

บทที่ 2

แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้อง

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องของการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013 เพื่อใช้สำหรับเผยแพร่ข้อมูลบนเว็บไซต์ และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ข้อมูล ดังนี้

2.1 แนวคิด

2.2 ทฤษฎี

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ในการดำเนินงานเรื่องการวิเคราะห์ข้อมูลรายงานการฆาตกรรมปี 2011-2013 และการแสดงผลข้อมูลบนเว็บไซต์ ผู้วิเคราะห์ข้อมูลได้ศึกษาหลักการและทฤษฎีต่าง ๆ องค์ประกอบหนึ่งที่สำคัญคือการวิเคราะห์ข้อมูล ซึ่งมีรายละเอียด ดังนี้

2.1.1.1 การวิเคราะห์ข้อมูล เป็นการจัดการข้อมูลด้วยวิธีต่าง ๆ เช่น การคำนวณ การนำเสนอข้อมูล เป็นต้น เพื่อให้ได้ผลลัพธ์ตามวัตถุประสงค์ การวิเคราะห์ เป็นการแยกแยะสิ่งที่จะพิจารณาออกเป็นส่วนย่อยที่มีความสัมพันธ์กัน เพื่อทำความเข้าใจแต่ละส่วนให้แจ่มแจ้ง รวมทั้งการสืบค้นความสัมพันธ์ของส่วนต่าง ๆ เพื่อดูว่าส่วนประกอบปลีกย่อยนั้นสามารถเข้ากันได้หรือไม่ สัมพันธ์เกี่ยวเนื่องกันอย่างไร ซึ่งจะช่วยให้เกิดความเข้าใจต่อสิ่งหนึ่งสิ่งใดอย่างแท้จริง โดยพื้นฐานแล้ว การวิเคราะห์ข้อมูล (Data Analysis) จำแนกได้ 2 แบบ ดังนี้

- การวิเคราะห์ข้อมูลเบื้องต้น ใช้สถิติเชิงพรรณนา (Descriptive Statistics) คือสถิติที่ใช้เพื่ออธิบาย บรรยาย หรือสรุป ลักษณะของกลุ่มข้อมูลที่เป็นตัวเลข ที่เก็บรวบรวม

มาซึ่งไม่สามารถอ้างอิงลักษณะประชากรได้ (ยกเว้นมีการเก็บข้อมูลของประชากรทั้งหมด) ตัวอย่างสถิติเชิงพรรณนา เช่น การแจกแจงความถี่ การวัดค่ากลางของข้อมูล การวัดการกระจายของข้อมูล

- การวิเคราะห์ข้อมูลขั้นสูง ใช้สถิติเชิงอนุมาน (Inference Statistics) คือ สถิติที่ใช้ในการวิเคราะห์ข้อมูลกลุ่มตัวอย่างที่เป็นตัวแทนของประชากรซึ่งสามารถนำผลการวิเคราะห์นั้นไปสรุปอ้างอิงถึงประชากรได้โดยใช้ทฤษฎีความน่าจะเป็น สถิติเชิงอนุมาน ประกอบด้วย การประมาณค่า และการทดสอบสมมติฐาน

2.1.2 แนวคิดเกี่ยวกับเหมืองข้อมูล (Data Mining)

กระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยก ได้ดังนี้

- กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง
- การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือการนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการสังเกตการณ์ที่ทันสมัย
 - การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล
 - การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

ซึ่งเหตุผลที่ต้องมีการทบทวนเหมืองข้อมูลเพราะข้อมูลที่ถูกรวบรวมไว้ในฐานข้อมูลหากเก็บไว้เฉย ๆ ก็จะไม่เกิดประโยชน์ดังนั้นจึงต้องมีการสกัดสารสนเทศหรือการคัดเลือกข้อมูลออกมาใช้งานส่วนที่เราต้องการในอดีตได้ใช้คนเป็นผู้สืบค้นข้อมูลต่าง ๆ ในฐานข้อมูลซึ่งผู้สืบค้นจะทำการสร้างเงื่อนไขขึ้นมาตามภูมิปัญญาของผู้สืบค้น ในปัจจุบันการวิเคราะห์ข้อมูลจากฐานข้อมูล

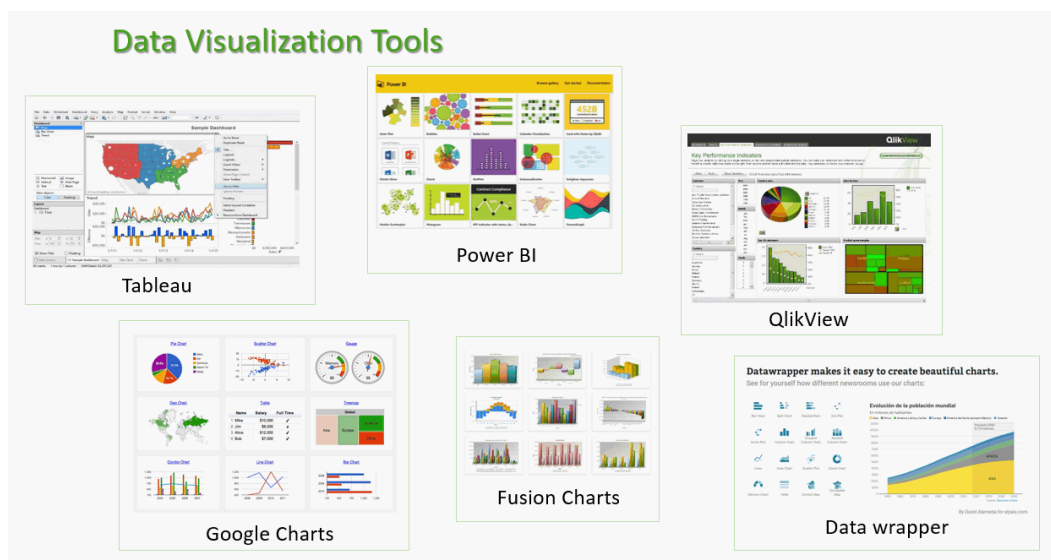
เดี๋ยวจริงๆไม่ให้ความรู้เพียงพอและลึกซึ้งซึ่งสำหรับการดำเนินงานภายใต้ภาวะที่มีการแข่งขันสูง และมีการเปลี่ยนแปลงที่รวดเร็วจึงจำเป็นที่จะต้องรวบรวมฐานข้อมูลหลาย ๆ ฐานข้อมูลเข้าด้วยกัน เรียกว่า “ คลังข้อมูล” (Data Warehouse) ดังนั้น จึงจำเป็นต้องใช้ Data Mining ในการดึงข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่ เพื่อนำข้อมูลนั้นมาใช้งานให้เกิดประโยชน์สูงสุด

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

Data Visualization หรือ Information Visualization คือ การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมีประสิทธิภาพ ซึ่งคำว่า “ประสิทธิภาพ” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้ว อาจทำให้เราไม่สามารถค้นพบรายละเอียดของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์เชื่อมโยงได้

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพในความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณ ให้สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

2.1.3.1 เครื่องมือที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts, Data wrapper และอื่นๆอีกมากมาย



ภาพที่ 2.1 เครื่องมือที่ใช้ทำ Data Visualization

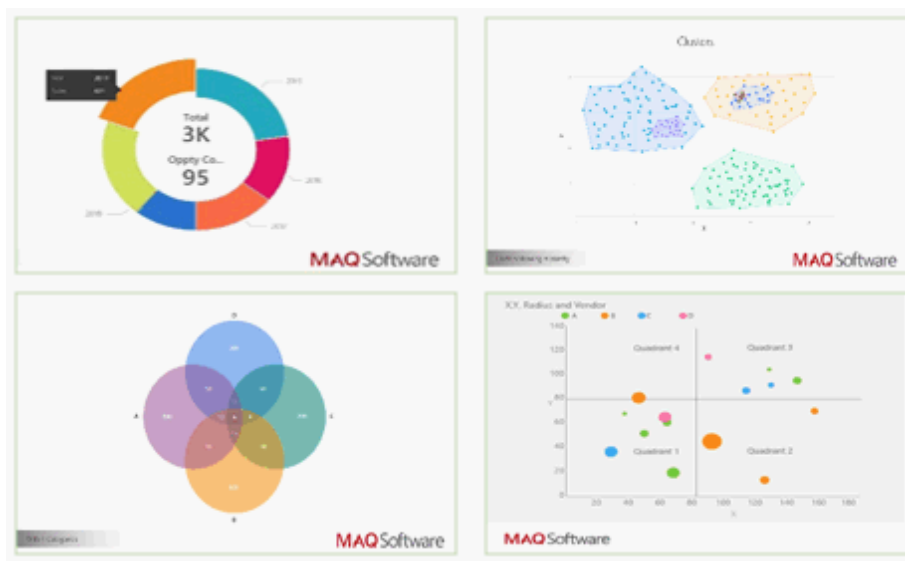
2.1.3.2 รูปแบบในการใช้ Data Visualization จำแนกได้ ดังนี้

2.1.3.2.1 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending) ใช้กราฟที่แสดงผลแบบทิศทางหรือแนวโน้ม เพื่อนำเสนอข้อมูลให้เห็นจำนวนข้อมูลที่เกิดขึ้นในแต่ละช่วงเวลา (period) รวมถึงเน้นข้อมูลที่ต้องการนำเสนอ เช่น Line Chart, Bar Chart, Radar Chart, Area Chart เป็นต้น



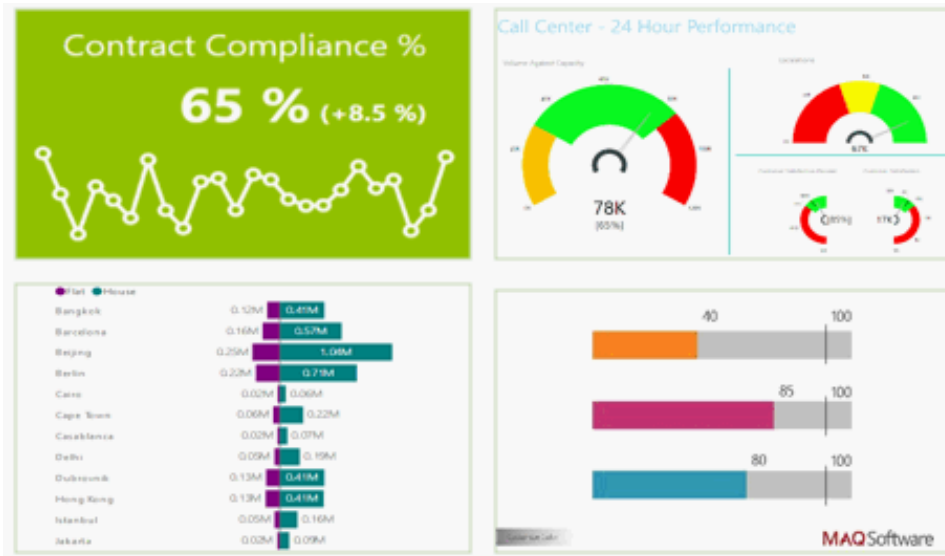
ภาพที่ 2.2 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

