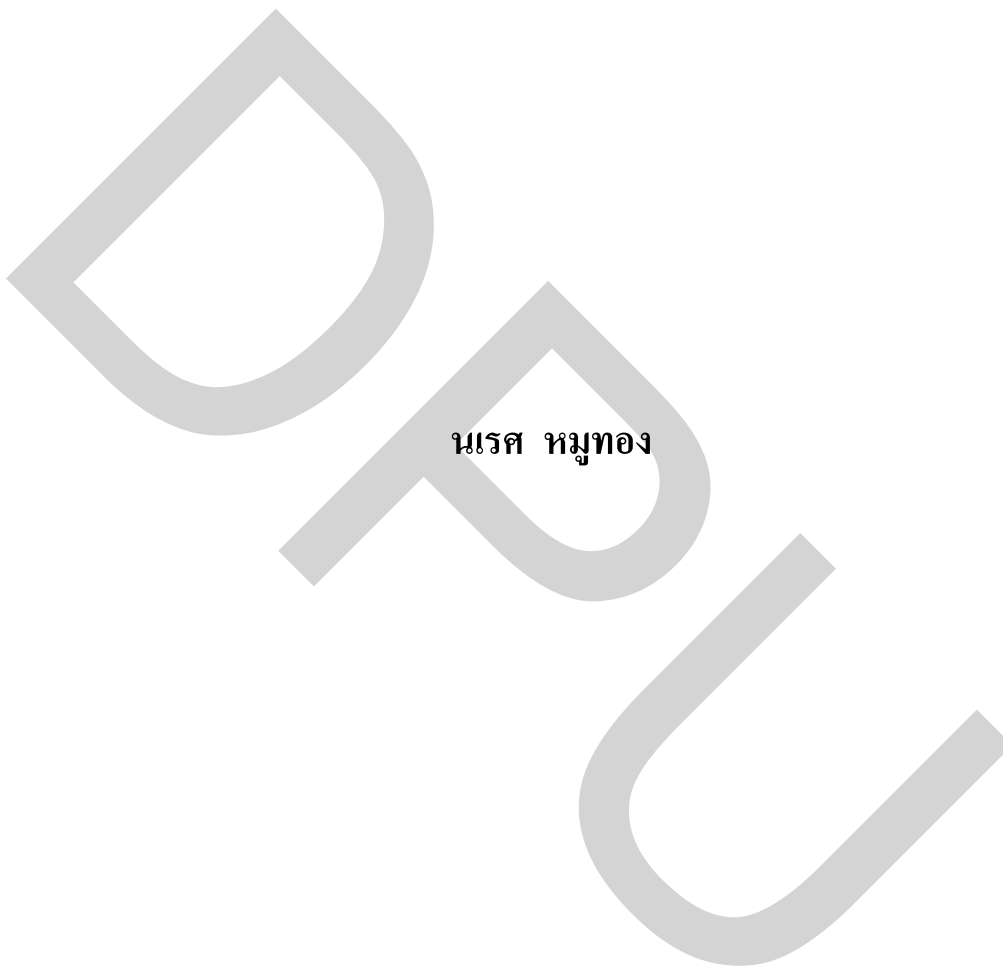


การใช้เหมืองข้อมูลช่วยในการตัดสินใจจ่ายจำนวน
คดีความให้กับผู้พิพากษา: กรณีศึกษาศาลจังหวัดปทุมธานี



งานค้นคว้าอิสระนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรวิทยาศาสตรมหาบัณฑิต
สาขาเทคโนโลยีคอมพิวเตอร์และการสื่อสาร คณะวิศวกรรมศาสตร์
มหาวิทยาลัยธุรกิจบัณฑิตย์
พ.ศ. 2558

**Using data mining to Assist in Decision Making on Distribution
Of case among judges: Case Study of Pathumthani Court**



Nares Moothong

An Independent Study Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science

Department of Computer and Communication Technology

Faculty of Engineering, Dhurakij Pundit University

2015

หัวข้องานค้นคว้าอิสระ

การใช้เหมืองข้อมูลช่วยในการตัดสินใจง่ายสำหรับ
ความให้กับผู้พิพากษา: กรณีศึกษาศาลจังหวัดปทุมธานี

ชื่อผู้เขียน

นเรศ หมูทอง

อาจารย์ที่ปรึกษา

ผู้ช่วยศาสตราจารย์ ดร.วรพล พงษ์เพชร

สาขาวิชา

เทคโนโลยีคอมพิวเตอร์และการสื่อสาร

ปีการศึกษา

2557

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อศึกษาและค้นคว้ารูปแบบการเกิดความสัมพันธ์ของแอตทริบิวต์หรือตัวแปรจากข้อมูลสำนวนคดีที่มีอยู่ และนำความรู้ที่ได้จากการศึกษาไปสร้างโมเดลที่สามารถทำนายผลลัพธ์ที่คาดว่าจะเกิดขึ้นในอนาคตเพื่อช่วยให้ผู้บริหารสามารถตัดสินใจพิจารณาการจ่ายสำนวนคดีให้กับผู้พิพากษาของศาลจังหวัดปทุมธานีให้มีประสิทธิภาพมากขึ้น

ในการศึกษานี้ใช้วิธีการจำลองข้อมูลขึ้นมาด้วยชุดคุณลักษณะของข้อมูลที่จัดเก็บไว้ภายในฐานข้อมูลจริง โดยใช้โปรแกรม Weka (Waikato Environment for Knowledge Analysis) ในการศึกษาเป็นหลักโดยใช้เทคนิคของเหมืองข้อมูลประเภทเทคนิคต้นไม้ตัดสินใจ (Decisions tree) หรือ โมเดล Classifier tree J48 และวิธีการเรียนรู้แบบอย่างง่าย (Naïve Bayes) หรือ โมเดล Classifier Naïve Bayes มาใช้ในการศึกษา

โดยผลจากการศึกษาค้นคว้าพบว่าเทคนิคต้นไม้ตัดสินใจให้ความถูกต้องจากการทำนายสูงคิดเป็นร้อยละ 93.52 สามารถจำแนกกลุ่มประเภทคดีง่ายและประเภทคดียาก ที่สามารถแบ่งกลุ่มได้ตามเงื่อนไขที่ชัดเจนที่สุด และนำกฎที่ได้มาเป็นตัวอย่างที่สำคัญสำหรับการพัฒนาโปรแกรมตรวจสอบข้อมูลทางคดีที่ใช้ประกอบในการตัดสินใจง่ายสำหรับผู้พิพากษาได้

Independent Study Title	Using Data Mining to Assist in Decision Making on Distribution of Cases among Judges: Case Study of Pathumthani Court
Author	Nares Moothong
Independent Study Advisor	Asst. Prof. Worapol Pongpech, Ph.D.
Department	Computer and Communication Technology
Academic Year	2014

ABSTRACT

This research aims to study and investigate the pattern of relations of attributes or variables in an existing case data, and to use the knowledge gained from the study to develop a model that can predict the consequences that is expected to occur in the future. This can help the administrators to decide how to distribute cases among judges at Pathumthani court more efficiently.

In this study, data were duplicated according to features of the data stored in the actual database mainly using Weka (Waikato Environment for Knowledge Analysis) program. Moreover, data mining techniques called the decisions tree technique or Classifier tree J48 model, and Classifier Naïve Bayes model were applied.

The results of the study showed that the decision tree technique could provide the predicted outcomes with high accuracy at 93.52 percent. It could classify cases into easy and difficult cases according to the most apparent criteria. Finally, the rule from the study could be used as an important example for developing a program to verify case data that is to be included in the decision on the distribution of cases among the judges.

กิตติกรรมประกาศ

งานค้นคว้าอิสระนี้สำเร็จได้ด้วยความกรุณาของผู้ช่วยศาสตราจารย์.ดร.วรพล พงษ์เพชร ผู้อำนวยการหลักสูตรมหาบัณฑิต สาขาเทคโนโลยีคอมพิวเตอร์และการสื่อสาร ที่ได้กรุณาให้คำปรึกษาแนะนำ ช่วยเหลือ ให้กำลังใจ และแก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่ เป็นอย่างยิ่งมาโดยตลอด ทำให้งานค้นคว้าอิสระฉบับนี้มีความสมบูรณ์มากยิ่งขึ้น ผู้ศึกษารู้สึกซาบซึ้งในความกรุณาของท่านเป็นอย่างยิ่ง ผู้ศึกษาขอกราบขอบพระคุณอาจารย์เป็นอย่างสูง

ขอขอบพระคุณ อาจารย์ ดร.ประศาสน์ จันทราทิพย์ และอาจารย์ ดร.สันห์ รัฐวิบูลย์ ที่กรุณาตรวจแก้ไข และให้คำแนะนำในการปรับปรุงผลลัพธ์ที่ได้จากการศึกษาครั้งนี้ ให้มีประสิทธิภาพมากยิ่งขึ้น และขอบคุณเจ้าหน้าที่ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยธุรกิจบัณฑิต ในการอำนวยความสะดวก ให้ความร่วมมือในการติดต่อประสานงาน สนับสนุนข้อมูลต่างๆ ทำให้การศึกษาครั้งนี้สำเร็จลุล่วงลงด้วยดี

สุดท้ายนี้ผู้ศึกษาขอกราบขอบพระคุณบิดา มารดา รวมทั้งเพื่อน พี่ น้อง สาขาเทคโนโลยีคอมพิวเตอร์ทุกท่านที่ได้ให้การสนับสนุนในทุกๆ ด้านและเป็นกำลังใจในการทำงานค้นคว้าอิสระฉบับนี้มาโดยตลอด

นเรศ หมูทอง

สารบัญ

	หน้า
บทคัดย่อภาษาไทย.....	ฉ
บทคัดย่อภาษาอังกฤษ.....	ง
กิตติกรรมประกาศ.....	จ
สารบัญตาราง.....	ช
สารบัญภาพ.....	ฉ
บทที่	
1. บทนำ.....	1
1.1 หลักการและเหตุผล.....	1
1.2 วัตถุประสงค์.....	3
1.3 ขอบเขตการศึกษา.....	3
1.4 ผลที่คาดว่าจะได้รับ.....	3
1.5 นิยามศัพท์เฉพาะ.....	3
2. แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง.....	5
2.1 ศาสนุติธรรมและศาลจังหวัดปทุมธานี.....	5
2.2 ทฤษฎีและแนวความคิดที่เป็นพื้นฐานในการพัฒนาตัวโมเดลต้นแบบ.....	8
2.3 ซอฟต์แวร์ที่ใช้ในการศึกษา.....	19
2.4 การวิเคราะห์และออกแบบระบบ.....	28
2.5 งานวิจัยที่เกี่ยวข้อง.....	30
3. ระเบียบวิธีวิจัย.....	33
3.1 ขั้นตอนการศึกษาค้นคว้า.....	33
3.2 เครื่องมือที่ใช้ในการศึกษา.....	34
3.3 ขั้นตอนการวิเคราะห์และออกแบบระบบ.....	35

สารบัญ (ต่อ)

บทที่	หน้า
3.4 ระยะเวลาในการดำเนินการศึกษาค้นคว้า.....	39
4. ผลการดำเนินการศึกษาค้นคว้า.....	41
4.1 การศึกษาและรวบรวมข้อมูล.....	41
4.2 การเตรียมข้อมูล (Data Preparation).....	42
4.3 การศึกษา ทดสอบ และเลือกเทคนิคที่เหมาะสม.....	46
4.4 การวิเคราะห์และสรุปผล.....	51
4.5 ขั้นตอนการพัฒนาระบบ.....	57
5. สรุปผลการศึกษาค้นคว้าและข้อเสนอแนะ.....	70
5.1 สรุปผลการศึกษาค้นคว้า.....	70
5.2 ปัญหาและแนวทางแก้ไข.....	71
5.3 ข้อเสนอแนะสำหรับพัฒนาต่อไป.....	72
บรรณานุกรม.....	73
ภาคผนวก.....	77
ประวัติผู้เขียน.....	85

สารบัญตาราง

ตารางที่	หน้า
2.1 แสดงข้อมูลสภาพอากาศย้อนหลัง 14 วัน.....	15
2.2 แสดงข้อมูลการคำนวณจากโมเดลของ Naïve Bayes	18
2.3 แสดงข้อมูลสภาพอากาศในวันปัจจุบัน ซึ่งเรายังไม่สามารถทราบคลาส คำตอบ.....	18
3.1 ตารางแสดงการจัดเก็บข้อมูลของฐานข้อมูลชื่อ data_case.....	38
3.2 ตารางแสดงข้อมูลเกี่ยวกับการจัดการข้อมูลบุคคลที่เกี่ยวข้องภายในระบบ.	38
3.3 แสดงระยะเวลาในการดำเนินการศึกษาค้นคว้า.....	39
4.1 แสดงคุณลักษณะ (Attribute) ที่ใช้ดำเนินการศึกษา.....	44
4.2 แสดงการแปลงค่าข้อมูล.....	46

