่มของข้อมูล โดยที่ในช่วงแรกจะมีการประมาณความแม่นยำในการจัดกลุ่มข้อมูลโดยการให้ชุดทดสอบ (test set) ของตัวอย่างที่รู้ว่าอยู่กลุ่มใดเป็นเครื่องทดสอบ โดยที่ตัวอย่างเหล่านี้จะถูกเลือกอย่างสุ่มและเป็นอิสระจากตัวอย่างที่นำมาสอน ความแม่นยำของโมเดลสำหรับชุดทดสอบเป็นเปอร์เซ็นต์ของตัวอย่างชุดทดสอบที่โมเดลสามารถจัดกลุ่มได้ถูกต้อง สำหรับแต่ละตัวอย่างที่นำมาทดสอบกลุ่มที่รู้อยู่แล้วของตัวอย่างจะถูกนำมาเปรียบเทียบกับกลุ่มที่โมเดลทำนาย และถ้าความแม่นยำในการทำนายของโมเดลอยู่ในเกณฑ์ที่ยอมรับได้จะสามารถนำโมเดลมาใช้เพื่อจัดกลุ่มข้อมูลในอนาคตโดยที่ไม่ทราบว่าคุณสมบัติในอนาคตอยู่ในกลุ่มใดได้

2.5.2 การเปรียบเทียบประสิทธิภาพของการจัดกลุ่ม

Han และ Kamber (2000) ได้อธิบายว่า ประสิทธิภาพของการจัดกลุ่มสามารถเปรียบเทียบและประเมินได้ด้วยเกณฑ์ดังต่อไปนี้

- 1.ความแม่นยำในการทำนาย (Predictive accuracy) คือ ความสามารถในการทำนายกลุ่มของข้อมูลใหม่หรือข้อมูลที่ไม่เคยเห็นมาก่อนได้ถูกต้อง
- 2.ความเร็ว (Speed) คือ ต้นทุนการคำนวณ (computational cost) ที่เกี่ยวข้องกับวิธีการนั้น
- 3.ความเสถียร (Robustness) คือ ความสามารถของโมเดลในการทำนายเมื่อข้อมูลเป็นข้อมูลที่มีค่าบางค่าหายไป หรือเป็นข้อมูลที่มีสิ่งปลอมปน (noisy data)
- 4.การขยายขนาดได้ (Scalability) คือ ความสามารถของโมเดลในการทำงานอย่างมีประสิทธิภาพเมื่อมีข้อมูลจำนวนมาก
- 5.การแปลความหมายได้ (Interpretability) คือ ระดับของความเข้าใจและความเข้าถึงที่โมเดลมีให้

จะเห็นว่าการประเมินระดับความน่าเชื่อถือของเว็บไซต์มีความสอดคล้องกับการจัดกลุ่มข้อมูล โดยการประเมินระดับความน่าเชื่อถือของเว็บไซต์จะใช้ชุดตัวอย่างเว็บไซต์เป็นชุดข้อมูลที่นำมาสอน (training data set) ให้กับโมเดล และให้ชุดเว็บไซต์ทดสอบ (test set) ในการประมาณความแม่นยำของโมเดล เพื่อให้โมเดลจัดกลุ่มของเว็บไซต์ตัวอย่างได้ ดังนั้น ในการประเมินระดับความน่าเชื่อถือของเว็บไซต์สำหรับงานวิจัยนี้ จึงนำเทคนิคการเรียนรู้ของเครื่องด้านการจัดกลุ่มมาใช้ในการประเมินระดับความน่าเชื่อถือของเว็บไซต์

เทคนิคสำหรับการจัดกลุ่มมีอยู่หลายเทคนิคและมีข้อเด่นข้อด้อยแตกต่างกัน ในงานวิจัยนี้จะใช้เทคนิคการจัดกลุ่มที่มีหลักการทำงานแตกต่างกัน ทั้งนี้เพื่อเปรียบเทียบประสิทธิภาพในการประเมินระดับความน่าเชื่อถือของเว็บไซต์ของแต่ละเทคนิค โดยเทคนิคการจัดกลุ่มข้อมูลที่นำมาใช้ในงานวิจัยนี้ได้แก่ ต้นไม้ตัดสินใจ (Decision tree) โครงข่ายประสาทเทียม (Artificial Neural Network) นาอิวเบย์ (Naïve Bayes) และตรรกศาสตร์คลุมเครือ (Fuzzy Logic)

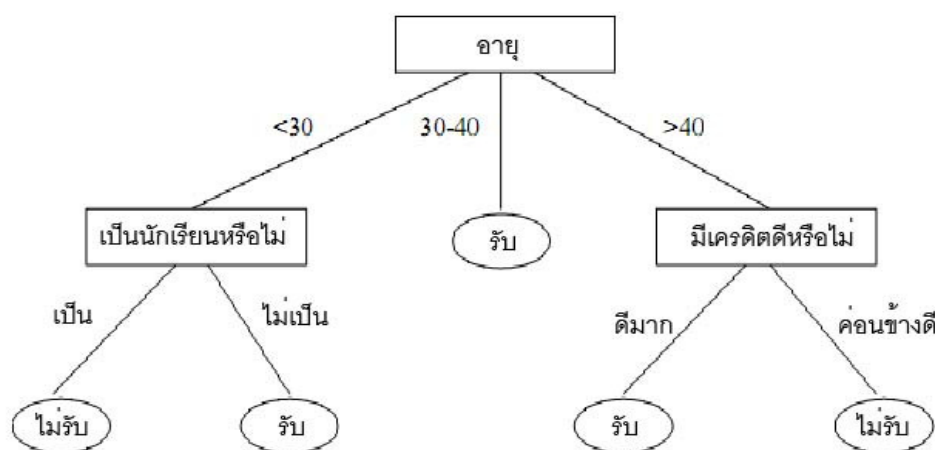
2.5.3 ต้นไม้ตัดสินใจ (Decision tree)

Dunham (2003) ได้อธิบายว่า ต้นไม้ตัดสินใจ เป็นเทคนิคการจัดกลุ่มที่ใช้หลักการของการแบ่งเพื่อเอาชนะ (divide and conquer) เพื่อแบ่งปัญหาใหญ่ออกเป็นปัญหาย่อยแล้วทำการแก้ปัญหาย่อยทีละปัญหา ต้นไม้ตัดสินใจเป็นโมเดลของการคำนวณที่ประกอบด้วย 3 ส่วนคือ

1. ต้นไม้ตัดสินใจ ที่ประกอบด้วย ราก (root) โหนดภายใน (internal node) กิ่ง (branch) และโหนดใบ (leaf node) โดยรากและโหนดภายใน ถูกกำหนดด้วยคุณลักษณะ (attribute) กิ่ง ถูกกำหนดด้วยค่าแต่ละค่าของคุณลักษณะ และโหนดใบเป็นกลุ่มของตัวอย่าง ตัวอย่างโครงสร้างต้นไม้ตัดสินใจ ดังแสดงในภาพที่ 2.4
2. อัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจ เกิดจากข้อมูลที่นำมาสอนให้กับโมเดล
3. อัลกอริทึมที่นำต้นไม้ตัดสินใจมาประยุกต์ใช้กับข้อมูลและทำนายผล

ภาพที่ 2.4

ตัวอย่างโครงสร้างของต้นไม้ตัดสินใจ



อัลกอริทึมแบบละโมภ (Greedy Algorithm)