2.1.3.2.2 การนำเสนอแบบกลุ่มข้อมูล (Classification) เป็นการนำเสนอโดยนำข้อมูลมาจัดเป็นกลุ่มๆ เช่น Donut Chart, Ring Chart, Pie Chart,



ภาพที่ 2.3 การนำเสนอแบบกลุ่มข้อมูล (Classification)

2.1.3.2.3 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison) เหมาะสำหรับการนำเสนอที่ต้องการเปรียบเทียบผลการดำเนินงาน เช่นเทียบกับปีที่แล้ว (YoY) เปรียบเทียบกับเป้าที่ตั้งไว้ (Target) ซึ่งกราฟที่เหมาะสมและมักนำมาใช้ เช่น KPI Indicator, Bullet Chart, Power BI Card with state เป็นต้น



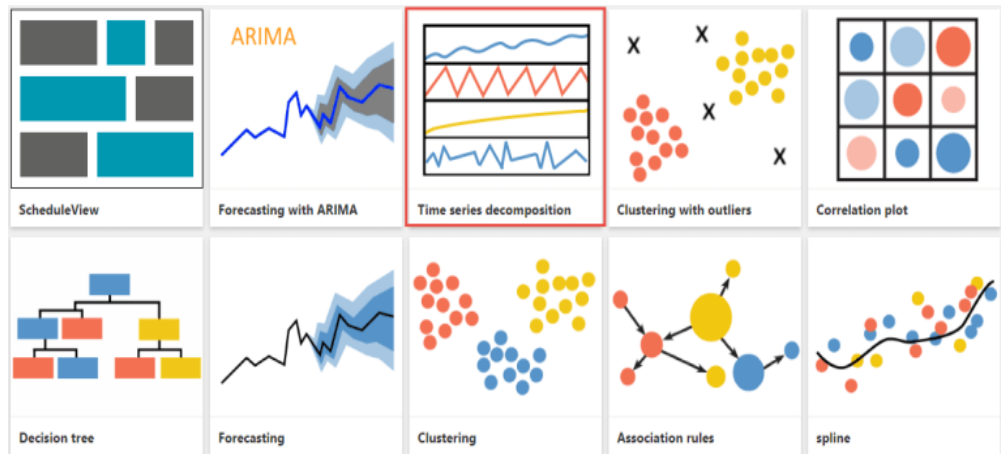
ภาพที่ 2.4 การนำเสนอเชิงเปรียบเทียบข้อมูล (Comparison)

2.1.3.2.4 การนำเสนอรูปแบบแผนที่ (Geographical) เหมาะสำหรับการนำเสนอข้อมูลบนแผนที่ โดยสามารถที่จะนำยอดขาย, รายได้, ความหนาแน่นของประชากร เพื่อ Focus กลุ่มผู้คนในแต่ละพื้นที่ เช่น Globe Map, Google Map, Flow Map เป็นต้น



ภาพที่ 2.5 การนำเสนอรูปแบบแผนที่ (Geographical)

2.1.3.2.5 กลุ่มที่ต้องพยากรณ์ล่วงหน้าและการวิเคราะห์ข้อมูลที่ซับซ้อน (Analytics) เราสามารถที่จะใช้ภาษา R หรือ Python ดึงข้อมูลในอดีตมาเพื่อวิเคราะห์เพื่อพยากรณ์อนาคต และนำเสนอข้อมูลที่ได้ในรูปของกราฟ เช่น Association Rules, Clustering, Forecasting Time series, Calculation plot เป็นต้น



ภาพที่ 2.6 กลุ่มที่ต้องพยากรณ์ล่วงหน้า

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

การพัฒนาอย่างรวดเร็วของเทคโนโลยีทำให้การเก็บข้อมูลที่มีปริมาณมหาศาลเป็นเรื่องที่เป็นไปได้ในปัจจุบัน ข้อมูลรูปแบบใหม่ที่จัดเก็บในปัจจุบันไม่เพียงแต่มีความใหญ่ด้านปริมาณเท่านั้น แต่มีความซับซ้อนและรูปแบบที่หลากหลายมากขึ้น เช่น เสียง ภาพนิ่ง ภาพเคลื่อนไหว เป็นต้น ซึ่งก็คือการใช้ประโยชน์จากเทคโนโลยีของอุปกรณ์ตรวจจับที่มีขนาดเล็กลง แต่มีประสิทธิภาพมากขึ้น ลักษณะข้อมูลขนาดใหญ่หรือ Big Data นี้จึงถูกให้คำนิยามว่ามีลักษณะ “The Four V’s” (IBM, Online, 2017) ซึ่งประกอบไปด้วย

- 1) ปริมาณ (Volume) การจัดเก็บข้อมูลในปัจจุบัน นอกจากจะมีความหลากหลายแล้วความถี่ในการจัดเก็บก็ยิ่งเพิ่มขึ้นด้วย ข้อมูลบางประเภทที่เคยจัดเก็บรายเดือน อาจถูกเปลี่ยนเป็นจัดเก็บรายวัน หรือในบางกรณี เช่น ข้อมูลตำแหน่งของยานพาหนะในกรณีบริษัทขนส่งอาจจัดเก็บในรูปแบบ Real time เหล่านี้ล้วนทำให้ปริมาณข้อมูลเพิ่มขึ้นเป็นอย่างมาก

2) ความหลากหลาย (Variety) นอกเหนือจากข้อมูลที่ถูกจัดเก็บประจำวัน ระหว่างกระบวนการทำงานปกติแล้ว ปัจจุบันองค์กรต่าง ๆ ยังเริ่มจัดเก็บข้อมูลที่มีรูปแบบ หลากหลายมากขึ้น เช่น ข้อมูลจาก Sensor ต่าง ๆ , ลักษณะ Posts บน Social network, รูปภาพหรือไฟล์วิดีโอ, ระยะเวลาการใช้งานบนเว็บไซต์ เป็นต้น

3) ความเร็ว (Velocity) ความเร็วในที่นี้หมายถึงความเร็วของข้อมูลที่ถูกผลิต ตัวอย่างที่เห็นได้ชัด ได้แก่ ข้อมูลราคาซื้อขายหุ้นของตลาดหลักทรัพย์ที่ราคาของหลักทรัพย์มีการแสดงผลจากธุรกรรมซื้อขายที่เกิดขึ้นในอัตราที่รวดเร็วเป็นอย่างมาก ทำให้ ความต้องการใช้ข้อมูลในลักษณะ Realtime เพื่อวิเคราะห์และตอบสนองต่อเหตุการณ์ที่ เกิดขึ้นในตลาดหลักทรัพย์เพิ่มสูงขึ้นอย่างมาก

4) ความไม่แน่นอน (Veracity) นิยามความไม่แน่นอนถูกกำหนดให้เป็น หนึ่งในคำนิยาม Big Data เนื่องจากข้อมูลล้นมีที่มาจากหลากหลายแหล่งไม่ว่าจะมาจาก ภายในหรือภายนอกองค์กร ข้อมูลภายในองค์กรหมายถึง ข้อมูลที่มาจากการประมวลผล ภายในหน่วยงาน ยกตัวอย่าง เช่น สถาบันการเงินจะมีข้อมูลจำนวนบัญชีเงินฝากจากสาขา บัญชีเงินกู้จากฝ่ายสินเชื่อ ประวัติการเดินบัญชี ส่วนแหล่งข้อมูลภายนอกองค์กร เช่น สภาพ การจราจร สภาพอากาศ ปริมาณนักท่องเที่ยว ด้วยสาเหตุนี้ข้อมูล Big Data จึงมีความไม่แน่นอนจากระดับความซับซ้อนและความหลากหลายของข้อมูล

2.2.1.2 เทคโนโลยีการรวบรวมและจัดเก็บข้อมูล (Data Collection and Data Storage)

เนื่องจากลักษณะข้อมูลของ Big Data มีลักษณะเฉพาะตัว การพัฒนา เทคโนโลยี จัดเก็บข้อมูลเพื่อให้เท่าทันกับลักษณะที่มีความหลากหลายและปริมาณข้อมูลที่มากขึ้น รวมทั้ง การคำนึงถึงความถูกต้องและปลอดภัยของข้อมูลจึงมีความสำคัญเป็นอย่างมาก การเก็บ ข้อมูลในอดีต (Traditional Database) มักจะถูกเก็บอยู่ใน Rational Database ซึ่งคือการจัดเก็บ ข้อมูลในรูปแบบของตารางซึ่งแต่ละตารางสามารถเชื่อมโยง ความสัมพันธ์ระหว่างกันได้ อย่างไรก็ตามโครงสร้างข้อมูล Rational Database จะต้องเป็นข้อมูล ที่กำหนดรูปแบบ (Normalization) ก่อนการจัดเก็บหรือที่เรียกว่า Schema on write ในการจัดการ ระบบ ฐานข้อมูลรายธุรกรรม (Database Transaction) เพื่อให้ชุดกระบวนการฐานข้อมูลมี ความ น่าเชื่อถือ (Data Integrity) และหลีกเลี่ยงข้อผิดพลาดที่จะเกิดขึ้นกับฐานข้อมูล จึงมี การ

กำหนดคุณลักษณะของข้อมูลที่จัดเก็บบน Rational Database 4 ประการภายใต้ชื่อ “ACID” โดยมีรายละเอียดดังนี้(Suphakit Annopornchai, Online, 2560)

1) Atomicity ความถูกต้องของฐานข้อมูลภายใต้แนวคิด “All or Nothing” นั่นคือ แต่ละธุรกรรมจะต้องถูกดำเนินการอย่างครบถ้วนสมบูรณ์ มิฉะนั้นแล้วธุรกรรมนั้น จะไม่ถูกดำเนินการใด ๆ เลย เช่น ธุรกรรมการถอนเงินหรือโอนเงิน

2) Consistency ข้อมูลที่จะเข้าสู่ฐานข้อมูลได้จะต้องมีความสอดคล้อง ถูกต้อง และสมเหตุสมผลกันเท่านั้น

3) Isolation ข้อมูลเป็นอิสระต่อกัน กล่าวคือ เมื่อมีการประมวลผลข้อมูลใดข้อมูลหนึ่ง ผลจากการประมวลผลนั้นจะไม่ส่งผลกระทบต่อข้อมูลอื่น ถึงแม้ว่าการประมวลผลข้อมูล นั้นจะผิดพลาดหรือหยุดชะงักก็ตาม

4) Durability คือ ข้อมูลแต่ละธุรกรรมจะต้องยังคงอยู่ครบถ้วน ถึงแม้จะเป็นเหตุให้ระบบใช้การไม่ได้ก็ตาม