สารบัญภาพ

ภาพที่	หน้า
2.1 แสดงกระบวนการค้นหาความรู้ในการขุดข้อมูล.....	11
2.2 โมเดลโครงสร้างต้นไม้ตัดสินใจ.....	16
2.3 แสดงหน้าต่าง Weka GUI Chooser.....	19
2.4 เมนูหลักในหน้าต่าง Weka Explorer.....	20
2.5 แสดงขั้นตอนการทำงาน PHP Script Request/Response	22
2.6 แสดงขั้นตอนการเข้าสู่โปรแกรม Adobe Dreamweaver CS5	26
2.7 ส่วนประกอบของ Welcom Screen.....	27
2.8 ส่วนประกอบของหน้าจอโปรแกรม.....	27
2.9 แสดงวงจรการพัฒนาระบบ System Development Life Cycle: SDLC.....	29
3.1 แสดงแผนภาพยูสเคสของโปรแกรมตรวจสอบข้อมูลทางคดี.....	35
3.2 แสดงแผนภาพการทำงานในส่วนของการเข้าสู่ระบบ.....	36
3.3 แสดงแผนภาพการทำงานในส่วนของการทำนายผลลัพธ์ที่เกิดขึ้น.....	37
3.4 แผนผังแสดงขั้นตอนการดำเนินการศึกษาวิเคราะห์ข้อมูลในโปรแกรม Weka	40
4.1 แสดงรายละเอียดของฐานข้อมูล.....	42
4.2 แสดงการโอนย้ายข้อมูลจากระบบจัดการฐานข้อมูลมาอยู่ในรูปแบบไฟล์ Microsoft Excel.....	43
4.3 แสดงการเปิดโปรแกรม Weka เวอร์ชัน 3.6.....	47
4.4 แสดงการเข้าหน้าจอโปรแกรม Weka เวอร์ชัน 3.6.....	47
4.5 แสดงการนำเข้าข้อมูลในโปรแกรม Weka เวอร์ชัน 3.6.....	48
4.6 หน้าจอแสดงข้อมูลที่นำเข้าในโปรแกรม Weka เวอร์ชัน 3.6.....	49

สารบัญภาพ (ต่อ)

ภาพที่	หน้า
4.7 การเลือกเทคนิคที่ใช้ในการศึกษาในการจำแนกข้อมูลแบบ Decsion Tree: J48	49
4.8 หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decsion Tree: J48.....	50
4.9 การเลือกเทคนิคที่ใช้ในการศึกษาในการจำแนกข้อมูลแบบ Naïve Bayes	50
4.10 หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Naïve Bayes.....	51
4.11 ผลลัพธ์จากการจำแนกกลุ่มแบบ Decsion Tree: J48.....	52
4.12 แสดงแผนภาพการจำแนกกลุ่มแบบ Decsion Tree: J48.....	52
4.13 ผลลัพธ์จากการจำแนกกลุ่มแบบ Naïve Bayes.....	53
4.14 แสดงขั้นตอนการใช้งาน phpMyAdmin Database Manager.....	58
4.15 แสดงการป้อน Username และ Password เพื่อเข้าสู่ระบบ.....	58
4.16 แสดงหน้าต่างของ phpMyAdmin.....	59
4.17 แสดงขั้นตอนการสร้างฐานข้อมูล.....	60
4.18 แสดงผลการสร้างฐานข้อมูล.....	60
4.19 แสดงโครงสร้างหน้าเว็บเพจของโปรแกรมตรวจสอบข้อมูลทางคดี.....	61
4.20 หน้าจอแสดงการล็อกอินเข้าสู่ระบบ.....	61
4.21 หน้าจอบันทึกข้อมูลทางคดีที่ต้องการวิเคราะห์.....	62
4.22 หน้าจอแสดงผลการวิเคราะห์.....	63
4.23 หน้าจอการจัดการเพิ่มข้อมูลบุคคล.....	64
4.24 แสดงข้อมูลบุคคลผู้มีสิทธิใช้งานในระบบ.....	64
4.25 แสดงเงื่อนไขในการเขียนโปรแกรมในส่วนของการทำนายผลลัพธ์.....	66

บทที่ 1

บทนำ

1.1 หลักการและเหตุผล

นับตั้งแต่กระทรวงยุติธรรมทำหน้าที่ดูแล รับผิดชอบงานธุรการของศาลยุติธรรมเป็นระยะเวลา 100 ปีเศษ จึงเกิดแนวคิดที่จะแยกศาลยุติธรรมออกจากกระทรวงยุติธรรมโดยมีวัตถุประสงค์เพื่อให้พ้นจากข้อระแวง ข้อสงสัย การแทรกแซงจากฝ่ายบริหาร จนกระทั่งรัฐธรรมนูญแห่งราชอาณาจักรไทย พ.ศ. 2540 มาตรา 275 วรรคหนึ่ง บัญญัติให้ศาลยุติธรรมมีหน่วยงานธุรการของศาลยุติธรรมที่เป็นอิสระ ให้เลขาธิการสำนักงานศาลยุติธรรมเป็นผู้บังคับบัญชาขึ้นตรงต่อประธานศาลฎีกา (ราชกิจจานุเบกษา เล่ม 144 ตอนที่ 55 ก 11 ตุลาคม 2540 หน้า 71) และ พระราชบัญญัติระเบียบบริหารราชการศาลยุติธรรม พ.ศ.2543 มาตรา 5 วรรคหนึ่ง บัญญัติให้มีสำนักงานศาลยุติธรรมเป็นส่วนราชการที่เป็นหน่วยงานอิสระมีฐานะเป็นนิติบุคคล (ราชกิจจานุเบกษา เล่ม 117 ตอนที่ 59 ก 21 มิถุนายน 2543,น.2) ดังนั้นจึงถือได้ว่าสำนักงานศาลยุติธรรมมีการแยกออกจากกระทรวงยุติธรรม นับตั้งแต่วันที่ 20 สิงหาคม 2543 เป็นต้นมา

สำนักงานศาลยุติธรรมมีโครงสร้างลักษณะการทำงาน แบ่งได้ 2 ส่วนใหญ่ๆ คือ ส่วนงานตุลาการมีอำนาจหน้าที่ในการพิจารณาพิพากษาอรรถคดีต่างๆ อันได้แก่ คดีแพ่ง คดีอาญา คดีล้มละลาย และคดีประเภทอื่นที่อยู่ในเขตอำนาจของศาลยุติธรรม ส่วนงานธุรการ มีสำนักงานศาลยุติธรรมทำหน้าที่ดูแลเกี่ยวกับงานธุรการของศาลยุติธรรม งานส่งเสริมงานตุลาการและงานวิชาการ เพื่อสนับสนุนและอำนวยความสะดวก ให้แก่ศาลยุติธรรม รวมทั้งเสริมสร้างให้การพิจารณาพิพากษาคดีเป็นไปโดยสะดวก รวดเร็ว และมีประสิทธิภาพ

ศาลยุติธรรมตามพระธรรมนูญศาลยุติธรรม พ.ศ.2543 มีอยู่ที่วราชอาณาจักร และระบบศาลยุติธรรมสามารถแบ่งออกได้เป็น 3 ชั้น คือ ศาลชั้นต้น ศาลอุทธรณ์ และศาลฎีกา มีจำนวนศาลรวมทั้งสิ้น 248 ศาลทั่วราชอาณาจักร (สำนักแผนและงบประมาณ ข้อมูล ณ วันที่ 1 เมษายน 2557)

ศาลจังหวัดปทุมธานี จัดอยู่ในศาลชั้นต้น ที่ทำหน้าที่รับคำฟ้องหรือคำร้อง ในชั้นเริ่มต้นไม่ว่าจะเป็นคดีแพ่ง คดีผู้บริ โภค คดีแรงงาน คดีภาษีอากร คดีล้มละลาย คดีทรัพย์สินทางปัญญาและการค้าระหว่างประเทศ หรือ คดีอาญา โดยดำเนินการกระบวนกรตัดสินชี้ขาดเป็นชั้นศาลแรก ทั้งมีอำนาจดำเนินการพิจารณาแทนศาลอุทธรณ์และศาลฎีกาในบางเรื่อง

ศาลจังหวัดปทุมธานีเป็นหนึ่งในสิบสี่ศาล ที่ได้รับคัดเลือกจากสำนักเทคโนโลยีสารสนเทศ สำนักงานศาลยุติธรรม ให้เป็นศาลต้นแบบในการนำระบบเทคโนโลยีสารสนเทศมาช่วยอำนวยความสะดวกด้านการบริหารจัดการกับข้อมูลทางคดี โดยเรียกชื่อโปรแกรมว่า “ระบบสารสนเทศสำนวนคดีศาลชั้นต้นและศาลอุทธรณ์ ในกลุ่มศาลต้นแบบ” ตั้งแต่เดือนตุลาคม พ.ศ.2552 เป็นต้นมา

ในการปฏิบัติงานประจำวันของเจ้าหน้าที่ ที่ทำหน้าที่รับคำฟ้องหรือคำร้อง ได้บันทึกข้อมูลทางคดีผ่าน โปรแกรมระบบสารสนเทศสำนวนคดีศาลชั้นต้นและศาลอุทธรณ์ ในกลุ่มศาลต้นแบบ โดยมีซอฟต์แวร์ที่ใช้ในการบริหารจัดการฐานข้อมูล (DBMS) ทางคดีชื่อ Microsoft SQL Server 2005 และพบว่าข้อมูลทางคดีที่บันทึกมีจำนวน 10-30 เรื่องต่อหนึ่งวันและมีปริมาณคดีเพิ่มขึ้นอยู่ทุกๆวัน

ข้อมูลทางคดีที่จัดเก็บในระบบจัดการฐานข้อมูล (DBMS) ณ ปัจจุบันทำหน้าที่ในการจัดเก็บข้อมูลทางคดีโดยไม่มีการนำข้อมูลทางคดีที่จัดเก็บมาใช้ให้เกิดประโยชน์แต่อย่างใด การสืบค้นความรู้ที่เป็นประโยชน์และน่าสนใจจากฐานข้อมูลขนาดใหญ่ (Knowledge Discovery from very large Database: KDD) หรือที่เรียกกันว่า Data Mining เป็นสาขาหนึ่งทางวิทยาศาสตร์คอมพิวเตอร์ ที่กำลังได้รับความสนใจอย่างสูงในปัจจุบัน ด้วยเทคนิคของ KDD ข้อมูลขนาดใหญ่ที่เราสนใจจะถูกวิเคราะห์และดึงเอาความรู้หรือสิ่งสำคัญออกมารวบรวมให้อยู่ในรูปของฐานความรู้ (Knowledge Base) เพื่อใช้สำหรับการสืบค้นสิ่งที่เราต้องการ ซึ่งไม่สามารถสืบค้นได้จากวิธีการของระบบจัดการฐานข้อมูล (DBMS) โดยทั่วไป (ธีระวัฒน์ พงษ์ศิริปริดา, การจัดสรรกฎหมายที่เหมาะสมให้กับคดีความ โดยอด โนมัติ, 2544, น.3)

ด้วยเหตุนี้จึงมีแนวคิดที่จะศึกษาข้อมูลที่จัดเก็บภายในฐานข้อมูลของศาลจังหวัดปทุมธานี โดยมุ่งเน้นการศึกษา ค้นคว้า ที่จะหารูปแบบ หาความสัมพันธ์ หาโครงสร้างที่เด่นชัดเพื่อทำนายผลลัพธ์ที่เกิดขึ้นจากระบบฐานข้อมูลที่จัดเก็บโดยใช้วิธีการจำลองข้อมูลเปรียบเทียบกับโครงสร้างของฐานข้อมูลที่ได้จัดเก็บจริง โดยใช้เทคนิคของเหมืองข้อมูลประเภทเทคนิคต้นไม้ตัดสินใจ (Decisions tree) และวิธีการเรียนรู้เบย์อย่างง่าย (Naïve Bayes) มาใช้ในการศึกษาและช่วยในการอธิบายรูปแบบ ลักษณะบางอย่างภายใต้ข้อมูลที่มีอยู่และนำผลลัพธ์ที่ได้จากการศึกษานำไปทำนายหรือนำไปใช้ประโยชน์ต่อการสนับสนุนการตัดสินใจ หรือสนับสนุนทำงานของผู้บริหารในการพิจารณาจ่ายสำนวนคดีความให้กับผู้พิพากษาประจำศาล ได้โดยสะดวกและรวดเร็ว

1.2 วัตถุประสงค์

1. เพื่อศึกษา ค้นคว้า รูปแบบการเกิดความสัมพันธ์ของแอททริบิวต์หรือตัวแปรจากข้อมูลทางคดีที่มีอยู่
2. เพื่อสามารถนำความรู้ที่ได้จากการศึกษาค้นคว้าไปสร้างโมเดล (Model) การทำนายผลลัพธ์ที่คาดว่าจะเกิดขึ้นภายในอนาคต
3. เพื่อช่วยให้ผู้บริหารสามารถตัดสินใจพิจารณาในการจ่ายสำนวนคดีให้กับผู้พิพากษาให้มีประสิทธิภาพมากขึ้น