Han และ Kamber (2000) อธิบายว่าอัลกอริทึมที่ใช้ในการสร้างต้นไม้ตัดสินใจจากบนลงล่างแบบวนซ้ำ (recursive) คืออัลกอริทึมแบบละโมภ โดยขั้นตอนในการสร้างต้นไม้ตัดสินใจมีดังนี้

1. สร้างต้นไม้จากโหนดหนึ่งโหนด โดยให้โหนดเป็นกลุ่มของตัวอย่างที่นำมาสอน
2. ถ้าตัวอย่างทุกตัวอย่างอยู่ในกลุ่มเดียวกัน โหนดนั้นจะกลายเป็นโหนดใบ และจะเรียกชื่อตามกลุ่ม
3. กรณีที่ตัวอย่างทุกตัวอย่างไม่อยู่ในกลุ่มเดียวกัน จะใช้การหาค่าด้วยวิธีเอนโทรปี (entropy-based) หรือเรียกอีกอย่างหนึ่งว่า ประโยชน์ของข้อมูล (information gain) ในการเลือกคุณลักษณะ (attribute) ที่เหมาะสมที่สุดในการแบ่งตัวอย่าง โดยใช้กึ่งซึ่งคือค่าแต่ละค่าของคุณลักษณะในการแบ่งตัวอย่างออกเป็นกลุ่มใหม่หรือโหนดใหม่
4. อัลกอริทึมจะวนซ้ำเพื่อสร้างต้นไม้ตัดสินใจต่อไป โดยจะหยุดการสร้างต้นไม้เมื่อเกิดเงื่อนไขข้อหนึ่งข้อใดต่อไปนี้
 - ตัวอย่างทุกตัวของโหนดอยู่ในกลุ่มเดียวกัน
 - ไม่มีคุณลักษณะเหลือสำหรับให้แบ่งต่อไป

การเลือกคุณลักษณะสำหรับการแบ่งกลุ่ม

Han และ Kamber (2000) อธิบายว่าคุณลักษณะ (attribute) ที่ประโยชน์ของข้อมูล (information gain) มีค่ามากที่สุดจะถูกเลือกเพื่อใช้เป็นคุณลักษณะสำหรับแบ่งตัวอย่างออกเป็นกลุ่มสำหรับโหนดปัจจุบัน

ถ้ากำหนดให้ S เป็นเซตของตัวอย่างข้อมูลที่มีจำนวน s ข้อมูล และให้ C_i เป็นเซตของกลุ่มที่มีจำนวน m กลุ่ม โดย $i = 1, \dots, m$ และให้ s_i เป็นจำนวนของตัวอย่างของ S ในกลุ่ม C_i ข้อมูลที่คาดหวังที่ใช้ในการจำแนกตัวอย่างจะหาได้จากสมการ 2.1

$$I(s_1, s_2, \dots, s_m) = - \sum_{i=1}^m p_i \log_2(p_i) \quad (2.1)$$

โดย p_i คือค่าความน่าจะเป็นที่ตัวอย่างจะเป็นสมาชิกของกลุ่ม C_i และประมาณค่าได้จาก s_i/s

ถ้ากำหนดให้คุณลักษณะ A มีค่าที่แตกต่างกัน v ค่า $\{a_1, a_2, \dots, a_v\}$ คุณลักษณะ A สามารถใช้แบ่ง S ออกได้เป็น v สับเซต $\{S_1, S_2, \dots, S_v\}$ โดยที่ S_i คือตัวอย่างใน S ซึ่งมีค่า a_i ของ

คุณลักษณะ A ในกรณีที่ A ถูกเลือกเป็นคุณลักษณะที่เหมาะสมที่สุดในการแบ่งตัวอย่างออกเป็นกลุ่ม สับเซตเหล่านี้จะสอดคล้องกับการที่กิ่งเติบโตมาจากโหนดที่มีเซต S อยู่ในโหนด ถ้ากำหนดให้ s_{ij} เป็นจำนวนตัวอย่างของกลุ่ม C_i ในสับเซต S_j ค่าเอนโทรปีหรือข้อมูลที่คาดหวังบนพื้นฐานของการแบ่งเป็นสับเซตด้วย A จะหาได้จากสมการ 2.2

$$E(A) = \sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{S} I(s_{1j}, \dots, s_{mj}) \quad (2.2)$$

โดย $\sum_{j=1}^v \frac{s_{1j} + \dots + s_{mj}}{S}$ คือค่าน้ำหนักของสับเซตลำดับที่ j และเป็นจำนวนของตัวอย่างในสับเซตหารด้วยจำนวนตัวอย่างทั้งหมดใน S ถ้าค่าเอนโทรปียิ่งมีค่าน้อยเท่าใดความชัดเจนในการแบ่งสับเซตก็ยิ่งมากขึ้น ค่าประโยชน์ของข้อมูลที่จะได้รับการแตกกิ่งของ A หาได้จากสมการ 2.3

$$Gain(A) = I(s_1, s_2, \dots, s_m) - E(A) \quad (2.3)$$

หรืออีกนัยหนึ่ง $Gain(A)$ เป็นการลดลงของเอนโทรปีที่เกิดจากการรู้ค่าของคุณลักษณะ A โดยอัลกอริทึมจะคำนวณค่าประโยชน์ของข้อมูลของแต่ละคุณลักษณะ ซึ่งคุณลักษณะที่มีค่าประโยชน์ของข้อมูลมากที่สุดจะถูกเลือกเป็นคุณลักษณะในการแบ่งกลุ่มตัวอย่าง

ในการจัดกลุ่มของต้นไม้ตัดสินใจ ข้อมูลแต่ละแถวของฐานข้อมูลจะถูกนำมาเปรียบเทียบกับต้นไม้ตัดสินใจตั้งแต่รากไปจนถึงโหนดใบ จำนวนครั้งสูงสุดของการเปรียบเทียบในแต่ละระดับขึ้นอยู่กับปัจจัยการแตกกิ่ง (branching factor) ของระดับนั้น ความซับซ้อนของต้นไม้ตัดสินใจจึงขึ้นอยู่กับผลคูณของจำนวนระดับขั้นและปัจจัยการแตกกิ่งสูงสุด (Dunham, 2003)

ข้อดีของต้นไม้ตัดสินใจ คือเป็นวิธีที่ใช้งานง่ายและมีประสิทธิภาพ สามารถสร้างกฎที่เข้าใจง่ายและแปลความหมายได้ง่าย ต้นไม้ตัดสินใจสามารถนำมาใช้กับฐานข้อมูลขนาดใหญ่ได้ เนื่องจากขนาดของต้นไม้ไม่ขึ้นกับขนาดของฐานข้อมูล ข้อมูลแต่ละแถวในฐานข้อมูลจะถูกกรองตลอดทางของต้นไม้ซึ่งทำให้เวลาที่ต้นไม้ตัดสินใจใช้เป็นสัดส่วนกับความสูงของต้นไม้ซึ่งเป็นเวลาที่แน่นอน นอกจากนี้ ต้นไม้ตัดสินใจสามารถนำมาใช้กับข้อมูลที่มีหลายคุณลักษณะ (Dunham, 2003)