ต่อมาเมื่อก้าวเข้าสู่ยุคข้อมูลขนาดใหญ่ (Big Data) ที่ข้อมูลเกิดขึ้นอย่างรวดเร็ว มีปริมาณมหาศาลและมีลักษณะเป็น Unstructured หรือ Semi-structure ซึ่งข้อมูลเหล่านี้ Relational Database ไม่สามารถรองรับได้ดีเพียงพอ (สมเกียรติ ปุ้ยสูงเนิน, ออนไลน์, 2557) เนื่องจากข้อมูลลักษณะ Unstructured หรือ Semi-structure ในทางปฏิบัติไม่สามารถกำหนดรูปแบบก่อนการจัดเก็บได้ ทำให้ Non-Relational Database เข้ามามีบทบาทมากขึ้นเนื่องจาก Non-Relational Database มีจุดเด่นด้านโครงสร้างการจัดเก็บข้อมูลที่ยืดหยุ่นเมื่อเปรียบเทียบกับ Relational Database โดยไม่จำเป็นต้องกำหนดรูปแบบข้อมูลก่อนการจัดเก็บและสามารถจัดเก็บ ข้อมูลประเภท รูปภาพ ภาพเคลื่อนไหว ไฟล์เสียงได้ กระบวนการกำหนดรูปแบบจะดำเนินการเมื่อ ผู้ใช้งานเรียกข้อมูลกลับมาใช้งานอีกครั้ง (Schema on read) และมีจุดเด่นเรื่อง Scalability และ Availability โดยฐานข้อมูลประเภทดังกล่าวมีคุณสมบัติเรียกโดยย่อว่า “BASE” (Charles Roe, 2012) มีรายละเอียดดังนี้

1) Basic Availability ระบบฐานข้อมูลพร้อมใช้งานและต้องมี การตอบสนองต่อคำร้องขอตลอดเวลาเพื่อให้เป็นไปตามทฤษฎี CAP : ซึ่งได้แก่ Consistency ความสมบูรณ์ของระบบ Availability ความพร้อมใช้งานของข้อมูล และ Partition Tolerance ความทนทานต่อการล้มเหลวของระบบ

2) Soft state สถานะของระบบสามารถเปลี่ยนแปลงได้ตลอดเวลา แม้จะ
ไม่มีการใส่ข้อมูล (Input)

3) Eventual consistency เมื่อมีการเพิ่มข้อมูลลงในระบบ ระบบ จะทยอย
ปรับปรุงข้อมูล Node ในเครือข่ายทั้งหมด ดังนั้น จะมีช่วงระยะเวลาสั้น ๆ ที่สถานะ ข้อมูลไม่
สอดคล้องกัน อย่างไรก็ตาม เมื่อผ่านไปสักกระยะหนึ่งข้อมูลที่ถูเก็บในระบบจะถูกปรับปรุง ให้
ตรงกัน

ระบบการเก็บข้อมูลและประมวลผลแบบ Non-Relational Database ปัจจุบันเป็นที่นิยม
สำหรับการเก็บข้อมูลในลักษณะ Unstructured หรือ Semi-structure รวมกับ การใช้บริการจาก
ระบบประมวลผลบน Public service ที่มีความสามารถในการประมวลผลค่อนข้างสูง และมี
ต้นทุนที่ต่ำอย่าง Hadoop หรือ Service บน Public cloud เป็นต้น

2.2.1.3 เทคโนโลยีจัดเก็บข้อมูลแบบกระจายศูนย์ (Distributed ledger
technology)

Distributed ledger technology (DLT) เป็นแนวคิดเทคโนโลยีใหม่ที่ ทำการกระจาย
ฐานข้อมูลให้ผู้ใช้งานทุกคนในเครือข่ายสามารถถือข้อมูลที่เหมือนกันหมดทั้งระบบโดย ไม่
จำเป็นต้องมีศูนย์กลางในการให้สิทธิประมวลผลหรือเก็บข้อมูล โดยบัญชีธุรกรรม
อิเล็กทรอนิกส์ หรือ Ledger จะถูกเก็บไว้ในเครือข่ายเซิร์ฟเวอร์ที่เรียกว่า Node ข้อมูลบัญชีนี้จะ
ถูกคัดลอกเก็บไว้ใน ทุก ๆ Node ของเครือข่าย โดยที่ไม่ระบุตัวตนว่าเป็นของใคร ทำให้ทุกคน
ในเครือข่ายสามารถเห็น 17 รายการเดินบัญชีการเงินของผู้ใช้รายอื่น ๆ ได้หากเกิดธุรกรรมใด
ขึ้น ข้อมูลการทำธุรกรรมนั้นจะถูก ส่งไปถึงทุกคนในเครือข่ายเพื่อช่วยยืนยันความถูกต้องก่อน
เพิ่มข้อมูลเข้าไปในระบบฐานข้อมูล

Blockchain เป็นเทคโนโลยีที่มีแนวคิดต้นแบบมาจาก DLT โดย เทคโนโลยีดังกล่าวช่วย
ให้ข้อมูลธุรกรรมต่าง ๆ จะถูกบันทึกในฐานข้อมูลเพียงชุดเดียวและกระจาย ฐานข้อมูลหรือ
บัญชี(Ledger) เหล่านั้นไปตามเครื่องคอมพิวเตอร์ต่าง ๆ โดยไม่ผ่านตัวกลางใด ๆ ทั้งสิ้น
ผู้ใช้งานเครือข่าย Blockchain ในแต่ละเครื่องจะเห็นข้อมูลทั้งหมดเป็นชุดเดียวกัน (Single
version of truth) ดังนั้นการปลอมแปลงข้อมูลภายใต้เทคโนโลยีนี้จึงเกิดขึ้นได้ยากเนื่องจากการ
บันทึก ตรวจสอบ และรับรองธุรกรรมต่าง ๆ ที่เกิดขึ้นในระบบบัญชีจะต้องผ่านกระบวนการ
ตรวจสอบและกระบวนการรับรองความถูกต้องของธุรกรรมข้อมูลจากผู้ใช้งานเครือข่ายทุก

เครื่องที่มี บัญชีบันทึกข้อมูลธุรกรรมเหล่านั้น จุดเด่นสำคัญที่ทำให้Blockchain แตกต่างจากระบบฐานข้อมูล ทั่วไปโดยสามารถสรุปได้ดังนี้(Business Insider, Online, 2016)

- 1) Shared Publicly : ทุก Node ในเครือข่าย Blockchain จะเห็นบัญชีธุรกรรมอิเล็กทรอนิกส์ที่เกิดขึ้นและสามารถตรวจสอบได้ว่าธุรกรรมนั้นถูกจัดเก็บใน Block ใด
- 2) Decentralized : ไม่มีศูนย์กลางในการจัดเก็บข้อมูล รวมถึงไม่มีหน่วยงานกลางในการกำหนดกฎเกณฑ์หรือสิทธิในการเข้าถึงข้อมูล
- 3) Secure : ข้อมูลที่จัดเก็บในระบบจะไม่สามารถเปลี่ยนแปลงแก้ไข หรือเอาค่าคืนกลับมาได้
- 4) Trusted : เมื่อมีธุรกรรมเกิดขึ้น ข้อมูลจะต้องได้รับการยอมรับจาก ผู้ใช้งานส่วนใหญ่ก่อนถูกบันทึกเข้าในระบบ ทำให้ระบบมีความน่าเชื่อถือ
- 5) Automated : การจัดเก็บธุรกรรมข้อมูลอยู่ในรูปแบบอัตโนมัติ โดย Software ที่ถูกเขียนขึ้นทำหน้าที่ป้องกันการเขียนข้อมูลที่ซ้ำซ้อนหรือขัดแย้งกัน

เทคโนโลยีBlockchain ได้รับการจับตามองอย่างมากทั้งจากภาครัฐ และภาคธุรกิจด้วยจุดเด่นในเรื่องของความโปร่งใส ความสามารถในการตรวจสอบความถูกต้องข้อมูล และโอกาสในการปลอมแปลงข้อมูลที่น้อยมาก ทำให้องค์กรทั่วโลกต่างตื่นตัวและพยายามศึกษาเพื่อนำเทคโนโลยีBlockchain ไปประยุกต์ใช้ให้เกิดประโยชน์เช่น รัฐบาลประเทศเอสโตเนียใช้Blockchain เพื่อเก็บข้อมูลของประชาชนในประเทศ โดยสามารถป้องกันการปลอมแปลงตัวตน และ ทำให้ประชาชนสามารถเข้าถึงข้อมูลได้อย่างปลอดภัย (จรัล งามวิโรจน์เจริญ, ออนไลน์, 2561)

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เนื่องด้วยปัจจุบันเป็นยุคที่ข้อมูลสารและสนเทศมีความสำคัญ การเผยแพร่และสื่อสารข้อมูล ข่าวสาร ที่ตรงกับความต้องการของผู้ใช้จึงเป็นสิ่งจำเป็น การประยุกต์เทคโนโลยีสารสนเทศเพื่อช่วยในการสื่อสารข้อมูลจำนวนมากให้แก่ผู้ใช้ เช่น การให้บริการเว็บไซต์เพื่อเผยแพร่ข้อมูลข่าวสารและแลกเปลี่ยนความรู้ จึงเป็นเครื่องมือที่สำคัญในการสื่อสารข้อมูลถึงผู้ใช้งานจำนวนมาก ดังนั้นการศึกษาเกี่ยวกับพฤติกรรมของผู้ใช้บริการเว็บไซต์ จะช่วยให้องค์กรสามารถนำข้อมูลมาใช้ในการวางแผนพัฒนาเว็บไซต์ ให้ตรงกับความต้องการใช้งานหรือใช้ในการวางแผนกลยุทธ์ เพื่อสร้างความได้เปรียบทางการแข่งขัน

2.2.2.1 การทำเหมืองข้อมูล (Data Mining)

การทำเหมืองข้อมูล (Data Mining) คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและความสัมพันธ์ ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไปประยุกต์ใช้ในงานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และการแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

การทำเหมืองข้อมูลเปรียบเสมือนวิวัฒนาการหนึ่งในการจัดเก็บและตีความหมาย ข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่าย ๆ มาสู่การจัดเก็บในรูปแบบข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล

2.2.2.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

- Data Cleaning เป็นขั้นตอนสำหรับการคัดข้อมูลที่ไม่เกี่ยวข้องออกไป
- Data Integration เป็นขั้นตอนการรวมข้อมูลที่มีหลายแหล่งให้เป็นข้อมูลชุดเดียวกัน
- Data Selection เป็นขั้นตอนการดึงข้อมูลสำหรับการวิเคราะห์จากแหล่งที่บันทึกไว้
- Data Transformation เป็นขั้นตอนการแปลงข้อมูลให้เหมาะสมสำหรับการใช้งาน
- Data Mining เป็นขั้นตอนการค้นหารูปแบบที่เป็นประโยชน์จากข้อมูลที่มีอยู่
- Pattern Evaluation เป็นขั้นตอนการประเมินรูปแบบที่ได้จากการทำเหมืองข้อมูล
- Knowledge Representation เป็นขั้นตอนการนำเสนอความรู้ที่ค้นพบโดยใช้เทคนิคในการนำเสนอเพื่อให้เข้าใจ

2.2.2.3 ประเภทข้อมูลที่ใช้ทำเหมืองข้อมูล

- Relational Database เป็นฐานข้อมูลที่จัดเก็บอยู่ในรูปแบบของตาราง โดยในแต่ละตารางจะประกอบไปด้วยแถวและคอลัมน์ ความสัมพันธ์ของข้อมูลทั้งหมดสามารถแสดงได้โดย Entity Relationship Model

