บทที่ 2

แนวคิดทฤษฎีเครื่องมือและวรรณกรรมที่เกี่ยวข้อง

โครงงานเรื่องการวิเคราะห์ข้อมูลนักศึกษาเพื่อสนับสนุนการตัดสินใจในการวางแผนประชาสัมพันธ์ โดยใช้การนำเทคนิคเหมืองข้อมูล (Data mining) การจำแนกข้อมูล (Data classification) การค้นหากฎความสัมพันธ์ (Association rule) และรูปแบบของแผนภูมิต้นไม้ (Decision tree) เพื่อวิเคราะห์หารูปแบบที่เกิดขึ้นจากข้อมูลการยืนยันสิทธิ์เคลียริงเฮาส์ (Clearing house) สามารถนำรูปแบบ (Model) ได้มาวางแผนการคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS ของมหาวิทยาลัยเทคโนโลยีราชมงคลล้านนา นำไปสู่การวางแผนประชาสัมพันธ์ต่อไปในอนาคต

ในบทนี้เป็นการนำเสนอเกี่ยวกับ แนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่ เกี่ยวข้องของการวิเคราะห์ข้อมูลนักศึกษาเพื่อวางแผนการคัดเลือกนักศึกษาใหม่ในระบบ TCAS นำไปสู่การวางแผนประชาสัมพันธ์และการแสดงผลข้อมูลบนเว็บไซต์ ซึ่งได้รวบรวมการศึกษาเอกสารงานวิจัยที่เกี่ยวข้อง กับการวิเคราะห์ข้อมูล เพื่อใช้เป็นแนวทางการศึกษาประกอบด้วยรายละเอียดตามลำดับดังนี้

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่ง CSS และ Bootstrap

2.2.5 ทฤษฎีเกี่ยวกับการ Visualization

2.3 เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 ระบบสนับสนุนการตัดสินใจ (Decision Support System)

2.3.2 การตัดสินใจแบบโครงสร้างต้นไม้ (Decision Tree)

2.3.3 การจำแนกประเภทแบบ Random forest

2.3.4 ซอฟต์แวร์ Weka

2.3.5 ซอฟต์แวร์ KNIME

2.4 วรรณกรรมที่เกี่ยวข้อง

2.5 บทสรุป

2.1 แนวคิด

2.1.1 แนวคิดเกี่ยวกับการวิเคราะห์ข้อมูล (Data analytic)

ความหมายของการวิเคราะห์ข้อมูล เอื้อมพร หลินเจริญ (2555) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพ นับเป็นขั้นตอนที่สำคัญในกระบวนการวิจัย วิธีการหลักที่ใช้ใน การวิเคราะห์ข้อมูลเชิงคุณภาพ เป็นวิธีการสร้างข้อสรุปจากการศึกษาจากข้อมูลจำนวนหนึ่ง ซึ่งมักไม่ใช้สถิติในการวิเคราะห์หรือถ้าใช้สถิติก็ไม่ได้ถือว่าสถิติเป็นวิธีการวิเคราะห์หลัก แต่จะ ถือเป็นข้อมูลเสริมในการวิเคราะห์ ข้อมูลเชิงคุณภาพนั้น คณะผู้จัดทำนับเป็นผู้ที่มีบทบาท สำคัญยิ่งและควรมีความรอบรู้ในเรื่องแนวคิดทฤษฎีอย่างกว้างขวาง มีความเป็นสหวิทยากร อยู่ในตัวเอง มีความสามารถทางภาษาสามารถเชื่อมโยงข้อความและสร้างข้อสรุปเป็นกรอบแนวคิดและตีความหมายของข้อมูลได้หลายๆแบบทั้งนี้การวิเคราะห์ข้อมูลเชิงคุณภาพนั้น อาจไม่จำเป็นต้องใช้เฉพาะการวิจัยเชิงคุณภาพเท่านั้น แต่อาจใช้กับการวิจัยเชิงปริมาณที่มีการ เก็บรวบรวมข้อมูลเชิงคุณภาพ เช่น แบบสอบถามปลายเปิด การสัมภาษณ์ การสังเกต จด บันทึกมาทำการวิเคราะห์เพื่อให้ข้อมูลที่รวบรวมมามีความหมายและตอบคำถามหรือจุดมุ่งหมายของการวิจัยสำหรับสาระในการวิเคราะห์ข้อมูลเชิงคุณภาพโดยเทคนิคการ วิเคราะห์ข้อมูลเชิงคุณภาพที่สำคัญ ๆ และนักวิจัย นิยมใช้ ได้แก่ การจำแนกหรือการจัดกลุ่ม ข้อมูล การเปรียบเทียบเหตุการณ์ การวิเคราะห์ส่วนประกอบ การวิเคราะห์แบบอุปนัย และการวิเคราะห์ข้อมูลที่เป็นเอกสาร

หัสพร ทองแดง, พิศุทธิภา เมธีกุล, สมคิด ทุมวงศ์ และแพรว สมบัติใหม่ (2559) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเป็นการจำแนกข้อมูล ให้เป็นระบบ เพื่อทำให้เกิดความเข้าใจในความหลากหลาย ความหมายและความสัมพันธ์ของข้อมูลในบริบทของสังคมและวัฒนธรรมหนึ่ง ๆ ในการวิเคราะห์ข้อมูลเชิงคุณภาพนี้หลังจากที่มีการแยกข้อมูลออกได้เป็นหมวดหมู่แล้ว นักวิจัยต้องใช้วิธีการนำเสนอโดยการพรรณนาให้เห็นสภาพเงื่อนไข กระบวนการ ขั้นตอนการสัมพันธ์ต่าง ๆ รวมทั้งพยายามหาความหมายทางวัฒนธรรมของปรากฏการณ์ในทัศนะของบุคคลหรือกลุ่มตัวอย่างซึ่งเป็นผู้ให้ข้อมูล และอาจมีความเข้าใจและตีความแตกต่างกันออกไปหรือแตกต่างไปจากของผู้วิจัยเอง การวิเคราะห์ข้อมูลสำหรับการวิจัยเชิงคุณภาพ จึงเป็นเสมือนการทำความเข้าใจในแบบแผนและความหมายของพฤติกรรมมากกว่าที่จะมุ่งหาระดับความมากน้อยหรือระดับความเข้มข้นของความสัมพันธ์ของปรากฏการณ์จึงเป็นการวิเคราะห์ข้อมูลที่อาศัยเครื่องมือทางความคิด (Conceptual tools) มากกว่าที่จะใช้เครื่องมือทางสถิติ และที่สำคัญต้องอาศัยความสามารถของนักวิจัยที่จะอธิบายให้เห็นความสัมพันธ์และความหมายของปรากฏการณ์บนพื้นฐานความเข้าใจในวัฒนธรรมและสังคมที่เป็นบริบทของปรากฏการณ์ที่ศึกษา

จุมพล หนิมพานิช และวรวลัญช์ โรจนพล (2561) ได้ให้ความหมายการวิเคราะห์ข้อมูลเชิงคุณภาพไว้ว่า การวิเคราะห์ข้อมูลเชิงคุณภาพโดยทั่วไปหมายถึง การสรุปเนื้อหาสาระของข้อมูล และหาแบบแผนของความสัมพันธ์ระหว่างข้อมูลหรือตัวแปร หรือหมายถึงการสรุปพรรณนาและการวิเคราะห์ความสัมพันธ์ระหว่างข้อมูล หรือหมายถึง การนำเอาข้อมูลที่ได้จากการค้นคว้าวิจัยมาจัดกระทำให้เป็นระบบและหาความหมาย แยกแยะ องค์ประกอบ รวมทั้งเชื่อมโยงและหาความสัมพันธ์ของข้อมูลเพื่อให้สามารถนาไปสู่ความเข้าใจต่อการดำรงอยู่และการเปลี่ยนแปลงของปรากฏการณ์ที่ศึกษา หรือหมายถึง การแยกแยะ การตีความหมาย การเปรียบเทียบ การหาความสัมพันธ์เกี่ยวกับปรากฏการณ์ การหาแบบแผนการอธิบายและการสรุปเกี่ยวกับกิจกรรมหรือพฤติกรรมในสภาพสังคมหรือปรากฏการณ์ทางสังคมที่ศึกษา

2.1.2 แนวคิดเกี่ยวกับการประยุกต์ใช้ข้อมูลทางสถิติ

ความหมายของการประยุกต์ใช้ข้อมูลทางสถิติจากเอกสารประกอบการสอนการวิจัยสำหรับครู ได้ให้ความหมายเกี่ยวกับสถิติไว้ ว่า คำว่าสถิติ (Statistics) มาจากภาษาเยอรมันว่า Statistics มีรากศัพท์มาจาก Stat หมายถึง ข้อมูลหรือสารสนเทศ ซึ่งจะอำนวยประโยชน์ต่อ

การบริหารประเทศในด้านต่าง ๆ เช่น การทำ สำมะโนครัวเพื่อจะทราบจำนวนพลเมืองในประเทศทั้งหมด ในสมัยต่อมา คำว่า สถิติ ได้ หมายถึง ตัวเลขหรือข้อมูลที่ได้จากการเก็บรวบรวม เช่น จำนวนผู้ประสบอุบัติเหตุบนท้องถนน อัตราการเกิดของเด็กทารก ปริมาณน้ำฝนในแต่ละปี เป็นต้น สถิติในความหมายที่กล่าวมานี้ เรียกอีกอย่างหนึ่งว่า ข้อมูลทางสถิติ (Statistical data)

1) ค่ากึ่งกลาง (Median) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางวิธีหนึ่งที่ใช้การเรียงค่า การสะท้อนของจุดภาพจากค่าน้อยที่สุดไปหาค่ามากที่สุด โดยค่ากึ่งกลางเป็นค่าที่อยู่ใน ตำแหน่งกึ่งกลางของข้อมูลทั้งชุด ค่ากึ่งกลางจึงเป็นตัวแทนค่าการสะท้อนของจำนวนจุดภาพ ทั้งหมดในช่วงคลื่นหนึ่งๆ ที่แสดงให้ทราบว่ามีจำนวนจุดภาพที่มีค่าการสะท้อนมากกว่าและ น้อยกว่าค่ากึ่งกลางอยู่ประมาณร้อยละ 50

2) ค่าฐานนิยม (Mode) เป็นการวัดแนวโน้มเข้าสู่ส่วนกลางอีกวิธีหนึ่ง โดยดูจาก จำนวนความถี่ของค่าการสะท้อนซึ่งมีความถี่สูงที่สุด นิยมนำมาใช้กับข้อมูลที่เป็นนามบัญญัติ

เช่น ค่าของประเภทข้อมูลหลังจากการจำแนกประเภทแล้ว ถือเป็นค่าการสะท้อนที่แสดงการใช้ ที่ดินประเภทต่าง ๆ ไม่ใช่ค่าการสะท้อนของวัตถุอีกต่อไป

3) ความเบี่ยงเบนมาตรฐาน (Standard Deviation : S.D.) เป็นการวัดการกระจาย ที่นิยมใช้มากที่สุดการคำนวณใช้วิธียกกำลังสองของผลต่างระหว่างค่าการสะท้อนของทุกจุดภาพในแต่ละช่วงคลื่นกับค่าเฉลี่ยเลขคณิตของช่วงคลื่นนั้น

ภัทรธิดา ผลงาม (2558) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติ คือ ตัวเลขต่าง ๆ ที่ได้มีการรวบรวมขึ้นเพื่อบอกข้อเท็จจริงเกี่ยวกับคุณสมบัติหรือลักษณะบางสิ่งบางอย่างที่สามารถแสดงออกเป็นตัวเลขได้สถิติแบ่งเป็นประเภทใหญ่ ๆ ได้ 2 ประเภท ดังนี้ คือ

1) สถิติพรรณนา (Descriptive statistics) คือสถิติที่บรรยายถึงลักษณะของข้อมูลเฉพาะกลุ่มนั้น ๆ โดยไม่สรุปอ้างอิงไปยังประชากรกลุ่มอื่น ๆ สถิติประเภทนี้นิยมศึกษาในกลุ่มเล็กหรือกลุ่มใหญ่ก็ได้ สถิติประเภทนี้เป็นสถิติที่บรรยายลักษณะของข้อมูล เช่น ค่าเฉลี่ย ร้อยละ มัธยฐาน พิสัย ส่วนเบี่ยงเบนมาตรฐาน และการวัดความสัมพันธ์ต่าง ๆ เช่น สหสัมพันธ์

2) สถิติอ้างอิงหรือสถิติอนุมาน (Inferential or inductive statistics) คือ สถิติที่นำค่าสถิติพรรณนามาสรุปอ้างอิงไปยังประชากร หรือเป็นสถิติที่ศึกษาจากกลุ่มตัวอย่างแล้วนำ