ข้อเสียของต้นไม้ตัดสินใจ คือ การที่ไม่สามารถจัดการกับข้อมูลต่อเนื่องได้ง่าย โดยคุณลักษณะที่เป็นค่าต่อเนื่องจะต้องถูกแบ่งเป็นช่วง การจัดการกับข้อมูลที่หายไปทำได้ยาก เนื่องจากไม่สามารถไปตามกิ่งที่ถูกต้องได้ และเนื่องจากต้นไม้ตัดสินใจสร้างขึ้นมาจากข้อมูลที่นำมาสอน ดังนั้นทำให้เกิดปัญหาเรื่องความพหุพอดีมากเกินไป (overfitting) ซึ่งจะเหมาะกับข้อมูลเพื่อการเรียนรู้ แต่ไม่เหมาะสำหรับการทำนายข้อมูลทำให้เกิดประโยชน์จากการใช้ต้นไม้ตัดสินใจ จึงต้องมีการปรับแต่งโครงสร้างของต้นไม้ ซึ่งสามารถแก้ไขได้โดยเอากิ่งที่ไม่จำเป็นออกบางส่วน (Pruning) หรือการหยุดไม่ให้ต้นไม้โต (Dunham, 2003)

ลักษณะของปัญหาที่เหมาะสมกับต้นไม้ตัดสินใจ

ปกรณ จารุตระกูลชัย (2550) กล่าวว่าโครงสร้างของต้นไม้ตัดสินใจเหมาะสมกับปัญหาที่มีลักษณะดังนี้คือ

1. ข้อมูลมีลักษณะเป็นคุณลักษณะและมีค่าของคุณลักษณะเป็นทางเลือกของการตัดสินใจ ตัวอย่างของคุณลักษณะ ได้แก่ อุณหภูมิ ส่วนตัวอย่างค่าของคุณลักษณะที่สอดคล้องกับคุณลักษณะอุณหภูมิ ได้แก่ ความร้อน ความเย็น ซึ่งเป็นได้ทั้งค่าที่ต่อเนื่องและไม่ต่อเนื่อง
2. มีแอตทริบิวต์ที่เป็นทั้งค่าต่อเนื่องและไม่ต่อเนื่อง
3. ต้นไม้ตัดสินใจเป็นวิธีที่มีความทนทานต่อข้อผิดพลาด (Robustness) ซึ่งถ้าข้อมูลที่ใส่ให้แก่ระบบเพื่อการเรียนรู้มีความผิดพลาด ระบบยังสามารถทำงานต่อไปได้
4. ถึงแม้ข้อมูลบางข้อมูลของคุณลักษณะหรือค่าคุณลักษณะจะมีการหายไป แต่ต้นไม้ตัดสินใจก็ยังสามารถใช้ในการแก้ปัญหาได้

2.5.4 โครงข่ายประสาทเทียม (Artificial Neural Network)

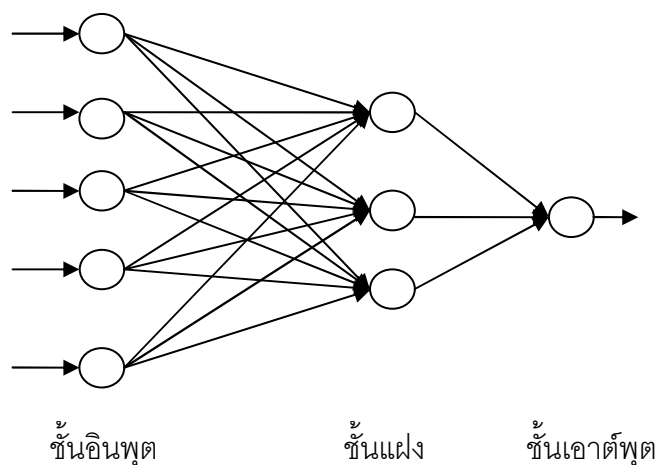
ปกรณ จารุตระกูลชัย (2550) อธิบายว่าโครงข่ายประสาทเทียมเป็นการเลียนแบบการทำงานของโครงข่ายประสาทของมนุษย์ ซึ่งประกอบด้วย เซลล์ประสาทที่แต่ละเซลล์ประกอบด้วยปลายประสาทหรือเดนไดรต์ (Dendrite) เพื่อรับกระแสประสาทหรือข้อมูลอินพุต (Input) แล้วนำมาประมวลผลในนิวรอน จากนั้นจึงส่งข้อมูลเอาต์พุต (Output) ไปยังแกนประสาทหรือแอกซอน (Axon) โดยนิวรอนจะทำงานขนานกัน

โครงข่ายประสาทเทียมหลายชั้นแบบไปข้างหน้า (Multilayer Feed-Forward Neural Network)

Han และ Kamber (2000) อธิบายว่าโครงข่ายประสาทเทียมหลายชั้นประกอบด้วย ชั้นอินพุต (input layer) ชั้นแฝง (hidden layer) และชั้นเอาต์พุต (output) ดังแสดงในภาพที่ 2.5 โดยอินพุตแต่ละค่าหมายถึงคุณลักษณะ (attribute) แต่ละคุณลักษณะของตัวอย่างที่นำมาสอนให้แก่โครงข่ายประสาทเทียม อินพุตทุกค่าจะถูกนำเข้าสู่ระบบพร้อมกันทำให้เกิดเป็นชั้นอินพุต เอาต์พุตที่มีการปรับน้ำหนักแล้วจากชั้นอินพุตจะถูกส่งพร้อมกันไปยังชั้นต่อไปหรือที่เรียกว่าชั้นแฝง จากนั้นเอาต์พุตที่มีการปรับน้ำหนักแล้วจากชั้นแฝงจะถูกส่งไปยังชั้นแฝงชั้นต่อไป ซึ่งจำนวนชั้นแฝงมีได้หลายชั้นแล้วแต่จะกำหนด แต่โดยทั่วไปมักจะใช้หนึ่งชั้น เอาต์พุตที่มีการปรับน้ำหนักแล้วจากชั้นแฝงชั้นสุดท้ายจะถูกส่งไปยังชั้นเอาต์พุตซึ่งจะทำการคำนวณค่าผลลัพธ์ของการทำนายตัวอย่าง โครงข่ายประสาทเทียมที่ไม่มีการส่งน้ำหนักกลับไปยังหน่วยอินพุตหรือหน่วยเอาต์พุตของชั้นที่อยู่ก่อนหน้านั้นจะเรียกว่า โครงข่ายประสาทเทียมแบบไปข้างหน้า (Feed-forward neural networks)

ภาพที่ 2.5

ตัวอย่างโครงสร้างของโครงข่ายประสาทเทียม



อัลกอริทึมแบ็กพรอพาเกชัน (Backpropagation Algorithm)

Han และ Kamber (2000) อธิบายว่าอัลกอริทึมแบ็กพรอพาเกชัน (Backpropagation Algorithm) คือ อัลกอริทึมที่โครงข่ายประสาทเทียมใช้ในการเรียนรู้ การเชื่อมต่อกันของอินพุตและเอาต์พุตของโครงข่ายประสาทเทียมจะมีความสัมพันธ์กับค่าน้ำหนัก (weight) โดยโครงข่ายประสาทเทียมจะเรียนรู้โดยการปรับค่าน้ำหนักเพื่อให้สามารถทำนายกลุ่มของตัวอย่างอินพุตได้ถูกต้อง