1.3 ขอบเขตของการศึกษา

1. ผู้ดำเนินการศึกษาใช้วิธีการจำลองของมูลแอททริบิวต์จากฐานข้อมูลจริงที่จัดเก็บในปัจจุบัน
2. ผู้ดำเนินการศึกษาใช้วิธีการศึกษาค้นคว้าโดยใช้เทคนิคของเหมืองข้อมูลประเภทเทคนิคต้นไม้ตัดสินใจ (Decisions tree) และวิธีการเรียนรู้แบบอย่างง่าย (Naïve Bayes)

1.4 ผลที่คาดว่าจะได้รับ

1. ผู้ศึกษาคาดว่าจะได้รูปแบบการเกิดความสัมพันธ์ของแอททริบิวต์ เพื่อนำไปใช้ประโยชน์ในการนำเสนอบทสรุปต่อผู้บริหารในลำดับต่อไป
2. สามารถทราบปัจจัยของมูลเหตุคดีในการจัดกลุ่มประเภทคดีและได้รูปแบบ โมเดลที่ชัดเจนสามารถนำไปใช้งานได้จริง
3. สามารถนำผลลัพธ์การศึกษาด้วยเทคนิคของเหมืองข้อมูลและคัดเลือกวิธีที่ดีที่สุดนำไปเขียนโปรแกรมเพื่อประยุกต์ใช้งานในการนำเสนอให้กับผู้บริหารได้

1.5 นิยามศัพท์เฉพาะ

เหมืองข้อมูล (Data Mining) เป็นกระบวนการในการค้นหาความสัมพันธ์ ค้นหา รูปแบบ ค้นหาแนวโน้ม จากความรู้ที่เก็บรวบรวมอยู่ไว้เป็นจำนวนมาก โดยอาศัยเทคนิคต่างๆ เพื่อนำความรู้ที่ได้ มาสนับสนุนกระบวนการตัดสินใจ จากข้อมูลที่ได้เพื่อนำไปทำนายผลลัพธ์ที่คาดว่าจะเกิดขึ้นภายในอนาคตอันใกล้

แอททริบิวต์ (Attribute) หมายถึง คุณลักษณะของสิ่งที่ต้องการอธิบายความสัมพันธ์ที่ต้องการจัดเก็บ

โครงสร้างต้นไม้ (Decisions tree) หมายถึง โครงสร้างที่มีลักษณะการเรียงเป็นกิ่งก้านสาขาแตกแขนงออกไป ที่มีความสัมพันธ์กันระหว่างโหนด โดยแต่ละโหนดจะลดความสำคัญลงไป

วิธีการเรียนรู้แบบอย่างง่าย (Naïve Bayes) หมายถึงการจำแนกประเภทโดยใช้หลักสถิติในการพยากรณ์ความน่าจะเป็นของสมาชิก

คดีจัดการพิเศษ (Special case management) หมายถึง คดีที่ไม่มีความยุ่งยากซับซ้อน และมีแนวโน้มที่จะพิจารณาให้แล้วเสร็จได้ภายในนัดเดียวหรือในวันหนึ่งสามารถพิจารณาคดีให้แล้วเสร็จได้หลายคดี หรือ สามารถส่งเอกสารแทนการสืบพยานได้

คดีสามัญ (General special case) หมายถึงคดีแพ่งหรือคดีอาญาที่ไม่ใช่คดีจัดการพิเศษ หรือต้องสืบพยานหลักฐานของคู่ความต่อไปและไม่สามารถนั่งพิจารณาคดีเสร็จได้ภายในเวลาอันรวดเร็วรวมถึงคดีที่ศาลสูงย้ายสำนวนให้ศาลชั้นต้นสืบพยานและพิพากษาคดีใหม่ คดีรื้อฟื้นคดีอาญา คดีจำเลยถอนคำให้การรับสารภาพเป็นการปฏิเสธ คดีแพ่งที่จำเลยขาดนัดคำให้การและศาลอนุญาตให้ยื่นคำให้การได้ หรือคดีประเภทอื่นที่ผู้บริหารศาลเห็นสมควรให้ดำเนินการคดีอย่างคดีสามัญ

คดีสามัญพิเศษ (Special case) หมายถึง คดีสามัญที่มีความยุ่งยากสลับซับซ้อน มีพยานหลักฐานที่จะต้องนำสืบจำนวนมาก ไม่สามารถนัดสืบต่อเนื่องกันไปจนเสร็จ จำเป็นต้องกำหนดวันนัดสืบพยานต่อเนื่อง ช่วงละ 2-4 วันเป็นช่วงๆ โดยกำหนดให้แล้วเสร็จภายใน 6 เดือน นับแต่วันกำหนดสืบพยาน แต่ไม่ว่ากรณีใดๆ ต้องไม่เกิน 1 ปี นับแต่วันกำหนดสืบพยาน โดยต้องได้รับอนุญาตจากผู้บริหารศาลชั้นต้น

คดีง่าย (Easy) หมายถึง ประเภทคดีที่จัดอยู่ในประเภทคดีที่ง่าย ไม่มีความซับซ้อนในการพิจารณาพิพากษาอรรถคดี

คดียาก (Difficulty) หมายถึง ประเภทคดีที่จัดอยู่ในประเภทคดีที่ยาก มีความซับซ้อนในการพิจารณาพิพากษาอรรถคดี

บทที่ 2

แนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้อง

ในการศึกษาค้นคว้าหาเพื่อค้นหารูปแบบการเกิดความสัมพันธ์หรือตัวแปรที่มีอยู่ด้วยเทคนิคของเหมืองข้อมูล ได้ดำเนินการศึกษา ทบทวนทฤษฎี แนวความคิด และผลงานวิจัยที่เกี่ยวข้องกับการพัฒนาโมเดล โดยดำเนินการรวบรวมข้อมูลที่จะเป็นประโยชน์ต่อการกำหนดแนวทางและระเบียบวิธีการวิจัย ซึ่งได้แบ่งการศึกษาออกเป็น 5 กลุ่มด้วยกันคือ

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

2.1 ศาลยุดิธรรมและศาลจังหวัดปทุมธานี

ศาลยุดิธรรมตามพระธรรมนูญศาลยุดิธรรม พ.ศ. 2543 มีอยู่ทั่วประเทศจัดระบบศาลยุดิธรรมสามารถแบ่งออกได้ เป็น 3 ชั้น คือ ศาลชั้นต้น ชั้นอุทธรณ์ ชั้นฎีกา รวมทั้งสิ้น จำนวน 248 ศาล ท้วราชอาณาจักร

ศาลชั้นต้น

ศาลชั้นต้น เป็นศาลซึ่งรับคำฟ้อง หรือคำร้องในชั้นเริ่มต้นคดีไม่ว่าเป็นคดีแพ่ง คดีผู้บริโภค คดีแรงงาน คดีภาษีอากร คดีล้มละลาย คดีทรัพย์สินทางปัญญาและการค้าระหว่างประเทศ หรือคดีอาญา โดยดำเนินการกระบวนการตัดสินชี้ขาดเป็นชั้นศาลแรก ทั้งมีอำนาจดำเนินการพิจารณาแทนศาลอุทธรณ์และศาลฎีกาในบางเรื่องด้วย

ศาลชั้นต้นทั่วประเทศมีจำนวนทั้งหมด 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. ศาลชำนาญพิเศษ 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

เป็นศาลที่มีอำนาจพิจารณาพิพากษาคดีที่อุทธรณ์คำพิพากษาหรือคำสั่งของศาลชั้นต้น ตามบทบัญญัติแห่งกฎหมายว่าด้วยการอุทธรณ์ รวมทั้งมีหน้าที่ในการพิจารณาคำสั่งอื่นๆ เช่นมี คำสั่งเกี่ยวกับการประกันตัวในคดีอาญาและ การขอทุเลาการบังคับคดีในคดีแพ่ง การพิจารณาคดี ของศาลอุทธรณ์มีลักษณะเป็นการตรวจสอบหรือทบทวน ในชั้นอุทธรณ์ประกอบด้วยศาลอุทธรณ์ กลางจำนวน 1 ศาล และศาลอุทธรณ์ภาค 1-9 จำนวน 9 ศาล รวม 10 ศาล

ชั้นฎีกา

เป็นศาลสูงสุดมีอำนาจพิจารณาพิพากษาคดีที่อุทธรณ์คำพิพากษาหรือคำสั่งของศาล อุทธรณ์ตามบทบัญญัติแห่งกฎหมายว่าด้วยการฎีกา และมีอำนาจวินิจฉัยชี้ขาดคดีที่ศาลฎีกามีอำนาจ วินิจฉัยได้ตามกฎหมายเฉพาะ แต่หากคดีใดมีปัญหาสำคัญไม่ว่าจะเป็นปัญหาข้อเท็จหรือปัญหาข้อ กฎหมาย ประธานศาลฎีกาเห็นว่าควรให้ วินิจฉัยโดยที่ประชุมใหญ่ของศาลฎีกาประธานศาลฎีกามี อำนาจสั่งให้นำปัญหาดังกล่าวเข้าสู่การวินิจฉัยโดยที่ประชุมใหญ่ของศาลฎีกา ศาลฎีกามีเพียงศาล เดียวตั้งอยู่ที่กรุงเทพมหานคร

ประวัติความเป็นมาของศาลจังหวัดปทุมธานี

ในปี ร.ศ.117 (พ.ศ.2442) พระเจ้าบรมวงศ์เธอ กรมพระนเรศวรฤทธิ ซึ่งขณะนั้นดำรงหลวงอารักษ์ประธาราชภูริ ข้าหลวงรักษากรุงเทพมหานครจัดซื้อที่ดินบริเวณปากคลองรังสิต ประยูรศักดิ์แขวงเมืองปทุมธานีซึ่งเห็นว่าเป็นที่ราษฎรไปมามากกว่าบริเวณอื่น เพื่อใช้เป็นที่สร้างศาลากลางเมืองปทุมธานีกับศาลสำหรับพิจารณาคดี แต่การจัดซื้อที่ดินครั้งนั้นมีปัญหา เพราะราษฎรชาวมอญเจ้าของที่ 6 ราย ไม่ยอมขายที่ให้ โดยอ้างว่าได้อยู่อาศัยมาแต่สมัยปู่ ย่า ตา ยาย และทั้งได้ปลูกต้นไม้ผลไม้มั่งในเนื้อที่นั้นแล้ว เนื่องจากชาวมอญ 5 คน ในจำนวนนั้นเป็นหมู่ทหารเรือพระเจ้าบรมวงศ์เธอ กรมพระนเรศวรฤทธิ จึงมีหนังสือทูลชี้แจงไปยังพระวรวงศ์เธอ กรมหมื่นปราบปรบักษ์ ผู้บัญชาการกรมทหารเรือให้ช่วยชี้แจงแก่ชาวมอญหมู่ทหารเรือซึ่งอยู่ในบังคับบัญชาตามหนังสือ ลงวันที่ 12 กรกฎาคม รัตนโกสินทรศก 117 ตอนหนึ่งว่า

“ หม่อมฉันขอลูกให้ทรงทราบที่ตำบลวัดเทียนถวาย ปากคลองรังสิตริมแม่น้ำนี้เป็นที่สมควรจะสร้างที่ว่าการเมืองปทุมธานี และศาลสำหรับเมือง เพราะจะได้ทำทางริมคลองรังสิตเข้าหารถไฟนครราชสีมา อันจะเดินทางได้ทั้งทางบกแลน้ำ จึงจำต้องตั้งที่ว่าการเมืองแลศาลตำบลนี้ เมื่อราษฎรเจ้าของที่ไม่ต้องการเงินจะต้องการที่โดยจะหาซื้อที่โดยลำพังเป็นการลำบาก ที่หลวงที่เมืองปทุมธานีมีอยู่ริมแม่น้ำเหมือนที่ราษฎรตั้งอยู่ตามที่ต้องการนี้ หม่อมฉันจะจัดเปลี่ยนให้ได้ แลทั้งเรือขลบ้านเรือนไปปลูกใหม่ให้ด้วย เพื่อการสำเร็จจะได้แล้วไปไม่ขัดข้องแก่ความเจริญของราชการ ขอท่านได้ทรงพระเมตตาโปรดให้เจ้าพนักงานชี้แจงแก่เจ้าของที่ให้เป็นการตกลงกัน ”

การก่อสร้างที่ว่าการเมืองปทุมธานี ศาลเมืองปทุมธานีครั้งนั้นใช้เวลาประมาณ 1 ปี ได้มีพิธีเปิดเมื่อวันที่ 10 มกราคม รัตนโกสินทรศก 118 (พ.ศ.2443) ต่อมาพระบาทสมเด็จพระมงกุฎเกล้าเจ้าอยู่หัว ทรงพระกรุณาโปรดเกล้าฯ ให้เปลี่ยนคำเรียก “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 อยู่ห่างไกลกันมากประมาณ 100 เส้นเศษ หนทางที่จะไปมาถึงกัน ถ้าใช้เรือแจวพายแล้วต้องใช้เวลาราวชั่วโมงเศษ จึงขอย้ายศาลจังหวัดปทุมธานีไปปลูกสร้าง ณ ที่ทำ