ผลอ้างอิงไปยังกลุ่มประชากร เช่น การศึกษาความคิดเห็นต่อการเลือกตั้งของประชาชนไม่จำเป็นต้องศึกษาจากประชาชนทุกคน แต่สามารถเลือกศึกษาจากประชาชนบางกลุ่มซึ่งจะเป็นตัวแทนของประชาชนทั้งหมด แล้วจึงสรุปว่าประชาชนมีความคิดเห็นอย่างไรต่อการเลือกตั้งได้ ดังนั้นในการใช้สถิติอ้างอิงนี้จึงจำเป็นต้องเลือกลุ่มตัวอย่างที่เหมาะสม สถิติอ้างอิงนี้ก่อนนำไปอ้างอิงกลุ่มประชากรต้องมีการทดสอบทางสถิติก่อนทุกครั้งจึงสามารถอ้างอิงประชากรได้อย่างมีประสิทธิภาพ

สุทิน ชนะบุญ (2555) ได้ให้ความหมายเกี่ยวกับสถิติไว้ว่า สถิติประกอบด้วย 2 องค์ประกอบ ดังนี้ สถิติเชิงพรรณนา (Descriptive Statistics) เป็นวิธีการทางสถิติเพื่อใช้ในการพรรณนาหรือบรรยายลักษณะของสิ่งที่ศึกษา เพื่อให้เข้าใจถึงลักษณะของข้อมูลที่เก็บรวบรวมมาได้ จะพรรณนาภายในขอบเขตของข้อมูลที่เก็บรวบรวมมาเท่านั้น ไม่สามารถจะคาดคะเนลักษณะต่าง ๆ ออกไปนอกเหนือจากข้อมูลที่มีอยู่ได้ หรือไม่มีการอ้างอิงหรืออนุมานไปถึงกลุ่มอื่น หากผู้วิจัยสามารถศึกษาทุกหน่วยของประชากรได้ ก็จะใช้สรุปหรือบรรยายลักษณะของประชากรที่ศึกษา หากผู้วิจัยไม่สามารถศึกษาสมาชิกทุกหน่วยของประชากรได้ สุ่มสมาชิกเพียงบางส่วนมาศึกษา(ศึกษาจากกลุ่มตัวอย่าง) ก็จะใช้สรุปหรือบรรยายเฉพาะกลุ่มตัวอย่างที่สุ่มมาศึกษาเท่านั้น สถิติเชิงอ้างอิงหรือสถิติอนุมาน (Inferential Statistics) เป็นวิธีการทางสถิติที่ใช้ในการสรุปลักษณะของประชากร จากผลการศึกษาข้อมูลในกลุ่ม ตัวอย่าง โดยอาศัย

ทฤษฎีความน่าจะเป็นในการอนุมานลักษณะประชากร โดยศึกษาจากกลุ่มตัวอย่าง แต่อ้างอิง หรืออนุมานไปถึงประชากร ประกอบด้วย 2 องค์ประกอบ คือ การประมาณค่าพารามิเตอร์ (Parametric Estimation) และการทดสอบสมมติฐาน (Testing Hypothesis)

ทฤษฎีความน่าจะเป็นในการอนุมานลักษณะประชากร โดยศึกษาจากกลุ่มตัวอย่าง แต่อ้างอิง หรืออนุมานไปถึงประชากร ประกอบด้วย 2 องค์ประกอบ คือ การประมาณค่าพารามิเตอร์ (Parametric Estimation) และการทดสอบสมมติฐาน (Testing Hypothesis)

2.1.3 แนวคิดเกี่ยวกับการแสดงผลข้อมูล (Data visualization)

ความหมายการแสดงผลข้อมูล

กานต์ ยงศิริวิทย์ และภาคภูมิ ชัยศิริประเสริฐ (2560) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) ผลลัพธ์ที่ได้จากการวิเคราะห์

ข้อมูลควรนำเสนอในรูปแบบที่น่าสนใจ และเข้าใจง่าย ดังนั้นเครื่องมือควรที่จะสามารถนำเสนอข้อมูลในรูปแบบของกราฟและแสดงผลลัพธ์บนแผนที่ในขั้นพื้นฐานได้ โดยไม่ควรที่จะต้องมีการติดตั้งซอฟต์แวร์หรือตัวช่วยเพิ่มเติม

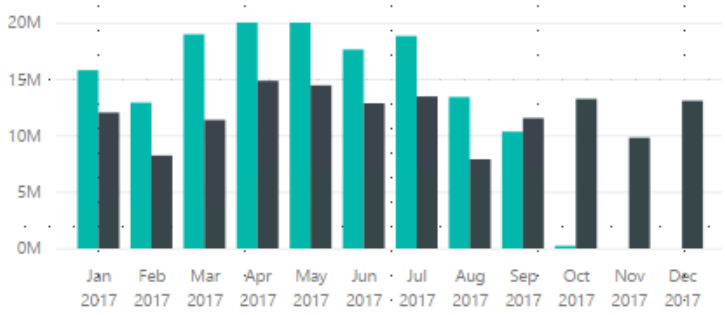
ชนาธิป ชื่นมนัส (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การ แสดงผลข้อมูล (Data Visualization) คือวิธีการที่ใช้ในการนำข้อมูลที่เป็นนามธรรมมาแสดงให้ เป็นภาพเพื่อเพิ่มประสิทธิผลในการติดต่อสื่อสาร หรือทำให้เข้าใจข้อมูลได้ง่ายขึ้น การใช้ เทคนิค Visualization มีหลายวิธี ซึ่ง แต่ละวิธีเหมาะสมที่จะใช้ในการแสดงข้อมูลที่แตกต่างกัน การเลือกใช้เทคนิคใดนั้นขึ้นกับองค์ประกอบหลายอย่าง เช่น ชนิดของข้อมูลที่จะนำมาแสดง วิธี ที่ต้องการแสดงข้อมูล เป็นต้น ในปัจจุบันมีงานวิจัยที่นำเอาเทคนิค Visualization มาใช้โดยมีจุดประสงค์เพื่อเพิ่มประสิทธิผลในการแสดงข้อมูลการติดต่อสื่อสารภายในเครือข่ายคอมพิวเตอร์ และช่วยในการวิเคราะห์ความผิดปกติที่เกิดขึ้นในเครือข่าย

อาทิตย์ สิทธิบรรเจิด (2552) ได้ให้ความหมายของการแสดงผลข้อมูลไว้ว่า การแสดงผลข้อมูล (Data Visualization) เป็นส่วนประกอบสำคัญใน Cognitive System ซึ่งเป็นส่วนในการแสดงข้อมูลหรือผลลัพธ์ต่าง ๆ ในระหว่างคอมพิวเตอร์และผู้ใช้งานในรูปแบบของภาพ โดยผู้ใช้สามารถเรียนรู้และจดจำข้อมูลผ่านการมองได้มากกว่าการใช้ประสาทสัมผัสอื่น ๆ หรือจะกล่าวว่า Visualization ก็คือ การสร้างมโนภาพของสิ่งต่าง ๆ ที่เราสนใจขึ้นมาในใจ ซึ่งต่อมาได้กลายเป็นการนำภาพมาใช้กับการนำเสนอหรือนามาเป็นกรอบความคิด ซึ่งได้นำไปใช้ในการสนับสนุนการตัดสินใจ

การเลือกรูปแบบ Visualization ให้เหมาะสมกับข้อมูล ในปัจจุบันเป็นยุคเทคโนโลยีเข้าถึงทุกคน ทำให้การรับรู้ข่าวสาร ข้อมูลต่าง ๆ เป็นไปได้ง่าย และรวดเร็วมากขึ้น คนที่นำเสนอข้อมูลจึงต้องนำเสนอข้อมูลให้น่าสนใจ เข้าใจง่าย และรวดเร็ว จึงเกิดการสร้าง Data Visualization ขึ้นมาData Visualization เป็นการใช้ภาพเพื่อแสดงข้อมูลในเชิงปริมาณที่วัดได้ ซึ่งอาจนำเสนอออกมาในรูปแบบ แผนภูมิ กราฟ กราฟิก และอื่นๆ อีกมากมาย เพื่อให้เข้าใจได้

1) แผนภูมิแท่ง (Bar chart) คือกราฟที่ประกอบด้วยรูปสี่เหลี่ยมผืนผ้า (เรียกว่า แท่ง) จำนวนหนึ่ง โดยมีลักษณะเป็นแท่งสูงหรือยาวที่เปลี่ยนแปลงตามขนาด แต่มีความกว้างเท่ากันหมด เราอาจเรียงแท่งเหล่านี้ในทางตั้งหรือทางนอนก็ได้ โดยเว้นระยะช่องว่างตามสมควรและจะต้องเขียนโครงเรื่องจำแนกแต่ละแท่งให้ชัดเจนด้วย แผนภูมิแท่งอาจจะมีการระบายสีหรือแร

เงาเพื่อให้ดูเด่น และในกรณีที่มีการเปรียบเทียบกันหลายแท่ง เช่น แผนภูมิแท่งซับซ้อนหรือเชิงประกอบ จำเป็นจะต้องระบายสีหรือแรเงา เพื่อจำแนกความแตกต่างของแผนภูมิแต่ละชุดที่นำมาเปรียบเทียบกันนั้น วัตถุประสงค์ของการนำเสนอข้อมูลในรูปแผนภูมิแท่งเพื่อแสดงความเปลี่ยนแปลงและการเปรียบเทียบข้อมูล



ภาพที่ 2.1 แผนภูมิแท่ง (Bar Chart)

ที่มา : blog.1moby.com(ม.ป.ป.)

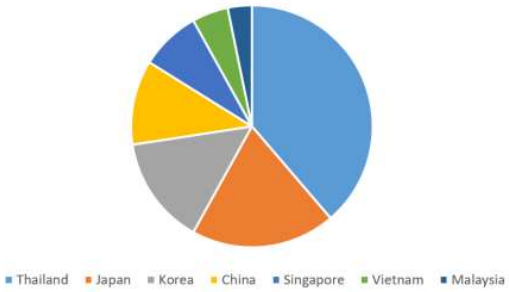
2) แผนภูมิเส้น (Line Charts) แผนภูมิเส้น มีลักษณะคล้ายแผนภูมิแท่ง ซึ่ง ประกอบด้วยแกนตั้งและนอน เพียงแต่เปลี่ยนจากแท่งข้อมูลเป็นจุดบนแผนภูมินั่นเองแผนภูมิ ประเภทนี้เหมาะกับการนำเสนอข้อมูลตัวเลขที่มีความสัมพันธ์กับข้อมูลที่มีลักษณะเป็นช่วงใช้ แสดงให้เห็นการเปลี่ยนแปลงที่เกิดขึ้นตามเวลาเพื่อดูแนวโน้ม รวมถึงสามารถใช้พยากรณ์ แนวโน้มในอนาคตได้เช่น ข้อมูลของยอดขายในแต่ละปี หรือไตรมาส และนำมาวิเคราะห์เพื่อดู แนวโน้ม เป็นต้น

ภาพที่ 2.2 แผนภูมิเส้น (Line Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

3) แผนภูมิวงกลม (Pie Charts) แผนภูมิวงกลมเหมาะกับการนำเสนอข้อมูลที่มีส่วนประกอบย่อยที่รวมกันเป็นส่วนใหญ่ มีการแบ่งส่วนให้ดูง่าย และสวยงามแต่ในทางกลับกันอาจจะดูยากในเรื่องของการประมาณขนาดของแต่ละชิ้น ยิ่งถ้ามีจำนวนชิ้นมาก จะยิ่งแยกยาก

เพราะต้องใช้หลายสี ในการนำเสนอข้อมูล เช่น ส่วนแบ่งทางการตลาด (Market Share), ข้อมูลแสดงส่วนผสมต่าง ๆ เป็นต้น



ภาพที่ 2.3 แผนภูมิวงกลม (Pie Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

4) แผนภูมิโดนัท (Doughnut Charts) แผนภูมิโดนัทมีหลักการออกแบบเช่นเดียวกับแผนภูมิวงกลมแต่สามารถแสดงชุดข้อมูลได้มากกว่า 1 ชุด โดยนำเสนอข้อมูลเป็นวงกลมซ้อนกันหลายๆ ชั้น นั่นเอง

ภาพที่ 2.4 แผนภูมิโดนัท (Doughnut Charts)

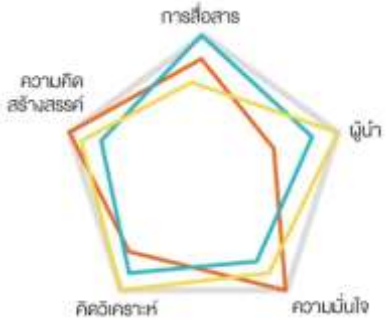
ที่มา : blog.1moby.com(ม.ป.ป.)

