บทที่ 2
แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในการศึกษาค้นคว้าหาเพื่อค้นหารูปแบบการเกิดความสัมพันธ์หรือตัวแปรที่มีอยู่ด้วยเทคนิคของเหมืองข้อมูล ได้ดำเนินการศึกษา ทบทวนทฤษฎี แนวความคิด และผลงานวิจัยที่เกี่ยวข้องกับการพัฒนาโมเดล โดยดำเนินการรวบรวมข้อมูลที่อาจจะเป็นประโยชน์ต่อการกำหนดแนวทางและระเบียบวิธีการวิจัย ซึ่งได้แบ่งการศึกษาออกเป็น 5 กลุ่มด้วยกันคือ

1. ศาลยุติธรรมและศาลจังหวัดปทุมธานี
2. ทฤษฎีและแนวความคิดที่เป็นพื้นฐานในการพัฒนาตัวโมเดลต้นแบบ
3. ซอฟต์แวร์ที่ใช้ในการศึกษา
4. การวิเคราะห์และออกแบบระบบ
5. งานวิจัยที่เกี่ยวข้อง

2.1 ศาลยุติธรรมและศาลจังหวัดปทุมธานี

ศาลยุติธรรมตามพระธรรมนูญศาลยุติธรรม พ.ศ. 2543 มีอยู่ทั่วราชอาณาจักรระบบศาลยุติธรรมสามารถแบ่งออกได เป็น 3 ชั้น คือ ศาลชั้นต้น ชั้นอุทธรณ์ ชั้นฎีกา รวมทั้งสิ้น จำนวน 248 ศาล ทั่วราชอาณาจักร

ศาลชั้นต้น

ศาลชั้นต้น เป็นศาลซึ่งรับคำฟ้อง หรือคำร้องในชั้นเริ่มต้นคดีไม้ว่าเป็นคดีแพ่ง คดีผู้บริโภค คดีแรงงาน คดีภาษีอากร คดีล้มละลาย คดีทรัพย์สินทางปัญญาและการค้าระหว่างประเทศหรือคดีอาญา โดยดำเนินกระบวนการตัดสินชี้ขาดเป็นชั้นศาลแรก ทั้งมีอำนาจดำเนินกระบวนพิจารณาแทนศาลอุทธรณ์และศาลฎีกาในบางเรื่องด้วย

ศาลชั้นต้นทั่วราชอาณาจักรมีจำนวนทั้งหมด 233 ศาล ประกอบด้วยศาลต่างๆดังนี้

1. ศาลชั้นต้นพิจารณาคดีแพ่งและคดีอาญาในกรุงเทพมหานคร จำนวน 6 ศาล ประกอบด้วย

กลุ่มศาลแพ่ง ได้แก่ ศาลแพ่ง ศาลแพ่งกรุงเทพใต้ ศาลแพ่งธนบุรี

กลุ่มศาลอาญา ได้แก่ ศาลอาญา ศาลอาญากรุงเทพใต้ ศาลอาญาธนบุรี

2. ศาลจังหวัดและศาลจังหวัดสาขามีจำนวน 114 ศาล ประกอบด้วย

ศาลจังหวัดที่ตั้งอยู่ในภาค 1-9 จำนวน 111 ศาล

ศาลจังหวัดสาขา จำนวน 3 ศาลสาขา

3. ศาลแขวง มีจำนวน 26 ศาล ประกอบด้วย

ศาลแขวง จำนวน 26 ศาล

4. ศาลพิเศษ (ศาลเยาวชนและครอบครัว) มีจำนวน 78 ศาลประกอบด้วย

ศาลเยาวชนและครอบครัวกลาง จำนวน 1 ศาล

ศาลเยาวชนและครอบครัวสาขา (สาขามีนบุรี) จำนวน 1 ศาลสาขา

ศาลเยาวชนและครอบครัวจังหวัด จำนวน 76 ศาล

5. ศาลชำนัญพิเศษ มีจำนวน 13 ศาล ประกอบด้วย

ศาลแรงงานกลาง จำนวน 1 ศาล และศาลแรงงานภาค จำนวน 9 ศาล

ศาลทรัพย์สินทางปัญญาและการค้าระหว่างประเทศกลาง จำนวน 1 ศาล

ศาลล้มละลายกลาง จำนวน 1 ศาล

ศาลภาษีอากรกลาง จำนวน 1 ศาล

ชั้นศาลอุทธรณ์

เป็นศาลที่มีอำนาจพิจารณาพิพากษาคดีที่อุทธรณ์คำพิพากษาหรือคำสั่งของศาลชั้นต้นตามบทบัญญัติแห่งกฎหมายว่าด้วยการอุทธรณ์ รวมทั้งมีหน้าที่ในการพิจารณาคำสั่งอื่นๆ เช่นมีคำสั่งเกี่ยวกับการประกันตัวในคดีอาญาและ การขอทุเลาการบังคับคดีในคดีแพ่ง การพิจารณาคดีของศาลอุทธรณ์มีลักษณะเป็นการตรวจสอบหรือทบทวน ในชั้นอุทธรณ์ประกอบด้วยศาลอุทธรณ์กลางจำนวน 1 ศาล และศาลอุทธรณ์ภาค 1-9 จำนวน 9 ศาล รวม 10 ศาล

ชั้นฎีกา

เป็นศาลสูงสุดมีอำนาจพิจารณาพิพากษาคดีที่อุทธรณ์คำพิพากษาหรือคำสั่งของศาลอุทธรณ์ตามบทบัญญัติแห่งกฎหมายว่าด้วยการฎีกา และมีอำนาจวินิจฉัยชี้ขาดคดีที่ศาลฎีกามีอำนาจวินิจฉัยได้ตามกฎหมายเฉพาะ แต่หากคดีใดมีปัญหาสำคัญไม่ว่าจะเป็นปัญหาข้อเท็จหรือปัญหาข้อกฎหมาย ประธานศาลฎีกาเห็นว่าควรให้ วินิจฉัยโดยที่ประชุมใหญ่ของศาลฎีกาประธานศาลฎีกามีอำนาจสั่งให้นำปัญหาดังกล่าวเข้าสู่การวินิจฉัยโดยที่ประชุมใหญ่ของศาลฎีกา ศาลฎีกามีเพียงศาลเดียวตั้งอยู่ที่กรุงเทพมหานคร

ประวัติความเป็นมาของศาลจังหวัดปทุมธานี

ในปี ร.ศ.117 (พ.ศ.2442) พระเจ้าบรมวงศ์เธอ กรมพระนเรศวรฤทธิ์ ซึ่งขณะนั้นดำรงหลวงอารักษ์ประชาราษฎร์ ข้าหลวงรักษาการเมืองปทุมธานีจัดซื้อที่ดินบริเวณปากคลองรังสิตประยูรศักดิ์แขวงเมืองปทุมธานีซึ่งเห็นว่าเป็นที่ราษฎรไปมามากกว่าบริเวณอื่น เพื่อใช้เป็นที่สร้างศาลากลางเมืองปทุมธานีกับศาลสำหรับพิจารณาคดี แต่การจัดซื้อที่ดินครั้งนั้นมีปัญหา เพราะราษฎรชาวมอญเจ้าของที่ 6 ราย ไม่ยอมขายที่ให้ โดยอ้างว่าได้อยู่อาศัยมาแต่สมัยปู่ ย่า ตา ยาย และทั้งได้ปลูกต้นไม้ผลไม้ลงในเนื้อที่นั้นแล้ว เนื่องจากชาวมอญ 5 คน ในจำนวนนั้นเป็นหมู่ทหารเรือพระเจ้าบรมวงศ์เธอ กรมพระนเรศวรฤทธิ์ จึงมีหนังสือทูลชี้แจงไปยังพระวรวงศ์เธอ กรมหมื่นปราบปรปักษ์ ผู้บัญชาการกรมทหารเรือให้ช่วยชี้แจงแก่ชาวมอญหมู่ทหารเรือซึ่งอยู่ในบังคับบัญชา ตามหนังสือ ลงวันที่ 12 กรกฎาคม รัตนโกสินทรศก 117 ตอนหนึ่งว่า

“ หม่อมฉันขอทูลให้ทรงทราบว่า ที่ตำบลวัดเทียนถวาย ปากคลองรังสฤษ์ริมแม่น้ำนี้เป็นที่สมควรจะสร้างที่ว่าราชการเมืองปทุมธานี และศาลสำหรับเมือง เพราะจะได้ทำทางริมคลองรังสิตเข้าหารถไฟนครราชสีห์มา อันจะเดินทางได้ทั้งทางบกแลน้ำ จึงจำต้องตั้งที่ว่าราชการเมืองแลศาลตำบลนี้ เมื่อราษฎรเจ้าของที่ไม่ต้องการเงินจะต้องการที่โดยจะหาซื้อที่โดยลำพังเป็นการลำบาก ที่หลวงที่เมืองปทุมธานีมีอยู่ริมแม่น้ำเหมือนที่ราษฎรตั้งอยู่ตามที่จะต้องการนี้ หม่อมฉันจะจัดเปลี่ยนให้ได้ แลทั้งรื้อขลบ้านเรือนไปปลูกใหม่ให้ด้วย เพื่อการสำเร็จจะได้แล้วไปไม่ขัดข้องแก่ความเจริญของราชการ ขอท่านได้ทรงพระเมตตาโปรดให้เจ้าพนักงานชี้แจงแก่เจ้าของที่ให้เป็นการตกลงกัน ”

การก่อสร้างที่ว่าการเมืองปทุมธานี ศาลเมืองปทุมธานีครั้งนั้นใช้เวลาประมาณ 1 ปี ได้มีพิธีเปิดเมื่อวันที่ 10 มกราคม รัตนโกสินทรศก 118 (พ.ศ.2443) ต่อมาพระบาทสมเด็จพระมงกุฎเกล้าเจ้าอยู่หัว ทรงพระกรุณาโปรดเกล้าฯให้เปลี่ยนคำเรียก “เมือง” เป็น “จังหวัด” ศาลยุติธรรมในหัวเมืองทั่วไปจึงเปลี่ยนจาก “ศาลเมือง” เป็น “ศาลจังหวัด” ด้วย ดังนั้นตามประกาศกระทรวงยุติธรรม ลงวันที่ 15 มิถุนายน พ.ศ.2459 ศาลเมืองปทุมธานีจึงเปลี่ยนเรียกเป็น **“ศาลจังหวัดปทุมธานี”** ตั้งแต่นั้นมา

ในปี พ.ศ.2461 ศาลากลางจังหวัดปทุมธานีได้ย้ายที่ตั้งใหม่ไปอยู่ ณ บริเวณซึ่งปัจจุบันใช้เป็นที่ว่าการอำเภอเมืองปทุมธานี ทำให้ศาลจังหวัดปทุมธานีกับศาลากลางจังหวัดปทุมธานีต้องอยู่ไกลกัน ปรากฏตามหนังสือของพระยาโบราณราชธานินทร์ อุปราชมณฑล กรุงเก่าถึงเสนาบดีกระทรวงยุติธรรม ที่ 45/3424 ลงวันที่ 10 มิถุนายน พ.ศ.2461 ว่าศาลจังหวัดปทุมธานีเดิมกับศาลากลางจังหวัดปทุมธานีที่ย้ายไป อยู่หางไกลกันมากประมาร 100 เส้นเศษ หนทางที่จะไปมาถึงกัน ถ้าใช้เรือแจวพายแล้วต้องใช้เวลาราวชั่วโมงเศษ จึงขอย้ายศาลจังหวัดปทุมธานีไปปลูกสร้าง ณ ที่ทำ