การจังหวัดใหม่ โดยให้สร้างต่อจากศาลากลางจังหวัดปทุมธานีลงมาทางใต้ กระทรวงยุติธรรมดำริแล้วเห็นชอบด้วย

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

เมื่อ พ.ศ.2516 นางกสิณ กุชยกานนท์ และครอบครัว ได้ทูลเกล้าฯถวายที่ดิน โฉนดที่ 5693 และ โฉนดที่ 2651 บางส่วนรวมเนื้อที่ 16 ไร่ อยู่ริมถนนสายกรุงเทพฯ-ปทุมธานี ให้จังหวัดปทุมธานีใช้เป็นที่ตั้งศาลากลางจังหวัดปทุมธานี ทางจังหวัดจึงขอให้ศาลจังหวัดปทุมธานีมาปลูกสร้างในบริเวณเดียวกันเพื่อจะได้เป็นศูนย์ราชการต่อไป กระทรวงยุติธรรมพิจารณาแล้วเห็นชอบด้วย กำหนดแล้วเสร็จภายในวันที่ 22 ธันวาคม พ.ศ. 2520 ซึ่งเป็นที่ตั้งปัจจุบันของศาลจังหวัดปทุมธานี

2.2 ทฤษฎีและแนวความคิดที่เป็นพื้นฐานในการพัฒนาตัวโมเดลต้นแบบ

2.2.1 นิยาม ความหมายของเหมืองข้อมูล

เหมืองข้อมูล คือ กระบวนการค้นหาสารสนเทศหรือข้อความรู้ที่อยู่ในฐานข้อมูลขนาดใหญ่ที่ซับซ้อน เพื่อนำข้อความรู้ที่ได้ไปใช้ประโยชน์ในการตัดสินใจ สารสนเทศที่ได้อาจนำมาสร้างการพยากรณ์หรือสร้างตัวแบบสำหรับการจำแนกหน่วยหรือกลุ่ม หรือแสดงความสัมพันธ์ระหว่างหน่วยงานต่างๆ หรือให้ข้อสรุปของสาระในฐานข้อมูล (สุชาติ กิระนันท์, 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) คือการสำรวจและวิเคราะห์ข้อมูลจำนวนมากให้อยู่ในรูปแบบที่มีความหมายและอยู่ในรูปของกฎ โดยความสัมพันธ์เหล่านี้ แสดงให้เห็นถึงความรู้อย่างไร (ชนาคม จุ้ยศิริ และ เนื่องวงศ์ ทวยเจริญ, 2556, pp.879-882)

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

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

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

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

เหมืองข้อมูล คือกระบวนการสกัดความรู้ที่น่าสนใจจากข้อมูลปริมาณมาก ซึ่งความรู้ที่ได้จากกระบวนการนี้ เป็นความรู้ที่ไม่ปรากฏให้เห็นเด่นชัด ความรู้ที่บ่งบอกเป็นนัย ความรู้ที่ไม่ทราบมาก่อน ที่มีศักยภาพในการนำไปใช้ประโยชน์ (ชนวัฒน์ ศรีสอาน, 2550, p. 207)

โดยสรุป เหมืองข้อมูล เป็นกระบวนการในการค้นหาความสัมพันธ์ ค้นหาแบบค้นหาแนวโน้ม จากความรู้ที่เก็บรวบรวมอยู่ไว้เป็นจำนวนมาก โดยอาศัยเทคนิคต่างๆ เพื่อนำความรู้ที่ได้ มาสนับสนุนกระบวนการตัดสินใจ จากข้อมูลที่ได้เพื่อนำไปทำนายผลลัพธ์ที่คาดว่าจะเกิดขึ้นภายในอนาคตอันใกล้

2.2.2 เทคนิคการวิเคราะห์ข้อมูลด้วยดาต้า ไม่นิ่ง

เทคนิคการวิเคราะห์ข้อมูลด้วยดาต้า ไม่นิ่ง 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 ซึ่งโมเดลในที่นี่อาจจะเป็นสมการทางคณิตศาสตร์ 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 แต่ในปัจจุบันเราก้าวเข้าสู่ยุคของบิก ดาต้า (Big Data) ซึ่งมีข้อมูลจำนวนมากมหาศาล แต่ข้อมูลเหล่านี้มักจะไม่มีการจัดโครงสร้าง (Unstructured Data) เช่น ข้อความต่างๆที่อยู่ในรูปของอีเมล (E-mail) หรือ เครือข่ายสังคมออนไลน์ (Social Network) ต่างๆ ถ้าเราต้องการนำข้อมูลที่ไม่มีการจัดโครงสร้างมาดำเนินการวิเคราะห์ เราต้องทำการแปลงข้อมูลเหล่านี้ให้เป็นข้อมูลที่มีโครงสร้างก่อน หรือแปลงให้อยู่ในรูปแบบของตารางก่อนนำข้อมูลที่ได้ไปดำเนินการวิเคราะห์ต่อไป

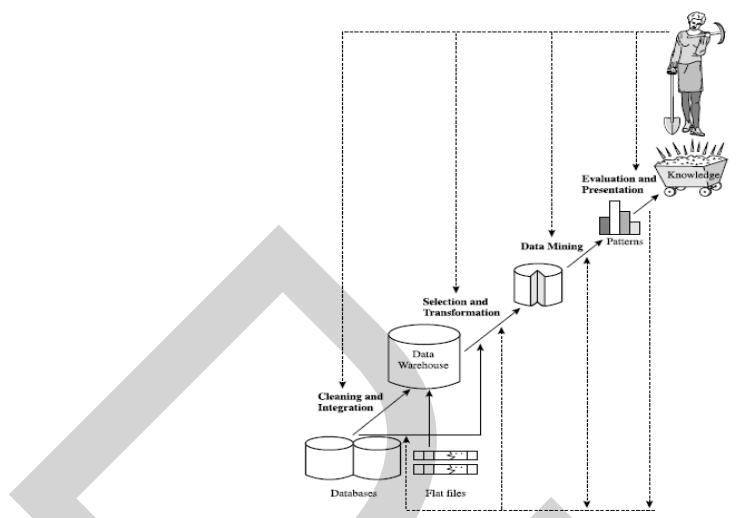
2.2.3.1 ข้อมูลแบบมีโครงสร้าง (Structure data)

ข้อมูลแบบมีโครงสร้างทั่วไปที่เรามักจะพบเห็นในรูปแบบของตาราง เช่น ข้อมูลที่เกี่ยวข้องรายละเอียดยของสมาชิก หรือการซื้อขายสินค้า โดยปกติข้อมูลเหล่านี้จะเก็บอยู่ในไฟล์ประเภท Microsoft Word หรือ Microsoft Excel หรือในฐานข้อมูลต่างๆ ซึ่งในการวิเคราะห์ข้อมูลโดยใช้เทคนิคเหมืองข้อมูล ส่วนใหญ่นั้นเราจะเรียกข้อมูลแต่ละแถวว่า “ตัวอย่าง (Example)” หรือ “อินสแตนซ์ (Instance)” และเรียกข้อมูลแต่ละคอลัมน์ว่า “แอตทริบิวต์ (Attribute)” และข้อมูลในตารางบรรทัดที่หนึ่ง ก็คือชื่อของแต่ละแอตทริบิวต์เสมอ

2.2.3.2 ข้อมูลแบบไม่มีโครงสร้าง (Unstructured data)

ข้อมูลส่วนใหญ่ที่พบ คือ ข้อมูลประเภทรูปภาพ ข้อความ ต่างๆ ซึ่งข้อมูลเหล่านี้มีก็ความสำคัญ ดังนั้นถ้าเราจำเป็นต้องข้อมูลประเภทเหล่านี้ไปใช้งาน เราต้องดำเนินการแปลงข้อมูลเหล่านี้ให้เป็นข้อมูลที่อยู่ในรูปแบบของตาราง โดยใช้กระบวนการแปลงข้อมูล (Data information)

2.2.4 ขั้นตอนการทำเหมืองข้อมูล หรือ ขั้นตอนการค้นหาคำความรู้



ภาพที่ 2.1 แสดงกระบวนการค้นหาคำความรู้ในการขุดค้นข้อมูล

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

ขั้นตอนการทำเหมืองข้อมูล ประกอบไปด้วยขั้นตอนหลักดังต่อไปนี้

1. การทำความเข้าใจปัญหา (Problem Understanding) ในขั้นตอนนี้เป็นการทำความเข้าใจถึงปัญหาและสามารถระบุปัญหาหรือโอกาส จากนั้นทำการแปลงโจทย์ที่ได้ ให้อยู่ในรูปแบบที่เหมาะสมต่อการนำมาวิเคราะห์ข้อมูลทางดาต้าไมน์นิ่ง

2. การทำความเข้าใจข้อมูล (Data Understanding) ขั้นตอนนี้เริ่มจากกระบวนการเก็บรวบรวมข้อมูล หลังจากนั้นจะดำเนินการตรวจสอบข้อมูลที่ได้รวบรวมไว้เพื่อตรวจสอบความถูกต้องของข้อมูลรวมถึงควรพิจารณาด้วยว่าเป็นข้อมูลที่ได้อาจมาจากแหล่งข้อมูลที่ต้องการและมีความน่าเชื่อถือ ประกอบกับข้อมูลที่ได้มีปริมาณมากพอ มีรายละเอียดเพียงพอต่อการนำไปใช้ในการวิเคราะห์

3. การเตรียมข้อมูล (Data Preparation) เป็นขั้นตอนที่ใช้เวลานานที่สุด เนื่องจากโมเดลที่ได้จากการทำดาต้าไมน์นิ่งจะให้ผลลัพธ์ที่ถูกต้องหรือไม่ ขึ้นอยู่กับคุณภาพของข้อมูลที่ใช้ กล่าวคือถ้าข้อมูลที่ใช้นั้นไม่ถูกต้อง มีผิดพลาด ขอมสะทอนถึงผลลัพธ์ที่ได้ 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) เป็นขั้นตอนการเตรียมข้อมูลให้อยู่ในรูปแบบที่พร้อมนำไปใช้ในการวิเคราะห์ตามอัลกอริทึมของค้ำาไมน์นิ่งที่เลือกใช้

4. การสร้างตัวแบบ (Modeling) เป็นขั้นตอนการวิเคราะห์ข้อมูลด้วยเทคนิคค้ำาไมน์นิ่ง ได้แก่ การสร้างตัวทำนาย (prediction model) ในบางครั้งพบว่ามีกรนำเทคนิคค้ำาไมน์นิ่งหลายเทคนิคมาใช้ในการวิเคราะห์ข้อมูล เพื่อให้ได้ผลลัพธ์ที่ดีที่สุด ดังนั้นเมื่อทำขั้นตอนนี้แล้ว อาจมีการย้อนกลับไปทำขั้นตอน data preparation เพื่อแปลงข้อมูลบางส่วนให้เหมาะสมกับแต่ละเทคนิคด้วย นอกจากนี้ยังมีการประเมินโมเดลวิเคราะห์ข้อมูลที่ได้ ในรูปแบบความถูกต้องของโมเดล เพื่อเป็นตัวบ่งชี้ความน่าเชื่อถือของโมเดลที่ได้

5. การประเมินผล (Evaluation) ในขั้นตอนนี้เป็นการประเมินประสิทธิภาพของผลลัพธ์จากโมเดลวิเคราะห์ข้อมูลว่าครอบคลุมและสามารถตอบ โจทย์ที่ตั้งไว้ในขั้นตอนแรกหรือไม่ ในกรณีที่มีการสร้างโมเดลวิเคราะห์ข้อมูลหลายโมเดล ในขั้นตอนนี้จะทำการประเมินแต่ละโมเดลด้วยว่า มีส่วนดีส่วนด้อยอย่างไร และควรเลือกใช้โมเดลใด การทำงานในส่วนนี้ต้องอาศัยทักษะในการวิเคราะห์ข้อมูลเพื่อช่วยให้การวิเคราะห์ทำได้สะดวกและรวดเร็วขึ้น จึงมีการใช้เครื่องมือทางด้านกราฟฟิค เช่น การแสดงผลการวิเคราะห์ด้วยกราฟ รายงานรูปแบบต่างๆ หรือ Dashboard

6. การนำไปใช้งาน (Deployment) เมื่อ ได้ตัวต้นแบบหรือ โมเดลหรือองค์ความรู้ที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิคค้ำาไมน์นิ่งที่มีความถูกต้องก็สามารถนำตัวต้นแบบหรือ โมเดลหรือองค์ความรู้ไปใช้งานและตรวจสอบว่าบรรลุเป้าหมายที่ได้ตั้งไว้หรือไม่

2.2.5. เทคนิคในการทำเหมืองข้อมูล สามารถแบ่งได้ดังนี้

2.2.5.1 การจำแนกประเภทข้อมูล (Classification)