5) แผนภูมิพื้นที่ (Area Charts) มีหน้าตาคล้ายแผนภูมิเส้น แต่มีการแรเงาพื้นที่ใต้ เส้นข้อมูล หรือระหว่าง 2 เส้นเพื่อแสดงให้เห็นปริมาณความแตกต่างระหว่างเส้น เหมาะ สำหรับเน้นความสำคัญของการเปลี่ยนแปลงในช่วงเวลาแสดงให้เห็นผลรวมของความแตกต่างระหว่างข้อมูล เช่น ข้อมูลของการซื้อสินค้าในห้างสรรพสินค้าในแต่เดือน ตามหมวดหมู่ต่าง ๆ ไลไป เครื่องสำอางค์ เสื้อผ้าแฟชั่น อาหาร ตามลำดับ

ภาพที่ 2.5 แผนภูมิพื้นที่ (Area Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

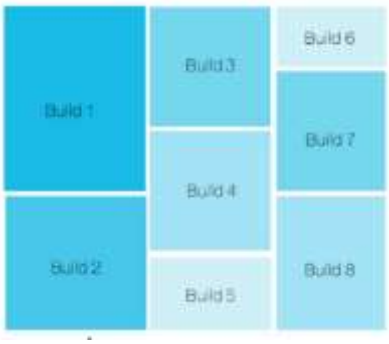
6) แผนภูมิเรดาร์ (Radar Charts) มีลักษณะคล้ายแผนภูมิเส้นที่มีการแสดงผลแบบวงกลมจำนวนเหลี่ยมของเรดาร์เท่ากับจำนวนหัวข้อของข้อมูล แผนภูมินี้ไม่ได้บอกถึงความต่อเนื่องของข้อมูลแต่เหมาะสำหรับการนำเสนอข้อมูลเป็นหัวข้อ แล้วนำมาวิเคราะห์หาจุดอ่อนจุดแข็งของข้อมูลเช่น นำเสนอข้อมูลเกี่ยวกับคุณสมบัติของการรับพนักงานใหม่ เพื่อดูจุดอ่อนจุดแข็งของแต่ละคน เป็นต้น



ภาพที่ 2.6 แผนภูมิเรดาร์ (Radar Charts)

ที่มา : blog.1moby.com(ม.ป.ป.)

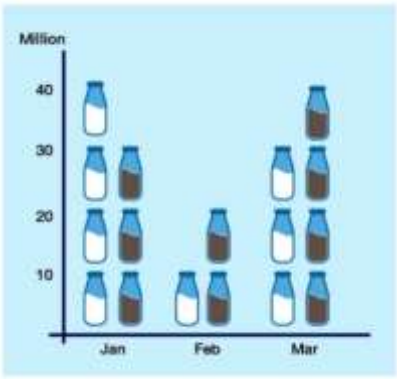
7) แผนภูมิต้นไม้ (Tree Maps) คือการนำเสนอข้อมูลแบบแสดงให้เห็นพื้นที่ แสดง ผลได้ในแบบลำดับชั้น เหมือนแบบโครงสร้างต้นไม้ อาจจะนำเสนอข้อมูลที่ต้องการให้เห็นถึง เขตพื้นที่ แสดงพื้นที่สีที่แตกต่างกันได้



ภาพที่ 2.7 แผนภูมิต้นไม้ (Tree Maps)

ที่มา : blog.1moby.com(ม.ป.ป.)

8) แผนภูมิรูปภาพ (Picture Graph) เป็นแผนภูมิที่ประกอบไปด้วยแกนนอน และ แกนตั้ง แต่เลือกใช้รูปภาพ หรือไอคอนแทนจำนวนของสิ่งของนั้น ๆ เช่น การแสดงผลจำนวน ของนมที่ขายได้ในแต่ละเดือน โดยนำเสนอทั้งนมรสจืด รสช็อกโกแลต เปรียบเทียบในแต่ละ เดือน ซึ่งมีการนำเสนอลักษณะคล้ายกับกราฟแท่ง แต่เปลี่ยนจากแท่งเป็นรูปกราฟิกของนม 2 รสชาติแทน ก็ทำให้การนำเสนอข้อมูลน่าสนใจมากยิ่งขึ้นซึ่งแนวทางการนำเสนอข้อมูลลักษณะนี้ต้องอาศัยความคุ้นชินของคนดู เพื่อแทนสัญลักษณ์ภาพลงไป เช่น เมื่อพูดถึงจำนวนคนอาจจะแทนด้วยภาพไอคอนคนหรือเมื่อพูดถึงจำนวนเงิน ควรแทนภาพเป็นเหรียญเงิน หรือแบงค์แทน ก็จะทำให้คนดูเข้าใจง่ายจากสัญลักษณ์ภาพที่คุ้นเคยอยู่แล้วและยังดึงดูดความสนใจได้มากกว่าการใช้กราฟแท่งสี่เหลี่ยมอีกด้วย



ภาพที่ 2.8 แผนภูมิรูปภาพ (Picture Graph)

ที่มา : blog.1moby.com(ม.ป.ป.)

2.2 ทฤษฎี

2.2.1 ทฤษฎีเกี่ยวกับการจัดการข้อมูลขนาดใหญ่

2.2.1.1 ข้อมูลขนาดใหญ่

ข้อมูลขนาดใหญ่ หมายถึงข้อมูลที่มีขนาดใหญ่ เร็ว หรือซับซ้อนจนยาก หรือเป็นไปไม่ได้ที่จะประมวลผลโดยใช้วิธีการแบบเดิม การเข้าถึงและจัดเก็บข้อมูลจำนวนมาก เพื่อทำการวิเคราะห์มีมานานแล้ว แต่แนวคิดเกี่ยวกับข้อมูลขนาดใหญ่เป็นที่แพร่หลายในช่วง ต้นปีค.ศ. 2000 เมื่อดัก ลานีย์ นักวิเคราะห์อุตสาหกรรมได้ให้คำจำกัดความที่เป็นที่เข้าใจกันใน ขณะนี้ว่า ข้อมูลขนาดใหญ่ประกอบด้วยSVs ดังนี้

Volume (ปริมาณ):องค์กรต่าง ๆ รวบรวมข้อมูลจากหลากหลายแหล่ง ซึ่งรวมถึงธุรกรรมของธุรกิจ อุปกรณ์อัจฉริยะ (IoT) อุปกรณ์อุตสาหกรรม วิดีโอ โซเชียลมีเดีย และอื่น ๆ ในอดีต การจัดเก็บข้อมูลมือเป็นปัญหาใหญ่ – แต่เมื่อค่าใช้จ่ายในการจัดเก็บบนแพลตฟอร์มต่าง ๆ เช่น พื้นที่จัดเก็บข้อมูลส่วนกลาง (Data Lake) และ Hadoop ลดลง ภาระนี้จึงบรรเทาลง

Velocity (ความเร็ว):ด้วยการเติบโตของ Internet of Things ข้อมูลจะถูก ส่งไปยังธุรกิจต่าง ๆ ด้วยความเร็วที่ไม่เคยมีมาก่อนและต้องได้รับการจัดการในเวลาที่เหมาะสม แท็ก RFID, เซ็นเซอร์ และสมาร์ทมิเตอร์ช่วยผลักดันความต้องการในการจัดการกับกระแสข้อมูลเหล่านี้ในแบบเรียลไทม์

Variety (ความหลากหลาย):ข้อมูลมีในทุกรูปแบบ นับตั้งแต่ข้อมูลที่มีโครงสร้าง ตัวเลขในฐานข้อมูลแบบดั้งเดิม ไปจนถึงเอกสารข้อความ อีเมล วิดีโอ เสียง ข้อมูลหุ้น และธุรกรรมทางการเงิน

การวิเคราะห์ Big Data ช่วยให้องค์กรควบคุมข้อมูลของพวกเขาและใช้เพื่อ ระบุโอกาสใหม่ ๆ ในทางกลับกันนำไปสู่การเคลื่อนไหวทางธุรกิจที่ชาญฉลาดเพื่อการดำเนินงานที่มีประสิทธิภาพมากขึ้น ผลกำไรที่สูงขึ้นและลูกค้าที่มีความสุขมากขึ้น ใน รายงาน Big Data ใน บริษัท ขนาดใหญ่โดยผู้อำนวยการฝ่ายวิจัยของ Tom Davenport ให้ สัมภาษณ์ว่ามากกว่า 50 ธุรกิจใช้และทำความเข้าใจว่าพวกเขาใช้ Big Data อย่างไร และ พบว่าสามารถช่วยเหลือธุรกิจได้ดังต่อไปนี้

ลดต้นทุน:เทคโนโลยีข้อมูลขนาดใหญ่เช่น Hadoop และการวิเคราะห์บนคลาวด์นำมาซึ่งความได้เปรียบด้านต้นทุนอย่างมีนัยสำคัญเมื่อพูดถึงการจัดเก็บข้อมูลจำนวนมากรวมทั้งสามารถระบุวิธีการทำธุรกิจที่มีประสิทธิภาพมากขึ้น

เร็วกว่าและตัดสินใจดีกว่า:ด้วยความเร็วของ Hadoop และการวิเคราะห์ในหน่วยความจำรวมกับความสามารถในการวิเคราะห์แหล่งข้อมูลใหม่ๆ ของธุรกิจจะสามารถสร้างข้อมูลได้ทันทีและสามารถดำเนินการต่อได้ทันทีจากการวิเคราะห์นั้น ๆ

ผลิตภัณฑ์และบริการใหม่:ด้วยความสามารถในการวัดความต้องการและความพึงพอใจของลูกค้าผ่านการวิเคราะห์นำมาซึ่งสิ่งที่ลูกค้าต้องการ ซึ่งยกตัวอย่างโดยดาเวนพอร์ท สามารถชี้ให้เห็นว่าด้วยการวิเคราะห์ข้อมูลขนาดใหญ่ของบริษัทต่าง ๆ จะสามารถสร้างผลิตภัณฑ์ใหม่ ๆ เพื่อตอบสนองความต้องการของลูกค้าได้เป็นอย่างดี

2.2.1.2 การจัดการข้อมูลขนาดใหญ่

การรวบรวมข้อมูลของ Big Data เป็นการรวบรวมข้อมูลของจากหลากหลายทั้งที่มาและการใช้งานที่แตกต่างกันอย่างมากมาย ซึ่งกลไกและเทคโนโลยีแบบดั้งเดิม ETL (extract, transform, and load) ไม่สามารถทำได้ ซึ่ง Big Data หรือ ข้อมูลขนาดใหญ่ต้องการเทคนิค วิธีการ และเทคโนโลยีใหม่ในการรวบรวมข้อมูลขนาด เทราไบต์ และอาจจะเป็นระดับเพธาไบต์เลยก็มี ในการรวบรวมข้อมูลนั้นต้องมีการประมวลผล จัดรูปแบบ ให้เหมาะสำหรับการใช้ในการวิเคราะห์หรือใช้งานสำหรับธุรกิจหรือวัตถุประสงค์นั้น ๆ

การจัดการข้อมูลข้อมูลขนาดใหญ่ หรือ Big Data นั้นมีความต้องการสถานที่จัดเก็บขนาด ใหญ่ การจัดเก็บข้อมูลมูลขนาดใหญ่จะเป็นชนิดใดก็ได้ไม่ว่าจะเป็นแบบ on premises หรือ แบบ cloud ขึ้นกับความต้องการหรือความสะดวกในการใช้ ซึ่งเราสามารถใช้และประเมินผลได้ เช่นเดียวกัน บางครั้งก็มีความจำเป็นที่ต้องจัดเก็บไว้ใกล้กับแหล่งข้อมูลหรือข้อมูลบางอัน ต้องการความยืดหยุ่นสูงและไม่ต้องการบริหารจัดการก็ใช้เป็นแบบ Cloud ซึ่งกำลังเป็นที่นิยม กันเป็นอย่างมาก

การวิเคราะห์การลงทุนสร้างข้อมูลขนาดใหญ่ หรือ Big data จะมีประโยชน์หรือคุ้มค่า ก็ต่อเมื่อคุณใช้และวิเคราะห์ข้อมูล การวิเคราะห์ข้อมูลทำให้เกิดความกระจ่างและชัดเจนในชุด ข้อมูลที่คุณมีอยู่ การสำรวจข้อมูลยังทำให้เราค้นพบสิ่งใหม่ แชร์สิ่งที่

ค้นพบใหม่ๆต่อคนอื่น สร้างรูปแบบจำลองข้อมูล ด้วยการเรียนรู้ของเครื่องจักรและปัญญาประดิษฐ์ AI และนำข้อมูล เหล่านั้นไปใช้งาน

2.2.1.3 เทคโนโลยีการประมวลผลข้อมูล

การประมวลผลข้อมูลที่เป็น Big Data จะมีทั้งการวิเคราะห์ข้อมูลที่เป็น business intelligence (BI) เพื่อที่จะดึงข้อมูลมานำเสนอ หรือการทำ Predictive Analytics โดยใช้ หลักการของ Data Science ความยากของการประมวลผลคือต้องการความเร็วในการประมวลผลข้อมูลที่นอกจากมีขนาดใหญ่แล้วบางครั้งยังเป็นข้อมูลที่ไม่มีโครงสร้าง ดังนั้นต้องจึงมีการนำเทคโนโลยีหรือภาษาต่าง ๆ มาเพื่อให้สามารถประมวลผลข้อมูลได้ ซึ่งในบางครั้งหน่วยงานอาจต้องพิจารณาต้องเลือกใช้