ข้อดีของโครงข่ายประสาทเทียมได้แก่ ความทนทาน (tolerance) ต่อข้อมูลที่มีสิ่งปลอมปน (noisy data) และความสามารถในการจำแนกกลุ่มที่ระบบไม่ได้เรียนรู้มาก่อน

ข้อเสียของโครงข่ายประสาทเทียมได้แก่ การใช้ระยะเวลานานสำหรับการเรียนรู้ ต้องมีการปรับค่าพารามิเตอร์เพื่อหาค่าที่เหมาะสมที่สุดสำหรับการทดลองในทุกๆ ครั้ง และค่าน้ำหนักที่ระบบเรียนรู้แปลเป็นความหมายให้เข้าใจได้ยาก

ปรกรณ์ จารุตระกูลชัย (2550) ได้อธิบายถึงขั้นตอนในการปรับค่าน้ำหนัก ดังต่อไปนี้

1. กำหนดค่าอัตราการเรียนรู้ และค่าโมเมนตัม
2. ปรับค่าน้ำหนักของแต่ละเส้นเชื่อมระหว่างโหนดตามสมการ 2.4

$$\Delta w_{ji}(n+1) = \eta \delta(n) \cdot y_j(n) + \alpha \Delta w_{ji}(n) \quad (2.4)$$

โดยที่

x_i คือ ค่าข้อมูลด้านเข้าที่โหนด i

w_i คือ ค่าน้ำหนักที่โหนด i

Δw_{ji} คือ ค่าปรับแก้ค่าน้ำหนักระหว่างโหนด i และ j

η คือ ค่าอัตราการเรียนรู้

α คือ ค่าโมเมนตัม

δ_j คือ ค่าผลต่างระหว่างค่าจริงกับค่าที่ได้จากการคำนวณในรูปของอนุพันธ์ของฟังก์ชันถ่ายโอน (transfer function) ของโหนด j

y_i คือ ค่าผลลัพธ์ของแบบจำลองที่โหนด j

$n, n+1$ คือ ค่าที่แสดงถึงรอบของการปรับแก้ที่ n หรือ $n+1$

2.5.5 นาอีฟเบย์ (Naïve Bayes)

Han และ Kamber (2000) ได้อธิบายว่านาอีฟเบย์ใช้หลักการตามทฤษฎีของเบย์ โดยสมมติว่าผลกระทบของค่าคุณลักษณะค่าหนึ่งต่อกลุ่มที่กำหนดให้เป็นอิสระจากค่าของคุณลักษณะอื่นๆ ข้อสมมตินี้เรียกว่าการเป็นอิสระของกลุ่มแบบมีเงื่อนไข (class conditional independence) และจากการศึกษาการเปรียบเทียบอัลกอริทึมของการจัดกลุ่ม พบว่าการจัดกลุ่มด้วยวิธีเบย์เซียนแบบง่าย (simple baysean classification) หรือนาอีฟเบย์ มีประสิทธิภาพในการจัดกลุ่มที่สามารถเปรียบเทียบได้กับต้นไม้ตัดสินใจและโครงข่ายประสาทเทียม นอกจากนี้ การจัดกลุ่มแบบเบย์เซียนเป็นการจัดกลุ่มที่ให้ความแม่นยำสูงและมีความรวดเร็วมากเมื่อนำไปใช้กับฐานข้อมูลขนาดใหญ่

ตามทฤษฎีของเบย์ ถ้ากำหนดให้ B เป็นตัวอย่างข้อมูลที่ไม่ทราบกลุ่ม และกำหนดให้ A เป็นสมมติฐานที่ B เป็นสมาชิกของกลุ่ม C ในการจัดกลุ่มข้อมูลจะต้องหาค่า $P(A|B)$ หรือค่าความน่าจะเป็นของสมมติฐาน A เมื่อมีเงื่อนไขของตัวอย่างข้อมูล B

กำหนดให้ $P(A)$ คือความน่าจะเป็นเริ่มแรก (prior probability) ของสมมติฐาน A และ $P(A|B)$ คือความน่าจะเป็นภายหลัง (posterior probability) ของสมมติฐาน A เมื่อมีเงื่อนไขของตัวอย่างข้อมูล B ในทำนองเดียวกัน ให้ $P(B)$ คือความน่าจะเป็นเริ่มแรก (prior probability) ของตัวอย่างข้อมูล B และ $P(B|A)$ คือความน่าจะเป็นภายหลัง (posterior probability) ของตัวอย่างข้อมูล B เมื่อมีเงื่อนไขของสมมติฐาน A

$P(A)$ $P(B)$ และ $P(B|A)$ สามารถประมาณค่าได้จากข้อมูลที่ให้มา ซึ่งทฤษฎีของเบย์มีประโยชน์สำหรับการคำนวณหาความน่าจะเป็นภายหลังหรือ $P(A|B)$ เมื่อรู้ค่าของ $P(A)$ $P(B)$ และ $P(B|A)$ โดยทฤษฎีของเบย์แสดงได้ดังสมการ 2.5

$$P(A|B) = [P(B|A) \times P(A)] / P(B) \quad (2.5)$$

ขั้นตอนการทำงานของนาอีฟเบย์

Han และ Kamber (2000) ได้อธิบายถึงขั้นตอนการทำงานของนาอีฟเบย์ดังต่อไปนี้

1. กำหนดให้ตัวอย่างข้อมูลแต่ละค่าเขียนแทนด้วยเวกเตอร์ $X = (x_1, x_2, \dots, x_n)$ โดยที่ n เป็นค่าคุณลักษณะของตัวอย่าง n ลักษณะ
2. สมมติว่ามีกลุ่มจำนวน m กลุ่ม คือ $C_1, C_2, \dots, C_m$ และมีการกำหนดตัวอย่างข้อมูล X ซึ่งไม่ทราบกลุ่มการทำนายว่า X เป็นสมาชิกของกลุ่มใดจะพิจารณาจากค่าความน่าจะเป็น

ภายหลังที่มีค่าสูงสุดเมื่อมีเงื่อนไขของตัวอย่างข้อมูล X นั่นคือนาอ็ฟเบย์จะกำหนดให้ตัวอย่างข้อมูล X เป็นสมาชิกของกลุ่ม C_i ก็ต่อเมื่อ

$$P(C_i|X) > P(C_j|X) \text{ สำหรับ } 1 \leq j \leq m, j \neq i \quad (2.6)$$

ดังนั้น จากทฤษฎีของเบย์

$$P(C_i|X) = [P(X | C_i) \times P(C_i)] / P(X) \quad (2.7)$$

3. เนื่องจาก $P(X)$ เป็นค่าคงที่สำหรับทุกกลุ่ม ดังนั้นมีเพียง $P(X | C_i) \times P(C_i)$ ที่ต้องทำให้มีค่ามากที่สุด แต่ถ้าไม่ทราบค่าความน่าจะเป็นภายหลังของ $P(X | C_i)$ อาจจะสามารถหาค่าความน่าจะเป็นของกลุ่มว่า $P(C_1) = P(C_2) = P(C_m)$ แล้วจึงทำให้ $P(X | C_i)$ มีค่ามากที่สุด