การจังหวัดใหม่ โดยให้สร้างต่อจากศาลากลางจังหวัดปทุมธานีลงมาทางใต้ กระทรวงยุติธรรมดำริแล้วเห็นชอบด้วย

ที่ทำการศาลจังหวัดปทุมธานีได้เริ่มเปิดทำการในวันที่ 25 ตุลาคม พ.ศ.2462 เวลา 11.00 น. หลังพิธีเปิดมีการพิจารณาคดีเรื่องคนร้ายลักทรัพย์ จำเลยให้การรับสารภาพ จึงพิพากษาคดีเสร็จไปในวันนั้น นอกจากการก่อสร้างที่ทำการศาลหลังใหม่แล้วยังได้สร้างบ้านพักผู้พิพากษาจ่าศาลด้วยอีกอย่างละ 1 หลัง ตั้งอยู่หลังที่ทำการศาลจังหวัดปทุมธานี

เมื่อ พ.ศ.2516 นางกฐิน กุยยกานนท์ และครอบครัว ได้ทูลเกล้าฯถวายที่ดินโฉนดที่ 5693 และโฉนดที่ 2651 บางส่วนรวมเนื้อที่ 16 ไร่ อยู่ริมถนนสายกรุงเทพฯ-ปทุมธานี ให้จังหวัดปทุมธานีใช้เป็นที่ตั้งศาลกลางจังหวัดปทุมธานี ทางจังหวัดจึงขอให้ศาลจังหวัดปทุมธานีมาปลูกสร้างในบริเวณเดียวกันเพื่อจะได้เป็นศูนย์ราชการต่อไป กระทรวงยุติธรรมพิจารณาแล้วเห็นชอบด้วย กำหนดแล้วเสร็จภายในวันที่ 22 ธันวาคม พ.ศ. 2520 ซึ่งเป็นที่ตั้งปัจจุบันของศาลจังหวัดปทุมธานี

2.2 ทฤษฎีและแนวความคิดที่เป็นพื้นฐานในการพัฒนาตัวโมเดลต้นแบบ

2.2.1 นิยาม ความหมายของเหมืองข้อมูล

เหมืองข้อมูล คือ กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยงานต่างๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล (สุชาดา กีระนันทน์, 2545)

เหมืองข้อมูล คือ เทคนิคการสืบค้น ความรู้บนฐานข้อมูลขนาดใหญ่ (KDD: Knowledge discovery in database) เพื่อค้นหารูปแบบ (pattern) และความสัมพันธ์ (Associations) ที่ซ่อนอยู่ในชุดข้อมูลนั้น (U.fayyad,G.piatetsky-shapiro and P.smyth,1997, pp.37-54)

การทำเหมืองข้อมูล (Data Mining) คือการสำรวจและวิเคราะห์ข้อมูลจำนวนมากให้อยู่ในรูปแบบที่มีความหมายและอยู่ในรูปของกฎโดยความสัมพันธ์เหล่านี้ แสดงให้เห็นถึงความรู้ต่างๆที่มี (ธนาคม จุ้ยสิริ และ เนื่องวงศ์ ทวยเจริญ, 2556, pp.879-882)

การทำเหมืองข้อมูล คือ กระบวนการค้นหาความรู้ที่เก็บอยู่ในฐานข้อมูลขนาดใหญ่ (สิริธร เจริญรัตน์ และชฎารัตน์ พิพัฒนนันท์, 2556, pp.131-151)

การทำเหมืองข้อมูล หมายถึง การสำรวจและวิเคราะห์ข้อมูลที่มีขนาดใหญ่ เพื่อค้นหารูปแบบ (pattern) หรือกฎเกณฑ์ (Rule) ที่แฝงอยู่ในข้อมูลจำนวนมากนั้นและนำความรู้ (Knowledge) ที่ค้นพบไปใช้ประโยชน์ ในการพัฒนาองค์กร รวมถึงเป็นการค้นหาความสัมพันธ์

(Relation) และแนวโน้ม (Trend) ของพฤติกรรมต่างๆ โดยอาศัยเทคนิคการสร้างแบบแผน เพื่อให้ได้มาซึ่งสารสนเทศที่มีประโยชน์ในด้านต่างๆ ตามที่ต้องการมาใช้สนับสนุนการตัดสินใจ (Pang-Ning Tan,Michael steinbanch,and Vipin kumar, 2006)

เหมืองข้อมูล คือ การค้นหาความสัมพันธ์ และรูปแบบทั้งหมด ซึ่งมีอยู่จริงในฐานข้อมูล แต่ได้ถูกซ่อนไว้ในข้อมูลจำนวนมาก เหมืองข้อมูลจะทำการสำรวจวิเคราะห์ข้อมูลดังกล่าว เพื่อให้ได้สารสนเทศที่อยู่ในรูปแบบที่เต็มไปด้วยความหมายและในรูปของกฎสารสนเทศที่ได้แสดงให้เห็นถึงความรู้ต่างๆที่มีประโยชน์ มีความถูกต้อง และสามารถนำไปใช้ได้จริง (ปณิธิ แก้วสวัสดิ์, 2553, pp. 51-60)

เหมืองข้อมูล คือกระบวนการสกัดความรู้ที่น่าสนใจจากข้อมูลปริมาณมาก ซึ่งความรู้ที่ได้ จากกระบวนการนี้ เป็นความรู้ที่ไม่ปรากฏให้เห็นเด่นชัด ความรู้ที่บ่งบอกเป็นนัย ความรู้ที่ไม่ทราบมาก่อน ที่มีศักยภาพในการนำไปใช้ประโยชน์ (ชนวัฒน์ ศรีสอ้าน, 2550, p. 207)

โดยสรุป เหมืองข้อมูล เป็นกระบวนการในการค้นหาความสัมพันธ์ ค้นหารูปแบบ ค้นหาแนวโน้ม จากความรู้ที่เก็บรวบรวมอยู่ไว้เป็นจำนวนมาก โดยอาศัยเทคนิคต่างๆ เพื่อนำความรู้ที่ได้ มาสนับสนุนกระบวนการตัดสินใจ จากข้อมูลที่ได้เพื่อนำไปทำนายผลลัพธ์ที่คาดว่าจะเกิดขึ้นภายในอนาคตอันใกล้

2.2.2 เทคนิคการวิเคราะห์ข้อมูลด้วยดาต้า ไมน์นิ่ง

เทคนิคการวิเคราะห์ข้อมูลด้วยดาต้าไมน์นิ่ง นั้นสามารถจำแนกได้ 2 ประเภทหลักๆ ด้วยกันคือ

2.2.2.1 เทคนิคการเรียนรู้แบบไม่มีผู้สอน (Unsupervised learning)

เทคนิคประเภทนี้จะเน้นการพิจารณาข้อมูลเป็นหลัก เช่น พิจารณาว่าข้อมูลมีความสัมพันธ์กันในลักษณะใดได้บ้าง ซึ่งเทคนิคประเภทนี้สามารถแบ่งย่อยได้อีก คือ เทคนิคการค้นหากฎความสัมพันธ์ (Association rule discovery) และการแบ่งกลุ่มข้อมูล (Clustering)

2.2.2.2 เทคนิคการเรียนรู้แบบมีผู้สอน (Supervised learning)

เทคนิคในประเภทนี้จะเน้นการเรียนรู้จากข้อมูลที่มีอยู่ในอดีต เพื่อนำมาสร้างโมเดลสำหรับการทำนายหรือคาดการณ์ในสิ่งที่คาดว่าจะเกิดขึ้นในอนาคต ซึ่งโมเดลในที่นี้อาจจะเป็นสมการทางคณิตศาสตร์ หรือกฎต่างๆ ซึ่งเทคนิคการเรียนรู้ประเภทนี้สามารถแบ่งย่อยได้อีก คือ การจำแนกประเภทข้อมูล (Classification) และเทคนิคการประมาณค่าข้อมูล (Regression) ซึ่งทั้งสองเทคนิคนี้มีลักษณะที่คล้ายคลึงกันมาก แต่มีความแตกต่างกันที่ผลลัพธ์ที่ต้องการทำนาย ซึ่งการจำแนกประเภทข้อมูลการทำนายข้อมูลที่มีค่าเป็น นอมินอล (Nominal) หรือ ค่าที่ไม่ใช่ตัวเลข ส่วนการประมาณค่าข้อมูลจะใช้กับข้อมูลที่เป็นตัวเลขเท่านั้น

2.2.3 ประเภทข้อมูลที่นำมาใช้ในวิเคราะห์

โดยปกติแล้วข้อมูลที่สามารถนำมาวิเคราะห์ได้จะต้องเป็นข้อมูลแบบมีโครงสร้าง (Structure data) หรืออยู่ในรูปแบบของตาราง เช่น ข้อมูลสมาชิก หรือ ข้อมูลการซื้อขายสินค้าต่างๆ แต่ในปัจจุบันเราก้าวเข้าสู่ยุคของบิ๊ก ดาต้า (Big Data) ซึ่งมีข้อมูลจำนวนมหาศาล แต่ข้อมูลเหล่านี้มักจะไม่มีโครงสร้าง (Unstructured Data) เช่น ข้อความต่างๆที่อยู่ในรูปของอีเมล์ (E-mail) หรือเครือข่ายสังคมออนไลน์ (Social Network) ต่างๆ ถ้าเราต้องการนำข้อมูลที่ไม่มีโครงสร้างมาดำเนินการวิเคราะห์ เราต้องทำการแปลงข้อมูลเหล่านี้ให้เป็นข้อมูลที่มีโครงสร้างก่อน หรือแปลงให้อยู่ในรูปแบบของตารางก่อนนำข้อมูลที่ได้ไปดำเนินการวิเคราะห์ต่อไป

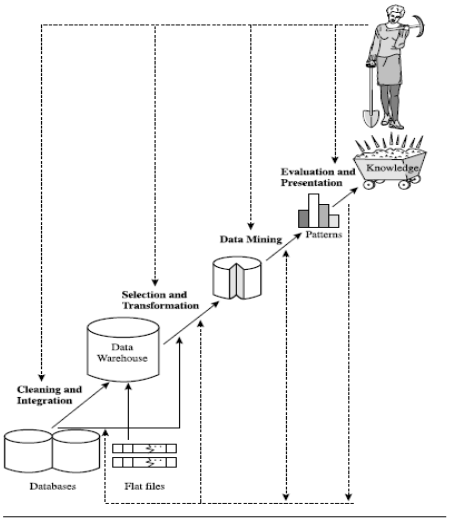
2.2.3.1 ข้อมูลแบบมีโครงสร้าง (Structure data)