2.2.2 ทฤษฎีเกี่ยวกับการทำเหมืองข้อมูล

เหมืองข้อมูล คือกระบวนการที่กระทำกับข้อมูลจำนวนมากเพื่อค้นหารูปแบบและ ความสัมพันธ์ที่ซ่อนอยู่ในชุดข้อมูลนั้น ในปัจจุบันการทำเหมืองข้อมูลได้ถูกนำไประยุกต์ใช้ใน งานหลายประเภท ทั้งในด้านธุรกิจที่ช่วยในการตัดสินใจของผู้บริหาร ในด้านวิทยาศาสตร์และ การแพทย์รวมทั้งในด้านเศรษฐกิจและสังคม

2.2.2.1 การทำเหมืองข้อมูล (Data Mining) เปรียบเสมือนวิวัฒนาการหนึ่งในการ จัดเก็บและตีความหมายข้อมูล จากเดิมที่มีการจัดเก็บข้อมูลอย่างง่ายๆ มาสู่การจัดเก็บในรูปแบบฐานข้อมูลที่สามารถดึงข้อมูลสารสนเทศมาใช้จนถึงการทำเหมืองข้อมูลที่สามารถค้นพบความรู้ที่ซ่อนอยู่ในข้อมูล หรือจะแยกๆ เป็นข้อๆ ได้ดังนี้

1) กระบวนการหรือการเรียงลำดับของการค้นข้อมูลจำนวนมากและเก็บข้อมูลที่เกี่ยวข้อง

2) การนำมาใช้โดยหน่วยงานทางธุรกิจและนักวิเคราะห์ทางการเงินหรือ การนำมาใช้งานในด้านวิทยาศาสตร์เพื่อเอาข้อมูลขนาดใหญ่ที่สร้างโดยวิธีการทดลองและการ สังเกตการณ์ที่ทันสมัย

3) การสกัดหรือแยกข้อมูลที่เป็นประโยชน์จากข้อมูลขนาดใหญ่หรือฐานข้อมูล

4) การวางแผนทรัพยากรขององค์กรโดยสามารถวิเคราะห์ทางสถิติและ ตรรกะของข้อมูลขนาดใหญ่เป็นการมองหารูปแบบที่สามารถช่วยการตัดสินใจได้

2.2.2.2 ขั้นตอนการทำเหมืองข้อมูล

ประกอบด้วยขั้นตอนการทำงานย่อยที่จะเปลี่ยนข้อมูลดิบให้กลายเป็นความรู้ ประกอบด้วยขั้นตอนดังนี้

1) Business Understanding เน้นไปที่การทำความเข้าใจในงาน ระบุโอกาส และหาปัญหาที่จะเกิดขึ้น กำหนดขอบเขตของข้อมูลที่จะนำวิเคราะห์ ซึ่งต้องสามารถระบุ ผลลัพธ์ที่มีได้

2) Data Understanding ทำความเข้าใจข้อมูลโดยการรวบรวมข้อมูลที่เกี่ยวข้อง คัดเลือกให้เหลือเพียงข้อมูลที่มีความถูกต้องและสำคัญต่องานมาทำการวิเคราะห์

3) Data Preparation ทำการแปลงข้อมูล (Raw Data) ให้กลายเป็นข้อมูล ที่สามารถนำมาช่วยในการวิเคราะห์ต่อไปได้ ขั้นตอนนี้จะใช้เวลามากที่สุดในทุกขั้นตอน เพราะ คุณภาพของงานที่ได้จะดีเพียงใดขึ้นอยู่กับคุณภาพข้อมูลที่จัดเตรียมในขั้นนี้ การเตรียมข้อมูล ประกอบด้วย การคัดเลือกข้อมูล การกลั่นกรองข้อมูล และแปลงรูปแบบของข้อมูล

4) Modeling การสร้างแบบจำลองเพื่อวิเคราะห์ข้อมูลที่ได้จากขั้นตอนที่ 3 พร้อมทดสอบผลลัพธ์แบบจำลองเพื่อให้ได้คำตอบที่ดีที่สุด บางครั้งอาจมีการย้อนกลับไปปรับ การเตรียมข้อมูลเพื่อให้ได้แบบจำลองที่เหมาะสมที่สุด

5) Evaluation การประเมินผลลัพธ์ที่ได้ก่อนที่จะนำไปใช้จริง ว่าตรงกับวัตถุประสงค์หรือเป้าหมายที่ได้ตั้งไว้หรือมีความน่าเชื่อถือมากน้อยเพียงใด หากไม่ได้ผลลัพธ์ตามวัตถุประสงค์ต้องย้อนกลับไปปรับปรุงแก้ไขการดำเนินงานในขั้นตอนก่อนหน้า

6) Deployment การนำเอาข้อมูลที่เป็นผลลัพธ์จากทั้งหมด มาใช้ประโยชน์ตามวัตถุประสงค์ และทำการประเมินผลลัพธ์ที่ได้ว่ามีประสิทธิภาพตรงตามวัตถุประสงค์ที่ตั้งไว้หรือไม่

2.2.3 ทฤษฎีเกี่ยวกับการออกแบบเว็บไซต์

2.2.3.1 หลักในการออกแบบเว็บไซต์

หน้าเว็บเป็นสิ่งแรกที่ผู้ใช้จะได้เห็นขณะที่เปิดเข้าสู่เว็บไซต์ และยังเป็นสิ่ง แรกที่แสดงถึงประสิทธิภาพในการออกแบบเว็บไซต์อีกด้วย หน้าเว็บจึงเป็นสิ่งสำคัญมาก เพราะเป็นสื่อกลางให้ผู้ชมสามารถใช้ประโยชน์จากข้อมูลของระบบงานของเว็บไซต์นั้นได้ โดยปกติหน้าเว็บจะประกอบด้วย รูปภาพ ตัวอักษร สีพื้น ระบบเนวิเกชั่น และองค์ประกอบอื่น ๆ ที่ช่วยสื่อความหมายของเนื้อหาและอำนวยความสะดวกต่อการใช้งานหลักสำคัญในการออกแบบหน้าเว็บก็คือ การใช้รูปภาพและองค์ประกอบต่าง ๆ ร่วมกันเพื่อสื่อความหมายเกี่ยวกับเนื้อหาหรือลักษณะสำคัญของเว็บไซต์ โดยมีเป้าหมายสำคัญเพื่อการสื่อความหมายที่ชัดเจนและน่าสนใจ บนพื้นฐานของความเรียบง่ายและความสะดวกของผู้ใช้ การออกแบบเว็บไซต์ ต้องคำนึงถึง

1) ความเรียบง่าย ได้แก่ มีรูปแบบที่เรียบง่าย ไม่ซับซ้อน และใช้งานได้สะดวก ไม่มีกราฟิกหรือตัวอักษรที่เคลื่อนไหวอยู่ตลอดเวลา ชนิดและสีของตัวอักษรไม่มากจนเกินไปทำให้วุ่นวาย

2) ความสม่ำเสมอ ได้แก่ ใช้รูปแบบเดียวกันตลอดทั้งเว็บไซต์ เช่น รูปแบบของหน้า สไตล์ของกราฟิก ระบบเนวิเกชันและโทนสี ควรมีความคล้ายคลึงกันตลอดทั้งเว็บไซต์

3) ความเป็นเอกลักษณ์ การออกแบบเว็บไซต์ควรคำนึงถึงลักษณะของ องค์กร เพราะรูปแบบของเว็บไซต์จะสะท้อนถึงเอกลักษณ์และลักษณะขององค์กรนั้น ๆ เช่น ถ้าเป็นเว็บไซต์ของทาง ราชการ จะต้องดูน่าเชื่อถือไม่เหมือนสวนสนุก ฯลฯ

4) เนื้อหาที่มีประโยชน์ เนื้อหาเป็นสิ่งที่สำคัญที่สุดในเว็บไซต์ ดังนั้นควร จัดเตรียมเนื้อหาและข้อมูลที่ผู้ใช้ต้องการให้ถูกต้อง และสมบูรณ์ มีการปรับปรุงและเพิ่มเติมให้ ทันเหตุการณ์อยู่เสมอ เนื้อหาไม่ควรซ้ำกับเว็บไซต์อื่น จึงจะดึงดูดความสนใจ

5) ระบบเนวิเกชันที่ใช้งานง่าย ต้องออกแบบให้ผู้ใช้เข้าใจง่ายและใช้งาน สะดวก ใช้กราฟิกที่สื่อความหมายร่วมกับคำอธิบายที่ชัดเจน มีรูปแบบและลำดับของรายการที่ สม่ำเสมอ เช่น วางไว้ ตำแหน่งเดียวกันของทุกหน้า

6) ลักษณะที่น่าสนใจ หน้าตาของเว็บไซต์จะต้องมีความสัมพันธ์กับคุณภาพขององค์ประกอบต่าง ๆ เช่น คุณภาพของกราฟิกที่จะต้องสมบูรณ์ การใช้สี การใช้

ตัวอักษรที่อ่านง่าย สบายตา การใช้โทนสีที่เข้ากันลักษณะหน้าตาที่น่าสนใจนั้นขึ้นอยู่กับความชอบของแต่ละบุคคล

7) การใช้งานอย่างไม่จำกัด ผู้ใช้ส่วนใหญ่สามารถเข้าถึงได้มากที่สุด เลือกใช้บราวเซอร์ชนิดใดก็ได้ในการเข้าถึงเนื้อหาสามารถแสดงผลได้ทุกระบบปฏิบัติการและความละเอียดหน้าจอต่าง ๆ กันอย่างไม่มีปัญหาเป็นลักษณะสำคัญสำหรับผู้ใช้ที่มีจำนวนมาก

8) คุณภาพในการออกแบบ การออกแบบและเรียบเรียงเนื้อหาอย่างรอบคอบ สร้างความรู้สึกว่าเว็บไซต์มีคุณภาพ ถูกต้อง และเชื่อถือได้

9) ลิงค์ต่าง ๆ จะต้องเชื่อมโยงไปหน้าที่มีอยู่จริงและถูกต้อง ระบบการทำงานต่าง ๆ ในเว็บไซต์จะต้องมีความแน่นอนและทำหน้าที่ได้อย่างถูก

2.2.3.3 โครงสร้างเว็บไซต์

1) เว็บที่มีโครงสร้างแบบเรียงลำดับ (Sequential Structure) เป็นโครงสร้างแบบธรรมดาที่ใช้กันมากที่สุดเนื่องจากง่ายต่อการจัดระบบข้อมูล ข้อมูลที่นิยม จัดด้วยโครงสร้างแบบนี้มักเป็นข้อมูลที่มีลักษณะเป็นเรื่องราวตามลำดับของเวลา เช่น การเรียงลำดับตามตัวอักษร ดรรชนี สารานุกรม หรืออภิธานศัพท์ โครงสร้างแบบนี้ เหมาะกับเว็บไซต์ที่มีขนาดเล็ก เนื้อหาไม่ซับซ้อนใช้การลิงก์ (Link) ไปทีละหน้า ทิศทางของการเข้าสู่เนื้อหา (Navigation) ภายในเว็บจะเป็นการดำเนินเรื่องในลักษณะเส้นตรง โดยมี ปุ่มเดินหน้าถอยหลังเป็นเครื่องมือหลักในการกำหนดทิศทาง ข้อเสียของโครงสร้างระบบนี้คือ ผู้ใช้ไม่สามารถกำหนดทิศทางการเข้าสู่เนื้อหาของตนเองได้ ทำให้เสียเวลาเข้าสู่เนื้อ

ภาพที่ 2.9 เว็บที่มีโครงสร้างแบบเรียงลำดับ

ที่มา : sites.google.com(ม.ป.ป.)

2) เว็บที่มีโครงสร้างแบบลำดับขั้น (Hierarchical Structure) เป็นวิธีที่ดีที่สุดวิธีหนึ่งในการจัดระบบโครงสร้างที่มีความซับซ้อนของข้อมูล โดยแบ่งเนื้อหา ออกเป็นส่วนต่างๆ และมีรายละเอียดย่อยๆ ในแต่ละส่วนลดหลั่นกันมาในลักษณะแนวคิดเดียวกับ แผนภูมิ

องค์กร จึงเป็นการง่ายต่อการทำความเข้าใจกับโครงสร้างของเนื้อหาในเว็บลักษณะนี้ ลักษณะเด่นเฉพาะของ เว็บประเภทนี้คือการมีจุดเริ่มต้นที่จุดร่วมจุดเดียว นั่นคือ โฮมเพจ (Homepage) และเชื่อมโยงไปสู่เนื้อหา ในลักษณะเป็นลำดับจากบนลงล่าง

ภาพที่ 2.10 เว็บที่มีโครงสร้างแบบลำดับขั้น

ที่มา : sites.google.com(ม.ป.ป.)