- Data Warehouses เป็นการเก็บรวบรวมข้อมูลจากหลายแหล่งมาเก็บไว้ในรูปแบบเดียวกันและรวบรวมไว้ในที่ ๆ เดียวกัน
- Transactional Database ประกอบด้วยข้อมูลที่แต่ละทรานแซกชันแทนด้วยเหตุการณ์ในขณะใดขณะหนึ่ง เช่น ใบเสร็จรับเงิน จะเก็บข้อมูลในรูปแบบชื้อลูกค้าและรายการสินค้าที่ลูกค้ารายชื้อ
- Advanced Database เป็นฐานข้อมูลที่จัดเก็บในรูปแบบอื่น ๆ เช่น ข้อมูลแบบ Object-Oriented ข้อมูลที่เป็น Text File ข้อมูลมัลติมีเดีย ข้อมูลในรูปแบบของ Web

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3.1 หลักการออกแบบเว็บไซต์

ในการออกแบบเว็บไซต์นั้นประกอบด้วยกระบวนการต่าง ๆ มากมาย เช่น การออกแบบ โครงสร้าง ลักษณะหน้าตาหรือการเขียนโปรแกรม แต่มีหลายคนที่พัฒนาเว็บไซต์โดยขาดการ วางแผนและทำงานไม่เป็นระบบ ตัวอย่าง เช่น การลงมือออกแบบโดยการใช้โปรแกรมช่วยสร้างเว็บ เนื้อหาและรูปแบบก็เป็นไปตามที่นึกขึ้นได้ขณะนั้น

การออกแบบเว็บไซต์อย่างถูกต้องจะช่วยลดความผิดพลาดเหล่านี้ และช่วยลดความเสี่ยงที่จะ ทำให้เว็บประสบความล้มเหลว การออกแบบเว็บไซต์ที่ดีต้องอาศัยการออกแบบและจัดระบบข้อมูล อย่างเหมาะสม

- องค์ประกอบของการออกแบบเว็บไซต์การออกแบบเว็บไซต์ที่มีประสิทธิภาพนั้นต้องคำนึงถึง องค์ประกอบสำคัญดังต่อไปนี้

1) ความเรียบง่าย (Simplicity) หมายถึงการจำกัดองค์ประกอบเสริมให้เหลือเฉพาะ องค์ประกอบหลัก กล่าวคือในการสื่อสารเนื้อหากับผู้ใช้นั้น เราต้องเลือกเสนอสิ่งที่เรา ต้องการนำเสนอจริง ๆ ออกมาในส่วนของกราฟิก สี สัน ตัวอักษรและภาพเคลื่อนไหว ต้องเลือกให้พอเหมาะ ถ้าหากมีมากเกินไปจะรบกวนสายตาและสร้างความรำคาญต่อ ผู้ใช้ตัวอย่างเว็บไซต์ที่ได้รับการออกแบบที่ดี ได้แก่เว็บไซต์ของบริษัทใหญ่ ๆ อย่างเช่น Apple Adobe Microsoft หรือ Nokia ที่มีการออกแบบเว็บไซต์ในรูปแบบที่เรียบง่ายไม่ซับซ้อนและใช้งานอย่างสะดวก

2) ความสม่ำเสมอ (Consistency) หมายถึง การสร้างความสม่ำเสมอให้เกิดขึ้นตลอดทั้งเว็บไซต์โดยอาจเลือกใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ก็ได้ เพราะถ้าหากว่าแต่ละ หน้าในเว็บไซตนั้นมี ความแตกต่างกันมากเกินไป อาจทำให้ผู้ใช้เกิดความสับสนและ ไม่แน่ใจว่ากำลังอยู่ในเว็บไซต์เดิมหรือไม่ เพราะฉะนั้นการออกแบบเว็บไซต์ใน

แต่ละ หน้าควรที่จะมีรูปแบบ สไตล์ของกราฟิก ระบบเนวิเกชัน (Navigation) และโทนสีที่มีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ (Identity) ในการออกแบบเว็บไซต์ต้องคำนึงถึงลักษณะขององค์กร เป็นหลัก เนื่องจากเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กร การเลือกใช้ 10 ตัวอักษร ชุดสี รูปภาพหรือกราฟิก จะมีผลต่อรูปแบบของเว็บไซต์เป็นอย่างมาก ตัวอย่างเช่น ถ้าเราต้องออกแบบเว็บไซต์ของธนาคารแต่เรากลับเลือกสีส้มและกราฟิก มากมาย อาจทำให้ผู้ใช้คิดว่าเป็นเว็บไซต์ของสวนสนุกซึ่งส่งผลต่อความเชื่อถือขององค์กรได้

4) เนื้อหา (Useful Content) ถือเป็นสิ่งสำคัญที่สุดในเว็บไซต์ เนื้อหาในเว็บไซต์ต้องสมบูรณ์ และได้รับการปรับปรุงพัฒนาให้ทันสมัยอยู่เสมอ ผู้พัฒนาต้องเตรียมข้อมูลและเนื้อหาที่ ผู้ใช้ต้องการให้ถูกต้องและสมบูรณ์ เนื้อหาที่สำคัญที่สุดคือเนื้อหาที่ทีมผู้พัฒนา สร้างสรรค์ขึ้นมาเอง และไม่ไปซ้ำกับเว็บอื่น เพราะจะถือเป็นสิ่งที่ดึงดูดผู้ใช้ให้เข้ามา เว็บไซต์ได้เสมอ แต่ถ้าเป็นเว็บที่ดึงข้อมูลจากเว็บอื่น ๆ มาเมื่อใดก็ตามที่ผู้ใช้ทราบว่าข้อมูลนั้นมาจากเว็บใด ผู้ใช้ก็ไม่จำเป็นต้องกลับมาใช้งานลิงค์เหล่านั้นอีก

5) ระบบเนวิเกชัน (User-Friendly Navigation) เป็นส่วนประกอบที่มีความสำคัญต่อเว็บไซต์ มาก เพราะจะช่วยไม่ให้เกิดความสับสนระหว่างดูเว็บไซต์ ระบบเนวิเกชันจึง เปรียบเสมือนป้ายบอกทาง ดังนั้นการออกแบบเนวิเกชัน จึงควรให้เข้าใจง่าย ใช้งานได้ สะดวก ถ้ามีการใช้กราฟิก ก็ควรสื่อความหมาย ตำแหน่งของการวางเนวิเกชันก็ควรวางให้สม่ำเสมอ เช่น อยู่ตำแหน่งบนสุดของทุกหน้า เป็นต้น ซึ่งถ้าจะให้ดีเมื่อมีเนวิเกชันที่เป็นกราฟิกก็ควรเพิ่มระบบเนวิเกชันที่เป็นตัวอักษรไว้ส่วนล่างด้วยเพื่อช่วยอำนวยความสะดวกให้กับผู้ใช้ที่ยกเลิกการแสดงผลภาพกราฟิกบนเว็บเบราว์เซอร์

6) คุณภาพของสิ่งที่ปรากฏให้เห็นในเว็บไซต์ (Visual Appeal) ลักษณะที่น่าสนใจของ เว็บไซต์นั้นขึ้นอยู่กับความชอบส่วนบุคคลเป็นสำคัญ แต่โดยรวมแล้วก็สามารถสรุปได้ ว่าเว็บไซต์ที่น่าสนใจนั้นส่วนประกอบต่าง ๆ ควรมีคุณภาพ เช่น กราฟิกควรสมบูรณ์ไม่มีรอยหรือขอบขรุขระให้เห็นชนิดตัวอักษรอ่านง่ายสบายตามีการเลือกใช้โทนสีที่เข้ากันอย่างสวยงาม เป็นต้น

7) ความสะดวกของการใช้ในสภาพต่าง ๆ (Compatibility) การใช้งานของเว็บไซต์นั้นไม่ ควรมีข้อจำกัดกล่าวคือต้องสามารถใช้งานได้ดีในสภาพแวดล้อมที่หลากหลายไม่มีการบังคับให้ผู้ใช้ต้องติดตั้งโปรแกรมอื่นใดเพิ่มเติม นอกเหนือจากเว็บเบราว์เซอร์ ควรเป็น เว็บที่แสดงผลได้ดีในทุกระบบปฏิบัติการ สามารถแสดงผลได้ในทุกความ

ละเอียดหน้าจอบ ซึ่งหากเป็นเว็บไซต์ที่มีผู้ให้บริการมากและกลุ่มเป้าหมายหลากหลายควรให้ความสำคัญกับเรื่องนี้ให้มาก

8) ความคงที่ในการออกแบบ (Design Stability) ถ้าต้องการให้ผู้ใช้งานรู้สึกว่าเป็นเว็บไซต์ที่มี คุณภาพถูกต้องและเชื่อถือได้ควรให้ความสำคัญกับการออกแบบเว็บไซต์เป็นอย่างมาก ต้องออกแบบวางแผนและเรียบเรียงเนื้อหาอย่างรอบคอบ ถ้าเว็บไซต์จัดทำขึ้นอย่างลวก ๆ ไม่มีมาตรฐานการออกแบบและระบบการจัดการข้อมูลถ้ามีปัญหาเกิดขึ้นอาจส่งผลให้เกิดปัญหาและทำให้ผู้ใช้หมดความเชื่อถือ

9) ความคงที่ของการทำงาน (Function Stability)ระบบการทำงานต่าง ๆ ในเว็บไซต์ควรมี ความถูกต้องแน่นอน ซึ่งต้องได้รับการออกแบบสร้างสรรค์และตรวจสอบอยู่เสมอ ตัวอย่างเช่น ลิงค์ต่าง ๆ ในเว็บไซต์ ต้องตรวจสอบว่ายังสามารถลิงค์ข้อมูลได้ถูกต้อง หรือไม่ เพราะเว็บไซต์อื่นอาจมีการเปลี่ยนแปลงได้ตลอดเวลา ปัญหาที่เกิดจากลิงค์ก็คือ ลิงค์ขาด ซึ่งพบได้บ่อยเป็นปัญหาที่สร้างความรำคาญกับผู้ใช้เป็นอย่างมาก

- กำหนดเป้าหมายของเว็บไซต์ ขั้นตอนแรกของการออกแบบเว็บไซต์คือการกำหนดเป้าหมายของเว็บไซต์ให้แน่ชัดเสียก่อน เพื่อจะได้ออกแบบการใช้งานได้ตรงกับเป้าหมายที่ได้ตั้งเอาไว้ โดยทั่วไปมักจะเข้าใจว่าการทำเว็บไซต์มีจุดมุ่งหมายเพื่อบริการข้อมูลของหน่วยงานหรือองค์กรเท่านั้น แต่ในความเป็นจริงแล้ว เว็บไซต์แต่ละแห่งก็จะมีเป้าหมายของตนเองแตกต่างกันออกไป