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

$$\underbrace{P_1, P_2, \dots, P_n}_{\text{Attribute Condition}} \rightarrow \underbrace{C_i}_{\text{Class}} \quad \begin{array}{l} P_n = \text{เงื่อนไขของกฎ} \\ C_i = \text{คลาสของกฎ} \end{array}$$

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

ตัววัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล 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

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

4. Accuracy เป็นการวัดความถูกต้องของโมเดล โดยพิจารณารวมทุกคลาส คือจำนวน True Positive ของทุกคลาสรวมกัน

การแบ่งข้อมูลเพื่อใช้ในการวัดประสิทธิภาพของโมเดลการจำแนกประเภทข้อมูล

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 ซึ่งเป็นจุดประสงค์หลักที่ใช้ในการแก้กับปัญหา คือการสร้างตัวต้นแบบ เมื่อมีข้อมูลใหม่เข้ามาสู่ระบบจะสามารถทำนายได้โดยการนำข้อมูลที่ได้รับมาทำการเปรียบเทียบกับตัวต้นแบบ โมเดลที่เราจำแนกได้จากขั้นตอนที่ 1 เพื่อวิเคราะห์ เพื่อตัดสินใจในความเป็นไปได้ของข้อมูลนั้นๆ

การแบ่งกลุ่มข้อมูลด้วยวิธี Decision Tree

เป็นโครงสร้างที่ใช้แสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล โดยต้นไม้ช่วยในการตัดสินใจจะมีลักษณะคล้ายกับโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณลักษณะ (Attribute) แต่ละกิ่งแสดงเงื่อนไขในการทดสอบและโหนดปลาย แสดงกลุ่มที่กำหนดไว้

การแทนต้นไม้การตัดสินใจ (Decision Tree Representation)

1. โหนดภายใน (internal node) คือคุณสมบัติต่างๆของข้อมูล ซึ่งเมื่อข้อมูลใดๆตกลงมาที่โหนด จะใช้คุณสมบัตินี้เป็นตัวตัดสินใจว่าข้อมูลที่ได้จะไปในทิศทางใด โดยโหนดภายในที่เป็นจุดเริ่มต้นของต้นไม้เราเรียกว่าโหนดราก

2. กิ่ง (Branch, Link) เป็นค่าคุณสมบัติของคุณสมบัติในโหนดภายในที่แตกกิ่งนี้ออกมา ซึ่งโหนดภายในจะแตกกิ่งเป็นจำนวนเท่ากับค่าคุณสมบัติของโหนดภายในนั้นๆ

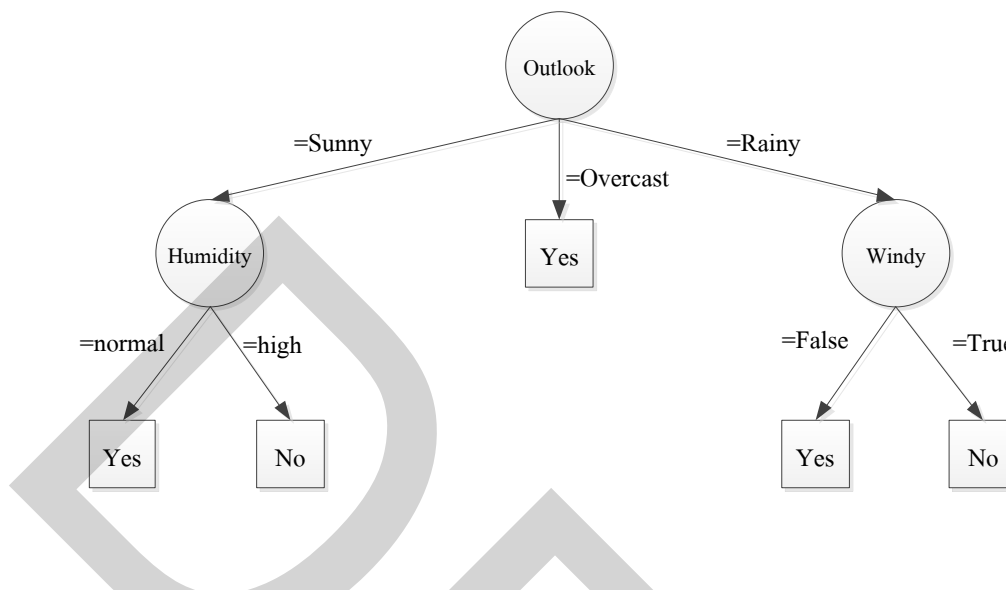
3. โหนดใบ (Leaf node) คือกลุ่มต่างๆซึ่งเป็นผลลัพธ์ในการแยกแยะข้อมูล

เทคนิคการจำแนกประเภทข้อมูลด้วยวิธี Decision Tree เป็นวิธีการหนึ่งที่นิยมใช้กันอย่างแพร่หลาย เนื่องจากเป็นโมเดลที่สามารถแปลความหมายและเข้าใจได้ง่าย 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 ไปใช้งานเราจะเริ่มพิจารณาจากโหนดบนสุด(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 เป็นอีกวิธีหนึ่งที่ได้รับคความนิยม 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

$P(A \cap B)$ คือ ค่า joint probability หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ A และเหตุการณ์ B เกิดขึ้นร่วมกัน

$P(B)$ คือ ค่าความน่าจะเป็นที่เกิดเหตุการณ์ B เกิดขึ้น

ในลักษณะเดียวกันเราจะเขียน $P(B|A)$ หรือค่าความน่าจะเป็นที่เกิดเหตุการณ์ A เกิดขึ้นก่อนและเหตุการณ์ B เกิดขึ้นตามมาทีหลังได้เป็น

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

จากสมการทั้ง 2 แบบจะเห็นว่าค่า $P(A \cap B)$ ที่เหมือนกันอยู่ดังนั้นเราสามารถเขียนสมการของ $P(A \cap B)$ 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 แล้วสามารถคำนวณได้จากค่าความน่าจะเป็นของแอตทริบิวต์ A ที่มีคลาส C ในเทรนนิ่ง 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 $A = a_1 \cap a_2 \dots \cap a_M$ โดยที่ M คือจำนวนแอตทริบิวต์ใน training data

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

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

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

ซึ่งในการคำนวณเราสามารถคำนวณแยกทีละแอตทริบิวต์ได้ดังสมการข้างล่างนี้

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

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	?

$$\begin{aligned}
 P(\text{Play}=\text{yes}|A) &= P(\text{Outlook}=\text{sunny}|\text{Play}=\text{yes}) \times P(\text{Temperature}=\text{hot}|\text{Play}=\text{yes}) \\
 &\quad \times P(\text{Humidity}=\text{high}|\text{Play}=\text{yes}) \times P(\text{Windy}=\text{false}|\text{Play}=\text{yes}) \times P(\text{Play}=\text{yes}) \\
 &= (2/9) \times (2/9) \times (3/9) \times (6/9) \times (9/14) \\
 &= 0.22 \times 0.22 \times 0.33 \times 0.67 \times 0.64 \\
 &= 0.0068
 \end{aligned}$$

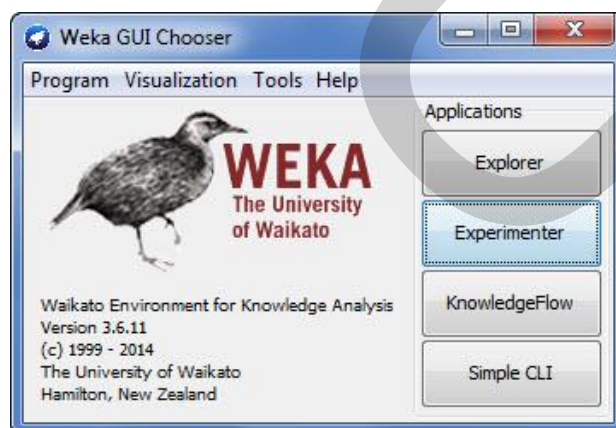
$$\begin{aligned}
 P(\text{Play}=\text{no}|A) &= P(\text{Outlook}=\text{sunny}|\text{Play}=\text{no}) \times P(\text{Temperature}=\text{hot}|\text{Play}=\text{no}) \\
 &\quad \times P(\text{Humidity}=\text{high}|\text{Play}=\text{no}) \times P(\text{Windy}=\text{false}|\text{Play}=\text{no}) \times P(\text{Play}=\text{no}) \\
 &= (3/5) \times (2/5) \times (4/5) \times (2/5) \times (5/14) \\
 &= 0.6 \times 0.4 \times 0.8 \times 0.4 \times 0.36 \\
 &= 0.0276
 \end{aligned}$$

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

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

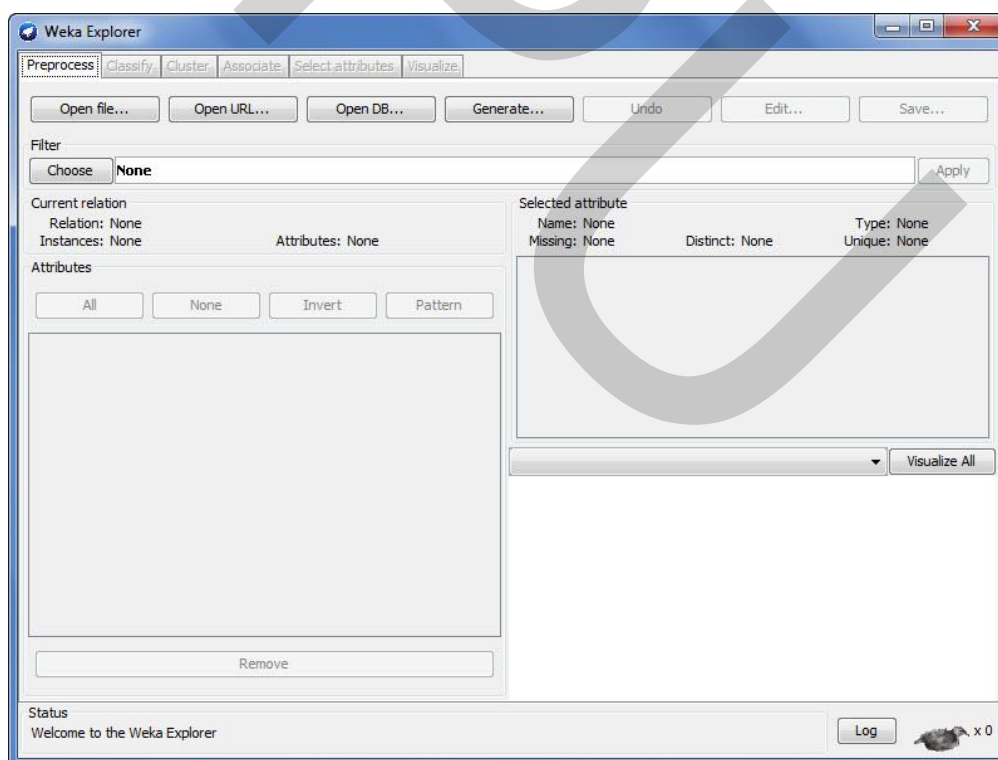
Explorer เป็นส่วนที่ผู้ใช้สามารถใช้ฟังก์ชันการทำงานต่างๆ ของ Weka ผ่านทางหน้าจอ GUI ซึ่งเป็นส่วนที่เหมาะสมสำหรับผู้เริ่มต้นใช้งาน Weka เพราะผู้ใช้จะสามารถเรียกฟังก์ชันการทำงานต่างๆ ได้เพียงแค่คลิกและเปลี่ยนแปลงพารามิเตอร์ในหน้าฟอร์มเท่านั้น

Experimenter เป็น GUI ที่ยอมให้ผู้ใช้สามารถเปลี่ยนแปลงเทคนิคในการวิเคราะห์ข้อมูลและค่าพารามิเตอร์ต่างๆ ที่เกี่ยวข้องได้หลากหลายรูปแบบ จนได้ผลลัพธ์ที่น่าพอใจ ซึ่งการใช้ Explorer นั้น ไม่สะดวกเนื่องจากต้องมาทำการเปลี่ยนแปลงค่าต่างๆ

Knowledge Flow เป็น GUI อีกส่วนหนึ่งที่ยอมให้ผู้ใช้สามารถนำเทคนิคต่างๆ ของ Weka มาเรียงต่อกันเพื่อช่วยให้การวิเคราะห์ข้อมูลให้ทำงานได้ตามที่ต้องการ

Simple CLI เป็นส่วนที่ผู้ใช้สามารถเรียกฟังก์ชันการทำงานของ Weka มาใช้ผ่านทาง command line ได้ ซึ่งการเรียกใช้ฟังก์ชันผ่านทาง command line นี้จะช่วยให้ผู้ใช้เข้าใจการเรียกฟังก์ชันต่างๆ เบื้องหลังหน้าจอ GUI ของ Weka ได้ และสามารถนำไปประยุกต์ใช้ในการเขียนโปรแกรมเพื่อเรียก Weka ในการใช้งานได้อีกด้วย

ในส่วนนี้จะนำเสนอหน้าต่างในส่วนของ Explorer เป็นหลัก โดยในหน้าต่าง Weka Explorer สามารถแบ่งพื้นที่ได้ดังนี้