3) เว็บที่มีโครงสร้างแบบตาราง (Grid Structure) โครงสร้างรูปแบบนี้มีความซับซ้อนมากกว่ารูปแบบที่ผ่านมา การออกแบบเพิ่มความยืดหยุ่น ให้แก่การเข้าสู่เนื้อหาของผู้ใช้ โดยเพิ่มการเชื่อมโยงซึ่งกันและกันระหว่างเนื้อหาแต่ละส่วน เหมาะแก่ การแสดงให้เห็นความสัมพันธ์กันของเนื้อหา การเข้าสู่เนื้อหาของผู้ใช้จะไม่เป็นลักษณะเชิงเส้นตรงเนื่องจากผู้ใช้สามารถเปลี่ยนทิศทางการเข้าสู่เนื้อหาของตนเองได้

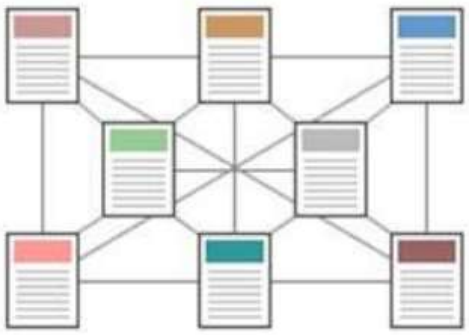
ภาพที่ 2.11 เว็บที่มีโครงสร้างแบบตาราง

ที่มา : sites.google.com(ม.ป.ป.)

ในการจัดระบบโครงสร้างแบบนี้ เนื้อหาที่นำมาใช้แต่ละส่วนควรมีลักษณะที่เหมือนกัน และ สามารถใช้รูปแบบร่วมกัน หลักการออกแบบคือนำหัวข้อทั้งหมดมาบรรจุลงในที่เดียวกันซึ่งโดยทั่วไป จะเป็นหน้าแผนภาพ (Map Page) ที่แสดงในลักษณะเดียวกับ

โครงสร้างของเว็บ เมื่อผู้ใช้คลิกเลือก หัวข้อใด ก็จะเข้าไปสู่หน้าเนื้อหา (Topic Page) ที่แสดงรายละเอียดของหัวข้อนั้นๆ และภายในหน้านั้น ก็จะมีการเชื่อมโยงไปยังหน้ารายละเอียดของหัวข้ออื่นที่เป็นเรื่องเดียวกัน นอกจากนี้ยังสามารถนำ โครงสร้างแบบเรียงลำดับและแบบลำดับขั้นมาใช้ร่วมกันได้อีกด้วย ถึงแม้โครงสร้างแบบนี้ อาจจะสร้างความยุ่งยากในการเข้าใจได้ และอาจเกิดปัญหาการคงค้าง ของหัวข้อ (Cognitive Overhead) ได้ แต่จะเป็นประโยชน์ที่สุดเมื่อผู้ใช้ได้เข้าใจถึงความสัมพันธ์ ระหว่างเนื้อหา ในส่วนของการออกแบบจำเป็นจะต้องมีการวางแผนที่ดี เนื่องจากมีการเชื่อมโยงที่เกิดขึ้น ได้หลายทิศทาง นอกจากนี้การปรับปรุงแก้ไขอาจ เกิดความยุ่งยากเมื่อต้องเพิ่มเนื้อหาในภายหลัง

4) เว็บที่มีโครงสร้างแบบใยแมงมุม (Web Structure) โครงสร้างประเภท นี้จะมีความยืดหยุ่นมากที่สุด ทุกหน้าในเว็บสามารถจะเชื่อมโยงไปถึงกัน ได้หมด เป็นการสร้าง รูปแบบการเข้าสู่เนื้อหาที่เป็นอิสระ ผู้ใช้สามารถกำหนดวิธีการเข้าสู่เนื้อหาได้ด้วยตนเอง การ เชื่อมโยงเนื้อหาแต่ละหน้าอาศัยการโยงใยข้อความที่มีมโนทัศน์ (Concept) เหมือนกัน ของแต่ ละหน้าในลักษณะของไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย โครงสร้างลักษณะนี้จัดเป็นรูปแบบที่ ไม่มี โครงสร้างที่แน่นอนตายตัว (Unstructured) นอกจากนี้การเชื่อมโยงไม่ได้จำกัดเฉพาะเนื้อหา ภายในเว็บนั้น ๆ แต่สามารถเชื่อมโยงออกไปสู่เนื้อหาจากเว็บภายนอกได้



ภาพที่ 2.12 เว็บที่มีโครงสร้างแบบใยแมงมุม

ที่มา : sites.google.com(ม.ป.ป.)

ลักษณะการเชื่อมโยงในเว็บนั้น นอกเหนือจากการใช้ไฮเปอร์เท็กซ์หรือไฮเปอร์มีเดีย กับข้อความที่มีมโนทัศน์ (Concept) เหมือนกันของแต่ละหน้าแล้ว ยังสามารถใช้ลักษณะการเชื่อมโยง จากรายการที่รวบรวมชื่อหรือหัวข้อของเนื้อหาแต่ละหน้าไว้ ซึ่งรายการนี้

จะปรากฏอยู่บริเวณใด บริเวณหนึ่งในหน้าจอ ผู้ใช้สามารถคลิกที่หัวข้อใดหัวข้อหนึ่งในรายการเพื่อเลือกที่จะเข้าไปสู่หน้าใด ๆ ก็ได้ตามความต้องการ ข้อดีของรูปแบบนี้คือง่ายต่อผู้ใช้ในการท่องเที่ยวบนเว็บ โดยผู้ใช้สามารถกำหนดทิศทาง การเข้าสู่เนื้อหาได้ด้วยตนเอง แต่ข้อเสียคือถ้ามีการเพิ่มเนื้อหาใหม่ๆ อยู่เสมอจะเป็นการยากในการ ปรับปรุง นอกจากนี้การเชื่อมโยงระหว่างข้อมูลที่มีมากมายนั้นอาจทำให้ผู้ใช้เกิดการสับสนและ เกิดปัญหาการคงค้างของหัวข้อ (Cognitive Overhead) ได้

2.2.3.3 การใช้สีในการออกแบบเว็บไซต์

การสร้างสีสันบนหน้าเว็บเป็นสิ่งที่สื่อความหมายของเว็บไซต์ได้อย่างชัดเจน การเลือกใช้สีให้เหมาะสม กลมกลืน ไม่เพียงแต่จะสร้างความพึงพอใจให้กับผู้ใช้ แต่ยังสามารถทำให้เห็นถึงความแตกต่างระหว่างเว็บไซต์ได้ สีเป็นองค์ประกอบหลักสำหรับการตกแต่งเว็บ จึงจำเป็นอย่างยิ่งที่จะต้องทำความเข้าใจเกี่ยวกับการใช้สีระบบสีที่แสดงบนจอคอมพิวเตอร์ มีระบบการแสดงผลผ่านหลอดลำแสงที่เรียกว่า CRT (Cathode ray tube) โดยมีลักษณะระบบสีแบบบวก อาศัยการผสมของของแสงสีแดง สีเขียว และสีน้ำเงิน หรือระบบสี RGB สามารถกำหนดค่าสีจาก 0 ถึง 255 ได้ จากการรวมสีของแม่สีหลักจะทำให้เกิดแสงสีขาว มีลักษณะเป็นจุดเล็ก ๆ บนหน้าจอไม่สามารถมองเห็นด้วยตาเปล่าได้ จะมองเห็นเป็นสีที่ถูกผสมเป็นเนื้อสีเดียวกันแล้ว จุดแต่ละจุดหรือพิกเซล (Pixel) เป็นส่วนประกอบของภาพบนหน้าจอคอมพิวเตอร์ โดยจำนวนบิตที่ใช้ในการกำหนดความสามารถของการแสดงสีต่าง ๆ เพื่อสร้างภาพบนจอนั้นเรียกว่า บิตเด็ป (Bit depth) ในภาษา HTML มีการกำหนดสีด้วยระบบเลขฐานสิบหก ซึ่งมีเครื่องหมาย (#) อยู่ด้านหน้าและตามด้วยเลขฐานสิบหกจำนวนอักษรอีก 6 หลัก โดยแต่ละไบต์ (byte) จะมีตัวอักษรสองตัว แบ่งออกเป็น 3 กลุ่ม เช่น #FF12AC การใช้ตัวอักษรแต่ละไบต์นี้เพื่อกำหนดระดับความเข้มของแม่สีแต่ละสีของชุดสี RGB โดย 2 หลักแรกแสดงถึงความเข้มของสีแดง 2 หลักต่อมา แสดงถึงความเข้มของสีเขียว 2 หลักสุดท้ายแสดงถึงความเข้มของสีน้ำเงิน

สีมีอิทธิพลในเรื่องของอารมณ์การสื่อความหมายที่เด่นชัด กระตุ้นการรับรู้ทางด้านจิตใจมนุษย์ สีแต่ละสีให้ความรู้สึก อารมณ์ที่ไม่เหมือนกัน สีบางสีให้ความรู้สึกสงบ บางสีให้ความรู้สึกตื่นเต้นรุนแรง สีจึงเป็นปัจจัยสำคัญอย่างยิ่งต่อการออกแบบเว็บไซต์ ดังนั้นการเลือกใช้โทนสีภายในเว็บไซต์เป็นการแสดงถึงความแตกต่างของสีที่แสดงออกทาง

อารมณ์ มีชีวิตชีวาหรือเศร้าโศก รูปแบบของสีที่สายตาของมนุษย์มองเห็น สามารถแบ่งออกเป็น 3 กลุ่ม คือ

1) สีโทนร้อน (Warm Colors) เป็นกลุ่มสีที่แสดงถึงความสุข ความปลอบโยน ความอบอุ่น และดึงดูดใจ สีกลุ่มนี้เป็นกลุ่มสีที่ช่วยให้หายจากความเฉื่อยชา มีชีวิตชีวามากยิ่งขึ้น

2) สีโทนเย็น (Cool Colors) แสดงถึงความที่ดูสุภาพ อ่อนโยน เรียบร้อย เป็นกลุ่มสีที่มีคนชอบมากที่สุด สามารถโน้มนาวในระยะไกลได้

3) สีโทนกลาง (Neutral Colors) สีที่เป็นกลาง ประกอบด้วย สีดำ สีขาว สี เทา และสีน้ำตาล กลุ่มสีเหล่านี้คือ สีกลางที่สามารถนำไปผสมกับสีอื่น ๆ เพื่อให้เกิดสีกลางขึ้นมา

2.2.4 ทฤษฎีเกี่ยวกับชุดคำสั่งภาษาของการออกแบบเว็บไซต์

2.2.4.1 CSS (ย่อมาจาก Cascading Style Sheet) มักเรียกโดยย่อว่า "สไตล์ชีต" คือภาษาที่ใช้เป็นส่วนของการจัดรูปแบบการแสดงผลเอกสาร HTML โดยที่ CSS กำหนดกฎเกณฑ์ในการระบุรูปแบบ (หรือ "Style") ของเนื้อหาในเอกสาร อันได้แก่ สีของข้อความ สีพื้นหลัง ประเภทตัวอักษร และการจัดวางข้อความ ซึ่งการกำหนดรูปแบบ หรือ Style นี้ใช้หลักการของการแยกเนื้อหาเอกสาร HTML ออกจากคำสั่งที่ใช้ในการจัดรูปแบบการแสดงผล กำหนดให้รูปแบบของการแสดงผลเอกสาร ไม่ขึ้นอยู่กับเนื้อหาของเอกสาร เพื่อให้ง่ายต่อการจัดรูปแบบการแสดงผลลัพธ์ของเอกสาร HTML โดยเฉพาะในกรณีที่มีการเปลี่ยนแปลงเนื้อหาเอกสารบ่อยครั้ง หรือต้องการควบคุมให้รูปแบบการแสดงผลเอกสาร HTML มีลักษณะของความสม่ำเสมอทั่วกันทุกหน้าเอกสารภายในเว็บไซต์เดียวกัน โดยกฎเกณฑ์ในการกำหนดรูปแบบ (Style) เอกสาร HTML ถูกเพิ่มเข้ามาครั้งแรกใน HTML 4.0 เมื่อปีพ.ศ. 2539 ในรูปแบบของ CSS level 1 Recommendations ที่กำหนดโดย องค์กร World Wide Web Consortium หรือ W3C

2.2.4.2 Bootstrap คือชุดคำสั่งที่ประกอบด้วยภาษา CSS, HTML และ Javascript เป็นชุดคำสั่งที่ถูกพัฒนาขึ้นมาเพื่อกำหนดกรอบหรือรูปแบบการพัฒนาเว็บไซต์ในส่วนของการปฏิสัมพันธ์กับผู้ใช้งานเว็บไซต์ (User Interface) เราจึงสามารถเรียก Bootstrapว่า

เป็น Front-end framework คือใช้สำหรับ พัฒนาเว็บไซต์ส่วนการแสดงผล ซึ่งแตกต่างจากภาษาประเภท Server Side Script อย่าง PHP, Python หรือภาษาอื่น ๆ