- กำหนดกลุ่มผู้ใช้เป้าหมาย ผู้ออกแบบเว็บไซต์จำเป็นต้องทราบกลุ่มผู้ใช้เป้าหมายที่เข้ามาใช้บริการเว็บไซต์เพื่อที่จะได้ตอบสนองความต้องการของผู้ใช้ได้อย่างชัดเจน ตัวอย่างเช่นเว็บไซต์ที่มีกลุ่มผู้ใช้หลากหลาย เช่น เซิร์ชเอ็นจิน เว็บไซต์ และเว็บไดเรกทอรี แต่เว็บไซต์ส่วนใหญ่จะตอบสนองความต้องการเฉพาะ กลุ่มเท่านั้น ไม่สำหรับทุกคน เพราะคุณไม่สามารถตอบสนองความต้องการของคนที่หลากหลายได้ ในเว็บไซต์เดียว

- สิ่งที่ผู้ใช้ต้องการจากเว็บหลังจากที่ได้เป้าหมายและกลุ่มเป้าหมายของเว็บไซต์แล้ว ลำดับต่อไปคือการออกแบบเว็บไซต์เพื่อดึงดูดผู้ใช้งานให้นานที่สุดด้วยการสร้างสิ่งที่น่าสนใจเพื่อดึงดูดผู้ใช้โดยทั่วไปแล้ว สิ่งที่ผู้ใช้ คาดหวังจากการเข้าชมเว็บไซต์หนึ่ง

- ข้อมูลหลักที่ควรมีอยู่ในเว็บไซต์เมื่อเราทราบถึงความต้องการที่ผู้ใช้ต้องการได้รับเมื่อเข้าชมเว็บไซต์หนึ่ง ๆ แล้วก็ออกแบบเว็บไซต์ให้มีข้อมูลที่ผู้ใช้ต้องการ

2.2.4 ทฤษฎีเกี่ยวกับการ visualization

Visualization เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้ในรูปแบบของภาพ โดยผู้ใช้

สามารถ เรียนรู้และจดจำข้อมูลผ่านการมองเห็นได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวได้ว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการ นำภาพมาใช้กับการนำเสนอหรือนำมาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการ ตัดสินใจ

ข้อดีของ Visualization มีดังนี้

- ช่วยในการแสดงข้อมูลที่มีปริมาณสูง
- ช่วยในการรับรู้หรือคาดคะเนสิ่งที้อาจจะเกิดขึ้นในอนาคต
- Visualization ไม่เพียงแต่แสดงรายละเอียดข้อมูลในตัวเองเท่านั้นยังสามารถแสดงปัญหา ที่เกิดขึ้นได้ด้วย

สามารถแสดงปัญหา ที่เกิดขึ้นได้ด้วย

- ช่วยให้ผู้ใช้สามารถเข้าใจหรือวิเคราะห์ข้อมูลได้สะดวกขึ้นไม่ว่าขอบเขตข้อมูลนั้นจะมี ขนาดใหญ่หรือเล็ก

- ช่วยให้ผู้ใช้สามารถสันนิษฐานข้อมูลได้สะดวกขึ้น (Ware, 2004)

ซึ่ง Visualization เป็นมากกว่า วิธีการทาง Computer Visualization เป็นการนำเสนอข้อมูล แบบหนึ่งที่ทำให้การแสดงข้อมูลให้อยู่ในรูปแบบของ VisualForm ซึ่งอาจจะเป็นการแสดงในรูปแบบ ของรูปภาพ ,กราฟ หรือ แผนภาพ ซึ่ง ผลลัพธ์ของการทำ Visualization คือ การนำข้อมูลต่าง ๆ ที่ ซ่อนอยู่ในตัวของข้อมูลเองออกมาให้ผู้ใช้สามารถสัมผัสได้ ซึ่งกระบวนการต่าง ๆ ในการแสดงหรือ นำข้อมูลต่าง ๆ ออกมา อาจจะถูกแอบซ่อนอยู่หรือผู้ใช้ไม่ทันสังเกต แต่อย่างไรก็ตาม Visualization เป็นสิ่งที่จำเป็นในการค้นหาข้อมูล หรือ ใช้ในการวิเคราะห์ข้อมูล โดยจุดมุ่งหมายของ Visualization ก็คือ การถ่ายทอดข้อมูลไปสู่ระบบการรับรู้โดยภาพของผู้ใช้ระบบ (Diehl, 2007) เพื่อช่วยในการ ลดช่องว่างระหว่างผู้ใช้และตัวข้อมูล และยังช่วยผู้ใช้สามารถเรียนรู้ ได้อย่างมีประสิทธิภาพ (Keim, 2002)

2.2.4.1 วิธีการของ Visualization

การใช้ Visual เพื่อทำการค้นหาข้อมูลนั้นผู้ใช้จะทำขั้นตอนหลักๆอยู่ 3 ขั้นตอน คือ Overview First, Zoom and Filter และ Detail on Demand โดยอันดับแรก ผู้ใช้ต้องการที่จะดูข้อมูลภาพรวมทั้งหมดซึ่งหลังจากดูภาพรวมทั้งหมดแล้วผู้ใช้ก็จะทำการตัดสินใจเลือกรูปแบบหรือกลุ่มข้อมูลที่สนใจซึ่งก็จะมาถึงขั้นตอนในการวิเคราะห์ข้อมูลผู้ใช้ก็จะทำการเจาะลึกถึงข้อมูลในรายละเอียด ซึ่ง Visualization Technology ก็จะอ้างอิงหรือพัฒนาจากขั้นตอนเหล่านี้ซึ่งVisualization Technique มีประโยชน์มากในการแสดงภาพรวมหรือแสดงข้อมูลย่อยที่ผู้ต้องการโดยอาจจะใช้หลายๆวิธีการรวมกันเพื่อให้ได้ข้อมูลที่ผู้ต้องการ ซึ่งช่วยลดช่องว่างของกิจกรรมที่ใช้ในการดึงข้อมูลต่าง ๆ ไปใช้ ซึ่งลักษณะของข้อมูลที่สามาร

นำมาผ่านกระบวนการของ Visualization มีลักษณะต่าง ๆ มากมายดังนี้ ข้อมูล 1D ได้แก่ ข้อมูลที่เกี่ยวข้องกับช่วงเวลาใดเวลาหนึ่ง, ข้อมูล 2D ได้แก่ ข้อมูลที่เกี่ยวข้องกับแผนที่ภูมิศาสตร์, Multi Dimensional Data ได้แก่ Relation Table, Text และ Hypertext ได้แก่ ข้อมูลหัวข้อความต่าง ๆ และ Web Document, Hierarchies และ Graph ได้แก่ หมายเลขโทรศัพท์ และ Web Document , Algorithms และ Software ได้แก่ Debugging Operation ซึ่งแต่ละข้อมูลก็จะมีวิธีการที่ช่วยในการจัดการแสดงผลข้อมูลที่หลากหลายๆ เช่น

- ประเภทที่แสดงเป็น 2D/3D เช่น แผนภูมิแท่ง และ แผนภูมิ xy
- ประเภทที่แสดงข้อมูลในลักษณะภูมิศาสตร์ เช่น Parallel Coordinates และ ภาพ Landscape
- ประเภทที่แสดงเป็น Icon-Base เช่น Needle Icon และ Star Icon
- ประเภทที่แสดงข้อมูลแบบ Dense Pixel เช่น Recursive Pattern และ Circle Segment
- ประเภทที่แสดงข้อมูลแบบ Stack เช่น Tree Stamp และ Dimension Stacking
- (Dykes, MacEachren & Kraak, 2005, Keim ,2002)

2.2.5 ทฤษฎีเกี่ยวกับต้นไม้ตัดสินใจ

เทคนิคต้นไม้ตัดสินใจเป็นเทคนิคหนึ่งของการจำแนกประเภท (Classification) ซึ่งเป็นวิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูลโดยการจำแนกประเภทนั้นเป็นเทคนิคหนึ่งของการ

2.2.5.1 เหมือนข้อมูลคือการวิเคราะห์ข้อมูลเพื่อแยกประเภท จำแนกรูปแบบและความสัมพันธ์ของข้อมูลจากฐานข้อมูลที่มีขนาดใหญ่หรือคลังข้อมูลโดยมีเทคนิคต่าง ๆ หลายวิธีซึ่งรูปแบบการทำเหมืองข้อมูลนั้นได้รวมความรู้จากหลายแขนงเข้าไว้ด้วยกันที่ประกอบด้วยระบบการเรียนรู้ของเครื่องจักรร่วมกับวิทยาศาสตร์สารสนเทศ สถิติและ ฐานข้อมูล โดยทั่วไปแล้วเทคนิคที่นำมาใช้ส่วนใหญ่มี 5 ประเภท (ศุภชัย, 2551)

1) เทคนิคการจำแนกข้อมูล (Classification) เป็นการจำแนกกลุ่มข้อมูลด้วยคุณลักษณะต่าง ๆ ที่ได้มีการกำหนดไว้แล้วเทคนิคประเภทนี้เหมาะกับการสร้างแบบจำลองเพื่อการพยากรณ์ค่าข้อมูล (Predictive Modeling) ในอนาคตจากการที่ได้จำแนกกลุ่มข้อมูลตัวอย่างไว้แล้วซึ่งในลักษณะดังกล่าวนี้เรียกว่าการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) เทคนิคการการจำแนกข้อมูล มี 2 รูปแบบ ได้แก่ แบบต้นไม้ตัดสินใจ และแบบโครงข่ายประสาทเทียม และเป็นกระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่ใน

กลุ่มที่กำหนดตัวอย่างเช่น การแบ่งประเภทลูกค้าว่าเชื่อถือได้หรือไม่ซึ่งเป็นการสร้างแบบจำลองโดยการเรียนรู้จากข้อมูลที่ได้กำหนดกลุ่มไว้เรียบร้อยแล้ว

2) เทคนิคค้นหาความสัมพันธ์ (Association Rule Discovery) เป็นการค้นหาความสัมพันธ์ของฐานข้อมูลที่มีขนาดใหญ่ เพื่อที่จะทำการวิเคราะห์ข้อมูลและหาสิ่งที่น่าสนใจในข้อมูลนั้น เช่น การวิเคราะห์ข้อมูลการขายในซูเปอร์มาร์เก็ต เพื่อทำการจัดการวางแผนเพื่อจัดการส่งเสริมการขายและเตรียมการวางแผนการเรียงชั้นวางสินค้า เช่น การวางน้ำอัดลมกับข้าวโพดคั่ว ใกล้เคียงกัน

3) เทคนิคจัดหมวดหมู่ (Clustering) เป็นการลดขนาดของข้อมูลด้วยการรวมกลุ่มตัวแปรที่มีลักษณะเดียวกันไว้ด้วยกัน ทำให้สามารถค้นหาข้อมูลที่ถูกละเลยไปได้เทคนิคนี้กับถูกใช้เป็นขั้นตอนเบื้องต้นในการทำเหมืองข้อมูลและเหมาะกับข้อมูลที่ยังไม่มีกลุ่มอย่างชัดเจนจึงทำการจัดหมวดหมู่เพื่อหากกลุ่มต่าง ๆ ของข้อมูลโดยจำนวนกลุ่มของข้อมูลมักจะใช้ตัว k แทนซึ่งผู้ใช้เทคนิคนี้จะเป็นผู้กำหนดจำนวนกลุ่มวิธีนี้อาจเรียกอีกอย่างว่าเคมีน (K-mean)