ภาพที่ 2.4 เมนูหลักในหน้าต่าง Weka Explorer

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

ส่วนที่อยู่ตรงกลางซึ่งจะเปลี่ยนไปตามการกดแท็บต่างๆ ส่วนนี้เป็นส่วนของการเลือก option ต่างๆ ในการวิเคราะห์ข้อมูล และส่วนการแสดงผลหลังจากทำการวิเคราะห์ข้อมูลเสร็จ 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 ซึ่งมีความยาวไม่มากนักและสามารถทำงานได้ดั่งกับเว็บไซต์) PHP เป็นภาษาสคริปต์ที่เป็น Server Side Scrip และเป็น Open Source ที่ผู้ใช้ทั่วไปสามารถ Download Source Code ได้ฟรี จุดประสงค์ที่สำคัญของภาษา PHP คือการช่วยให้นักพัฒนาเว็บเพจสามารถเขียนเว็บเพจที่เป็นแบบไดนามิกได้อย่างรวดเร็ว ภาษา PHP จะทำงานร่วมกันกับเอกสาร HTML โดยการสร้างโค้ดแทรกระหว่าง Tag HTML และสร้างเป็นไฟล์ที่มีนามสกุล .php .php3 หรือ php4 และไวยากรณ์ที่ใช้ใน PHP เป็นการนำรูปแบบของภาษาต่างๆ มารวมกัน เช่น ภาษา C Perl และ Java ทำให้ผู้ใช้ที่มีพื้นฐานของภาษาเหล่านี้สามารถใช้งานได้

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

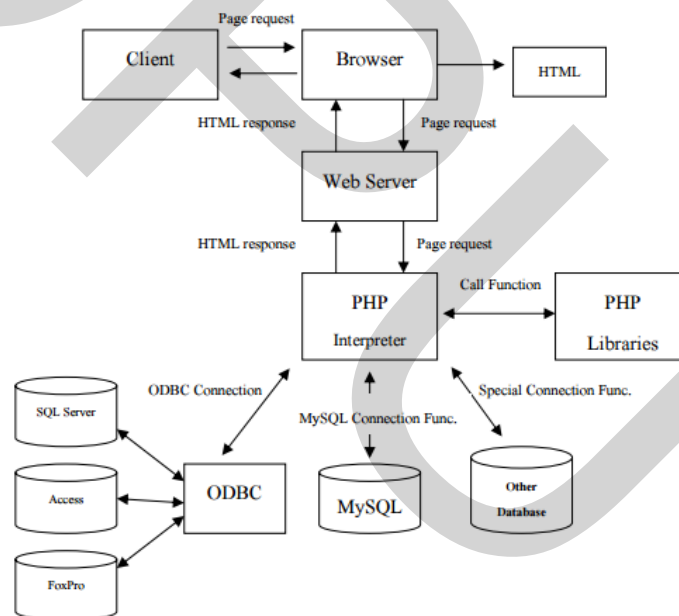
1. จากไคลเอนต์จะเรียกไฟล์ php script ผ่านทางโปรแกรมบราวเซอร์ (Internet Explore)
2. บราวเซอร์จะส่งคำร้อง (Request) ไปยังเว็บเซิร์ฟเวอร์ผ่านทางเครือข่ายอินเทอร์เน็ต

3. เมื่อเว็บเซิร์ฟเวอร์รับคำร้องขอจากบราวเซอร์แล้วก็จะนำสคริปต์ php ที่เก็บอยู่ในเซิร์ฟเวอร์มาประมวลผลด้วยโปรแกรมแปลภาษา PHP ที่เป็น อินเทอร์เน็ต

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

5. Web Service ส่งไฟล์ HTML ที่ได้ผ่านการแปลแล้วกลับไปยังบราวเซอร์ที่ร้องขอผ่านทางเครือข่ายอินเทอร์เน็ต

6. บราวเซอร์รับไฟล์ HTML ที่เว็บเซิร์ฟเวอร์ส่งมาให้แปล HTML แสดงผลออกมาทางจอภาพเป็นเว็บเพจ โดยใช้ตัวแปลภาษา HTML ที่อยู่ในบราวเซอร์ซึ่งเป็นอินเทอร์เน็ต เช่นเดียวกัน



ภาพที่ 2.5 แสดงขั้นตอนการทำงาน PHP Script Request/Response

ความสามารถของภาษา PHP

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

เป็นภาษาที่ทำความเข้าใจและใช้งานง่ายไม่เหมือนกับ JAVA หรือ C++ และมีส่วนที่สนับสนุนการทำงานได้กับทุกเว็บไซต์

เป็น Open Source ผู้ใช้สามารถดาวน์โหลด และนำ source code ของ PHP ไปใช้ได้โดยไม่เสียค่าใช้จ่าย

เป็นสคริปต์แบบเซิร์ฟเวอร์ไซด์ดังนั้นจึงทำงานบนเว็บเซิร์ฟเวอร์ไม่ส่งผลกระทบต่อการทำงานของเครื่องไคลเอนต์โดย PHP จะอ่านโค้ด และทำงานที่เซิร์ฟเวอร์จากนั้นจึงส่งผลลัพธ์ที่ได้จากการประมวลผลมาที่เครื่องของผู้ใช้ในรูปแบบของเอกสาร HTML ซึ่งอ่านโค้ดของ PHP ผู้ใช้ไม่สามารถมองเห็นได้

PHP สามารถทำงานได้ในระบบปฏิบัติการที่ต่างชนิดกัน เช่น Unix, Windows, Mac, OS หรือ Risc OS อย่างมีประสิทธิภาพเนื่องมาจาก PHP เป็นสคริปต์ที่ต้องทำงานบนเซิร์ฟเวอร์ดังนั้นคอมพิวเตอร์ที่ใช้ สำหรับเรียกคำสั่ง PHP จึงจำเป็นต้องติดตั้งโปรแกรมประเภทเว็บเซิร์ฟเวอร์ไว้ด้วยเพื่อให้สามารถประมวลผล PHP ได้ซึ่งเป็นเหตุผลที่ทำให้ PHP สามารถทำงานได้กับหลายระบบปฏิบัติการหลายชนิด

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

สนับสนุนการเขียนสคริปต์ที่ใช้หลักของ Object Orientation

PHP สามารถสร้างเว็บไซต์ที่บรรจุข้อมูลรูปแบบต่าง ๆ ลงในเว็บ เช่น รูปภาพ ไฟล์ PDF หรือ Flash Movie

คุณสมบัติที่สำคัญอีกประการหนึ่งของ PHP คือความสามารถในการทำงานร่วมกับระบบจัดการฐานข้อมูลที่หลากหลายซึ่งระบบการจัดการฐานข้อมูลที่สนับสนุนการทำงาน of 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, mSQL และ MySQL

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

ผู้ใช้สามารถเขียน โค้ด 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 หรือ Windows 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 จะเป็นส่วนที่ทำหน้าที่บริหารจัดการระบบฐานข้อมูลในที่นี้ก็หมายถึงตัว MySQL Server นั่นเองและเป็นที่จัดเก็บข้อมูลทั้งหมดข้อมูลที่เก็บไว้จะมีข้อมูลที่จำเป็นสำหรับการทำงานกับระบบฐานข้อมูลและข้อมูลที่เกิดจากการที่ผู้ใช้แต่ละคนสร้างขึ้นมา

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

หลักการทำงานในลักษณะ Client/ Server มีดังนี้

1. ที่ฝั่งของ Server จะมีโปรแกรมหรือระบบสำหรับจัดการฐานข้อมูลทำงานรออยู่เพื่อเตรียมหรือรอคอยการร้องขอการให้บริการจาก Client

2. เมื่อมีการร้องขอการให้บริการเข้ามา Server จะทำการตรวจสอบตามวิธีการของตน เช่น อาจจะมีการให้ผู้ให้บริการระบุชื่อและรหัสผ่าน และสำหรับ MySQL สามารถกำหนดได้ว่าจะอนุญาตหรือปฏิเสธ Client ใดๆ ในระบบที่จะเข้าใช้บริการอีกด้วย

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

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

2.3.4 โปรแกรม Adobe Dreamweaver CS5

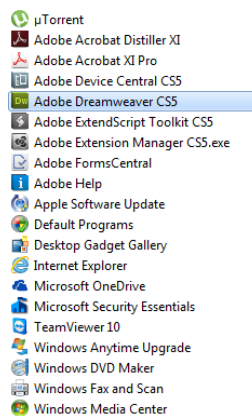
Adobe Dreamweaver CS5 เป็นโปรแกรมสำเร็จรูปที่ใช้สำหรับการพัฒนาเว็บไซต์ โดยมีคุณสมบัติในการออกแบบและสร้างเว็บเพจ จนถึงการพัฒนาแอปพลิเคชันเบื้องต้น ซึ่งในโปรแกรมตัวนี้มีเครื่องมือสำหรับการวางข้อความ ภาพกราฟิก ตาราง แบบฟอร์ม พร้อมทั้งมีลิตมีเดียต่างๆ เพื่อแสดงให้ผู้พัฒนาเว็บไซต์ใช้งานได้ง่าย โดยไม่ต้องรู้จักภาษา HTML, JavaScript, CSS หรือภาษาสคริปต์อื่นๆ ซึ่งเมื่อออกแบบและพัฒนาเว็บไซต์ในโปรแกรม แล้วนำมาแสดงผลทาง browser ก็จะเห็นเป็นลักษณะที่ได้จัดวางไว้ 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

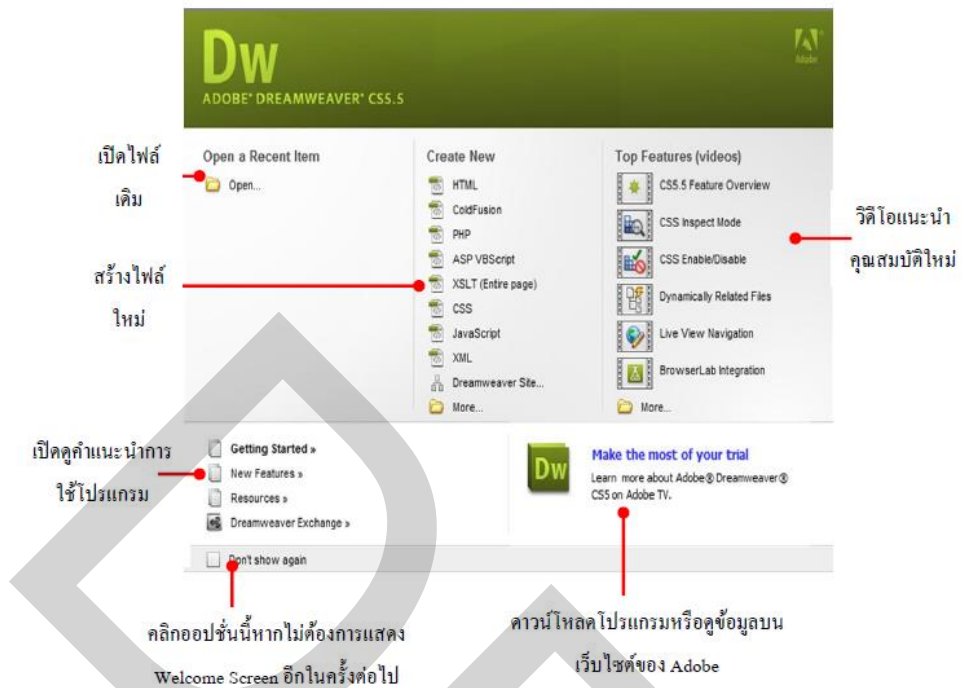
Welcome Screen

Welcome Screen เป็นเครื่องมือสำหรับช่วยเลือกขั้นตอนเริ่มต้นในการทำงาน โปรแกรม โดยตัวเลือกจะแบ่งออกเป็นกลุ่มๆ ดังภาพต่อไปนี้

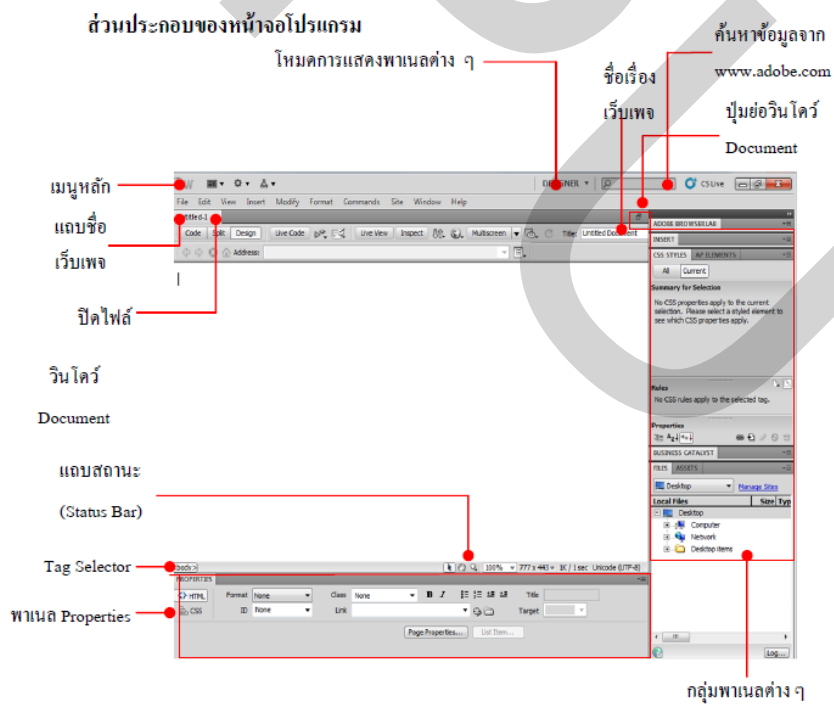
Open a Recent Item เปิดไฟล์ที่เคยใช้ โดยคลิกเลือกจากรายชื่อที่แสดงอยู่ (เรียงลำดับจากที่เคยเปิดล่าสุดเป็นต้นไปหรือคลิก Open เพื่อเปิดไฟล์อื่นๆ