ข้อมูลแบบมีโครงสร้างทั่วๆไปที่เรามักจะพบเห็นในรูปแบบของตาราง เช่น ข้อมูลที่เก็บรายละเอียดของสมาชิก หรือการซื้อขายสินค้า โดยปกติข้อมูลเหล่านี้จะเก็บอยู่ในไฟล์ประเภท Microsoft Word หรือ Microsoft Excel หรือในฐานข้อมูลต่างๆ ซึ่งในการวิเคราะห์ข้อมูลโดยใช้เทคนิคเหมืองข้อมูล ส่วนใหญ่นั้นเราจะเรียกข้อมูลแต่ละแถวว่า “ตัวอย่าง (Example)” หรือ “อินสแตนซ์ (Instance)” และเรียกข้อมูลแต่ละคอลัมน์ ว่า “แอตทริบิวต์ (Attribute)” และข้อมูลในตารางบรรทัดที่หนึ่ง คือชื่อของแต่ละแอตทิบิวต์เสมอ

2.2.3.2 ข้อมูลแบบไม่มีโครงสร้าง (Unstructured data)

ข้อมูลส่วนใหญ่ที่พบ คือ ข้อมูลประเภทรูปภาพ ข้อความ ต่างๆ ซึ่งข้อมูลเหล่านี้มีก็ความสำคัญ ดังนั้นถ้าเราจำเป็นต้องข้อมูลประเภทเหล่านี้ไปใช้งาน เราต้องดำเนินการแปลงข้อมูลเหล่านี้ให้เป็นข้อมูลที่อยู่ในรูปแบบของตารางโดยใช้กระบวนการแปลงข้อมูล (Data information)

2.2.4 ขั้นตอนการทำเหมืองข้อมูล หรือ ขั้นตอนการค้นหาความรู้



ภาพที่ 2.1 แสดงกระบวนการค้นหาความรู้ในการขุดค้นข้อมูล
ที่มา: Jiawei Han and Micheline Kamber. (2011, p.17)

ขั้นตอนการทำเหมืองข้อมูล ประกอบไปด้วยขั้นตอนหลักดังต่อไปนี้

1. การทำความเข้าใจปัญหา (Problem Understanding) ในขั้นตอนนี้เป็นการทำความเข้าใจถึงปัญหาและสามารถระบุปัญหาหรือโอกาส จากนั้นทำการแปลงโจทย์ที่ได้ ให้อยู่ในรูปแบบที่เหมาะสมต่อการนำมาวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่ง

2. การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนนี้เริ่มจากกระบวนการเก็บรวบรวมข้อมูล หลังจากนั้นจะดำเนินการตรวจสอบข้อมูลที่ได้รวบรวมไว้เพื่อตรวจเช็คความถูกต้องของข้อมูลรวมถึงควรพิจารณาด้วยว่าเป็นข้อมูลที่ได้มาจากแหล่งข้อมูลที่ถูกต้องและมีความน่าเชื่อถือ ประกอบกับข้อมูลที่ได้มีปริมาณมากพอ มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์

3. การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ใช้เวลานานที่สุด เนื่องจากโมเดลที่ได้จากการทำดาต้าไมน์นิ่งจะให้ผลลัพทธ์ที่ถูกต้องหรือไม่นั้น ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้ กล่าวคือถ้าข้อมูลที่ใช้นั้นไม่ถูกต้อง มีผิดพลาด ย่อมสะท้อนถึงผลลัพธ์ที่ได้ ซึ่งอาจทำให้ตีความผลลัพธ์ได้คลาดเคลื่อนเช่นกัน โดยการเตรียมข้อมูลนั้น สามารถแบ่งออกได้เป็น 3 ขั้นตอนย่อยดังนี้

3.1 ทำการคัดเลือกข้อมูล (Data Selection) เราควรกำหนดเป้าหมายก่อนว่าเราจะทำการวิเคราะห์อะไร แล้วจึงเลือกใช้เฉพาะข้อมูลที่เกี่ยวข้องกับสิ่งที่เราจะทำการวิเคราะห์

3.2 การกลั่นกรองข้อมูล (Data Cleaning) ในบางกรณีอาจพบข้อมูลที่ไม่ถูกต้อง อันเนื่องมาจากปัญหาในระหว่างการจัดเก็บข้อมูล เช่น การกรอกข้อมูลไม่ครบ กรอกข้อมูลซ้ำซ้อน ในขั้นตอนนี้เราจะทำการกรองข้อมูลที่ไม่ถูกต้องหรือซ้ำซ้อนออก หรืออาจทำการซ่อมแซมข้อมูลที่ขาดหายไปด้วยวิธีการบางอย่าง

3.3 การแปลงรูปข้อมูล (Data Transformation) เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ตามอัลกอริทึมของดาต้าไมน์นิ่งที่เลือกใช้

4. การสร้างตัวแบบ (Modeling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิ่ง ได้แก่ การสร้างตัวทำนาย (prediction model) ในบางครั้งพบว่ามีการนำเทคนิคดาต้าไมน์นิ่งหลายเทคนิคมาใช้ในการวิเคราะห์ข้อมูล เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด ดังนั้นเมื่อทำขั้นตอนนี้แล้ว อาจมีการย้อนกลับไปที่ขั้นตอน data preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย นอกจากนี้ยังมีการประเมินโมเดลวิเคราะห์ข้อมูลที่ได้ ในรูปแบบความถูกต้องของโมเดล เพื่อเป็นตัวบ่งชี้ความน่าเชื่อถือของโมเดลที่ได้

5. การประเมินผล (Evaluation) ในขั้นตอนนี้เป็นการประเมินประสิทธิภาพของผลลัพธ์จากโมเดลวิเคราะห์ข้อมูลว่าครอบคลุมและสามารถตอบโจทย์ที่ตั้งไว้ในขั้นตอนแรกหรือไม่ ในกรณีที่มีการสร้างโมเดลวิเคราะห์ข้อมูลหลายโมเดล ในขั้นตอนนี้จะทำการประเมินแต่ละโมเดลด้วยว่า มีส่วนดีส่วนด้อยอย่างไร และควรเลือกใช้โมเดลใด การทำงานในส่วนนี้ต้องอาศัยทักษะในการวิเคราะห์ข้อมูลเพื่อช่วยให้การวิเคราะห์ทำได้สะดวกและรวดเร็วขึ้น จึงมีการใช้เครื่องมือทางด้านกราฟฟิก เช่น การแสดงผลการวิเคราะห์ด้วยกราฟ รายงานรูปแบบต่างๆ หรือ Dashboard

6. การนำไปใช้งาน (Deployment) เมื่อได้ตัวต้นแบบหรือโมเดลหรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคดาต้าไมน์นิ่งที่มีความถูกต้องก็สามารถนำตัวต้นแบบหรือโมเดลหรือองค์ความรู้ไปใช้งานและตรวจสอบว่าบรรลุเป้าหมายที่ได้ตั้งไว้หรือไม่

2.2.5. เทคนิคในการทำเหมืองข้อมูล สามารถแบ่งได้ดังนี้

2.2.5.1 การจำแนกประเภทข้อมูล (Classification)

เทคนิคการจำแนกประเภทข้อมูลเป็นกระบวนการสร้างโมเดลจัดการข้อมูลให้อยู่ในกลุ่มที่กำหนดมาให้ จากกลุ่มตัวอย่างข้อมูลที่เรียกว่าข้อมูลสอนระบบ (Training data) ที่แต่ละแถวของข้อมูลประกอบด้วยฟิลด์หรือแอตทริบิวต์จำนวนมาก แอตทริบิวต์นี้อาจเป็นค่าต่อเนื่อง (continuous) หรือค่ากลุ่ม (categorical) โดยจะมีแอตทริบิวต์แบ่ง (classifying attribute) ซึ่งเป็นตัวบ่งชี้คลาสของข้อมูล จุดประสงค์ของการจำแนกประเภทข้อมูลคือการสร้างโมเดลการแยกแอตทริ-บิวต์หนึ่งโดยขึ้นกับแอตทริบิวต์อื่น โมเดลที่ได้จากการจำแนกประเภทข้อมูลจะทำให้สามารถพิจารณาคลาสในข้อมูลที่ยังมิได้แบ่งกลุ่มในอนาคตได้ โดยมีรูปแบบการเขียนกฎจำแนกได้ดังนี้

$\underbrace{P_1, P2 \ldots\ldots P_n}_{\text{Attribute Condition}} \rightarrow \underbrace{C_i}_{\text{Class}}$

P_n = เงื่อนไขของกฎ

C_i = คลาสของกฎ

หรือ

IF<Conditions> THEN <Class>

“ถ้า<เงื่อนไข>แล้ว<คลาส>”

ในขั้นตอนของการจำแนกข้อมูลประกอบไปด้วย 3 ขั้นตอน คือ

ขั้นตอนที่ 1 การสร้างตัวแบบโมเดล เป็นการนำชุดข้อมูลฝึกสอนให้ระบบเกิดการเรียนรู้ ซึ่งผลลัพธ์ที่ได้จะอยู่ในรูปของตัวแบบโมเดล เช่น โครงสร้างต้นไม้ตัดสินใจ

ขั้นตอนที่ 2 การทดสอบประสิทธิภาพของโมเดล เป็นขั้นตอนที่เราจะต้องทำการวัดประสิทธิภาพของโมเดลที่ได้สร้าง โดยการวัดประสิทธิภาพของโมเดลสามารถแบ่งได้เป็น

<u>ตัววัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล</u> โดยทั่วไปแล้วจะมีตัววัดที่นิยมใช้ อยู่ 4 ค่า ดังนี้

1. Precision เป็นการวัดความแม่นยำของโมเดล โดยพิจารณาแยกทีละคลาส

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}$$

True Positive (TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่

False Positive (FP) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งกำลังสนใจอยู่

2. Recall เป็นการวัดความถูกต้องของโมเดล โดยพิจารณาแยกทีละคลาส

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

True Positive (TP) คือ จำนวนข้อมูลที่ทำนายถูกว่าเป็นคลาสซึ่งกำลังสนใจอยู่

False Negative (FN) คือ จำนวนข้อมูลที่ทำนายผิดมาเป็นคลาสซึ่งไม่ได้สนใจอยู่

3. F-measure เป็นการวัดค่า Precision และ Recall พร้อมกันของโมเดล โดยพิจารณาแยกทีละคลาส

$$\text{F-measure}=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}}$$

4. Accuracy เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส คือจำนวน True Positive ของทุกคลาสรวมกัน

<u>การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล</u>

เป็นการแบ่งข้อมูลเพื่อมาทดสอบ แบ่งได้ 3 วิธีการใหญ่ๆคือ

1. วิธี Self-Consistency Test หรือเรียกว่า Use Training set เป็นวิธีที่ง่ายที่สุด เนื่องจากข้อมูลที่ใช้ในการสร้างโมเดลและข้อมูลที่ใช้ในการทดสอบโมเดลเป็นข้อมูลชุดเดียวกัน

2. วิธี Split Test เป็นการแบ่งข้อมูลด้วยการสุ่มออกเป็น 2 ส่วน เช่น 70% ต่อ 30% หรือ 80% ต่อ 20% โดยข้อมูลส่วนที่หนึ่งใช้ในการสร้างโมเดล และข้อมูลส่วนที่สองใช้ในการทดสอบประสิทธิภาพของโมเดล