4) เทคนิคหาค่าที่แตกต่างไปจากค่ามาตรฐาน (Deviation Detection) เป็นการหาค่าที่แตกต่างไปจากค่ามาตรฐาน หรือค่าที่คาดคิดไว้ว่าต่างไปมาเล็กน้อยเพียงใดโดยทั่วไปมักใช้วิธีการทางสถิติหรือการแสดงให้เห็นภาพสำหรับเทคนิคนี้ใช้ในการตรวจสอบลายเซ็นปลอมหรือบัตรเครดิตปลอม เป็นต้น

5) เทคนิคการวิเคราะห์ลำดับ (Sequential Analysis) เป็นการวิเคราะห์ลำดับเพื่อค้นพบรูปแบบของการปรากฏของข้อมูล ซึ่งปรากฏในรายการที่แยกออกมา เช่น ถ้าผู้ซื้อ ซื้อสินค้า A แล้วเขาจะซื้อสินค้า B ในภายหลัง เทคนิคนี้จะแตกต่างจากเทคนิคค้นหาความสัมพันธ์เพราะคำนึงถึงลำดับการซื้อด้วย

2.2.5.2 วิธีการแบ่งประเภทหรือแยกหมวดหมู่ข้อมูลการจำแนกประเภท

คือกระบวนการสร้างแบบจำลองเพื่อจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดเป็นการสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่ได้กำหนดไว้ล่วงหน้าและสามารถพยากรณ์กลุ่มของข้อมูลที่ยังไม่เคยนำมาจัดหมวดหมู่ได้แบบจำลองที่ได้ อาจอยู่ในรูปแบบของต้นไม้ตัดสินใจหรือแบบโครงข่ายประสาทเทียมในการจัดหมวดหมู่จำเป็นต้องชุดของข้อมูลเพื่อใช้สำหรับการเรียนรู้ (Training Data) และสร้างแบบจำลอง (Model Construction) ซึ่งสามารถทดสอบโดยชุดของข้อมูลสำหรับการทดสอบ (Testing Data) เพื่อประเมินความ

ถูกต้องของแบบจำลอง(Model Evaluation) อีกทั้งใช้ชุดข้อมูลที่ไม่เคยเห็นมาก่อน(Unseen Data) เพื่อทำการกำหนดประเภทให้กับข้อมูลใหม่ที่ได้ทำนายค่าออกมาตามที่ต้องการ เช่น การจัดหมวดหมู่ของผู้ขอเครดิตเป็นระดับต่ำระดับกลางและระดับสูงของความเสี่ยงที่จะได้รับหรือการอนุมัติบุคคลเข้ารับทำงานในลักษณะงานต่าง ๆ

2.2.5.3 เทคนิคต้นไม้ตัดสินใจคือแบบจำลองทางคณิตศาสตร์เพื่อการหาทางเลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ซึ่งมีการเรียนรู้ข้อมูลแบบมีผู้สอนสามารถสร้างแบบจำลองการจัดหมวดหมู่ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้าโดยอัตโนมัติและสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้วผลลัพธ์” เช่น

“If Income= High and Married = No THEN Risk = Poor”

“If Income = High and Married = Yes THEN Risk = Good”

2.2.5.3.1 ส่วนประกอบของต้นไม้ตัดสินใจประกอบด้วย

- โหนด(Node) คือคุณสมบัติต่าง ๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทางใดซึ่งโหนดที่อยู่สูงสุดเรียกว่าโหนดราก(Root Node)
- กิ่ง(Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมาโดยจำนวนของกิ่งจะเท่ากับคุณสมบัติของโหนด
- ใบ(Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดงส่วนประกอบของต้นไม้ตัดสินใจ แสดงตัวอย่างของต้นไม้ตัดสินใจได้

2.2.5.3.2 การสร้างต้นไม้ตัดสินใจหลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจเป็นการสร้างในลักษณะจากบนลงล่าง(Top-Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่งไปจนถึงใบโดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้ (ฮานและแคมเบอร์, 2001)

- 1) ต้นไม้เริ่มต้นโดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก
- 2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่จะต้องวัดค่าเกน(Gain) ของแต่ละแอททริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกนมากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือแอททริบิวต์ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้

4) กิ่งของต้นไม้ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่าง ๆ ที่สร้างขึ้น

5) ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกนมากที่สุดสำหรับข้อมูลที่ถูกแบ่งแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์นี้มาสร้างเป็นโหนดตัดสินใจต่อไปโดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อ ๆ ไป

6) ทำการวนซ้ำ เพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้ เป็นจริง

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 โปรแกรม RapidMiner

ซอฟต์แวร์ RapidMiner สามารถจัดการข้อมูลได้ตั้งแต่การเตรียมข้อมูล (Data Preparation) สร้างโมเดล (Model & Validate) ไปจนถึงนำไปใช้งานใน production (Operationalize) ซึ่งจะแยกส่วนได้ดังนี้

- RapidMiner Radoop เป็นเวอร์ชันที่ทำงานบน Hadoop (ซึ่งเป็นการนคอมพิวเตอรืหลาย ๆ เครื่องมาช่วยประมวลผล) ทำให้สามารถรองรับการทำงานกับข้อมูลที่มีขนาดใหญ่ ๆ หลาย (ร้อย) ล้านเรคอร์ดได้

- RapidMiner Studio เป็นเวอร์ชันที่ทำงานบนเครื่องคอมพิวเตอร์ของเราเอง (อาจจะเป็น PC หรือ Notebook ก็ได้) เป็นตัวหลักในการออกแบบโปรเซส (process) หรือ workflow เพื่อใช้ในการวิเคราะห์ข้อมูลต่าง ๆ เช่น สร้างโปรเซสในการคาดการณ์ว่าลูกค้าคนใดจะยกเลิกการใช้บริการ (churn) ด้วยโมเดล Decision Tree

- RapidMiner Server เป็นเวอร์ชันที่ทำงานบนเครื่องคอมพิวเตอร์และรองรับการทำงานที่มีผู้ใช้งานหลาย ๆ คนพร้อมกัน โดยเวอร์ชันนี้สามารถสร้างกราฟในลักษณะของ BI

(Business Intelligence) ตั้งเวลาให้ทำงาน (scheduler) และสร้าง web service เพื่อให้โปรแกรมต่าง ๆ มาติดต่อได้ด้วย

2.3.2 โปรแกรม Tableau Public

Tableau คือ โปรแกรมที่ใช้ในการแสดงผลข้อมูลในรูปแบบของภาพ(Data Visualization) เพื่อให้มีความเข้าใจในข้อมูลมากขึ้น ทำให้สามารถนำข้อมูลจำนวนมากมาวิเคราะห์และมองออกได้อย่างง่ายขึ้น ซึ่งตัวโปรแกรม Tableau มีด้วยกันหลายเวอร์ชันด้วยกัน เช่น Tableau Desktop ,Tableau Online เป็นต้น โปรแกรม Tableau Public เป็นโปรแกรมที่ใช้งานฟรี และสามารถใช้งานได้ดี



ภาพที่ 2.7 โปรแกรม Tableau Public

คุณสมบัติและจุดเด่นของโปรแกรม Tableau

- 1) คุณสมบัติที่โดดเด่นอย่างเห็นได้ชัด คือ ผู้ใช้มีเครื่องมือที่ใช้วิเคราะห์ข้อมูลได้หลากหลายมิติ
- 2) รูปแบบการใช้งานที่ง่ายเพียงคลิกเมาส์ก็สามารถที่จะเปลี่ยนแปลงรายงานได้ตามความต้องการ
- 3) รองรับการเข้าถึงข้อมูลจากหลายฐานข้อมูล เช่น Microsoft Excel, Microsoft Access, Firebird 2.0, IBM DB2, MS SQL Server, Microsoft Power pivot เป็นต้น
- 4) รูปแบบการนำเสนอรายงานที่สวยงาม เข้าใจง่าย และง่ายในการนำเสนอต่อผู้บริหาร

5) สามารถเพิ่มทักษะและแนวคิดให้กับพนักงานระดับปฏิบัติการ ให้มีแนวคิดเชิงสถิติและการประยุกต์ใช้งานมากขึ้น

Tableau นั้นถูกพัฒนาอย่างต่อเนื่องเพื่อให้แสดงผลได้บนทุก Platform อย่างง่ายดายไม่ว่าจะเป็น Desktop, Mobile หรือแม้แต่ Browser โดยองค์กรสามารถเลือกติดตั้งใช้งานภายในองค์กรเองก็ได้ หรือเช่าใช้ผ่านบริการ Tableau Cloud ก็ได้เช่นกัน ทำให้การทำ Business Intelligence และ Data Analytics ไม่ใช่เรื่องยากอีกต่อไป

2.4 วรรณกรรมที่เกี่ยวข้อง

พัทธนันท์ รัตนวรเศวต (2563) ระยะเวลาที่มีการเกิดเหตุอาชญากรรมเกี่ยวกับบุคคลและทรัพย์สินสูงในซอยเสือใหญ่อุทิศนั้น จะมีการเกิดเหตุอาชญากรรมสูงมากในช่วงเวลา 00.01 – 04.00นาฬิกา รองลงมาคือช่วงเวลา 20.01 – 00.00นาฬิกา ซึ่งในช่วงเวลาที่มีการเกิดเหตุอาชญากรรมสูงดังกล่าวนี้เป็นช่วงเวลากลางคืน อันเป็นเวลาพักผ่อนของประชาชนโดยทั่วไป ร้านค้าและอาคารพาณิชย์ส่วนใหญ่จะปิด ทำให้มีประชาชนสัญจรไปมาในพื้นที่น้อย ส่งผลให้อาชญากรมีโอกาสก่ออาชญากรรมได้โดยสะดวก และพื้นที่เสี่ยงสูงต่อการเกิดอาชญากรรมสูงเกี่ยวกับบุคคลและทรัพย์สินในซอยเสือใหญ่อุทิศนั้น ส่วนใหญ่จะอยู่ในเขตพื้นที่ซอยรัชดา 36 แยกต่าง ๆ ซึ่งเป็นพื้นที่ที่มีที่พักอาศัยประเภทหอพัก อาคารชุดพักอาศัย และอาคารพาณิชย์เป็นส่วนใหญ่ เนื่องจากซอยรัชดา 36หรือซอยเสือใหญ่อุทิศนั้น เป็นย่านใจกลางเมือง อยู่ใกล้สถานศึกษาที่มีชื่อเสียง ห้างสรรพสินค้าหลายแห่ง มีบริษัท ห้างฯ ร้านค้ามากมาย มีผู้คนอพยพย้ายถิ่นฐานเข้ามาหางานทำและประกอบอาชีพมากมายในพื้นที่ ทำให้มีการเกิดอาชญากรรมสูงในบริเวณพื้นที่