Create New สร้างไฟล์ใหม่ โดยถ้าคลิก HTML จะเป็นการสร้างเว็บเพจพื้นฐาน แต่ถ้าคลิกหัวข้ออื่นจะเป็นการสร้างเว็บเพจหรือไฟล์ตามชนิดนั้นๆ

Top Features (videos) เป็นเส้นทางลัดสำหรับเข้าดูรายละเอียดและเทคนิคในการทำงานต่างๆ ของโปรแกรมผ่านทางเว็บไซต์ของ Adobe



ภาพที่ 2.7 ส่วนประกอบของ Welcome Screen



ภาพที่ 2.8 ส่วนประกอบของหน้าจอโปรแกรม

2.4 การวิเคราะห์และออกแบบระบบ

วงจรการพัฒนาระบบ (System Development Life Cycle :SDLC) คือกระบวนการหรือวงจรในการพัฒนาระบบด้านคอมพิวเตอร์หรือระบบสารสนเทศ โดยแปลจากความต้องการของผู้ใช้งานมาเป็นในรูปของแอปพลิเคชัน โดยมีการกำหนดกิจกรรมต่างๆที่เกิดขึ้นในแต่ละระยะของการพัฒนาระบบขึ้นอย่างชัดเจนตั้งแต่ระยะเริ่มแรกไปจนถึงหลังระยะสิ้นสุดการพัฒนาระบบมีอยู่ด้วยกัน 7 ขั้นด้วยกัน 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 สรุปขั้นตอนของระยะ

การนำระบบไปใช้

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 พงษ์ศิริปริดาและกฤษณะ ไวยมัย

งานวิจัยนี้ได้นำเสนอเทคนิค Data Classification มาใช้สร้างระบบจัดสรรกฎหมายที่เหมาะสมให้กับคดีความแบบอัตโนมัติ ซึ่งผลการทดลองที่ได้ยังไม่เป็นที่น่าพอใจ ซึ่งสามารถสรุปได้ ดังนี้

ประเด็นที่ 1 การตัดคำในขั้นตอนการเตรียมข้อมูลยังขาดประสิทธิภาพ ทำให้ผลการทำนายของระบบมีความถูกต้องน้อยมากหรืออยู่ในระดับต่ำ

ประเด็นที่ 2 ระบบไม่สามารถทำนายกฎหมายให้กับคดีความได้ครั้งละมากกว่าหนึ่งกฎหมาย

ประเด็นที่ 3 ระบบไม่สามารถทำนายกฎหมายแบบมีลำดับชั้นในเวลาเดียวกันได้

2.5.2 งานวิจัยเรื่อง การจัดสรรกฎหมายที่เหมาะสมให้กับคดีความอัตโนมัติโดยธีระวัฒน์ พงษ์ศิริปริดา ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

งานวิจัยนี้ได้นำเสนอเทคนิค การตัดคำโดยใช้เทคนิค Suffix Array เพื่อทำการสร้างพจนานุกรมวลีขึ้นมาใช้งาน แล้วสร้างตัวทำนายผลกฎหมายโดยใช้เทคนิค Association Rule Discovery มาใช้ร่วมกับเทคนิค Data Classification ซึ่งการจำแนกประเภทข้อมูลให้ความแม่นยำในการทำนายสูง เนื่องจากได้นำเทคนิคการหาความสัมพันธ์เข้ามาพร้อมกับเทคนิคการจำแนกประเภทข้อมูล แต่เนื่องจากวิธีการทำนายของเทคนิคดังกล่าวได้พิจารณาเพียงเฉพาะความสัมพันธ์เดียวในการทำนาย 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 รวมถึงการค้นหาดัชนีทฤษฎีความน่าจะเป็นในการวินิจฉัยคดี ซึ่งงานวิจัยชิ้นนี้ได้นำเสนอวิธีการวินิจฉัยคดีโดยประยุกต์ใช้เทคนิคต้นไม้ในการตัดสินใจเพื่อจำแนกความถูกต้อง ซึ่งสามารถอธิบายได้โดยชุดคุณลักษณะของข้อมูล ออกเป็นรายมาตราที่เหมาะสมกับคดีความนั้นๆ และเปรียบเทียบความถูกต้องของการจำแนกข้อมูลด้วยขั้นตอนวิธี 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 ได้นำเสนอเทคนิคต่างๆ ที่สำคัญของการทำเหมืองข้อมูลมาประยุกต์ใช้ในการสืบค้นสิ่งที่น่าสนใจจากข้อมูลนิติตคณะวิศวกรรมศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ โดยใช้เทคนิคที่สำคัญ 3 เทคนิค คือ กฎความสัมพันธ์ (Association rule discovery) เทคนิคการจำแนกข้อมูล (Data Classification) การพยากรณ์ข้อมูล (Data Prediction) มาประยุกต์ในการช่วยนิติตเลือกสาขาที่เหมาะสม และสามารถทำนายเกรดแต่ละรายวิชาในภาคการศึกษาต่อไป โดยใช้ต้นไม้ในการตัดสินใจ (Decision tree) ของการจำแนกข้อมูล ในการสร้างตัวต้นแบบโมเดล (Model)

ข้อดี

สามารถทำนายสาขาวิชาที่เหมาะสมที่สุดให้กับนักศึกษาได้

ข้อเสีย

ตัวแบบโมเดล (Model) จะทำนายแนวโน้มเชิงไปทางสาขาวิชาที่มีจำนวนนิติตนักศึกษามากเป็นผลทำให้ความถูกต้องของโมเดลที่ได้ค่อนข้างต่ำ

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

มีเปอร์เซ็นต์ความถูกต้องค่อนข้างสูงและมีปัญหาบางประการได้แก่ การจำแนกข้อมูลในบางสาขาวิชาที่มีปริมาณน้อยทำให้โมเดลตัวต้นแบบที่ได้ไม่มีความแม่นยำเท่าที่ควร

2.5.5 งานวิจัยเรื่อง การวัดประสิทธิภาพของขั้นตอนวิธี ตัวจำแนก C4.5, ADTree และ Naïve Bayes ในการจำแนกข้อมูลการชุกซ่อนสิ่งเสพติดสำหรับไปรษณีย์ระหว่างประเทศ โดยสิทธิโชค มุกดาสกุลภินาล นิสิตปริญญาโทสาขาวิทยาการคอมพิวเตอร์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

งานวิจัยฉบับนี้ได้นำเสนอการวัดความแม่นยำและประสิทธิภาพด้านความเร็วของตัวจำแนก 3 ชนิด คือ C4.5, ADTree และ Naïve Bayes เพื่อนำไปใช้งานในการตรวจสอบสิ่งของชุกซ่อนสิ่งเสพติดสำหรับไปรษณีย์ระหว่างประเทศ 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

บทที่ 3

ระเบียบวิธีวิจัย

3.1 ขั้นตอนการศึกษาค้นคว้า

ในบทนี้เป็นการอธิบายถึงวิธีในการดำเนินการศึกษาค้นคว้า และมีส่วนประกอบที่จำเป็นในการศึกษาหลายส่วน โดยประกอบไปด้วยขั้นตอนการดำเนินงานดังนี้

(1) การศึกษาและรวบรวมข้อมูล

ในส่วนของการศึกษาและรวบรวมข้อมูล ผู้ศึกษาได้ดำเนินการศึกษาโครงสร้างของฐานข้อมูลจริงจากโปรแกรมระบบจัดการฐานข้อมูล Microsoft SQL Server 2005 ที่ได้ใช้งานจริง รวมถึงรวบรวมข้อมูลต่างๆที่มีความเกี่ยวข้องกับการใช้งานในการศึกษาค้นคว้าครั้งนี้ เช่น รวบรวมข้อมูลงานวิจัย งานศึกษาค้นคว้า จากผู้มีประสบการณ์

(2) การเตรียมข้อมูล

หลังจากที่ได้ศึกษาและรวบรวมข้อมูล ในขั้นตอนนี้จะต้องมีการจัดเตรียมข้อมูลและคัดเลือกข้อมูลที่จะนำมาใช้งาน เนื่องจากข้อมูลที่เราจำลองขึ้นมาแล้วยังไม่สามารถนำไปใช้งานได้ ต้องผ่านกระบวนการทดสอบและวิเคราะห์ด้วยเครื่องมือของการทำเหมืองข้อมูล จึงต้องมีการจัดเตรียมข้อมูลให้เป็นไปตามมาตรฐานเดียวกัน และคัดเลือกข้อมูลที่มีความถูกต้อง สมบูรณ์มากที่สุด เพื่อให้ได้ผลการทดสอบที่มีประสิทธิภาพสูงสุด

(3) การกำหนดคุณลักษณะ (Attribute) และค่าที่เป็นไปได้ของคุณลักษณะ

จากการศึกษาโครงสร้างของฐานข้อมูลจริง และการจัดเตรียมข้อมูลนั้น เราจะได้คุณลักษณะที่คาดว่าจะนำมาใช้งานเพื่อใช้ในการหาความรู้ และได้มีการกำหนดค่าที่เป็นไปได้ของคุณลักษณะ

(4) การแปลงข้อมูลสำหรับการประมวลผล

ในส่วนของการแปลงข้อมูลสำหรับการประมวลผลนั้น พบว่าก่อนที่จะดำเนินการแปลง เราต้องมีการโอนย้ายข้อมูล (Data Transfer) การลดขนาดข้อมูล (Data Reduction) และการทำความสะอาดข้อมูลก่อน (Data Cleaning)

(5) การศึกษา การทดสอบ และเลือกเทคนิคที่เหมาะสมนำมาใช้งาน

ในขั้นตอนนี้เป็นการนำข้อมูลที่ได้คัดเลือกจากขั้นตอนที่ผ่านมา นำข้อมูลที่ได้มาดำเนินการทดสอบกับโปรแกรม Weka ซึ่งต้องมีการเลือกเทคนิคการจำแนกกลุ่มข้อมูลที่เหมาะสม

โดยเทคนิคแต่ละเทคนิคนั้นจะให้ผลลัพธ์ที่ต่างกันไป ดังนั้นต้องนำโมเดลที่ได้จากการศึกษามา คัดเลือกเทคนิคที่เหมาะสม เพื่อนำมาใช้วิเคราะห์การตรวจสอบข้อมูลทางคดี และนำไปพัฒนา ระบบสนับสนุนการตัดสินใจ ซึ่งมีการทดสอบจำนวน 2 เทคนิค โดยรายละเอียดของเทคนิคได้ กล่าวไว้ในรายละเอียดของบทที่ 2 และแสดงผลลัพธ์ที่ได้ในบทที่ 4 และผลลัพธ์ฉบับเต็มอยู่ที่ ภาคผนวก

(6) การวิเคราะห์ผลและการสรุปผลการทดสอบ

ในขั้นตอนนี้เป็นการวิเคราะห์ผลลัพธ์ที่ได้จากการทดสอบของเทคนิคของเหมือง ข้อมูลจำนวน 2 เทคนิค โดยนำผลลัพธ์ที่ได้ของแต่ละเทคนิคมาวิเคราะห์ความถูกต้องจากข้อมูลที่ นำมาทดสอบ เพื่อทำการเลือกกฎที่มีความน่าเชื่อถือที่สุด และมีค่าความเป็นจริงมากที่สุด เพื่อใช้ เื่อนใจในการพัฒนาระบบสนับสนุนการตัดสินใจจ่ายสำนวนคดีความให้กับผู้พิพากษา ในขั้นตอน ต่อไป

(7) การพัฒนาระบบสนับสนุนการตัดสินใจจ่ายสำนวนคดีความให้กับผู้พิพากษา

ในขั้นตอนนี้เป็นการนำกฎที่ได้จากการทำเหมืองข้อมูลมาเป็นเงื่อนไขในการพัฒนา โปรแกรมให้เป็นเครื่องมือสนับสนุนการตัดสินใจจ่ายสำนวนคดีความให้กับผู้พิพากษา