2.2.5 ทฤษฎีเกี่ยวกับการ visualization

Data Visualization หรือ Information Visualization คือ การนำข้อมูลในเชิงปริมาณ ทั้งที่จัดเก็บไว้ในรูปแบบของข้อมูลจำนวนน้อย และข้อมูลจำนวนมหาศาล (Big Data) มา ประมวลผล จากนั้นจึงนำมาแสดงผลในรูปแบบของกราฟ แผนภูมิอินฟอร์เมชันกราฟิก หรือ แม้กระทั่งอินเทอร์แอคทีฟกราฟิก ที่ผู้บริโภคสามารถคลิกหรือมีปฏิสัมพันธ์กับกราฟิกนั้น ๆ ได้ ซึ่งปัจจุบันสื่อหลายสำนักในประเทศไทย เริ่มมีการนำเสนอข่าวโดยการนำข้อมูลแบบประยุกต์ ด้วยภาพ (Data Visualization) การถ่ายทอดข้อมูลในเชิงปริมาณที่มีความซับซ้อนได้อย่างมี ประสิทธิผล ซึ่งคำว่า “ประสิทธิผล” ในที่นี้หมายถึงมีความชัดเจน (Clarity), มีความแม่นยำ (Precision), และมีประสิทธิภาพ (Efficiency) หากไม่มีการทำ Data Visualization แล้วอาจทำให้ เราไม่สามารถค้นพบนัยยะของข้อมูลในแง่ของแนวโน้ม, รูปแบบพฤติกรรม, และความสัมพันธ์ เชื่อมโยงได้

Visualization คือ การจินตนาการ หรือสร้างภาพขึ้นในความคิด ซึ่งเป็นกระบวนการ ทำให้เกิดความรู้ความเข้าใจสำหรับเรื่องที่จินตนาการยาก เข้าใจยาก วิธีการที่เป็นทางลัดก็คือ การสร้าง ภาพ ให้เป็น บันไดความคิด ไปสู่ การใช้ความคิดอีกระดับ บันไดนี้จะช่วยตัดปริมาณ ข้อมูล ช่วยลดภาระการคำนวณหรือการนำไปผ่านหลากหลายกระบวนการความคิด เพื่อ นำไปสู่คำตอบที่ต้องการได้โดยเร็วและถูกต้อง

Visualization System คือระบบ ซึ่งประกอบด้วยกระบวนการต่าง ๆ ที่ออกแบบมา เพื่อสร้าง รักษา นำไปใช้ และปรับปรุงทัศนสนเทศ เพื่อทำให้เกิดความรู้ความเข้าใจ อย่างถูกต้อง รวดเร็ว และได้ผลเป็นอย่างสูง

Data Visualization เป็นการนำข้อมูลมาผสมผสานกับจินตนาการ เพื่อสร้างภาพใน ความคิดขึ้นมา ซึ่งมีกระบวนการนำเสนอข้อมูลที่มีความซับซ้อนหรือข้อมูลเชิงปริมาณให้ สามารถเข้าใจได้ง่าย ในแบบของ กราฟ แผนภูมิ

2.2.5.1 โปรแกรมสำหรับการสร้างแดชบอร์ด (Data Visualization)

Tools ที่นิยมใช้ในตอนนี้ได้แก่ Tableau, Microsoft Power BI, Qlik View, Google Charts, Fusion Charts,Data wrapper และอื่น ๆ อีกมากมาย

ภาพที่ 2.13 Tools Data Visualization

ที่มา : autosoft.in.th(ม.ป.ป.)

2.2.5.2 รูปแบบในการใช้ Data Visualization

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

ภาพที่ 2.14 การนำเสนอแบบทิศทางหรือแนวโน้ม (Trending)

ที่มา : autosoft.in.th(ม.ป.ป.)

2.3. เครื่องมือในการออกแบบและวิเคราะห์ข้อมูล

2.3.1 คอนฟิวชั่น เมทริกซ์ (Confusion Matrix) ถือเป็นเครื่องมือสำคัญในการประเมินผลลัพธ์ของการทำนาย หรือ Prediction ที่ทำนายจากModelที่เราสร้างขึ้น ในMachine learning โดยมีไอเดียจากการวัดว่า สิ่งที่เราคิด (Model ทำนาย) กับ สิ่งที่เกิดขึ้นจริง

	Actually Positive (1)	Actually Negative (0)
Predicted Positive (1)	True Positives (TPs)	False Positives (FPs)
Predicted Negative (0)	False Negatives (FNs)	True Negatives (TNs)

ภาพที่ 2.15 คอนฟิวชั่น เมทริกซ์ (Confusion Matrix)

ที่มา : medium.com (2562)

True Positive (TP)= สิ่งที่ทำนาย ตรงกับสิ่งที่เกิดขึ้นจริง ในกรณี ทำนายว่าจริง และสิ่งที่เกิดขึ้นก็คือ จริง

True Negative (TN)= สิ่งที่ทำนายตรงกับสิ่งที่เกิดขึ้น ในกรณี ทำนายว่า ไม่จริง และสิ่งที่เกิดขึ้นก็คือ ไม่จริง

False Positive (FP)= สิ่งที่ทำนายไม่ตรงกับสิ่งที่เกิดขึ้น คือทำนายว่า จริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง

False Negative (FN)= สิ่งที่ทำนายไม่ตรงกับที่เกิดขึ้นจริง คือทำนายว่าไม่จริง แต่สิ่งที่เกิดขึ้น คือ จริง

โดย TP,TN,FP,FN ในตารางจะแทนด้วยค่าความถี่

เราสามารถใช้ Confusion Matrix มาคำนวน การประเมินประสิทธิภาพของการทำนายด้วยModel ของเรา ในรูปแบบค่าต่างๆได้หลายค่า ได้แก่

Accuracy (ความถูกต้องที่เราทายได้ตรงกับสิ่งที่เกิดขึ้นจริง)

Accuracy (ความถูกต้อง) = (TPs + TNs) / (TPs+TNs+FPs + FNs)

หรือกล่าวได้ว่า Accuracy = ผลรวมของตัวเลขบนเส้นทแยงมุมในตาราง Confusion Matrix / จำนวนobservationsทั้งหมด

โดย ความเป็นจริงแล้ว Confusion matrix ไม่จำเป็นต้องเป็นแบบ 2x2 หรือมีผลลัพธ์แค่ 2 แบบ เสมอไป โดยอาจเป็น 3x3,4x4, nxn ก็ได้ โดยวิธีการหาAccuracy ก็ใช้แบบเดิม คือ ผลรวมของตัวเลขบนเส้นทแยงมุมในตารางConfusion Matrix / จำนวน observationsทั้งหมด

Precision (ค่าความแม่นยำ) เป็นการเปรียบเทียบ การทำนายที่ถูกต้องว่า จริง และก็เกิดขึ้นจริง (TP) กับ การทำนายว่า จริง แต่สิ่งที่เกิดขึ้น คือ ไม่จริง (FP)

Precision = TPs / (TPs + FPs)

Recall (ความถูกต้องของการทำนายว่าจะเป็น “จริง” เทียบกับ จำนวนครั้งของเหตุการณ์ทั้งทำนาย และ เกิดขึ้น ว่า “เป็นจริง”)

Recall = TPs/(TPs+FNs)

F1 Score F1–Score เป็นค่าเฉลี่ยแบบ harmonic mean ระหว่าง precision และ recall จุดประสงค์ของการสร้าง F1 ขึ้นมา คือ เพื่อเป็น single metric ที่วัดความสามารถของโมเดล

F1 = 2 x (Precision x Recall)/(Precision + Recall)

ตัวอย่างเสริมการอธิบาย เพื่อให้เข้าใจมากขึ้น

ในกรณีที่เราอยากทราบว่า โมเดลของเราทายแม่นขนาดไหน คือ ทายถูกต้องว่าเป็น Spam จากการพยายามทำนายทั้งหมด เราต้องใช้ Precision ก็คือ Precision ของทำนายSpam = 20 / 32 = 0.625

แต่ถ้าเราต้องการทราบว่า โมเดลที่เราสร้างขึ้น สามารถตรวจจับ Spamได้ถูกต้องขนาดไหน จากSpam Email ทั้งหมด เราต้องใช้ Recall Recall ของการตรวจจับ Spam = 20 /38 = 0.526

แต่ในกรณีที่เราต้องการหาประสิทธิภาพของโมเดลการทำนายนี้ ที่ต้องมีทั้งการทายถูกต้องว่าสิ่งที่เจออันนั้นคือ Spam จริงๆ และในขณะเดียวกันก็ต้องตรวจจับ Spamได้ด้วย เราก็ต้องเลือกใช้ F1 score ก็คือ เป็นค่าเฉลี่ยของทั้ง Precision และ Recall

F1 ของ Model นี้ = 2 * ((0.625 * 0.526) / (0.625 + 0.526)) = 0.571

2.3.2 การตัดสินใจแบบโครงสร้างต้นไม้ (Decision Tree)

เทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นเทคนิคหนึ่งของ Classification ซึ่งเป็นวิธีการ แบ่งประเภทหรือแยกหมวดหมู่ข้อมูล โดย Classification นั้นเป็นเทคนิคหนึ่งของเหมืองข้อมูล (Data Mining)

2.3.2.1 เทคนิคต้นไม้ตัดสินใจ (Decision Tree) คือ แบบจำลองทางคณิตศาสตร์เพื่อการหาทาง เลือกที่ดีที่สุด โดยการนำข้อมูลมาสร้างแบบจำลองการพยากรณ์ในรูปแบบของโครงสร้างต้นไม้ ซึ่ง มีการเรียนรู้ข้อมูลแบบมีผู้สอน (Supervised Learning) สามารถสร้างแบบจำลองการจัดหมวดหมู่ (Clustering) ได้จากกลุ่มตัวอย่างของข้อมูลที่กำหนดไว้ล่วงหน้า (Training set) ได้โดยอัตโนมัติ และสามารถพยากรณ์กลุ่มของรายการที่ยังไม่เคยนำมาจัดหมวดหมู่ได้อีกด้วย โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้า เงื่อนไข แล้ว ผลลัพธ์” เช่น (พยน, 2548)

"If Income = High and Married = No THEN Risk = Poor"

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

ส่วนประกอบของต้นไม้ตัดสินใจประกอบด้วย (จิตตภู, 2550)

1) โหนด (Node) คือ คุณสมบัติต่างๆ เป็นจุดที่แยกข้อมูลว่าจะให้ไปในทิศทาง ใด ซึ่งโหนดที่อยู่สูงสุดเรียกว่า โหนดราก (Root Node)

2) กิ่ง (Branch) คือ คุณสมบัติของคุณสมบัติในโหนดที่แตกออกมา โดยจำนวน ของกิ่งจะเท่ากับคุณสมบัติของโหนด

3) ใบ (Leaf) คือ กลุ่มของผลลัพธ์ในการแยกแยะข้อมูล โดยสามารถแสดง ส่วนประกอบของต้นไม้ตัดสินใจ

สมการ Entropy คือ สมการในการหาค่าความไม่สมบูรณ์ของสารสนเทศของข้อมูล ดังสมการที่ 1

$$Entropy(S) = \sum_{i=1}^{C} -P_i \log 2\ P_i|$$

โดย S คือ แอทริบิวต์ที่นำมาวัดค่า

P_i คือ สัดส่วนของจำนวนสมาชิกในกลุ่ม i เทียบกับ จำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

สมการ Information Gain คือ สมการในการหาค่า สารสนเทศก่อนนำไปใช้ในการหาค่ามาตรฐานอัตราส่วน ดัง สมการที่ 2

$$Gain(S, A) = Entropy(S) - \sum_{v=Values(A)} \frac{|s_v|}{|s|} Entropy(s_v)$$

โดย A คือ แอทริบิวต์ A

$|s_v\ |$ สมาชิกของแอทริบิวต์ A

$|s|$ คือ จำนวนสมาชิกทั้งหมดของกลุ่มตัวอย่าง

สมการ Split Information คือ สมการที่ใช้ในการเพิ่ม ค่าสารสนเทศการแบ่งข้อมูล โดยจะบอกถึงลักษณะการกระ จายของข้อมูล เป็นการแก้ปัญหาความโน้มเอียง ดังสมการที่ 3

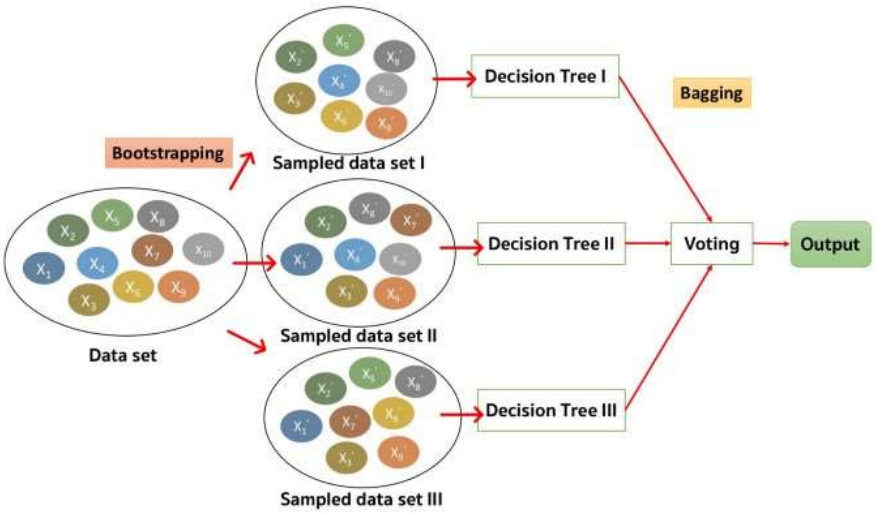
$$Split\ Information(S, A) = -\sum_{i=1}^{n} \frac{|s_i|}{|s|} log2 \frac{|s_i|}{|s|} s_i\ |$$