3. วิธี Cross-Validation Test เป็นวิธีที่นิยมใช้ในการทดสอบประสิทธิภาพของโมเดล เนื่องจากผลที่ได้มีความน่าเชื่อถือ การวัดวิธีนี้จะทำการแบ่งข้อมูลออกเป็นหลายส่วน เช่น 5-fold cross-validation คือการแบ่งข้อมูลออกเป็น 5 ส่วน โดยที่แต่ละส่วนมีจำนวนข้อมูลเท่ากัน หลังจากนั้นข้อมูลส่วนหนึ่งจะใช้เป็นตัวทดสอบประสิทธิภาพของโมเดล ทำวนเช่นนี้ไปจนครบจำนวนที่แบ่งไว้

ขั้นตอนที่ 3 การใช้ตัวแบบโมเดล เพื่อการทำนาย ซึ่งเป็นจุดประสงค์หลักที่ใช้ในการแก้ไขกับปัญหา คือการสร้างตัวต้นแบบ เมื่อมีข้อมูลใหม่เข้ามาสู่ระบบจะสามารถทำนายได้โดยการนำข้อมูลที่ได้รับมาทำการเปรียบเทียบกับตัวต้นแบบโมเดลที่เราจำแนกได้จากขั้นตอนที่ 1 เพื่อวิเคราะห์ เพื่อตัดสินใจในความเป็นไปได้ของข้อมูลนั้นๆ

การแบ่งกลุ่มข้อมูลด้วยวิธี Decision Tree

เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ช่วยในการตัดสินใจจะมีลักษณะคล้ายกับโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณลักษณะ (Attribute) แต่ละกิ่งแสดงเงื่อนไขในการทดสอบและโหนดปลาย แสดงกลุ่มที่กำหนดไว้

การแทนต้นไม้การตัดสินใจ (Decision Tree Representation)

1. โหนดภายใน (internal node) คือคุณสมบัติต่างๆของข้อมูล ซึ่งเมื่อข้อมูลใดๆตลลงมาที่โหนด จะใช้คุณสมบัตินี้เป็นตัวตัดสินใจว่าข้อมูลที่ได้จะไปในทิศทางใด โดยโหนดภายในที่เป็นจุดเริ่มต้นของต้นไม้เราเรียกว่าโหนดราก

2. กิ่ง (Branch, Link) เป็นค่าคุณสมบัติของคุณสมบัติในโหนดภายในที่แตกกิ่งนี้ออกมา ซึ่งโหลดภายในจะแตกกิ่งเป็นจำนวนเท่ากับค่าคุณสมบัติของโหนดภายในนั้นๆ

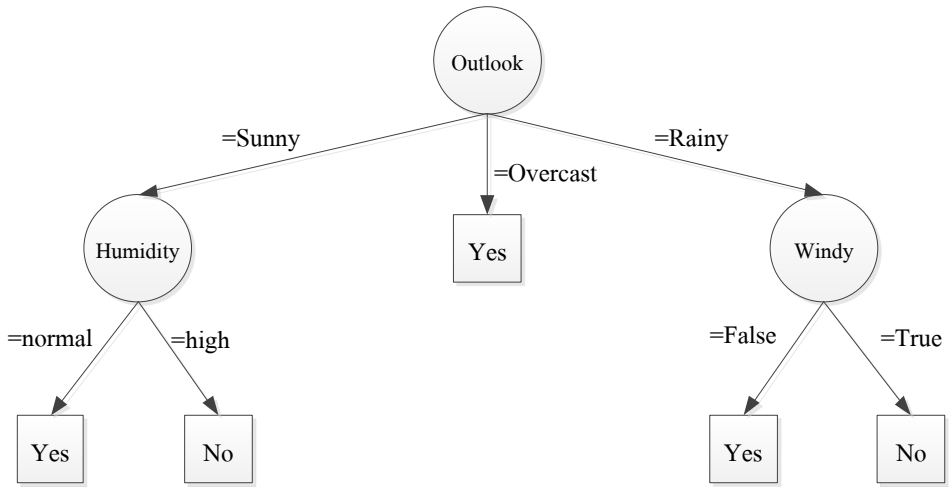
3. โหนดใบ (Leaf node) คือกลุ่มต่างๆซึ่งเป็นผลลัพธ์ในการแยกแยะข้อมูล

เทคนิคการจำแนกประเภทข้อมูลด้วยวิธี Decision Tree เป็นวิธีการหนึ่งที่นิยมใช้กันอย่างแพร่หลาย เนื่องจากเป็นโมเดลที่ได้สามารถแปลความหมายและเข้าใจได้ง่าย ซึ่งจะใช้ข้อมูลดังตารางที่ 2.1 แสดงข้อมูลสภาพอากาศย้อนหลัง 14 วัน มาดำเนินการศึกษารูปแบบ กระบวนการทำงานของการจำแนกประเภทข้อมูลด้วยวิธีการสร้างต้นไม้ในการตัดสินใจ

ตารางที่ 2.1 แสดงข้อมูลสภาพอากาศย้อนหลัง 14 วัน

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
Sunny	Hot	High	False	No
Sunny	Hot	High	True	No
Overcast	Hot	High	False	Yes
Rainy	Mild	High	False	Yes
Rainy	Cool	Normal	False	Yes
Rainy	Cool	Normal	True	No
Overcast	Cool	Normal	True	Yes
Sunny	Mild	High	False	No
Sunny	Mild	Normal	False	Yes
Rainy	Mild	Normal	False	Yes
Sunny	Mild	Normal	True	Yes
Overcast	Mild	High	True	Yes
Overcast	Hot	Normal	False	Yes
Rainy	Mild	High	True	No

จากข้อมูลในตารางที่ 2.1 เราสามารถนำมาสร้างเป็นโครงสร้างต้นไม้ช่วยในการตัดสินใจ ดังนี้



ภาพที่ 2.2 โมเดลโครงสร้างต้นไม้ตัดสินใจ

เมื่อนำโมเดลโครงสร้างต้นไม้ตัดสินใจ นำไปใช้งานเราจะเริ่มพิจารณาจากโหนดบนสุด(Root node) เช่น ถ้าแอตทริบิวต์ Outlook มีค่าเป็น Overcast แล้วคลาส Play คำตอบที่ได้เป็น Yes และยังสามารถนำมาสร้างเป็นกฎ IF-THEN ได้ ดังนี้

IF Outlook=Overcast THEN play=yes

IF Outlook=rainy AND Windy= TRUE THEN play=no

IF Outlook=rainy AND Windy= FALSE THEN play=yes

IF Outlook=sunny AND Humidity= high THEN play=no

IF Outlook= sunny AND Humidity = normal THEN play=yes

การจำแนกประเภทข้อมูลด้วยวิธี Naïve Bayes

การจำแนกประเภทข้อมูลด้วยวิธี Naïve Bayes เป็นอีกวิธีหนึ่งที่ได้รับความนิยม เนื่องจากเป็นการสร้างโมเดลที่ง่ายและไม่ซับซ้อน โดยจะอาศัยทฤษฎีความน่าจะเป็น (Probability) เป็นหลัก โดยสามารถแสดงสมการของความน่าจะเป็นได้ดังนี้

$$P(A|B)=\frac{P(A\cap B)}{P(B)}$$

โดยที่ P(A|B) คือ ค่า Conditional probability หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ B ขึ้นก่อนและจะมีเหตุการณ์ A ตามมา

P(A∩B) คือ ค่า joint probability หรือค่าความน่าจะเป็นที่เหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน

P(B) คือ ค่าความน่าจะเป็นที่เหตุการณ์ B เกิดขึ้น

ในลักษณะเดียวกันเราจะเขียน P(B|A) หรือค่าความน่าจะเป็นที่เหตุการณ์ A เกิดขึ้นก่อนและเหตุการณ์ B เกิดขึ้นตามมาทีหลังได้เป็น

$$P(B|A)=\frac{P(A\cap B)}{P(A)}$$

จากสมการทั้ง 2 แบบจะเห็นว่ามีค่า P(A∩B) ที่เหมือนกันอยู่ดังนั้นเราสามารถเขียนสมการของ P(A∩B) ได้เป็นดังนี้

$$P(A\cap B)=P(A|B)\times P(B)=P(B|A)\times P(A)$$

$$P(B|A)=\frac{P(A|B)\times P(B)}{P(A)}$$

จากสมการที่ได้เราเรียกว่า Bayes theorem หรือทฤษฎีของเบย์ ส่วนในการนำไปใช้งานจะขอเปลี่ยนสัญลักษณ์ A และ B ใหม่ให้เป็น A และ C โดยที่ A คือ แอตทริบิวต์ (attribute) และ C คือ ค่าคลาส (class) แสดงได้ดังสมการ

$$P(C|A)=\frac{P(A|C)\times P(C)}{P(A)}$$

จากสมการของ Bayes อธิบายได้ว่า ถ้าต้องการทำนายคลาส C เมื่อเราทราบแอตทริบิวต์ A แล้วสามารถคำนวณได้จาดความน่าจะเป็นของแอตทริบิวต์ A ที่มีคลาส C ในเทรนนิง ดาต้า และค่าความน่าจะเป็นของแอตทริบิวต์ A และคลาส C เพื่อให้เข้าใจง่ายจะแบ่งสมการของ Bayes ออกเป็น 3 ส่วน คือ

(1) Posterior probability หรือ P (C|A) คือ ค่าความน่าจะเป็นที่ข้อมูลที่มีแอตทริบิวต์เป็น A จะมีคลาส C

(2) Likelihood หรือ P (A|C) คือ ค่าความน่าจะเป็นที่ข้อมูล training data ที่มีคลาส C และมีแอตทริบิวต์ A โดยที่ $A = a_1 \cap a_2 \ldots \cap a_M$ โดยที่ M คือจำนวนแอตทริบิวต์ใน training data

(3) Prior probability หรือ P(C) คือ ค่าความน่าจะเป็นของคลาส C

แต่การที่แอตทริบิวต์ $A = a_1 \cap a_2 \ldots \cap a_M$ ที่เกิดขึ้นใน training data อาจจะมีจำนวนน้อยมากหรือไม่มีรูปแบบของแอตทริบิวต์แบบนี้เกิดขึ้นเลย ดังนั้นจึงได้ใช้หลักการที่ว่าแต่ละแอตทริบิวต์เป็น independent ต่อกันทำให้สามารถเปลี่ยนสมการ P(A|C) ได้เป็น

$$P(A|C)=P(a_1|C)\times P(a_2|C)\times\ldots\times P(a_M|C)$$

ซึ่งในการคำนวณเราสามารถคำนวณแยกทีละแอตทริบิวต์ได้ดังสมการข้างล่างนี้

$$P(C|A)=P(a_1|C)\times P(a_2|C)\times\ldots\times P(a_M|C)\times P(C)$$

จากสมการของ Bayes สามารถแสดงวิธีการคำนวณโดยใช้ข้อมูลจากตารางที่ 2.1 ได้ดังต่อไปนี้

ตารางที่ 2.2 แสดงข้อมูลการคำนวณจากโมเดลของ Naïve Bayes

Outlook			**Temperature**			**Humidity**			**Windy**			**Play**	
	Yes	**No**		**Yes**	**No**		**Yes**	**No**		**Yes**	**No**	**Yes**	**No**
Sunny	2	3	Hot	2	2	High	3	4	False	6	2	9	5
Overcast	4	0	Mild	4	2	Normal	6	1	True	3	3		
Rainy	3	2	Cool	3	1								
Sunny	2/9	3/5	Hot	2/9	2/5	High	3/9	4/5	False	6/9	2/5	9/14	5/14
Overcast	4/9	0/5	Mild	4/9	2/5	Normal	6/9	1/5	True	3/9	3/5		
Rainy	3/9	2/5	Cool	3/9	1/5								

ตารางที่ 2.3 แสดงข้อมูลสภาพอากาศในวันปัจจุบัน ซึ่งเรายังไม่สามารถทราบคลาสคำตอบ

Outlook	**Temperature**	**Humidity**	**Windy**	**Play**
sunny	hot	high	false	?