4. ถ้ากำหนดให้ชุดของตัวอย่างข้อมูลมีค่าคุณลักษณะหลายค่า การคำนวณหา $P(X | C_i)$ จะใช้ทรัพยากรทางด้านคอมพิวเตอร์สูงมาก ดังนั้นจึงมีการกำหนดข้อสมมติของการเป็นอิสระของกลุ่มแบบมีเงื่อนไข (class conditional independence) โดยที่

$$P(X | C_i) = \prod_{k=1}^n P(x_k | C_i) \quad (2.8)$$

ค่าความน่าจะเป็นของ $P(x_1 | C_i), P(x_2 | C_i), \dots, P(x_n | C_i)$ ประมาณได้จากตัวอย่างที่นำมาสอน

5.การจัดกลุ่มให้ตัวอย่างข้อมูล X จะคำนวณหาค่า $P(X | C_i) \times P(C_i)$ สำหรับแต่ละกลุ่ม C_i โดยที่ X จะถูกจัดว่าเป็นสมาชิกของกลุ่ม C_i ก็ต่อเมื่อ

$$P(X | C_i) \times P(C_i) > P(X | C_j) \times P(C_j) \text{ สำหรับ } 1 \leq j \leq m, j \neq i \quad (2.9)$$

2.5.6 ตรรกศาสตร์คลุมเครือ (Fuzzy Logic)

Zadeh (1965) กล่าวว่าเซตคลุมเครือเป็นกลุ่มของวัตถุที่มีระดับความเป็นสมาชิกต่อเนื่อง โดยลักษณะของเซตถูกกำหนดด้วยฟังก์ชันของความเป็นสมาชิก ระดับความเป็นสมาชิกของวัตถุในเซตจะมีค่าอยู่ระหว่าง 0 และ 1 และมีการกระทำระหว่างเซต เช่น ยูเนียน อินเตอร์เซกชัน คอมพลิเมนต์ ความสัมพันธ์ และอื่นๆ

เซตทวินัยและเซตคลุมเครือ (Crisp set and Fuzzy set)

Kecman (2001) อธิบายว่าเซตในเอกภพสัมพัทธ์ (universe) U สามารถแสดงได้หลายรูปแบบ ได้แก่

1.แสดงโดยรายการของสมาชิก เช่น

$$S = \{1,2,3,5,6,7,8,9,11\}$$

$$M = \{\text{ช้าง, กวาง, แกะ}\}$$

2.แสดงโดยคำจำกัดความของคุณสมบัติ เช่น

$$S = \{x \in N \mid x < 15\}$$

$$A = \{x \in R \mid \text{"x มีค่าน้อยกว่า 10"}\}$$

3.แสดงโดยฟังก์ชันของการเป็นสมาชิก เช่น

$$\mu_s(x) = 1 \quad \text{ถ้า } x < 5$$

$$\mu_s(x) = 0 \quad \text{ถ้า } x > 5$$

สำหรับเซตคลุมเครือ $\mu_s(x)$ ถ้าเอกภพสัมพัทธ์ U มีสมาชิกคือ $[0,1]$ ระดับของการเป็นสมาชิกของสมาชิก x บางค่าในเอกภพสัมพัทธ์ U สามารถเป็นค่าใดๆที่ $0 \leq \mu_s(x) \leq 1$

สุจินต์ เจริญขจรชัย (2543) อธิบายว่าเซตคลุมเครือสามารถแสดงค่าการเป็นสมาชิกของเซตในรูปของ (Member Value: μ) สำหรับเซตทวินัย กรณีที่เป็นสมาชิกจะกำหนดค่า $\mu = 1$ และกรณีที่ไม่เป็นสมาชิกจะกำหนดค่า $\mu = 0$ ตัวอย่างการเป็นสมาชิกของเซตทวินัยของปริมาณน้ำฝนที่ตกลงมาดังแสดงในภาพที่ 2.6 โดยที่

$$\mu_{\text{ฝนตก}} = 0 \quad \text{ถ้ามีปริมาณน้ำฝนน้อยกว่า 10 มม.}$$

$$\mu_{\text{ฝนตก}} = 1 \quad \text{ถ้ามีปริมาณน้ำฝนมากกว่า 10 มม.}$$

คำว่าฝนตกเป็นตัวแปรเชิงคำพูด (linguistic variable) ที่อธิบายคุณลักษณะของสมาชิกในเซตนั้น สำหรับเซตคลุมเครือ จะอ้างอิงค่าความเป็นสมาชิกได้หลายค่า ซึ่งจากเซตทวินัยของปริมาณน้ำฝนดังตัวอย่างด้านบนจะสามารถเขียนในรูปเซตคลุมเครือได้ดังแสดงในภาพที่ 2.7 โดยที่

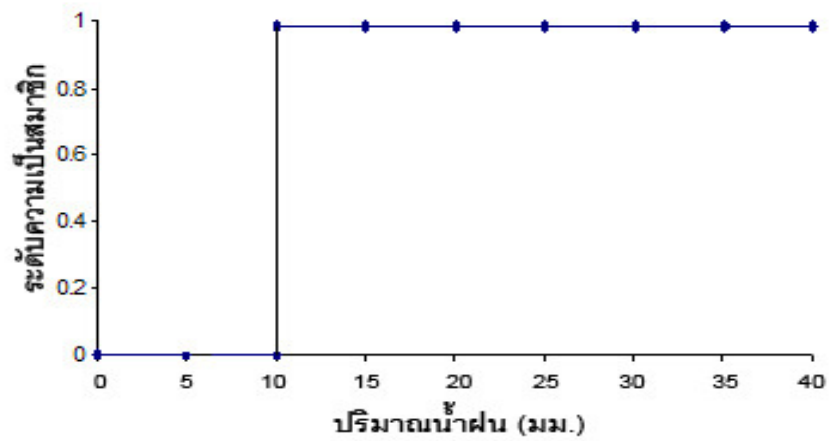
$$\mu_{\text{ฝนตก}} = 0 \quad \text{ถ้า } h < 5 \text{ มม.}$$

$$\mu_{\text{ฝนตก}} = (h-5)/5 \quad \text{ถ้า } 5 \leq h < 10 \text{ มม.}$$

$$\mu_{\text{ฝนตก}} = 1 \quad \text{ถ้า } h \geq 10 \text{ มม.}$$

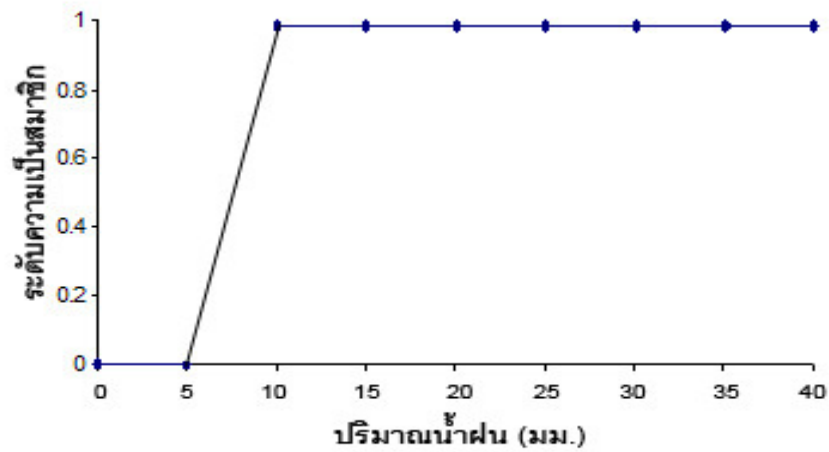
ภาพที่ 2.6

ตัวอย่างความเป็นสมาชิกของเซตทวินัย



ภาพที่ 2.7

ตัวอย่างความเป็นสมาชิกของเซตคลุมเครือ



ภาพที่ 2.7 จะเห็นว่าค่าของความเป็นสมาชิกจะเปลี่ยนแปลงได้ด้วยการกำหนดความคลุมเครือ เช่น ปริมาณฝนตกที่ 9 มม. สามารถเป็นสมาชิกของเซตคลุมเครือของฝนตกด้วยค่าความเป็นสมาชิก $\mu = 0.8$