ชำนาญ คนใจ (2562) ปัญหาอาชญากรรมในสังคมไทยในปัจจุบัน จากข้อมูลคดีอาญาทางข่าวสารหนังสือพิมพ์ โทรทัศน์หรือ สื่อสารทางอิเล็กทรอนิกส์ ต่าง ๆ ทว่าราชอาณาจักรมีแนวโน้มที่จะรุนแรงขึ้น ทั้งในด้านสถิติตัวเลข การก่ออาชญากรรม และวิธีการประกอบอาชญากรรมของผู้กระทำความผิด ที่สามารถมองเห็นได้ เช่น คดีลักทรัพย์ ชิงทรัพย์ ปล้นทรัพย์ ข่มขืนกระทำชำเรา คดียาเสพติด เป็นต้น และที่มองไม่เห็น เช่น คดีเกี่ยวกับอาชญากรรมคอมพิวเตอร์ แก๊งคอลเซ็นเตอร์ การหลอกลวงขายสินค้าทางอินเทอร์เน็ต เป็นต้น จากการนำเสนอภาคข่าวเกี่ยวกับการกระทำผิดผ่านสื่อต่าง ๆ ทั้งจากโทรทัศน์ วิทยุ หนังสือพิมพ์ และอินเทอร์เน็ต มีจำนวนผู้กระทำความผิด และผู้ที่เป็นเหยื่อของอาชญากรรม

เป็นจำนวนมาก ทำให้รัฐบาล จะต้องเพิ่มงบประมาณ ด้านกำลังพล ด้านวัสดุอุปกรณ์ ด้าน การสืบสวน สอบสวน การป้องกัน การปราบปรามอาชญากรรม การควบคุม การสร้าง สวัสดิการ บำบัดรักษา และการแก้ไขฟื้นฟู ดังนั้น ปัญหาอาชญากรรมในสังคมไทย จะหมดสิ้น ไปถ้าคนในสังคมร่วมมือสอดส่องดูแล ให้ความร่วมมือกันและกัน สนับสนุนรัฐบาลและรัฐบาล ก็สนับสนุนประชาชนในการรักษาความสงบเรียบร้อย การอยู่ดีกินดี ดำเนินชีวิตเศรษฐกิจ พอเพียง สังคมก็จะสงบสุข

อัชฌาพร กว่างสวาสดี, เพียงฤทัย หนูสวัสดิ์, วราลี คงเหมาะ, ปวีณา ทิพยากุลรักษ์ และบุษกร สังขนันท์ (2560) งานวิจัยนี้มีวัตถุประสงค์ 1) เพื่อสร้างโมเดลการทำนายระดับ ความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ (Decision tree) 2) เพื่อพัฒนาระบบทำนายระดับ ความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ (Decision tree) 3) เพื่อประเมินการยอมรับของ ระบบทำนายระดับความเครียด โดยมีเครื่องมือในการพัฒนาระบบ คือ ภาษา PHP ซึ่งการ สร้างโมเดลการทำนายระดับความเครียดโดยใช้เทคนิคการทำเหมืองข้อมูลแบบต้นไม้การ ตัดสินใจ ใช้กลุ่มข้อมูลตัวอย่างจากผู้ตอบแบบสอบถาม จำนวน 2,000 ตัวอย่าง โดยมีการวัด ประสิทธิภาพของกฎด้วยวิธีการแบ่งกลุ่มข้อมูลแบบ Cross-validation Test และนำกฎที่ได้มา พัฒนาเป็นระบบทำนายระดับความเครียด เมื่อได้ระบบทำนายระดับความเครียดแล้ว นำไป ประเมินการยอมรับของระบบ โดยเครื่องมือที่ใช้ในงานวิจัยคือ 1) ระบบทำนายระดับ ความเครียด 2) แบบสอบถามวัดการยอมรับของระบบ โดยมีการวัดการยอมรับจากการใช้งาน ของกลุ่มตัวอย่าง จำนวน 30 ตัวอย่าง สถิติที่ใช้ได้แก่ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน ผลการวิจัยพบว่า 1) จากการสร้างโมเดลทำนายระดับความเครียดทำให้ได้กฎการทำนาย จำนวน 120 กฎ ซึ่งมีค่าความถูกต้องของโมเดลอยู่ที่ร้อยละ 91.10 2) ผลการพัฒนาระบบ ทำนายระดับความเครียดผ่านเว็บไซต์ www.predictstress.com ซึ่งระบบสามารถวัดระดับ ความเครียดได้และเสนอแนะข้อปฏิบัติเป็นแนวทางแก้ไขความเครียดได้ 3) ผลการประเมินการ ยอมรับของระบบทำนายระดับความเครียด ด้วยเทคนิคต้นไม้การตัดสินใจ โดยมีการประเมิน 3 ด้าน คือด้านเนื้อหา ด้านการออกแบบ และด้านการใช้งาน ซึ่งภาพรวมยอมรับระดับมาก (4.21)

เด่นเดือน เลิศทยากุล, ธรา อังสกุล, วีรพงษ์ พลนิกรกิจและจิตติมนต์ อังสกุล (2558) ด้วยประสิทธิภาพของสื่ออินเทอร์เน็ตที่สามารถเสนอขายสินค้า และบริการได้ตลอด 24 ชั่วโมง เข้าถึงกลุ่มผู้บริโภคได้แบบเจาะจงเป็นรายบุคคล ประกอบกับอัตราค่าใช้จ่ายในการ แพร่กระจายสื่อโฆษณาทางสื่ออินเทอร์เน็ตนั้นยังต่ำกว่าสื่อโฆษณาประเภทอื่น จึงทำให้สื่อ

อินเทอร์เน็ตกลายเป็นสื่อโฆษณารูปแบบใหม่ที่ผู้ประกอบการพาณิชย์อิเล็กทรอนิกส์ และผู้ประกอบการทั่วไปให้ความสนใจมากขึ้น ซึ่งในการโฆษณาทางอินเทอร์เน็ตนั้น ผู้ประกอบการต้องคำนึงถึงเว็บไซต์ และหมวดหมู่ในเว็บไซต์ รวมไปถึงงบประมาณในการโฆษณาที่สามารถตอบสนองความต้องการของกลุ่มผู้บริโภคเป้าหมายได้ บทความนี้จึงนำเสนอระบบสนับสนุนการตัดสินใจเพื่อการเลือกโฆษณาทางเว็บไซต์ โดยคำนึงถึงปัจจัยสำคัญด้านลักษณะประชากรศาสตร์ของผู้บริโภคกลุ่มเป้าหมาย และหมวดหมู่ในเว็บไซต์ที่เหมาะสมกับกลุ่มผู้บริโภคเป้าหมาย โดยระบบได้นำเสนอการทำเหมืองข้อมูลโดยใช้แบบจำลองต้นไม้ประกอบการตัดสินใจ เพื่อใช้เลือกหมวดหมู่ในเว็บไซต์ที่เหมาะสม ซึ่งผลที่ได้จากการทดสอบพบว่า การเลือกหมวดหมู่ในการโฆษณาทางเว็บไซต์เหล่านั้นมีความถูกต้องตรงกับกลุ่มผู้บริโภคเป้าหมายได้สูงถึง 84.74% และจากการประเมินความสามารถในการใช้งานได้ของระบบจากกลุ่มตัวอย่าง 3 กลุ่ม ได้แก่ กลุ่มผู้ประกอบการพาณิชย์อิเล็กทรอนิกส์ กลุ่มผู้ประกอบการทั่วไป และกลุ่มผู้ให้บริการโฆษณาทางเว็บไซต์ จากการใช้แบบสอบถามออนไลน์โดยรวม พบว่า กลุ่มตัวอย่างมีความพึงพอใจอยู่ในระดับมาก โดยพึงพอใจมากด้านประสิทธิภาพของระบบในส่วนที่ระบบที่สามารถประมวลผลได้รวดเร็ว ผู้ใช้สามารถใช้งานได้ทันทีโดยไม่ต้องเรียนรู้ใหม่ เมื่อกลับมาใช้งานระบบอีกครั้ง ระบบสามารถวางแผนโฆษณาได้ตรงตามงบประมาณที่ผู้ใช้กำหนด สามารถคาดการณ์ค่าใช้จ่ายในการโฆษณาได้ ช่วยในการวางแผนการโฆษณาทางเว็บไซต์ได้

สุขุมมา อรุณจิตและวุฒิพล มั่นเหมาะ (2559) อาชญากรรมทางเพศเป็นปัญหาสังคมที่ร้ายแรงการศึกษานี้มีวัตถุประสงค์เพื่อศึกษาปัจจัยการเกิดอาชญากรรมทางเพศรวมถึงแนวทางการป้องกันอาชญากรรมทางเพศตามทฤษฎีสามเหลี่ยมอาชญากรรม ด้วยวิธีการสังเคราะห์เอกสารงานวิจัยทั้งหมด 10 เรื่อง ผลการศึกษาพบว่าปัจจัยการเกิดอาชญากรรมทางเพศสามารถแบ่งได้ ดังนี้ 1) ปัจจัยด้านผู้กระทำผิด/คนร้าย ได้แก่ ฐานะทางเศรษฐกิจต่ำ เสพยาเสพติด ได้รับสิ่งช่วยจากสื่อ ครอบครัวมีปัญหาหรือแตกแยก บุคลิกภาพเฉพาะ เช่น การควบคุมตนเองต่ำ ขาดความมั่นใจซึมเศร้า เก็บกด มีค่านิยมชายเป็นใหญ่ เป็นผู้ใช้ชีวิตที่เสี่ยงภัย และมีกลุ่มเพื่อนที่ช่วย 2) ปัจจัยด้านเหยื่อ ได้แก่ อยู่ในภาวะเสพของมีนเมารูปร่างหน้าตาดี แต่งกายล่อแหลม ครอบครัวไม่สมบูรณ์แตกแยกห่างเหิน มีช่วงวัยอยู่ในช่วงวัยรุ่น และมีฐานะทางเศรษฐกิจต่ำ 3) ปัจจัยด้านโอกาส ได้แก่ ความใกล้ชิดเนื่องจากเป็นคนรู้จักสนิทสนมกัน เหตุเกิดในเวลากลางคืน-ดึก และอยู่ในสภาพแวดล้อมเสี่ยง พื้นที่เปลี่ยว ลับตาคน