P(Play=yes|A)=P(Outlook=sunny|Play=yes)×P(Temperature=hot|Play=yes)
×P(Humidity=high|Play=yes)×P(Windy=false|Play=yes)×P(Play=yes)
= (2/9) × (2/9) × (3/9) × (6/9) × (9/14)
= 0.22 × 0.22 × 0.33 × 0.67 × 0.64
= 0.0068

P(Play=no|A)=P(Outlook=sunny|Play=no)×P(Temperature=hot|Play=no)
×P(Humidity=high|Play=no)×P(Windy=false|Play=no)×P(Play=no)
= (3/5) × (2/5) × (4/5) × (2/5) × (5/14)
= 0.6 × 0.4 × 0.8 × 0.4 × 0.36
= 0.0276

ดังนั้น ข้อมูลที่เราทำนายจากตารางที่ 2.11 โดยใช้โมเดลของ Naïve bayes พบว่าคลาส Play=no เนื่องจาก ค่า (Play=no|A) = 0.0276 ซึ่งมีค่ามากกว่า P(Play=yes|A) = 0.0068

2.3 ซอฟต์แวร์ที่ใช้ในการศึกษา

2.3.1 โปรแกรม WEKA

คำว่า Weka ย่อมาจากคำว่า Waikato Environment for Knowledge Analysis พัฒนาขึ้นโดยใช้ภาษา java และเป็นซอฟต์แวร์เสรีที่อยู่ภายใต้ข้อตกลงของ GNU (General Public License) ซึ่งภายในโปรแกรมจะมีเมนูให้เราเลือกใช้งานดังภาพที่ 2.12

ภาพที่ 2.3 แสดงหน้าต่าง Weka GUI Chooser

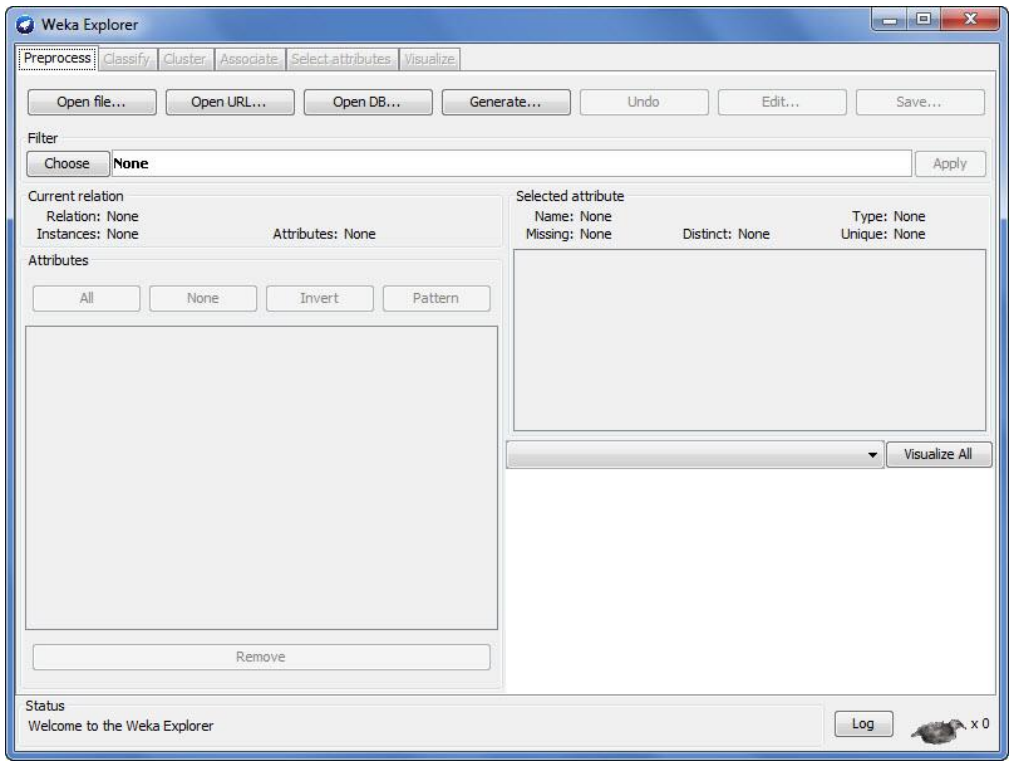
Explorer เป็นส่วนที่ผู้ใช้สามารถใช้ฟังก์ชันการทำงานต่างๆ ของ Weka ผ่านทางหน้าจอ GUI ซึ่งเป็นส่วนที่เหมาะสำหรับผู้ที่เริ่มต้นใช้งาน Weka เพราะผู้ใช้จะสามารถเรียกฟังก์ชันการทำงานต่างๆ ได้เพียงแค่คลิกและเปลี่ยนแปลงพารามิเตอร์ในหน้าฟอร์มเท่านั้น

Experimenter เป็น GUI ที่ยอมให้ผู้ใช้สามารถเปลี่ยนแปลงเทคนิคในการวิเคราะห์ข้อมูลและค่าพารามิเตอร์ต่างๆ ที่เกี่ยวข้องได้หลากหลายรูปแบบ จนได้ผลลัพธ์ที่น่าพอใจ ซึ่งการใช้ Explorer นั้นไม่สะดวกเนื่องจากต้องมาทำการเปลี่ยนแปลงค่าต่างๆ

Knowledge Flow เป็น GUI อีกส่วนหนึ่งที่ยอมให้ผู้ใช้สามารถนำเทคนิคต่างๆ ของ Weka มาเรียงต่อกันเพื่อช่วยให้การวิเคราะห์ข้อมูลให้ทำงานได้ตามที่ต้องการ

Simple CLI เป็นส่วนที่ผู้ใช้สามารถเรียกฟังก์ชันการทำงานของ Weka มาใช้ผ่านทาง command line ได้ ซึ่งการเรียกใช้ฟังก์ชันผ่านทาง command line นี้จะช่วยให้ผู้ใช้เข้าใจการเรียกฟังก์ชันต่างๆเบื้องหลังหน้าจอ GUI ของ Weka ได้และสามารถนำไปประยุกต์ใช้ในการเขียนโปรแกรมเพื่อเรียก Weka ในการใช้งานได้อีกด้วย

ในส่วนนี้จะนำเสนอหน้าต่างในส่วนของ Explorer เป็นหลักโดยในหน้าต่าง Weka Explorer สามารถแบ่งพื้นที่ได้ดังนี้



ภาพที่ 2.4 เมนูหลักในหน้าต่าง Weka Explorer

ส่วนบนสุดจะเป็นแท็บ (tab) ซึ่งมีด้วยกันทั้งหมด 6 แท็บวางเรียงกันอยู่ทางด้านบน ซึ่งแท็บต่างๆ เหล่านี้จะเป็นเมนูให้ผู้ใช้สามารถใช้งานเทคนิคต่างๆ ของ Weka ได้

ส่วนที่อยู่ตรงกลางซึ่งจะเปลี่ยนไปตามการกดแท็บต่างๆ ส่วนนี้เป็นส่วนของการเลือก option ต่างๆ ในการวิเคราะห์ข้อมูล และส่วนการแสดงผลลัพธ์หลังจากทำการวิเคราะห์ข้อมูลเสร็จ ส่วนนี้บางครั้งจะเรียกว่า Workspace

ส่วนที่อยู่ด้านล่างสุด จะเป็นส่วนที่บอกสถานะ (status) ของการทำงานในแต่ละขั้นตอน และปุ่มทางขวามือจะเป็นปุ่มสำหรับดู Log ไฟล์ เมื่อคลิกจะแสดงหน้าจอของ log ขึ้นมา

2.3.2 ภาษา PHP

PHP เป็นชื่อย่อของภาษาโปรแกรมมิ่งชนิดหนึ่งที่มีชื่อว่า “Professional Home pages” แต่ในปัจจุบันภาษาชนิดนี้ถูกพัฒนาต่อมาจนกลายเป็นภาษาโปรแกรมมิ่งชนิดใหม่ที่มีชื่อเรียกว่า “Personal Hypertext Processor: PHP” ภาษาชนิดใหม่นี้เป็นที่นิยมในการนำมาพัฒนาใช้เขียนสคริปต์ (ชุดคำสั่งควบคุมการทำงานของโปรแกรม ซึ่งมีความยาวไม่มากนักและสามารถทำงานได้ดีกับเว็บไซต์) PHP เป็นภาษาสคริปต์ที่เป็น Server Side Scrip และเป็น Open Source ที่ผู้ใช้ทั่วไปสามารถ Download Source Code ได้ฟรี จุดประสงค์ที่สำคัญของภาษา PHP คือการช่วยให้นักพัฒนาเว็บเพจสามารถเขียนเว็บเพจที่เป็นแบบไดนามิคได้อย่างรวดเร็ว ภาษา PHP จะทำงานร่วมกันกับเอกสาร HTML โดยการสร้างโค้ดแทรกระหว่าง Tag HTML และสร้างเป็นไฟล์ที่มีนามสกุล .php .php3 หรือ php4 และไวยากรณ์ที่ใช้ใน PHP เป็นการนำรูปแบบของภาษาต่างๆมารวมกัน เช่น ภาษา C Perl และ Java ทำให้ผู้ใช้ที่มีพื้นฐานของภาษาเหล่านี้สามารถใช้งานได้

เนื่องจาก PHP จะทำงานโดยมีตัวแปลและเอ็กซิคิวต์ที่ฝั่งเซิร์ฟเวอร์อาจจะเรียกการทำงานว่าเป็นเซิร์ฟเวอร์ไซด์ (Server Side) ส่วนการทำงานของบราวเซอร์ของผู้ใช้เรียนว่าไคลเอนไซด์ (Client Side) โดยการทำงานจะเริ่มตัวที่ผู้ใช้ส่งความต้องการผ่านเว็บบราวเซอร์ทาง HTTP (HTTP Request) ซึ่งอาจเป็นการกรอกแบบฟอร์ม หรือใส่ข้อมูลที่ต้องการ หรือแสดงโดยเรียกเอกสาร PHP (เอกสารนี้จะมีส่วนขยายเป็น php) เช่น test.php เมื่อเอกสาร PHP เข้ามาถึงเว็บเซิร์ฟเวอร์ก็จะถูกส่งต่อไปให้ PHP Interpreter เพื่อทำหน้าที่แปลคำสั่งแล้วเอ็กซิคิวต์คำสั่งตามบรรทัดที่ระบุบคำสั่งนั้นๆ จากนั้น PHP จะสร้างผลลัพธ์ในรูปแบบเอกสาร HTML ส่งกลับไปให้เว็บเซิร์ฟเวอร์เพื่อส่งต่อไปให้บราวเซอร์แสดงผลทางฝั่งผู้ใช้ต่อไป (HTTP Response) ดังรูปที่ 2.5 ตามกระบวนการดังนี้