สัดส่วนของจำนวนสมาชิกในกลุ่ม i

สมการ Gain Ratio คือ สมการการคำนวณหาค่า มาตรฐานอัตราส่วนของค่า Gain เพื่อลดความลำเอียงของข้อมูล ดังสมการที่ 4

$$Gain\ Ratio(S,A) = Gain(S,A)/(Split\ Information(S,A))$$

2.3.3 การจำแนกประเภทแบบ Random forest



ภาพที่ 2.16 การจำแนกประเภทแบบ Random forest

ที่มา : medium.com (2561)

หลักการของ Random Forest คือ สร้าง model จาก Decision Tree หลายๆ model ย่อยๆ (ตั้งแต่ 10 model ถึง มากกว่า 1000 model) โดยแต่ละ Model จะได้รับ Data Set ไม่เหมือนกัน ซึ่งเป็น Subset ของ Data Set ทั้งหมด ตอนทำ Prediction ก็ให้แต่ละ Decision Tree ทำ Prediction ของใครของมัน และคำนวณผล Prediction ด้วยการ Vote Output ที่ ถูกเลือกโดย Decision Tree มากที่สุด (กรณี classification) หรือ หาค่า Mean จาก Output ของแต่ละ Decision Tree (กรณี regression)

Decision Tree แต่ละ model ใน Random Forest ถือว่าเป็น weak learner ประมาณว่าเป็น model ที่ไม่เก่งเท่าไหร่ แต่พอนำเอาแต่ละ Decision Tree มาทำ prediction ร่วมกัน ก็จะได้ model รวมที่มีความเก่ง และแม่นยำมากกว่า Decision Tree ที่ทำ prediction แบบเดี่ยวๆ

Sample ข้อมูล (bootstrapping) จาก Data Set ทั้งหมด ให้ได้ข้อมูลออกมา n ชุด ที่ไม่เหมือนกัน ตามจำนวน Decision Tree ใน Random Forest เช่น data set ตั้งต้นมีอยู่ 10 feature (X1,X2,...,X10) แต่ละ Decision Tree จะได้ feature ไปไม่เหมือนกัน และ จะได้ข้อมูลไม่ครบทุก row ด้วยจาก data set ทั้งหมดด้วย (X1 –> X1',X2–>X2',...)

2.3.4 ซอฟต์แวร์ Weka

Weka ย่อมาจาก Waikato Environment for Knowledge Analysis (ขจรศักดิ์ ศรีอ่อน, 2552) ซึ่ง เป็นซอฟต์แวร์สำเร็จภาพประกอบประเภทซอฟแวร์ให้เปล่า (Free Ware) ที่สามารถดาวน์โหลดได้ จากเว็บไซต์ ซึ่งอยู่ภายใต้การควบคุมของ GPL License ซึ่งโปรแกรม Weka ได้ถูกพัฒนามาจาก ภาษาจาวาทั้งหมด ซึ่งเขียนมาโดยเน้นกับงานทางด้านการเรียนรู้ด้วยเครื่อง (Machine Learning) และ การทำเหมืองข้อมูลโปรแกรมประกอบไปด้วยโมดูลย่อยๆ สำหรับใช้ในการจัดการข้อมูล และ เป็นโปรแกรมที่สามารถใช้ Graphic User Interface (GUI) และใช้คำสั่งในการให้ซอฟต์แวร์ประมวลผล

ภาพที่ 2.17 ซอฟต์แวร์ Weka

ที่มา : glurgeek.com(2557)

2.3.4.1 โปรแกรมหลักของซอฟต์แวร์ Weka

2.3.4.1.1 Simple CLI(Command Line Interface) เป็นโปรแกรมรับคำสั่งการทำงานผ่านการพิมพ์

2.3.4.1.2 Explorer เป็นโปรแกรมที่ออกแบบในลักษณะ GUI

2.3.4.1.3 Experimenter เป็นโปรแกรมที่ออกแบบการทดลองและการทดสอบผล

2.3.4.1.4 KnowledgeFlow เป็นโปรแกรมออกแบบผังการไหลของความรู้

2.3.4.1.5 ArifViewer เป็นโปรแกรมที่ใช้สำหรับแก้ไขแฟ้มประเภท Aiff

2.3.4.1.6 Log เป็นโปรแกรมที่ใช้อ่านข้อความบันทึกเก็บระหว่างการทำงาน

2.3.4.2 เมนูหลักของ Explorer

2.3.4.2.1 Preprocess การเตรียมข้อมูล

2.3.4.2.2 Classify รวมโมดูลการทำเหมืองข้อมูลแบบจัดจำแนกประเภท

2.3.4.2.3 Cluster รวมโมดูลการทำเหมืองข้อมูลแบบการเกาะกลุ่ม

2.3.4.2.4 Associate รวมโมดูลการทำเหมืองข้อมูลแบบกฎเชื่อมโยง

2.3.4.2.5 Select attributes รวมโมดูลสำหรับการวิเคราะห์ความเกี่ยวพันของลักษณะประจำ

2.3.4.2.6 Visualize นำเสนอข้อมูลด้วยภาพนามธรรมสองมิติ

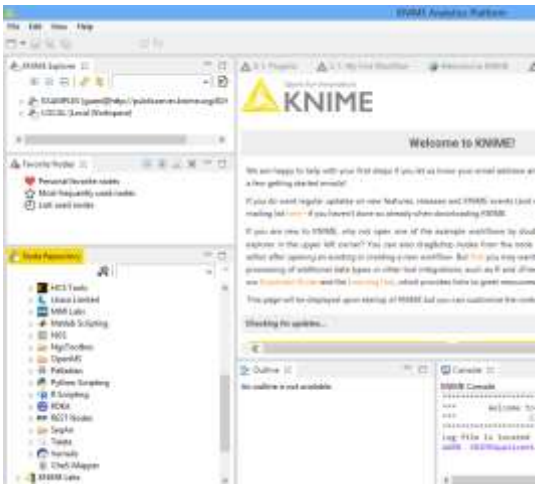
2.3.3.3 ประเภทของแฟ้มข้อมูลที่รับได้

2.3.4.3.1 แฟ้มข้อมูลที่รับต้องอยู่ในรูปแบบ ASCII อาจเป็น arft, csv, C45

2.3.4.3.2 ในกรณีแฟ้มข้อมูลอยู่ในเครือข่ายผู้ใช้สามารถเรียกใช้โดยอาศัย URL

2.3.4.3.3 หรืออาจใช้ข้อมูลที่อยู่ในฐานข้อมูลที่เชื่อมโยงผ่าน JDBC

2.3.5 ซอฟต์แวร์ Konstanz Information Miner (KNIME)



ภาพที่ 2.18 ซอฟต์แวร์ KNIME

ที่มา : thaidatascience.wordpress.com (2562)

KNIME เป็น Open Source ที่ได้รับการจัดอันดับจาก Gartner (ประกาศเมื่อเดือน Jan 2019) อยู่ในกลุ่ม Leader ด้าน Data Science & Machine Learning มาหลายปีติดกัน แล้วเหตุผลสำคัญที่ทำให้ KNIME เป็น Software ที่ครบเครื่องและได้รับความนิยมจาก Data Engineer & Data Scientist ทั่วโลก มีหลายประเด็น อาทิเช่น

1) KNIME มีผู้ใช้และ community จากหลากหลายวงการ เช่น การตลาด, การผลิต, อุตสาหกรรมเภสัช เคมีภัณฑ์ ชีววิทยา และนักวิชาการต่างๆ

2) KNIME มีแหล่งเรียนรู้ที่เผยแพร่ไว้อย่างครบถ้วน คุณสามารถดูตัวอย่างการประยุกต์ทำงานวิเคราะห์ข้อมูลได้หลากหลาย เช่น Customer Intelligence, Social Media, Finance, Manufacturing, Pharma/ Health Care, Retail, Industry & Government เป็นต้น

3) KNIME เป็น Open source ที่พร้อมใช้ มาแบบฟรี และไม่มีข้อจำกัดใดๆ ซึ่งมันน่าสนใจและน่าใช้มาก

4) KNIME เป็น Data driven software ที่ทำได้หลายภาษา เช่น Text mining, Image processing, สามารถเชื่อมกับภาษา Python, R, Spark H2O, Keras/TensorFlow for Deep Learning ได้อย่างลื่นไหล

Knime (K เงียบอยู่ดังนั้นจึงออกเสียงnīm) เป็นแพลตฟอร์มการวิเคราะห์ข้อมูลที่ได้รับการจัดอันดับสูงพร้อมการใช้งานที่หลากหลายและการผนวกรวมกับผลิตภัณฑ์อื่น ๆ เช่นฐานข้อมูล ภาษากรอบการเรียนรู้ของเครื่องและกรอบการเรียนรู้เชิงลึก ปรัชญาของ Knime คือการรวมและ "ผสมผสาน" ซอฟต์แวร์และแหล่งข้อมูลใด ๆ ที่คุณต้องการใช้

การสำรวจการสร้างแบบจำลองการสร้างภาพการรายงานและการพัฒนาบางส่วนของแพลตฟอร์มเป็นโอเพ่นซอร์สเช่นเดียวกับส่วนขยายชุมชน Knime Server ซึ่งให้ความร่วมมือการทำงานอัตโนมัติการจัดการและความสามารถในการปรับใช้เป็นเชิงพาณิชย์เช่นเดียวกับส่วนขยายของคู่ค้า แพลตฟอร์ม Knime Analytics และเซิร์ฟเวอร์ Knime พร้อมใช้งานสำหรับการติดตั้งในครั้งแรกและสำหรับ AWS และ Azure clouds

ในบทช่วยสอนนี้ฉันจะให้ความสำคัญกับแพลตฟอร์มโอเพ่นซอร์ส Knime Analytics และส่วนขยายโอเพนซอร์ซที่เลือกไว้ เป้าหมายของฉันคือการนำคุณไปสู่จุดที่คุณสามารถค้นหาเวิร์กโฟลว์ Knime ที่มีอยู่ซึ่งคุณสามารถใช้เป็นจุดเริ่มต้นสำหรับงานด้านวิทยาศาสตร์ข้อมูลของคุณเองและที่คุณเข้าใจเวิร์กโฟลว์ Knime ดีพอที่จะปรับแต่งมัน เพื่อให้บรรลุผลนั้นในพื้นที่ที่ จำกัด ฉันจะแนะนำคุณไปยังเนื้อหาบางส่วนของ Knime เพื่อกรอกรายละเอียด

2.4 วรรณกรรมที่เกี่ยวข้อง

2.4.1 การวิเคราะห์สารสนเทศเพื่อการพัฒนาระบบสนับสนุนการวางแผนการคัดเลือกบุคคลเข้าศึกษาในระบบ TCAS มหาวิทยาลัยมหาสารคาม

นโยบายการคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาระบบใหม่ (TCAS) เป็นนโยบายปฏิรูปการศึกษาโดยมีหลักการ สำคัญ 3 ประการ คือ นักเรียนควรอยู่ในห้องเรียนจนจบมัธยมศึกษาปีที่ 6, นักเรียนแต่ละคนมีเพียง 1 สิทธิ์ในการตอบรับใน สาขาวิชาที่เลือกเพื่อความเสมอภาค และสถาบันอุดมศึกษาในเครือข่ายที่ประชุมอธิการบดีแห่งประเทศไทย (ทปอ.) ทุกแห่งต้อง เข้าร่วมระบบเคลียริงเฮาส์เพื่อบริหาร 1 สิทธิ์ของนักเรียน จากนโยบายมหาวิทยาลัยมหาสารคามเป็นมหาวิทยาลัยของรัฐที่อยู่ ในเครือข่ายที่ประชุมอธิการบดีแห่งประเทศไทย ซึ่งต้องดำเนินการตามนโยบายของการคัดเลือกนิสิตในระดับปริญญาตรีระบบใหม่ จากปัญหาการคัดเลือกระบบใหม่พบว่าจำนวนการยืนยันสิทธิ์ (Clearing house) เข้าศึกษา