ส่วนข้อเสนอแนะในการป้องกันอาชญากรรมทางเพศแบ่งเป็น 3 ส่วนคือ ส่วนปัจจัยความเสี่ยงด้านโอกาส ได้แก่ หน่วยงานที่เกี่ยวข้องต้องสนับสนุนให้ประชาชนมีความรู้ทักษะการป้องกันตนเอง สถาบันครอบครัว สถาบันการศึกษาต้องให้ความรู้เกี่ยวกับการป้องกันอาชญากรรมทางเพศ การมีมาตรการ นโยบายตรวจตราความปลอดภัยของพื้นที่ให้มากขึ้น การอาศัยการมีส่วนร่วมของชุมชน รวมทั้งบูรณาการการปฏิบัติงานของทุกภาคส่วน ส่วนปัจจัยความเสี่ยงด้านผู้กระทำผิด ได้แก่ สถาบันทางสังคมควรอบรมสั่งสอนขัดเกลาให้บุคคลเป็นคนดี มีความประพฤติเหมาะสม สื่อมวลชนจะต้องมีความรับผิดชอบต่อสังคมในการนำเสนอสื่อต่าง ๆ การควบคุมสื่อ การมีมาตรการหรือกฎหมายควบคุมปัจจัยเสี่ยงต่อการเกิดอาชญากรรม ส่วนปัจจัยความเสี่ยงด้านเหยื่อ ได้แก่ หน่วยงานต่าง ๆ ต้องร่วมรณรงค์ ประชาสัมพันธ์ไม่ให้ประชาชนตกเป็นเหยื่อ ประชาชนเองต้องเรียนรู้วิธีการป้องกันตนเองจากอาชญากรรม การบรรจุความรู้เรื่องอาชญากรรมทางเพศในหลักสูตรการศึกษา และการกำหนดโครงการกิจกรรมควรครอบคลุมการป้องกันแก้ไขปัจจัยเสี่ยงการเกิดอาชญากรรม

สุวิมล สิริธาดา (2560) งานวิจัยนี้ได้ศึกษาคุณลักษณะของนักศึกษาคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี เชียงใหม่ ที่มีผลต่อการเรียนแคลคูลัส 1 สำหรับวิศวกร ด้วยเทคนิคเหมืองข้อมูลเพื่อจำแนกนักศึกษาออกเป็น 2 กลุ่ม คือ กลุ่มเสี่ยงที่จะไม่ผ่านรายวิชา และกลุ่มที่ไม่มีความเสี่ยง โดยการวิเคราะห์ข้อมูลได้เก็บรวบรวมข้อมูลคุณลักษณะทั้ง 3 ด้าน รวมถึงปัจจัยส่วนบุคคลอื่น ๆ ของนักศึกษาจำนวน 453 คน แล้วดำเนินการตามกระบวนการ CRISP-DM (Cross Industry Standard Process for Data Mining) โดยใช้เทคนิคโครงข่ายประสาทเทียมและต้นไม้ตัดสินใจในการจำแนกประเภทข้อมูลและคัดเลือกคุณลักษณะที่สำคัญด้วย Filter Ranker Method โดยคำนวณค่าน้ำหนักด้วย Chi-Square และ Gain Ratio จากผลการศึกษา พบว่า ความแม่นยำในการจำแนกข้อมูลจาก 50 คุณลักษณะนั้น วิธี ANN มีค่า 71.52% และ Decision Trees - J48 มีค่า 66.23% หลังจากทำการคัดเลือกคุณลักษณะแสดงให้เห็นว่าการจำแนกข้อมูลด้วยวิธี ANN จาก 5 คุณลักษณะแรกที่ได้จากการคัดเลือกโดยวิธี Filter Ranker Method ที่คำนวณค่าน้ำหนักด้วย Chi-Square ทำให้ได้ค่าความถูกต้องมีค่าสูงสุด คือ 80.13% และการจำแนกข้อมูลด้วย Decision Tree ก็ให้ผลไปในทางเดียวกัน โดยมีค่าความถูกต้องสูงสุดที่ 75.83% คุณลักษณะที่ผ่านการคัดเลือกนั้นแสดงถึงคุณลักษณะในด้านพื้นฐานความรู้ของนักศึกษาที่มีมาก่อน จากผลการวิจัยนี้แสดงให้เห็นว่า

เทคนิคเหมืองข้อมูลสามารถนำมาประยุกต์ใช้ทางด้านการศึกษาในการวิเคราะห์ข้อมูลเพื่อแก้ปัญหาเกี่ยวกับการเรียนการสอนได้

พรรณธิดา เพชรบุญมี, ดวงกมล โพธิ์นาคและมนต์ชัย เทียนทอง (2556) งานวิจัยนี้มีวัตถุประสงค์เพื่อพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคลป์ ตัวแบบที่พัฒนาขึ้นโดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจ ถูกสร้างและทดสอบตัวแบบด้วยวิธีการตรวจสอบไขว้ (K-fold cross-validation) และวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ (percentage split) ทำการวิเคราะห์ข้อมูลด้วยโปรแกรม WEKA ผลการทดลองพบว่าได้กฎการจำแนกข้อมูลทั้งสิ้นจำนวน 8 กฎ เมื่อวัดค่าประสิทธิภาพของตัวแบบ การพัฒนาตัวแบบด้วยวิธีการแบ่งข้อมูลแบบสุ่มด้วยการแบ่งร้อยละ 66.00 จะมีค่าความถูกต้องเท่ากับร้อยละ 85 ค่าความแม่นยำเท่ากับร้อยละ 88.20 ค่าความระลึกเท่ากับร้อยละ 85.00 และค่าความถ่วงดุลเท่ากับร้อยละ 85.10 ซึ่งมีประสิทธิภาพทุกค่าสูงกว่าวิธีการอื่น โดยงานวิจัยนี้สามารถนำไปใช้ในการพัฒนาตัวแบบการพยากรณ์รูปแบบการเรียนรู้ตามประสบการณ์ของเดวิด โคลป์ โดยใช้กฎการจำแนกเทคนิคต้นไม้ตัดสินใจที่มีความถูกต้องและแม่นยำในการทำนายรูปแบบการเรียนรู้ได้เป็นอย่างดี

อนันต์ชัย ชูติภาสเจริญและจรัญ แสนราช (2561) การวิจัยนี้มีวัตถุประสงค์เพื่อเปรียบเทียบประสิทธิภาพของอัลกอริทึมในการทำนายและคุณลักษณะที่มีต่อโอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปโดยทางการศึกษาข้อมูลการโอนเงินของบุคคลทั่วไปจำนวน 51,901ระเบียบนาการเก็บข้อมูลตั้งแต่ปี2559-2560 โดยใช้เทคนิคการจำแนกข้อมูลทั้งหมด3เทคนิคได้แก่เทคนิคต้นไม้ตัดสินใจ เทคนิคนาอีฟเบย์ และเทคนิคการค้นหาเพื่อนบ้านใกล้สุด ซึ่งทำการเปรียบเทียบประสิทธิภาพรูปแบบเทคนิคการทำนายระหว่างการใช้คุณลักษณะทั้งหมดการทดสอบประสิทธิภาพตัวแบบทำนายด้วยวิธีการCross Validation โดยใช้โปรแกรม RapidMiner Studio 8จากนั้นทำการทดลองเพื่อหาผลการทดสอบประสิทธิภาพที่มีค่าความถูกต้องที่สูงที่สุดผลการศึกษาพบว่าการใช้เทคนิคต้นไม้ตัดสินใจด้วยการเลือกคุณลักษณะทั้งหมดมีค่าความถูกต้องเท่ากับ99.90% เทคนิคการค้นหาเพื่อนบ้านใกล้สุดมีค่าความถูกต้องเท่ากับ99.55%และเทคนิคนาอีฟเบย์มีค่าความถูกต้องเท่ากับ96.71% จากผลการเปรียบเทียบประสิทธิภาพในครั้งนี้สามารถนาเทคนิคต้นไม้ตัดสินใจที่มีค่าความถูกต้องสูงสุดไปใช้ในการพยากรณ์โอกาสความสำเร็จในการโอนเงินข้ามประเทศของบุคคลทั่วไปต่อไป

อิทธิพล ดวงแก้วและสายัญญ สายยศ (2562) การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อวิเคราะห์ปัจจัยที่มีผลต่อพัฒนาการล่าช้าในแต่ละกลุ่มอายุของเด็กวัยก่อนเรียน 4 คนคือ 9, 18, 30 และ 42 เดือนโดยประเมินจากคู่มือการเฝ้าระวังและส่งเสริมการขยาย การพัฒนาเด็กปฐมวัย (DSPM) โดยขอข้อมูลสนับสนุนจากสำนักงานสาธารณสุขจังหวัดขอนแก่น ข้อมูลเริ่มตั้งแต่ตุลาคม 2559 ถึงกันยายน 2560 นักวิจัยศึกษาปัจจัยที่มีผลต่อพัฒนาการในวัยเด็กจากการวิจัยทางการแพทย์และสาธารณสุขก่อนหน้านี้ รับ 24 ปัจจัย แบบจำลองนี้สร้างขึ้นโดยใช้เทคนิคการขุดข้อมูลเช่นเทคนิค Decision Tree, K-Neighbour เพื่อนบ้านและเทคนิคNaive Bayes ทำการทดสอบโดยใช้วิธีการตรวจสอบความถูกต้องข้ามแบบ K-fold Cross validation ผลการวิจัยปรากฏว่าเทคนิคDecision Tree มีความถูกต้องแม่นยำมากที่สุดในการทำนายพัฒนาการทั้ง 4 ช่วง

เกรียงศักดิ์ รักภักดีและวชิระ โมราชาติ (2561) งานวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อศึกษาความต้องการและพฤติกรรมการตัดสินใจเลือกแหล่งท่องเที่ยวทางธรรมชาติของนักท่องเที่ยว 2) เพื่อพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวทางธรรมชาติในจังหวัดอุบลราชธานี ตัวอย่างที่ใช้ในการศึกษา คือ นักท่องเที่ยวที่เดินทางมาท่องเที่ยวหรือชาวบ้านที่อาศัยอยู่บริเวณแหล่งท่องเที่ยว จำนวน 1,840 คน จากแหล่งท่องเที่ยว จำนวน 23 แห่ง โดยแบ่งแหล่งท่องเที่ยวตามกิจกรรมการท่องเที่ยวออกเป็น 5 กิจกรรม การพัฒนาฐานข้อมูลแนะนำแหล่งท่องเที่ยวดำเนินการด้วยการใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) ซึ่งเทคนิคที่นำมาใช้ คือ การเรียนรู้ต้นไม้ตัดสินใจ(Decision Tree Learning) และนำเสนอระบบฐานข้อมูลของงานวิจัยในรูปแบบเว็บไซต์ ผลการศึกษาความต้องการและพฤติกรรม พบว่าการได้รับรู้ข้อมูลข่าวสารของแหล่งท่องเที่ยวจากแอปพลิเคชันแนะนำแหล่งท่องเที่ยว มีความสำคัญต่อการตัดสินใจที่จะเดินทางมาท่องเที่ยวเป็นอย่างมาก และผลการประเมินประสิทธิภาพของระบบฐานข้อมูลแนะนำแหล่งท่องเที่ยวในภาพรวมอยู่ในระดับดี

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นนั้น ผู้วิเคราะห์ข้อมูลได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ทำเหมืองข้อมูลแบบ Classification ด้วยการสร้างโมเดล Decision Tree โดยใช้โปรแกรม RapidMiner Studio 9.6 ในการสร้างโมเดล

Decision Tree เพื่อทำการเปรียบเทียบ จากนั้นนำข้อมูลสารสนเทศมาทำการแสดงผลแบบ visualization ในรูปแบบของภาพโดยใช้โปรแกรม Tableau Public เผยแพร่บน Web Browser