1. จากไคลเอนต์จะเรียกไฟล์ php script ผ่านทางโปรแกรมบราวเซอร์ (Internet Explore)

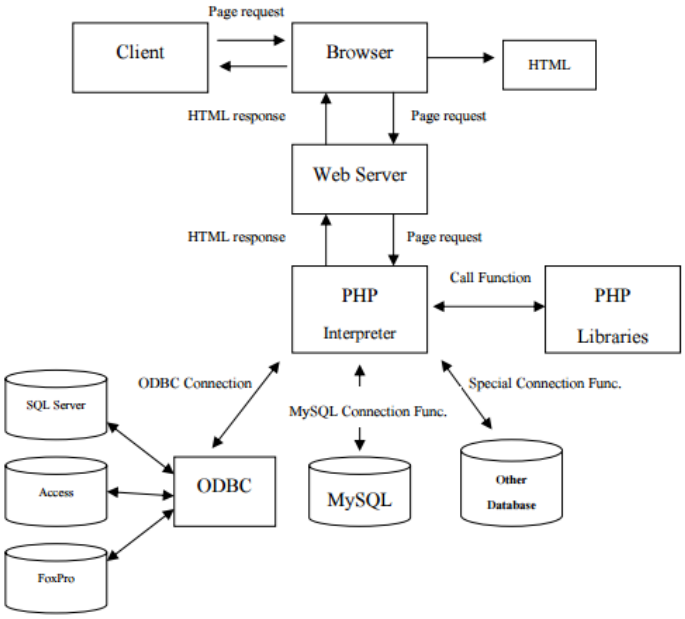
2. บราวเซอร์จะส่งคำร้อง (Request) ไปยังเว็บเซิร์ฟเวอร์ผ่านทางเครือข่ายอินเทอร์เน็ต

3. เมื่อเว็บเซิร์ฟเวอร์รับคำร้องขอจากบราวเซอร์แล้วก็จะนำสคริปต์ php ที่เก็บอยู่ในเซิร์ฟเวอร์มาประมวลผลด้วยโปรแกรมแปลภาษา PHP ที่เป็น อินเตอร์พรีเตอร์

4. กรณีที่ php script มีการเรียกใช้ข้อมูลก็จะติดต่อกับฐานข้อมูลต่าง ๆ ผ่านทาง ODBC Connection ถ้าเป็น ฐานข้อมูลกลุ่ม Microsoft SQL Server, Microsoft Access, FoxPro หรือใช้ Function Connection ที่มีอยู่ใน PHP Library ในการเชื่อมต่อฐานข้อมูลเพื่อดึงข้อมูลออกมา หลังจากแปลสคริปต์ PHP เสร็จแล้วจะได้รับไฟล์ HTML ใหม่ที่มีแต่แทกซ์ HTML ไปยัง Web Service

5. Web Service ส่งไฟล์ HTML ที่ได้ผ่านการแปลแล้วกลับไปยังบราวเซอร์ที่ร้องขอผ่านทางเครือข่ายอินเตอร์เน็ต

6. บราวเซอร์รับไฟล์ HTML ที่เว็บเซิร์ฟเวอร์ส่งมาให้แปล HTML แสดงผลออกมาทางจอภาพเป็นเว็บเพจ โดยใช้ตัวแปลภาษา HTML ที่อยู่ในบราวเซอร์ซึ่งเป็นอินเตอร์พรีเตอร์เช่นเดียวกัน



ภาพที่ 2.5 แสดงขั้นตอนการทำงาน PHP Script Request/Response

ความสามารถของภาษา PHP

ภาษา PHP เป็นภาษาที่พัฒนาขึ้นจากพื้นฐานของภาษาโปรแกรมมิ่งชนิดอื่นๆ เช่น C, C++และ Perl ทำให้มีลักษณะเด่นของภาษาต้นแบบแต่ละชนิดรวมกันอยู่ ความสามารถของภาษา PHPที่เห็นได้อย่างเด่นชัด สามารถจำแนกออกได้ดังนี้

เป็นภาษาที่ทำความเข้าใจและใช้งานง่ายไม่เหมือนกับ JAVA หรือ C++ และมีส่วนที่สนับสนุนการทำงานได้กับทุกเว็บไซต์

เป็น Open Source ผู้ใช้สามารถดาวน์โหลด และนำ source code ของ PHP ไปใช้ได้โดยไม่เสียค่าใช้จ่าย

เป็นสคริปต์แบบเซิร์ฟเวอร์ไซด์ดังนั้นจึงทำงานบนเว็บเซิร์ฟเวอร์ไม่ส่งผลกับการทำงานของเครื่องไคลเอนต์โดย PHP จะอ่านโค้ด และทำงานที่เซิร์ฟเวอร์จากนั้นจึงส่งผลลัพธ์ที่ได้จากการประมวลผลมาที่เครื่องของผู้ใช้ในรูปแบบของเอกสาร HTML ซึ่งอ่านโค้ดของ PHP ผู้ใช้ไม่สามารถมองเห็นได้

PHP สามารถทำงานได้ในระบบปฏิบัติการที่ต่างชนิดกัน เช่น Unix, Windows,Mac, OS หรือ Risc OS อย่างดีประสิทธิภาพเนื่องมาจาก PHP เป็นสริปต์ที่ต้องทำงานบนเซิร์ฟเวอร์ดังนั้นคอมพิวเตอร์ที่ใช้ สำหรับเรียกคำสั่ง PHP จึงจำเป็นต้องติดตั้งโปรแกรมประเภทเว็บเซิร์ฟเวอร์ไว้ด้วยเพื่อให้สามารถประมวลผล PHP ได้ซึ่งเป็นเหตุผลที่ทำให้ PHP สมารถทำงานได้กับหลายระบบปฏิบัติการหลายชนิด

PHP สามารถทำงานได้ในเว็บเซิร์ฟเวอร์หลายชนิด เช่น Personal Web Server (PWS), Apache, OmniHttpd, Microsoft Internet Information Server (IIS) เป็นต้น

สนับสนุนการเขียนสริปต์ที่ใช้หลักของ Object Orientation

PHP สามารถสร้างเว็บไซต์ที่บรรจุข้อมูลรูปแบบต่าง ๆ ลงในเว็บ เช่น รูปภาพ ไฟล์ PDF หรือ Flash Movie

คุณสมบัติที่สำคัญอีกประการหนึ่งของ PHP คือความสามารถในการทำงานร่วมกับระบบจัดการฐานข้อมูลที่หลากหลายซึ่งระบบการจัดการฐานข้อมูลที่สนับสนุนการทำงานของ PHP มีตัวอย่างดังนี้

1) ชนิด ORACLE เช่น Oracle (OC17 and OC18), AdabasD, Ingres, FilePro (read-only) และ Solid

2) ชนิด Access เช่น dBase, InterBase, Ovrimos Empress และ FrontBase

3) ชนิด SQL เช่น MS SQL, PostgreSQL, mSQLและMySQL

PHP อนุญาตให้ผู้ใช้สร้างเว็บไซต์ซึ่งทำงานผ่านโปรโตคอล (Protocol) ชนิดต่างๆ ได้ เช่น LDAP, IMAP, SNMP, NNTP, POP3, HTTP และ COM (สำหรับ Windows)

ผู้ใช้สามารถเขียนโค้ด PHP และอ่านข้อมูลในรูปแบบของ Extensible Markup Language (XML) ได้

2.3.3 ระบบจัดการฐานข้อมูล MySQL

MySQL (อ่านว่า “มาย-เอส-คิว-แอล”) จัดเป็นระบบจัดการฐานข้อมูลเชิงสัมพันธ์ (RDBMS: Relational Database Management System) ซึ่งเป็นที่นิยมกันมากในปัจจุบัน เพราะว่า MySQL เป็นฟรีแวร์ทางด้านฐานข้อมูลที่มีประสิทธิภาพสูง และเป็นทางเลือกใหม่จากผลิตภัณฑ์ระบบจัดการฐานข้อมูลในปัจจุบัน ผู้ที่เคยใช้ MySQL ต่างยอมรับในความสามารถ ความรวดเร็ว การรองรับจำนวนผู้ใช้และขนาดของข้อมูลจำนวนมหาศาล ทั้งยังสนับสนุนการใช้งานบนระบบปฏิบัติการมากมาย ไม่ว่าจะเป็น Unix, OS/2, Mac OS หรือWindows นอกจากนี้ MySQL ยังสามารถใช้งานร่วมกับ Web Development Platform ทั้งหลาย เช่น C, C++, Java, Perl, PHP, Python, Tcl หรือ ASP

สถาปัตยกรรมของ MySQL

สถาปัตยกรรม หรือ โครงสร้างภายในของ MySQL ก็คือ การออกแบบการทำงานในลักษณะของ Client/Server ซึ่งประกอบด้วยส่วนหลักๆ 2 ส่วน คือ ส่วนของผู้ให้บริการ (Server) และ ส่วนของผู้ใช้บริการ (Client) โดยในแต่ละส่วนจะมีโปรแกรมสำหรับการทำงานตามหน้าที่ของตน

ส่วนของผู้ให้บริการ หรือ Server จะเป็นส่วนที่ทำหน้าที่บริหารจัดการระบบฐานข้อมูล ในที่นี้ก็หมายถึงตัว MySQL Server นั่นเองและเป็นที่จัดเก็บข้อมูลทั้งหมดข้อมูลที่เก็บไว้นี้มีข้อมูลที่จำเป็นสำหรับการทำงานกับระบบฐานข้อมูลและข้อมูลที่เกิดจากการที่ผู้ใช้แต่ละคนสร้างขึ้นมา

ส่วนของผู้ใช้บริการ หรือ Client ก็คือผู้ใช้นั่นเอง โดยโปรแกรมสำหรับใช้งานในส่วนนี้ได้แก่ MySQL Client, Access, Web Development Platform ต่างๆ (เช่น Java, Perl, PHP, ASP)

หลักการทำงานในลักษณะ Client/ Server มีดังนี้

1. ที่ฝั่งของ Server จะมีโปรแกรมหรือระบบสำหรับจัดการฐานข้อมูลทำงานรออยู่ เพื่อเตรียมหรือรอคอยการร้องขอการใช้บริการจาก Client

2. เมื่อมีการร้องขอการใช้บริการเข้ามา Server จะทำการตรวจสอบตามวิธีการของตน เช่น อาจจะมีการให้ผู้ใช้บริการระบุชื่อและรหัสผ่าน และสำหรับ MySQL สามารถกำหนดได้ว่าจะอนุญาตหรือปฏิเสธ Client ใดๆ ในระบบที่จะเข้าใช้บริการอีกด้วย