ระบบคลุมเครืออัจฉริยะ (Fuzzy Expert System)

Klir และ Yuan (1995) อธิบายว่า ระบบคลุมเครืออัจฉริยะหรือการควบคุมด้วยความคลุมเครือ (fuzzy control) มีองค์ประกอบสำคัญของระบบ 3 ส่วน คือ พื้นความรู้ (knowledge base) ฐานข้อมูล (database) และ กลไกในการอนุมาน (inference engine)

การออกแบบระบบคลุมเครืออัจฉริยะ

โดยทั่วไประบบคลุมเครืออัจฉริยะประกอบด้วย 4 โมดูล คือ พื้นฐานของกฎความคลุมเครือ (fuzzy rule base) กลไกในการอนุมานแบบคลุมเครือ (fuzzy inference engine) การแปลงให้มีความคลุมเครือ (fuzzification) และการแปลงความคลุมเครือกลับ (defuzzification) โดยการออกแบบระบบคลุมเครืออัจฉริยะ มี 5 ขั้นตอน ดังนี้คือ

1. กำหนดตัวแปรอินพุต เอาท์พุต และช่วงของค่าอินพุตและเอาท์พุต ให้กับระบบ จากนั้นเลือกค่าที่มีความหมายเหมาะสมให้กับตัวแปรโดยอธิบายความหมายของค่าด้วยเซตคลุมเครือที่เหมาะสม

2. กำหนดฟังก์ชันคลุมเครือ (fuzzification function) ให้กับตัวแปรอินพุตเพื่อแสดงความไม่คงที่ของการวัด วัตถุประสงค์ของการกำหนดฟังก์ชันคลุมเครือ คือ เพื่อแปลความหมายการวัดค่าของอินพุต โดยฟังก์ชันคลุมเครือจะแสดงด้วยจำนวนจริง

3. นำความรู้ที่เกี่ยวข้องกับการแก้ปัญหา มาสร้างเป็นกฎอนุมานคลุมเครือ (fuzzy inference rules) ซึ่งทำได้ 2 ทาง คือ อาศัยประสบการณ์ของผู้เชี่ยวชาญ หรือ ข้อมูลจากการทดลองด้วยวิธีการเรียนรู้ที่เหมาะสม เช่น จากโครงข่ายประสาทเทียม เป็นต้น

4. ออกแบบกลไกในการอนุมาน โดยนำตัวแปรอินพุตมารวมกับกฎอนุมานคลุมเครืออย่างเหมาะสมเพื่ออนุมานตัวแปรเอาท์พุต

5. เลือกวิธีการแปลงความคลุมเครือกลับที่เหมาะสม เพื่อแปลงค่าที่ได้จากกลไกในการอนุมานซึ่งอยู่ในรูปของเซตคลุมเครือให้เป็นค่าของจำนวนจริงเพียงค่าเดียว

2.5.7 เทคนิคการจัดกลุ่มในการประเมินความน่าเชื่อถือของเว็บไซต์

มีงานวิจัยหลายสาขาที่นำเทคนิคการจัดกลุ่มมาใช้ ยกตัวอย่างเช่น การเปรียบเทียบประสิทธิภาพระหว่างโครงข่ายเบย์เซียนและโครงข่ายประสาทเทียมในการประเมินระยะเวลาตอบสนองสำหรับระบบการให้บริการเครือข่าย (Zhang & Bivens, 2007) การใช้โครงข่ายประสาทเทียมและโมเดลของเบย์เซียนในการทำนายสารคดีโฆษณาในสารประกอบเคมี (Bahler, Stone,

Wellington & Bristol, 2000) การใช้นาฬิกาชีพและต้นไม้ตัดสินใจในการตรวจจับการบุกรุกเครือข่าย (Amor, Benferhat & Elouedi 2004) และการจำแนกแหล่งกำเนิดของดินด้วยต้นไม้ตัดสินใจและโมเดลของเบย์เซียน (Zhou, Zhang & Wang, 2004)

สำหรับงานวิจัยที่เกี่ยวข้องกับเทคนิคการจัดกลุ่ม ได้แก่ งานวิจัยของ Ju, Liang และ Lui (2008) ที่ได้นำวิธีโครงข่ายประสาทเทียมมาประเมินประสิทธิภาพเว็บไซต์ของรัฐบาล (e-government) โดยให้ผู้เชี่ยวชาญประเมินความน่าเชื่อถือของเว็บไซต์จากองค์ประกอบ 54 องค์ประกอบ ของเว็บไซต์ 60 เว็บไซต์ โดยรายการขององค์ประกอบของเว็บไซต์ดังแสดงในตารางที่ 2.3 เมื่อนำองค์ประกอบของเว็บไซต์ 54 องค์ประกอบ มาจัดเป็นกลุ่มจะได้ 11 กลุ่มขององค์ประกอบ ดังแสดงในตารางที่ 2.4 ข้อมูลที่ผู้เชี่ยวชาญประเมินเว็บไซต์จำนวน 40 เว็บไซต์ ถูกนำมาเป็นตัวอย่างสอนให้กับโครงข่ายประสาทเทียม จากนั้นให้โครงข่ายประสาทเทียมทำการประเมินเว็บไซต์ 20 เว็บไซต์ที่เหลือ ผลการวิจัยพบว่า โครงข่ายประสาทเทียมสามารถประเมินระดับความน่าเชื่อถือของเว็บไซต์ได้ไม่แตกต่างจากที่ผู้เชี่ยวชาญประเมิน โดยเปอร์เซ็นต์ความแตกต่างของการระดับความน่าเชื่อถือของเว็บไซต์ที่ประเมินโดยผู้เชี่ยวชาญ และที่ประเมินโดยโครงข่ายประสาทเทียมดังแสดงในตารางที่ 2.5

ตารางที่ 2.3

องค์ประกอบของเว็บไซต์ 54 องค์ประกอบ เพื่อใช้สำหรับการประเมิน
ระดับความน่าเชื่อถือของเว็บไซต์ของรัฐบาล

	องค์ประกอบของเว็บไซต์
1	ความครบถ้วนของข้อมูลองค์กร (integrity of organization information)
2	ประสิทธิผลของข้อมูล (effectiveness of information)
3	ความเป็นวิทยาศาสตร์ของแคตาล็อก (science of categorical catalogue)
4	มีข้อมูลสำคัญครบถ้วน (available of important information)
5	ความหลากหลายของภาษา (diversity of language)
6	ความครบถ้วนของหน่วยปฏิบัติการ (integrity of operational unit)
7	ความสมบูรณ์ของข้อมูลปฏิบัติการ (operational information maturity)
8	ระดับความละเอียดของวิธีการติดต่อ (detailed degree of the way of contact)
9	มีข้อมูลการใช้งานอีเมลของประชาชนเพียงชุดเดียว (single logging of citizen email)