มีจำนวนน้อยกว่าแผนการรับ เข้าศึกษา เมื่อเทียบกับจำนวนนักเรียนที่สมัครและสนใจเข้าศึกษาที่มีจำนวนมาก จากปัญหาดังกล่าวผู้วิจัยได้มีแนวคิดในการนำ เทคนิคเหมืองข้อมูล (Data mining) เพื่อวิเคราะห์สารสนเทศที่เกิดขึ้นจากข้อมูลนักเรียนที่มีสิทธิ์ในการยืนยันสิทธิ์เข้าศึกษา นำรูปแบบที่ได้มาพัฒนาเป็นระบบสนับสนุนการวางแผนการคัดเลือกนิสิตใหม่ในระบบ TCAS เพื่อช่วยให้คณะกรรมการพิจารณา ผลคัดเลือกสามารถตัดสินใจจำนวนที่จะรับเข้ายืนยันยันสิทธิ์เพื่อให้จำนวนที่จะรับเข้าศึกษาใกล้เคียงกับแผนการรับเข้าศึกษา มากที่สุด จากการวิจัยพบว่า วิธีต้นไม้ตัดสินใจ และวิธีการค้นหากฎความสัมพันธ์ จากข้อมูลทดลองกลุ่มสาขาวิชามนุษยศาสตร์ และสังคมศาสตร์มีค่าความถูกต้องร้อยละ 82.85 สร้างเป็นกฎความสัมพันธ์ได้ 89 กฎ, กลุ่มสาขาวิชาวิทยาศาสตร์สุขภาพมีค่า ความถูกต้องร้อยละ 80.88 สร้างเป็นกฎความสัมพันธ์ได้ 85 กฎ และกลุ่มสาขาวิชาวิทยาศาสตร์เทคโนโลยีค่าความถูกต้องร้อยละ 78.85 สร้างเป็นกฎความสัมพันธ์ได้ 85 กฎ โดยสามารถนำผลการทดลองมาพัฒนาเป็นระบบสนับสนุนการวางแผนการคัดเลือก นิสิตใหม่ในระบบ TCAS ได้

2.4.2 การพัฒนาตัวแบบระบบสนับสนุนการตัดสินใจสำหรับการวางแผนก่อนการลงทะเบียนนักศึกษาระดับปริญญาตรีสาขาวิชาวิศวกรรมอุตสาหะการ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเชียงใหม่

นัยนา ยะสิงห์สาร (2557) จากการศึกษาค้นคว้าครั้งนี้ ได้จัดทำการรวบรวมข้อมูล ระบบฐานข้อมูล และสัมภาษณ์ความ ต้องการของผู้ใช้ แล้วนำมาวิเคราะห์ออกแบบระบบโดยใช้โปรแกรมพีเอชพิมายแอดมินในการสร้างฐานข้อมูลมายแอสคิวแอล และใช้ภาษาพีเอชพีในการพัฒนาโปรแกรมต้นแบบที่ ประกอบด้วยขอบเขตด้านผู้ใช้งาน ฐานข้อมูล และระบบงาน โดยจะกำหนดสิทธิ์เข้าถึงข้อมูลของ ผู้ใช้เป็น 3 กลุ่ม ได้แก่ นักศึกษา อาจารย์ที่ปรึกษา และเจ้าหน้าที่หลักสูตรผลการทดสอบโดยผู้ใช้งานทั้ง 3 กลุ่ม จำนวน 26 คน จากแบบสอบถามความพึงพอใจการใช้ งานระบบ ผลประเมินระบบพบว่า มีระดับความพึงพอใจเฉลี่ยเท่ากับ 4.11ซึ่งมีค่าแปลผลอยู่ในช่วง 3.50 –4.49 มีความหมายว่า ระดับความพอใจของผู้ใช้งานอยู่ในเกณฑ์มาก และลักษณะการใช้งาน ที่มีความพึงพอใจที่อยู่ในเกณฑ์มากที่สุด ได้แก่ความสะดวกและเข้าใจง่ายในการใช้งานความ ถูกต้องของข้อมูลในรายงานข้อมูลสารสนเทศตรงตามความต้องการใช้งานได้รับประโยชน์ที่ได้ จากการใช้งานระบบรองลงมาตามลำดับ คือ ความเหมาะสมของการวางเครื่องมือสำหรับใช้งาน ความสะดวกและง่ายต่อ

การใช้งานการค้นหาข้อมูลทำได้ง่ายและสะดวกคู่มือการใช้โปรแกรมมี ความชัดเจนเข้าใจง่าย และการแก้ไขปรับปรุงข้อมูลทำได้ง่ายและสะดวก

2.4.3 การใช้เทคนิคเหมืองข้อมูลในการเลือกสาขาวิชาเพื่อโอกาสในการเข้าศึกษาต่อระดับปริญญาตรี

อนันต์ ปินะเต (2557) มหาวิทยาลัยมหาสารคาม มีกระบวนการรับสมัครคัดเลือกบุคคลเข้าศึกษาในระดับปริญญาตรี ซึ่งในแต่ละปีการศึกษาจะมีผู้สมัคร เป็นจำนวนมาก ปัญหาที่สำคัญของการรับสมัครคัดเลือกบุคคลเข้าศึกษา คือการมีผู้สมัครจำนวนไม่น้อยที่ไม่มีสิทธิ์เข้าศึกษาใน สาขาวิชาที่สมัคร และมีบางสาขาวิชาที่มีผู้สมัครจำนวนน้อยกว่าแผนการรับเข้าศึกษา จากข้อมูลการรับเข้าศึกษาที่ผ่านมาพบว่ามีบางสาขาวิชาที่มีผู้สมัครเป็นจำนวนมากเมื่อเปรียบเทียบกับสัดส่วนแผนการรับเข้าศึกษาที่มีจำนวนน้อยซึ่งผู้สมัครส่วนใหญ่จะเลือกสมัครสาขาวิชาตามความชอบ ความรู้สึกของตนเอง โดยไม่ได้คำนึงถึงผลคะแนนของตนซึ่งเมื่อพิจารณาเข้าศึกษาตาม ผลคะแนนทำให้มีผู้สมัครจำนวนมากที่ไม่ผ่านคัดเลือกเนื่องจากมีผลคะแนนที่ต่ำเมื่อแข่งขันกับผู้สมัครที่มีคะแนนสูงกว่าในสาขา วิชาเดียวกัน งานวิจัยนี้ได้นำเสนอการใช้เทคนิคเหมืองข้อมูลด้วยวิธีต้นไม้ตัดสินใจ (Decision Tree) และการเรียนรู้เบย์อย่างง่าย (Naive Bayes) เพื่อเปรียบเทียบประสิทธิภาพแบบจำลองจากค่าความถูกต้อง (Accuracy) และนำแบบจำลองที่ได้ไปพัฒนา เป็นระบบสนับสนุนการตัดสินใจในการเลือกสาขาวิชาเพื่อโอกาสในการเข้าศึกษาต่อในระดับปริญญาตรีจากการวิจัยมีจำนวน 69 สาขาวิชาที่มีค่าความถูกต้องสูงสุดจากการทดลองด้วยวิธีต้นไม้ตัดสินใจ และมีจำนวน 1 สาขาวิชาที่มีค่าความถูกต้องสูงสุดด้วย วิธีการเรียนรู้เบย์อย่างง่าย

2.4.4 การพัฒนาระบบสนับสนุนการตัดสินใจในการเลือกสมัครในสาขาวิชาโดยใช้เทคนิคตนไมตัดสินใจ

อนันต์ ปินะเต (2558) การรับสมัครบุคคลเข้าศึกษาในระดับปริญญาตรี มหาวิทยาลัยมหาสารคาม มีกระบวนการในการคัดเลือกผู้สมัครที่มีคุณสมบัติ ตรงตามที่มหาวิทยาลัยกำหนด ซึ่งการแนะแนวการศึกษาจึงเป็นกระบวนการที่สำคัญ ในการประชาสัมพันธ์ข้อมูลการสมัครให้ กับผู้สมัครทราบ อาทิเช่น ข้อมูลการเลือกสาขาวิชาโดยทั่วไปผู้สมัครจะเลือกสาขาวิชาโดยยึดหลักตามความชอบ ความรู้สึก ของตนเองเป็นส่วนใหญ่ โดยไม่ได้คำนึงถึงความรู้และทักษะด้านต่างๆ ของตนเอง ส่งผลให้เมื่อเข้ามาศึกษาใน

สาขาวิชานั้นแล้ว เกิดปัญหาผลการเรียนที่ตกต่ำ ไม่ผ่านตามเกณฑ์ที่มหาวิทยาลัยกำหนด อันเป็นผลทำให้ต้องพ้นสภาพการเป็นนิสิตดังนั้นการ ประชาสัมพันธ์ข้อมูลเพื่อให้ผู้สมัครได้ใช้ประกอบการตัดสินใจในการเลือกสาขาวิชานั้นจึงเป็นสิ่งสำคัญในกระบวนการแนะแนวการศึกษาต่อ งานวิจัยนี้จึงได้นำเสนอการใช้เทคนิคต้นไม้ตัดสินใจ C4.5 เพื่อค้นหากฎการตัดสินใจในการเลือกสาขาวิชา และนำกฎการตัดสินใจที่ได้มาพัฒนาเป็นระบบสนับสนุนการตัดสินใจในการเลือกสาขาวิชา เพื่อให้ผู้ที่จะสมัครเข้าศึกษาในระดับ ปริญญาตรี ได้ทำการทดลองการเลือกสาขาวิชาก่อนการเลือกสมัครจริง ผู้สมัครจะได้ทราบถึงสาขาวิชาที่เหมาะสมกับความรู้ และทักษะของผู้สมัครเองเมื่อเข้ามาศึกษาในสาขาวิชานั้น จากการวิจัยสามารถสร้างเป็นกฎการตัดสินใจได้จำนวนทั้งสิ้น 333 กฎจากทั้งหมด 51 สาขาวิชาที่ทำการทดลองและสามารถสร้างระบบสนับสนุนการตัดสินใจในการเลือกสาขาวิชาได้จากบทความดังกล่าว

2.4.5 การศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีสาขาคอมพิวเตอร์ โดยใช้เทคนิคเหมืองข้อมูล

สำราญ วานนท์ (2561) การวิจัยครั้งนี้มีวัตถุประสงค์ 1) เพื่อศึกษาเทคนิคพยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญาตรีโดยใช้ เทคนิคเหมืองข้อมูลที่เหมาะสม 2) เพื่อเปรียบเทียบผลการวิเคราะห์พยากรณ์อาชีพสำหรับนักศึกษาระดับปริญญา ตรีโดยใช้เทคนิคเหมืองข้อมูล งานวิจัยนี้ได้นำข้อมูลภาวะการมีงานทำของบัณฑิต และข้อมูลระเบียนประวัติของ นิสิตระดับปริญญาตรีหลังสำเร็จการศึกษา จากสำนักงานคณะกรรมการการอุดมศึกษาย้อนหลัง 5 ปี คือปี 2555 – 2559 จำนวน 65,335 ระเบียน ในสาขาวิชาทางด้านคอมพิวเตอร์ และมีคุณลักษณะประกอบด้วย ผลการเรียน ความสามารถพิเศษ อาชีพของบิดามารดา รายได้ของบิดามารดา เพศ ตำแหน่งงาน ความสอดคล้องสาขา สาขาวิชา ทดลองวัดความแม่นยำด้วยเทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจ เทคนิคการจำแนกข้อมูลด้วยวิธีแรนดอม ฟอร์เรส และเทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิ่ง ผลการวิจัยพบว่า ความแม่นยำในการจำแนกประเภทข้อมูล จาก 3 เทคนิค 1) เทคนิคการจำแนกข้อมูลด้วยวิธีต้นไม้ตัดสินใจเท่ากับ 81.9196 2) เทคนิคการจำแนกข้อมูลด้วยวิธี แรนดอมฟอร์เรสเท่ากับ 84.29% และ 3) เทคนิคการจำแนกข้อมูลด้วยวิธีแบ็กกิ่งเท่ากับ 81.71% พบว่าเทคนิค แรนดอมฟอร์เรสให้ความถูกต้องในการจำแนกประเภทข้อมูลสูงที่สุด

2.5 บทสรุป

จากแนวคิด ทฤษฎี เครื่องมือและวรรณกรรมที่เกี่ยวข้องที่ได้กล่าวมาในข้างต้นทั้งหมดนั้น คณะผู้จัดทำได้เลือกใช้ขั้นตอนกระบวนการวิเคราะห์ข้อมูลด้วย CRISP-DM จากเทคนิคการทำเหมืองข้อมูลแบบ Data Classification ด้วยการสร้างโมเดล Decision Tree และโมเดล Random Forest ทำการทดสอบโมเดล เพื่อวัดประสิทธิภาพด้วยวิธี Self Consistency Test เพื่อดูแนวโน้มของโมเดลที่สร้างขึ้น ด้วยโปรแกรม Weka และโปรแกรม KNIME จากนั้นนำข้อมูลสารสนเทศมาทำการ แสดงผลแบบ Visualization ในรูปแบบของภาพโดยใช้โปรแกรม Power BI เผยแพร่บนเว็บไซต์ที่เป็นที่นิยมในยุคอินเตอร์เน็ตคือการเผยแพร่ทางสื่อออนไลน์ โดยใช้ภาษา HTML และ CSS ในการเขียนเว็บไซต์ขึ้นมา