3. ถ้าผ่านการตรวจสอบ Server ก็จะอนุมัติการให้บริการแก่ Client ที่ร้องขอการใช้บริการนั้นๆ ต่อไปและถ้าในกรณีที่ไม่ได้รับการอนุมัติ Server ก็จะส่งข่าวสารความผิดพลาดแจ้งกลับไปที่ Client ที่ร้องขอการใช้บริการนั้น

เครื่องคอมพิวเตอร์ที่ทำหน้าที่เป็น Client หรือ Server อาจจะอยู่บนเครื่องเดียวกัน หรือแยกเครื่องกันก็ได้ทั้งนี้ขึ้นอยู่กับลักษณะการทำงาน หรือการกำหนดของผู้บริหารระบบ ตามปกติถ้าเป็นการทำงานลักษณะ Web-based มีการใช้ฐานข้อมูลขนาดไม่ใหญ่นักตัว MySQL Server และ Client มักจะมีอยู่บนเครื่องเดียวกัน โดยเครื่องคอมพิวเตอร์ดังกล่าวจะต้องมีทรัพยากรเพื่อการทำงาน เช่น เนื้อที่ฮาร์ดดิสก์, RAM มากพอสมควร แต่สำหรับการทำงานจริง (Real-world Application) ก็มักจะแยก Client และ Server ออกเป็นคนละเครื่องกัน และสามารถรองรับงานได้ดีมากกว่า

2.3.4 โปรแกรม Adobe Dreamweaver CS5

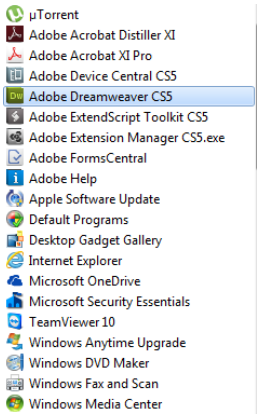
Adobe Dreamweaver CS5 เป็นโปรแกรมสำเร็จรูปที่ใช้สำหรับการพัฒนาเว็บไซต์ โดยมีคุณสมบัติในการออกแบบและสร้างเว็บเพจ จนถึงการพัฒนาแอพลิเคชั่นเบื้องต้น ซึ่งในโปรแกรมตัวนี้มีเครื่องมือสำหรับการวางข้อความ ภาพกราฟิก ตาราง แบบฟอร์ม พร้อมทั้งมัลติมีเดียต่างๆ เพื่อแสดงให้ผู้พัฒนาเว็บไซต์ใช้งานได้ง่าย โดยไม่ต้องรู้จักภาษา HTML, JavaScript, CSS หรือภาษาสคริปต์อื่นๆ ซึ่งเมื่อออกแบบและพัฒนาเว็บไซต์ในโปรแกรม แล้วนำมาแสดงผลทางบราวเซอร์ก็จะเห็นเป็นลักษณะที่ได้จัดวางไว้ หรือจะเรียกอีกอย่างว่า What You See Is What You Get (WYSIWYG)

เริ่มต้นใช้งาน Dreamweaver

การเรียกเปิดโปรแกรม Dreamweaver ขึ้นมาใช้งานมีขั้นตอนดังนี้

1.คลิก All Programs > Adobe Web Premium CS5 >Adobe Dreamweaver CS5

2.จะเข้าสู่หน้าจอโปรแกรม



ภาพที่ 2.6 แสดงขั้นตอนการเข้าสู่โปรแกรม Adobe Dreamweaver CS5

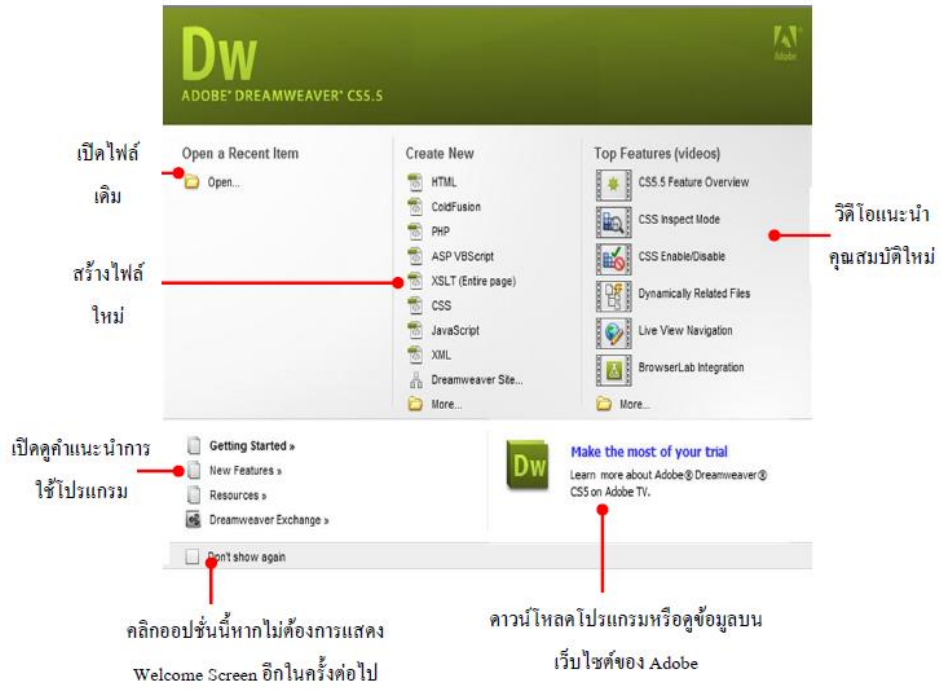
Welcome Screen

Welcome Screen เป็นเครื่องมือสาหรับช่วยเลือกขั้นตอนเริ่มต้นในการใช้งาน โปรแกรม โดยตัวเลือกจะแบ่งออกเป็นกลุ่มๆดังภาพต่อไปนี้

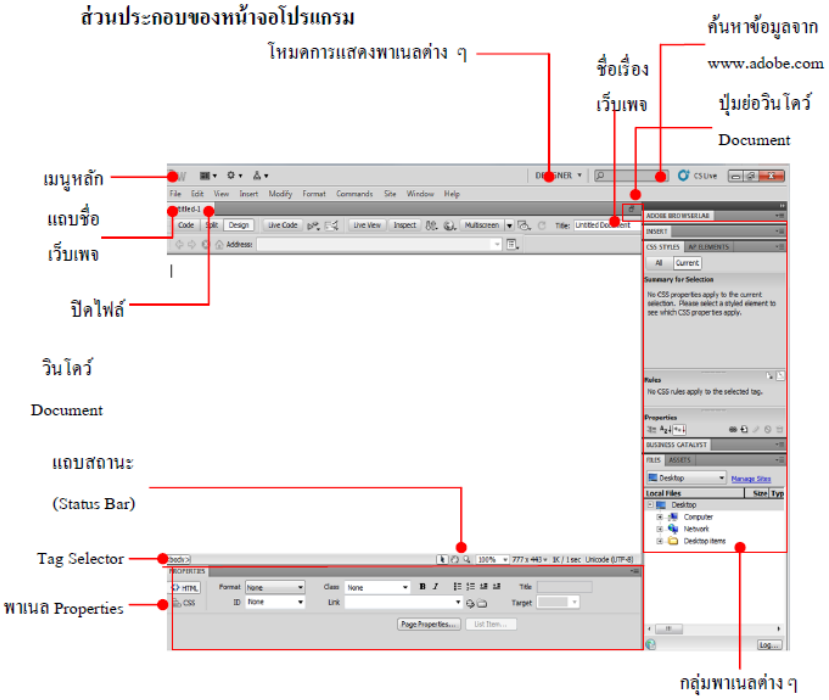
Open a Recent Item เปิดไฟล์ที่เคยใช้ โดยคลิกเลือกจากรายชื่อที่แสดงอยู่ (เรียงลาดับจากที่เคยเปิดหลังสุดเป็นต้นไปหรือคลิก Open เพื่อเปิดไฟล์อื่นๆ

Create New สร้างไฟล์ใหม่ โดยถ้าคลิก HTML จะเป็นการสร้างเว็บเพจพื้นฐานแต่ถ้าคลิกหัวข้ออื่นจะเป็นการสร้างเว็บเพจหรือไฟล์ตามชนิดนั้นๆ

Top Features (videos) เป็นเส้นทางลัดสาหรับเข้าดูรายละเอียดและเทคนิคในการใช้งานต่าง ๆ ของโปรแกรมผ่านทางเว็บไซต์ของ Adobe



ภาพที่ 2.7 ส่วนประกอบของ Welcome Screen



ภาพที่ 2.8 ส่วนประกอบของหน้าจอโปรแกรม

2.4 การวิเคราะห์และออกแบบระบบ

วงจรการพัฒนาระบบ (System Development Life Cycle :SDLC) คือกระบวนการหรือวงจรในการพัฒนาระบบด้านคอมพิวเตอร์หรือระบบสารสนเทศ โดยแปลจากความต้องการของผู้ใช้งานมาเป็นในรูปของแอพพลิเคชั่น โดยมีการกำหนดกิจกรรมต่างๆที่เกิดขึ้นในแต่ละระยะของการพัฒนาระบบขึ้นอย่างชัดเจนตั้งแต่ระยะเริ่มแรกไปจนถึงหลังระยะสิ้นสุดการพัฒนาระบบมีอยู่ด้วยกัน 7 ขั้น ด้วยกัน คือ

1. การกำหนดปัญหา (Problem Definition) สรุปขั้นตอนของระยะการกำหนดปัญหา
รับรู้สภาพของปัญหาที่เกิดขึ้น
ค้นหาต้นเหตุของปัญหา รวบรวมปัญหาของระบบงานเดิม
ศึกษาความเป็นไปได้ของโครงการพัฒนาระบบ
จัดเตรียมทีมงาน และกำหนดเวลาในการทำโครงการ
ลงมือดำเนินการ

2. วิเคราะห์ (Analysis) สรุปขั้นตอนของระยะการวิเคราะห์
วิเคราะห์ระบบงานปัจจุบัน
การกำหนดความต้องการ หรือเป้าหมายของระบบใหม่
วิเคราะห์ความต้องการเพื่อสรุปเป็นข้อกำหนด
สร้างแผนภาพ DFD และแผนภาพภาพ E – R

3. ออกแบบ (Design) สรุปขั้นตอนของระยะการออกแบบ
พิจารณาแนวทางในการพัฒนาระบบ
ออกแบบสถาปัตยกรรมระบบ
ออกแบบรายงาน
ออกแบบหน้าจออินพุตข้อมูล
ออกแบบผังงานระบบ
ออกแบบฐานข้อมูล
การสร้างต้นแบบ
การออกแบบโปรแกรม

4.การพัฒนา (Development) สรุปขั้นตอนของระยะการพัฒนา
พัฒนาโปรแกรม
เลือกภาษาโปรแกรมที่เหมาะสม
สามารถนำเครื่องมือมาช่วยพัฒนาโปรแกรมได้