ตารางที่ 2.3 (ต่อ)

องค์ประกอบของเว็บไซต์ 54 องค์ประกอบ เพื่อใช้สำหรับการประเมิน
ระดับความน่าเชื่อถือของเว็บไซต์ของรัฐบาล

	องค์ประกอบของเว็บไซต์
10	คอลัมน์ของการปรึกษาและร้องเรียน (column of consultation and appeal)
11	โทรศัพท์และกล่องจดหมายสำหรับร้องเรียน (telephone and mail box for appealing)
12	ให้คำแนะนำ (making suggestions)
13	ข้อมูลเป็นแบบใช้งานทั่วไป (information in common use)
14	ความหลากหลายของวิธีการค้นหา (diversity of search ways)
15	ฟังก์ชันของระบบนำร่องเว็บไซต์ (website's navigation function)
16	ความซับซ้อนของฟอร์มข้อมูล (form complicity of information)
17	สาธารณะประโยชน์ของการบริการข้อมูล (information service of public utilities)
18	บริการที่มีลักษณะเฉพาะ (characteristic service)
19	สัดส่วนโครงข่ายของหน่วยงานที่จัดการ (networking ratio of handling affairs)
20	อัตราการตอบสนองต่อข้อร้องเรียนของประชาชน (response rate to citizen's appeals)
21	ปัญหาเกี่ยวกับข้อมูลของหน่วยงาน (query of affairs information)
22	ระดับของการให้บริการ ณ จุดบริการเดียว (degree of one-stop service)
23	มีช่องทางการสื่อสารหลายช่องทาง (multi-outlets communications)
24	ระยะเวลาในการตอบกลับข้อมูล (feedback time of information)
25	บรรทัดฐานและความสัมพันธ์ของข้อมูล (norm and affinity of information)
26	บริการข้อมูลที่ยืดหยุ่น (customized information service)
27	ข้อมูลที่เป็นมิตรภาพ (amity information)
28	การรักษาความเป็นส่วนตัวและความปลอดภัย (privacy and safeties hint)
29	บริการสำหรับชุมชนพิเศษ (service of special community)
30	ความสะดวกของบริการ (convenience of the service)
31	ให้ความช่วยเหลือออนไลน์ (on-line supervision)

ตารางที่ 2.3 (ต่อ)

องค์ประกอบของเว็บไซต์ 54 องค์ประกอบ เพื่อใช้สำหรับการประเมิน
ระดับความน่าเชื่อถือของเว็บไซต์ของรัฐบาล

	องค์ประกอบของเว็บไซต์
32	ให้คำปรึกษาก่อนออกนโยบาย (consultation before the policy-making)
33	การอภิปรายเกี่ยวกับหัวข้อพิเศษ (discussion of special subject)
34	การสำรวจความคิดเห็นของสาธารณะ (opinion poll of the public)
35	คุณภาพในการอ่าน (reading quality)
36	ลำดับในเครื่องมือค้นหา (Ranks in search engine)
37	ลำดับในกลุ่มของเว็บไซต์เดียวกัน (rank in the same class website)
38	ดัชนีของลำดับหน้า (index of page rank)
39	การใช้งานได้หลากหลายของเว็บไซต์ (universal use of website)
40	ข้อมูลช่วยเหลือของเว็บไซต์ (help information of the website)
41	การควบคุมการสัมภาษณ์ผู้บริโภค (control of customer interview)
42	การกรองข้อมูลที่มีความอ่อนไหว (filtration of sensitive information)
43	ความปลอดภัยของข้อมูล (data safety)
44	ความคิดสร้างสรรค์ของข้อมูล (data instauration)
45	การตรวจติดตามความปลอดภัยของโครงข่าย (monitor of network safety)
46	บรรทัดฐานของชื่อโดเมน (norm of domain name)
47	ความเร็วในการลงทะเบียนเว็บไซต์ (register speed of the website)
48	ภาพเคลื่อนไหวของหน้าเว็บ (visual effect of web page)
49	ฟังก์ชันของลิงก์ของเว็บไซต์ (function of website link)
50	ลิงก์ของเว็บไซต์ใช้งานได้ (validity of website link)
51	การนับจำนวนผู้เข้าชมเว็บไซต์ (website counter)
52	ข้อมูลของผู้ซ่อมบำรุง (information of maintenance supervisor)
53	เลข ไอ ซี พี (ICP number)
54	การประเมินโดยสาธารณะ (public evaluation)

ตารางที่ 2.4

องค์ประกอบของเว็บไซต์ 11 กลุ่ม จากการจัดกลุ่มองค์ประกอบของเว็บไซต์ 54 องค์ประกอบ

	กลุ่มขององค์ประกอบ	องค์ประกอบที่
1	บริการพื้นฐานทางข้อมูล (basic service of information)	1 - 5
2	ขบวนการพื้นฐานทางธุรกิจ (basic processing of business)	6 - 7
3	การปฏิสัมพันธ์พื้นฐาน (basic interaction)	8 - 13
4	การพัฒนาการบริการทางข้อมูล (developmental service of information)	14 - 18
5	การพัฒนาขบวนการทางธุรกิจ (developmental processing of business)	19 - 22
6	การพัฒนาการทางปฏิสัมพันธ์ (developmental interaction)	23 - 25
7	ความสมบูรณ์ของการบริการทางข้อมูล (perfect service information)	26 - 28
8	ความสมบูรณ์ของการบริการทางธุรกิจ (perfect service of business)	29 - 31
9	ความสมบูรณ์ของปฏิสัมพันธ์ (perfect interaction)	32 - 40
10	ดัชนีจตุรัสที่สามของเทคนิค (the third square technique index)	41 - 53
11	แบบสอบถามสาธารณะ (public questionnaire)	54

ตารางที่ 2.5

การเปรียบเทียบคะแนนความน่าเชื่อถือของเว็บไซต์จากการประเมินโดยผู้เชี่ยวชาญ
และจากการทำนายโดยโครงข่ายประสาทเทียม

เว็บไซต์ที่	คะแนนจาก ผู้เชี่ยวชาญ	คะแนนจาก การทำนาย	เปอร์เซ็นต์ ความแตกต่าง
1	7.9241	7.8998	0.31
2	7.6385	7.4066	3.04
3	8.1904	7.7725	5.10
4	8.4547	8.3924	0.74
5	7.7822	7.5882	2.49
6	6.9453	6.7236	3.19
7	7.4124	6.9859	5.75
8	7.3508	6.8808	6.39
9	7.6805	7.7912	1.44
10	8.3231	8.448	1.50
11	8.1103	8.2756	2.04
12	8.1936	8.1340	0.73
13	7.9283	8.0341	1.33
14	7.8975	7.6305	3.38
15	8.1747	8.1278	0.57
16	7.4299	7.0152	5.58
17	7.7827	8.1394	4.58
18	7.3907	7.4162	0.35
19	8.0904	8.0471	0.54
20	6.9752	7.1280	2.19