5. การทดสอบ (Testing) สรุปขั้นตอนของระยะการพัฒนา

ทดสอบไวยากรณ์ภาษาคอมพิวเตอร์

ทดสอบความถูกต้องของผลลัพธ์ที่ได้

ทดสอบว่าระบบที่พัฒนาตรงตามความต้องการของผู้ใช้หรือไม่

สร้างเอกสารประกอบโปรแกรม

6. การนำระบบไปใช้ (Implementation Phase) หรือ Installation สรุปขั้นตอนของระยะการนำระบบไปใช้

ศึกษาสภาพแวดล้อมของพื้นที่ก่อนที่จะนำระบบไปติดตั้ง

ติดตั้งระบบให้เป็นไปตามสถาปัตยกรรมระบบที่ออกแบบไว้

จัดทำคู่มือระบบ

ฝึกอบรมผู้ใช้

ดำเนินการใช้ระยะงานใหม่

ประเมินผลการใช้งานของระบบใหม่

7. การบำรุงรักษา (Maintenance) สรุปขั้นตอนของระยะการบำรุงรักษา

กรณีเกิดข้อผิดพลาดขึ้นจากระบบ ให้ดำเนินการแก้ไขให้ถูกต้อง

อาจจำเป็นต้องเขียนโปรแกรมเพิ่ม กรณีที่ผู้ใช้มีความต้องการเพิ่มเติม

วางแผนรองรับเหตุการณ์ที่อาจเกิดขึ้นในอนาคต

บำรุงรักษาระบบงาน และอุปกรณ์

ภาพที่ 2.9 แสดงวงจรการพัฒนาระบบ System Development Life Cycle :SDLC

2.5 งานวิจัยที่เกี่ยวข้อง

2.5.1 งานวิจัยเรื่อง Using Data Mining technique to select the law and article for lawsuits โดยธีระวัฒน์ พงษ์ศิริปรีดาและกฤษณะ ไวยมัย

งานวิจัยนี้ได้นำเสนอเทคนิค Data Classification มาใช้สร้างระบบจัดสรรกฎหมายที่เหมาะสมให้กับคดีความแบบอัตโนมัติ ซึ่งผลการทดลองที่ได้ยังไม่เป็นที่น่าพอใจ ซึ่งสามารถสรุปได้ ดังนี้

ประเด็นที่ 1 การตัดคำในขั้นตอนการเตรียมข้อมูลยังขาดประสิทธิภาพ ทำให้ผลการทำนายของระบบมีความถูกต้องน้อยมากหรืออยู่ในระดับต่ำ

ประเด็นที่ 2 ระบบไม่สามารถทำนายกฎหมายให้กับคดีความได้ครั้งละมากกว่าหนึ่งกฎหมาย

ประเด็นที่ 3 ระบบไม่สามารถทำนายกฎหมายแบบมีลำดับชั้นในเวลาเดียวกันได้

2.5.2 งานวิจัยเรื่อง การจัดสรรกฎหมายที่เหมาะสมให้กับคดีความอัตโนมัติโดยธีระวัฒน์ พงษ์ศิริปรีดา ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

งานวิจัยนี้ได้นำเสนอเทคนิค การตัดคำโดยใช้เทคนิค Suffix Array เพื่อทำการสร้างพจนานุกรมวลีขึ้นมาใช้งาน แล้วสร้างตัวทำนายผลกฎหมายโดยใช้เทคนิค Association Rule Discovery มาใช้ร่วมกับเทคนิค Data Classification ซึ่งการจำแนกประเภทข้อมูลให้ความแม่นยำในการทำนายสูง เนื่องจากได้นำเทคนิคการหากฎความสัมพันธ์เข้ามาร่วมกับเทคนิคการจำแนกประเภทข้อมูล แต่เนื่องจากวิธีการทำนายของเทคนิคดังกล่าวได้พิจารณาเพียงเฉพาะกฎความสัมพันธ์เดียวในการทำนาย ซึ่งอาจจะทำให้เกิดความผิดพลาดได้ ซึ่งผลจากการทดสอบระบบนั้นสามารถจำแนกข้อดีและข้อเสีย ได้ดังต่อไปนี้

ข้อดี

(1) สามารถทำนายกฎหมายให้กับคดีความแบบอัตโนมัติได้

(2) สามารถทำนายกฎหมายได้ครบทุกมาตรา

(3) สามารถระบุจำนวนมาตราที่ต้องการเพื่อให้ระบบสามารถทำนายได้

ข้อเสีย

(1) ต้องแยกกฎเกณฑ์สำหรับการทำนายผลออกเป็น 2 ระดับ คือ ระดับกฎหมายและระดับมาตรา

(2) ร้อยละของความถูกต้องที่ได้จากระบบนั้นยังไม่สูงมากพอ

2.5.3 งานวิจัย เรื่องการวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ โดย ชัดชัยแก้วตา และอัจฉรา มหาวีรวัฒน์ ภาควิชาเทคโนโลยีสารสนเทศ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

งานวิจัยนี้ได้นำเสนอการวินิจฉัยคดีด้วยเทคนิคต้นไม้ตัดสินใจ โดยกล่าวถึงปัญหาในการวินิจฉัยคดี ซึ่งถ้าไม่มีความชำนาญก็จะทำให้เกิดความยากลำบากในการวินิจฉัยคดี รวมถึงการค้นหาตัวบทกฎหมายมาสนับสนุนในการวินิจฉัยคดี ซึ่งงานวิจัยชิ้นนี้ได้นำเสนอวิธีการวินิจฉัยคดีโดยประยุกต์ใช้เทคนิคต้นไม้ในการตัดสินใจเพื่อจำแนกความถูกผิด ซึ่งสามารถอธิบายได้โดยชุดคุณลักษณะของข้อมูล ออกเป็นรายมาตราที่เหมาะสมกับคดีความนั้นๆและเปรียบเทียบความถูกต้องของการจำแนกข้อมูลด้วยขั้นตอนวิธี ID3 และ C4.5 โดยวัดประสิทธิภาพตัวแบบในการทำนายด้วยวิธี K-fold Cross Validation จากการวินิจฉัยพบว่าขั้นตอนวิธี C4.5 ให้ค่าความถูกต้องสูงกว่าขั้นตอนวิธี ID3

2.5.4 งานวิจัยเรื่อง การใช้เทคนิคการทำเหมืองข้อมูล (Data Mining) เพื่อพัฒนาคุณภาพการศึกษาคณะวิศวกรรมศาสตร์ โดยกฤษณะ ไวยมัย ชิดชนก ส่งศิริ และธนาวินท์ รักธรรมานนท์ อาจารย์ประจำคณะวิศวกรรมศาสตร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ และนิสิตปริญญาโทวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ Published in NECTEC Technical Journal Vol.III, No. 11 page 134-142

งานวิจัยนี้ ได้นำเสนอเทคนิคต่างๆที่สำคัญของการทำเหมืองข้อมูลมาประยุกต์ใช้ในการสืบค้นสิ่งที่น่าสนใจจากข้อมูลนิสิตคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ โดยใช้เทคนิคที่สำคัญ 3 เทคนิค คือ กฎความสัมพันธ์ (Association rule discovery) เทคนิคการจำแนกข้อมูล (Data Classification) การพยากรณ์ข้อมูล (Data Prediction) มาประยุกต์ในการช่วยนิสิตเลือกสาขาที่เหมาะสม และสามารถทำนายเกรดแต่ละรายวิชาในภาคการศึกษาต่อไป โดยใช้ต้นไม้ในการตัดสินใจ (Decision tree) ของการจำแนกข้อมูล ในการสร้างตัวต้นแบบโมเดล (Model)

ข้อดี

สามารถทำนายสาขาวิชาที่เหมาะสมที่สุดให้กับนักศึกษาได้

ข้อเสีย

ตัวแบบโมเดล(Model) จะทำนายแนวโน้มเอียงไปทางสาขาวิชาที่มีจำนวนนิสิตนักศึกษามากเป็นผลทำให้ความถูกต้องของโมเดลที่ได้ค่อนข้างต่ำ

จากข้อเสียของโมเดลนั้น ผู้วิจัยจึงได้ใช้เทคนิคการค้นหากฎความสัมพันธ์มาช่วยในการทำนายแนวโน้มเกรด โดยหาความสัมพันธ์ของผลการเรียนในแต่ละสาขาวิชาที่ส่งผลต่อกัน ทำให้ทราบได้ว่าวิชาใดบ้างที่มีผลต่อวิชาที่ต้องการทำนายเกรดล่วงหน้า โดยได้กำหนดค่าสนับสนุนต่ำสุด (Minimum support) และค่าความมั่นใจต่ำสุด (Minimum confidence) เพื่อใช้ในการหาความสัมพันธ์ที่ได้จากนิสิตจำนวนมาก ทำให้สนับสนุนโมเดลมีความน่าเชื่อถือเป็นที่น่าพอใจ โดย

มีเปอร์เซ็นต์ความถูกต้องค่อนข้างสูงและมีปัญหาบางประการได้แก่ การจำแนกข้อมูลในบางสาขาวิชามีปริมาณน้อยทำให้โมเดลตัวต้นแบบที่ได้ไม่มีความแม่นยำเท่าที่ควร

2.5.5 งานวิจัยเรื่อง การวัดประสิทธิภาพของขั้นตอนวิธี ตัวจำแนก C4.5, ADTree และNaïve Bayes ในการจำแนกข้อมูลการซุกซ่อนสิ่งเสพย์ติดสำหรับไปรษณีย์ระหว่างประเทศ โดยสิทธิโชค มุกดาสกุลภิบาล นิสิตปริญญาโทสาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

งานวิจัยฉบับนี้ได้นำเสนอการวัดความแม่นยำและประสิทธิภาพด้านความเร็วของตัวจำแนก 3 ชนิด คือ C4.5, ADTree และ Naïve Bayes เพื่อนำไปใช้งานในการตรวจสอบสิ่งของซุกซ่อนสิ่งเสพย์ติดสำหรับไปรษณีย์ระหว่างประเทศ ของเจ้าหน้าที่ศุลกากร โดยได้ใช้ข้อมูลเก่าๆที่เจ้าหน้าที่ได้จัดเก็บ และยังใช้วิธีการเก็บข้อมูลใหม่ที่สุ่มเก็บเพิ่มเติมจากเจ้าหน้าที่ เพื่อให้ได้ปริมาณข้อมูลที่เพียงพอ โดยใช้วิธีการทดสอบกับตัวจำแนกทั้ง 3 ชนิด โดยทำการประเมินในแง่ของค่าความแม่นยำ ค่าส่วนเบี่ยงเบนมาตรฐาน และระยะเวลา และข้อมูลที่จะนำมาทำการวัดผลมีทั้งข้อมูลที่เป็นสองฉลาก และหลายฉลาก

ผลลัพธ์จากการศึกษาวิจัยพบว่าตัวจำแนกแต่ละตัวมีข้อดีและข้อเสียที่แตกต่างกัน ซึ่งในแง่ของความถูกต้อง โดยรวมพบว่า ADTree สามารถทำงานได้ดี แต่ในด้านความเร็วในการทำงานพบว่า ADTree ใช้เวลาในการประมวลผลนานสุดจากตัวจำแนกทั้งสาม ซึ่งพบว่า Naïve Bayes และ C4.5 ใช้เวลาในการประมวลผลที่ใกล้เคียงกัน และ C4.5 มีความสามารถที่ทำนายข้อมูลได้ดีที่สุด