Nefti, Meziane และ Kasiran (2005) ได้นำวิธีการทางตรรกศาสตร์คลุมเครือมาสร้างเป็นโมเดลสำหรับการประเมินความเชื่อถือของเว็บไซต์พาณิชย์อิเล็กทรอนิกส์ โดยประเมินความน่าเชื่อถือของเว็บไซต์จากองค์ประกอบของเว็บไซต์ 4 กลุ่ม คือ

1.การมีอยู่ (existence) ประกอบด้วย การมีอยู่ทางกายภาพ (physical existence) ได้แก่ เบอร์โทรศัพท์ เบอร์แฟกซ์ และที่อยู่ของผู้ค้า การให้มีการลงทะเบียน (mandatory registration) และการมีอยู่ของบุคคล (peoples' existence)

2.การได้รับการรับรอง (affiliation) ประกอบด้วย การรับรองจากบุคคลที่สาม (third party endorsement) การเป็นสมาชิก (membership) และการเป็นศูนย์รวม (portal)

3.นโยบาย (policy) ประกอบด้วย นโยบายความพึงพอใจของลูกค้า (customer satisfaction policy) ประโยคที่แจ้งความเป็นส่วนตัว (privacy statement) และนโยบายการรับประกัน (warranty policy)

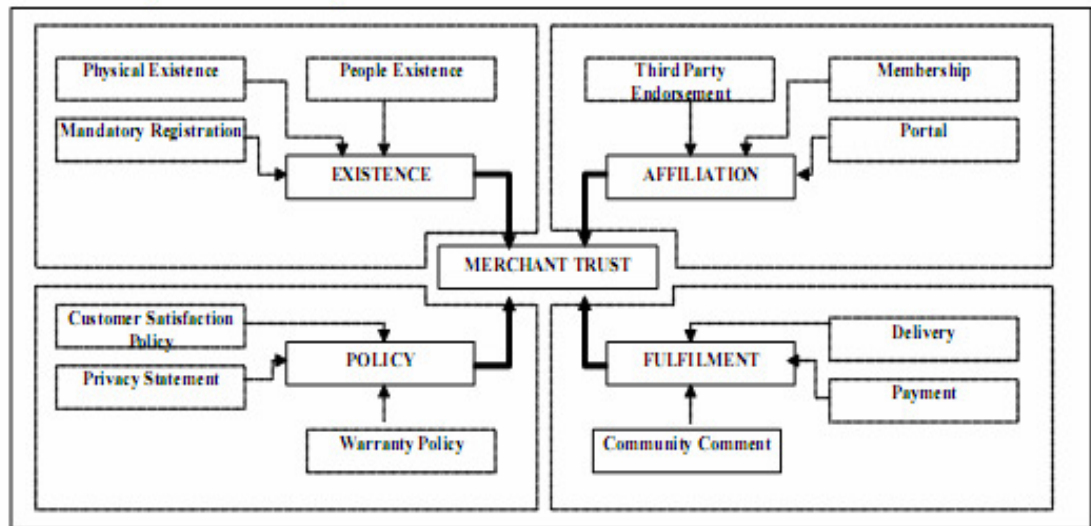
4.การบรรลุวัตถุประสงค์ (fulfillment) ประกอบด้วย การขนส่ง (delivery) การชำระเงิน (payment) และ คำแนะนำของชุมชน (community comment)

การประเมินระดับความน่าเชื่อถือของเว็บไซต์ในงานวิจัยนี้ทำการแบ่งการทดลองออกเป็น 5 โมดูล โดยที่ 4 โมดูลแรกเป็นการประเมินระดับความน่าเชื่อถือของเว็บไซต์จากองค์ประกอบแต่ละกลุ่ม สำหรับโมดูลที่ 5 เป็นโมดูลการตัดสินใจที่เกิดจากการนำผลของ 4 โมดูลแรกมาประเมิน โมเดลของการประเมินความเชื่อถือของเว็บไซต์ด้วยวิธีการทางตรรกศาสตร์คลุมเครือดังแสดงในภาพที่ 2.8

การแปลงข้อมูลอินพุตให้เป็นข้อมูลคลุมเครือ (fuzzification) ในงานวิจัยนี้ ใช้ฟังก์ชันการเป็นสมาชิก แบบกซ์เซียน โดยกำหนดให้ฟังก์ชันการเป็นสมาชิกของ 4 โมดูลแรก มี 2 ฟังก์ชัน และให้ชื่อว่า น้อย กับ มาก โดยนำฟังก์ชันมาใช้กับค่าตัวแปรที่เกี่ยวข้องกับแต่ละองค์ประกอบและเอาท์พุตของแต่ละโมดูล สำหรับโมดูลของการตัดสินใจ กำหนดให้ฟังก์ชันการเป็นสมาชิกมี 3 ฟังก์ชัน โดยให้ชื่อว่า ต่ำ ปานกลาง และสูง ซึ่งเอาท์พุตของโมดูลที่ 5 หรือโมดูลของการตัดสินใจแสดงถึงระดับความน่าเชื่อถือของเว็บไซต์ โมเดลของการประเมินระดับความน่าเชื่อถือของเว็บไซต์ด้วยวิธีตรรกศาสตร์คลุมเครือจากงานวิจัยนี้ สามารถนำมาประเมินเว็บไซต์จริงได้ โดยได้ทดลองนำมาประเมินเว็บไซต์ 2 เว็บไซต์ ที่มีอยู่จริง คือ <http://denimjunkies.com> และ <http://www.meshcomputers.com> ซึ่งเป็นเว็บไซต์ขายเสื้อผ้าออนไลน์ประเภทยีนส์ และเว็บไซต์ขายคอมพิวเตอร์และอุปกรณ์ ตามลำดับ

ภาพที่ 2.8

โมเดลที่ออกแบบมาสำหรับการประเมินระดับความน่าเชื่อถือของเว็บไซต์
ด้วยตรรกศาสตร์คลุมเครือ



ที่มา: “A Fuzzy Trust Model for E-Commerce,” โดย Nefti, S., Meziane, F., และ Kasiran, K, 2005, Proceeding of the Seventh IEEE International Conference on E-Commerce Technology (CEC 2005).

การศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้อง มีวัตถุประสงค์เพื่อให้เข้าใจองค์ประกอบสำคัญของเว็บไซต์ที่มีผลกระทบต่อความน่าเชื่อถือของเว็บไซต์ และเข้าใจหลักการทำงานของเทคนิคการเรียนรู้ของเครื่อง 4 เทคนิค คือ เทคนิคต้นไม้ตัดสินใจ โครงข่ายประสาทเทียม นาอ์ฟเบย์ และตรรกศาสตร์คลุมเครือ เพื่อให้สามารถเปรียบเทียบประสิทธิภาพของแต่ละเทคนิคในการประเมินระดับความน่าเชื่อถือของเว็บไซต์ โดยใช้การประเมินระดับความน่าเชื่อถือของเว็บไซต์ด้วยคนเป็นเกณฑ์ในการเปรียบเทียบได้ ในบทถัดไปจะกล่าวถึง การพิจารณาเลือกองค์ประกอบสำคัญของเว็บไซต์ และวิธีการนำเทคนิคการเรียนรู้ของเครื่องทั้ง 4 เทคนิค รวมทั้งความรู้จากทฤษฎีและงานวิจัยที่เกี่ยวข้องมาทำการดำเนินงานวิจัย เพื่อให้วิทยานิพนธ์ฉบับนี้บรรลุตามวัตถุประสงค์ที่ได้กำหนดไว้