(8) สรุปผลการศึกษาและข้อเสนอแนะ

ในขั้นตอนนี้เป็นการกล่าวถึงผลการศึกษาในการใช้เหมืองข้อมูลมาช่วยสนับสนุนการ ตัดสินใจจ่ายสำนวนคดีความให้กับผู้พิพากษา เพื่อช่วยให้ด้านการบริหารงานของผู้บริหารได้ สะดวก

(9) ดำเนินการเรียบเรียงงานค้นคว้าอิสระ

3.2 เครื่องมือที่ใช้ในการดำเนินการศึกษา

- (1) โปรแกรมสำเร็จรูป WEKA Version 3.6 เพื่อใช้ในการศึกษาในการทำเหมืองข้อมูล
- (2) ระบบปฏิบัติการ Windows 7 Professional 64 bit
- (3) โปรแกรมสำเร็จรูป Microsoft Excel version 2010 ใช้ในการเตรียมข้อมูลสำหรับ นำเข้าทดสอบในโปรแกรม WEKA
- (4) โปรแกรมสำเร็จรูป Microsoft Word version 2010 ใช้ในการจัดทำรายงานค้นคว้า อิสระและเอกสารอื่นๆ
- (5) โปรแกรมสำเร็จรูป Microsoft PowerPoint version 2010 ใช้ในการนำเสนอผล การศึกษางานค้นคว้าอิสระ
- (6) โปรแกรม Adobe Dreamweaver CS5 ใช้ในการพัฒนาโปรแกรม

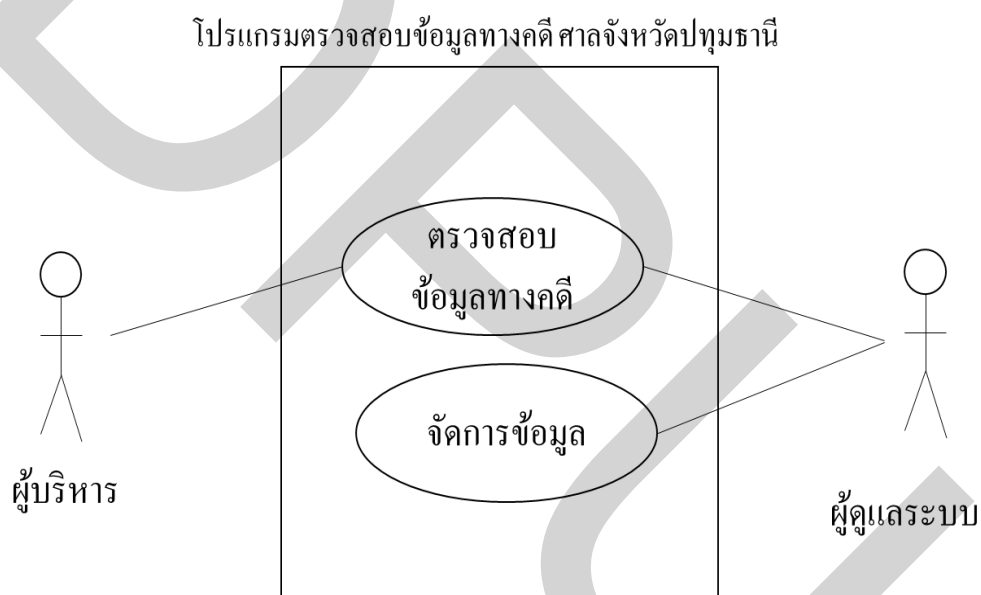
(7) โปรแกรม MySQL ใช้เป็นฐานข้อมูลในการใช้งาน

(8) ภาษา PHP

3.3 ขั้นตอนการวิเคราะห์และออกแบบ

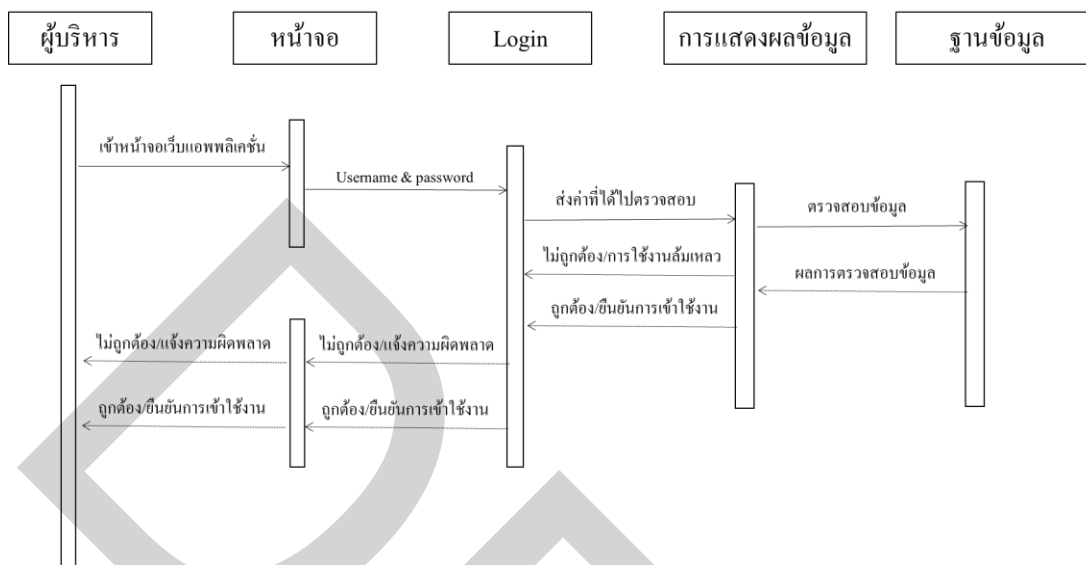
3.3.1 แผนภาพแสดง Use Case Diagram

แผนภาพแสดงข้อมูลยูสเคสของโปรแกรมจะแสดงภาพรวมของระบบที่มีการใช้งาน โดยจะมีผู้เกี่ยวข้องอยู่ด้วยกันสองส่วนคือ ผู้บริหารและผู้ดูแลระบบ ซึ่งผู้บริหารจะมีสิทธิแค่ตรวจสอบข้อมูลทางคดีเท่านั้น ส่วน ผู้ดูแลระบบมีสิทธิในการจัดการกับข้อมูลทุกอย่าง รวมถึงมีสิทธิในการตรวจสอบข้อมูลทางคดีได้ด้วย ดังแสดงในแผนภาพที่ 3.1



ภาพที่ 3.1 แสดงแผนภาพยูสเคสของโปรแกรมตรวจสอบข้อมูลทางคดี

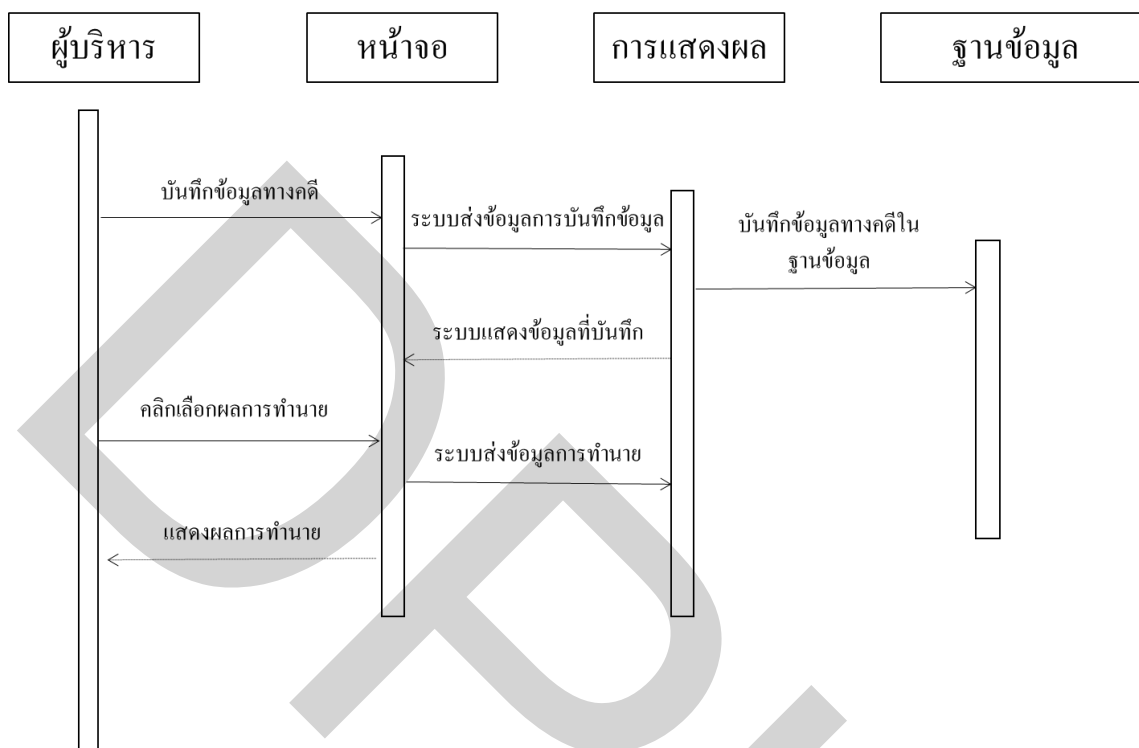
3.3.2 แผนภาพแสดง Sequence Diagram ที่แสดงขั้นตอนการทำงานของระบบ



ภาพที่ 3.2 แสดงแผนภาพการทำงานในส่วนของการเข้าสู่ระบบ

ในส่วนของหน้าจอในการล็อกอิน (Login) เข้าสู่ระบบนั้น ผู้บริหารต้องการล็อกอินเข้าสู่ระบบต้องดำเนินการเปิดเว็บแอปพลิเคชัน (Web Application) หน้าจอของเว็บแอปพลิเคชันจะตอบสนองให้ผู้บริหารดำเนินการกรอก Username และ Password เพื่อดำเนินการล็อกอินเข้าสู่ระบบ ระบบจะดำเนินการส่งค่าที่ได้ไปตรวจสอบภายในฐานข้อมูล ถ้าผลการตรวจสอบข้อมูลมีความถูกต้อง ระบบจะแสดงผลให้ผู้บริหารสามารถเข้าไปใช้งานได้ ถ้าผู้บริหารกรอก Username หรือ Password ใดๆอย่างหนึ่งผิดพลาด ระบบจะแสดงผลตอบสนองและแจ้งเตือนผิดพลาดที่เกิดขึ้นให้ทราบ

3.3.3 แผนภาพแสดง Sequence Diagram แสดงผลการทำนายผลลัพธ์ที่เกิดขึ้น



ภาพที่ 3.3 แสดงแผนภาพการทำงานในส่วนของการทำนายผลลัพธ์ที่เกิดขึ้น

เมื่อผู้บริหารล็อกอินเข้าสู่ระบบเป็นที่เรียบร้อยแล้ว ในขั้นตอนต่อมาเป็นการแสดงผลการบันทึกข้อมูลทางคดี โดยผู้บริหรสามารถบันทึกข้อมูลทางคดีในหน้าเว็บแอปพลิเคชันเมื่อดำเนินการบันทึกข้อมูลทางคดีเสร็จ ระบบจะส่งข้อมูลการบันทึกไปยังฐานข้อมูลเพื่อบันทึกข้อมูลที่ได้ลงฐานข้อมูล จากนั้นผู้บริหรจะต้องคลิกแท็บของหน้าเว็บแอปพลิเคชันในแท็บถัดไป ระบบจะแสดงผลข้อมูลที่ได้บันทึกไป จากนั้นให้ผู้บริหรคลิกผลการทำนายในการแบ่งกลุ่มประเภทคดี ซึ่งผลลัพธ์ที่ได้จากการแบ่งกลุ่มจะเป็นประเภทคดีง่ายหรือประเภทคดียาก

3.3.4 การออกแบบการจัดเก็บข้อมูลในระบบฐานข้อมูล

สำหรับการจัดเก็บข้อมูลในระบบฐานข้อมูลแสดงในตารางที่ 3.1 ถึงตารางที่ 3.2 โดยมีรายละเอียดดังต่อไปนี้

ตารางที่ 3.1 ตารางแสดงการจัดเก็บข้อมูลของฐานข้อมูลชื่อ Data_case

ลำดับ	ชื่อฟิลด์	ประเภท	ขนาด	คำอธิบาย	คีย์
1	id	int	11	ลำดับของข้อมูลที่จัดเรียงภายในฐานข้อมูล	PK
2	Gender	Varchar	255	เพศ	
3	Reasoning	Varchar	255	ประเภทคดี	
4	Have_rate	Varchar	255	ความเกี่ยวเนื่องของคดี	
5	Time	Varchar	255	ช่วงระยะเวลา	
6	Document	Varchar	255	ความครบถ้วนของเอกสาร	
7	Consider	Varchar	255	ผลการพิจารณา/ผลการตัดสิน	

ตารางที่ 3.2 ตารางแสดงข้อมูลเกี่ยวกับการจัดการข้อมูลบุคคลที่เกี่ยวข้องภายในระบบ

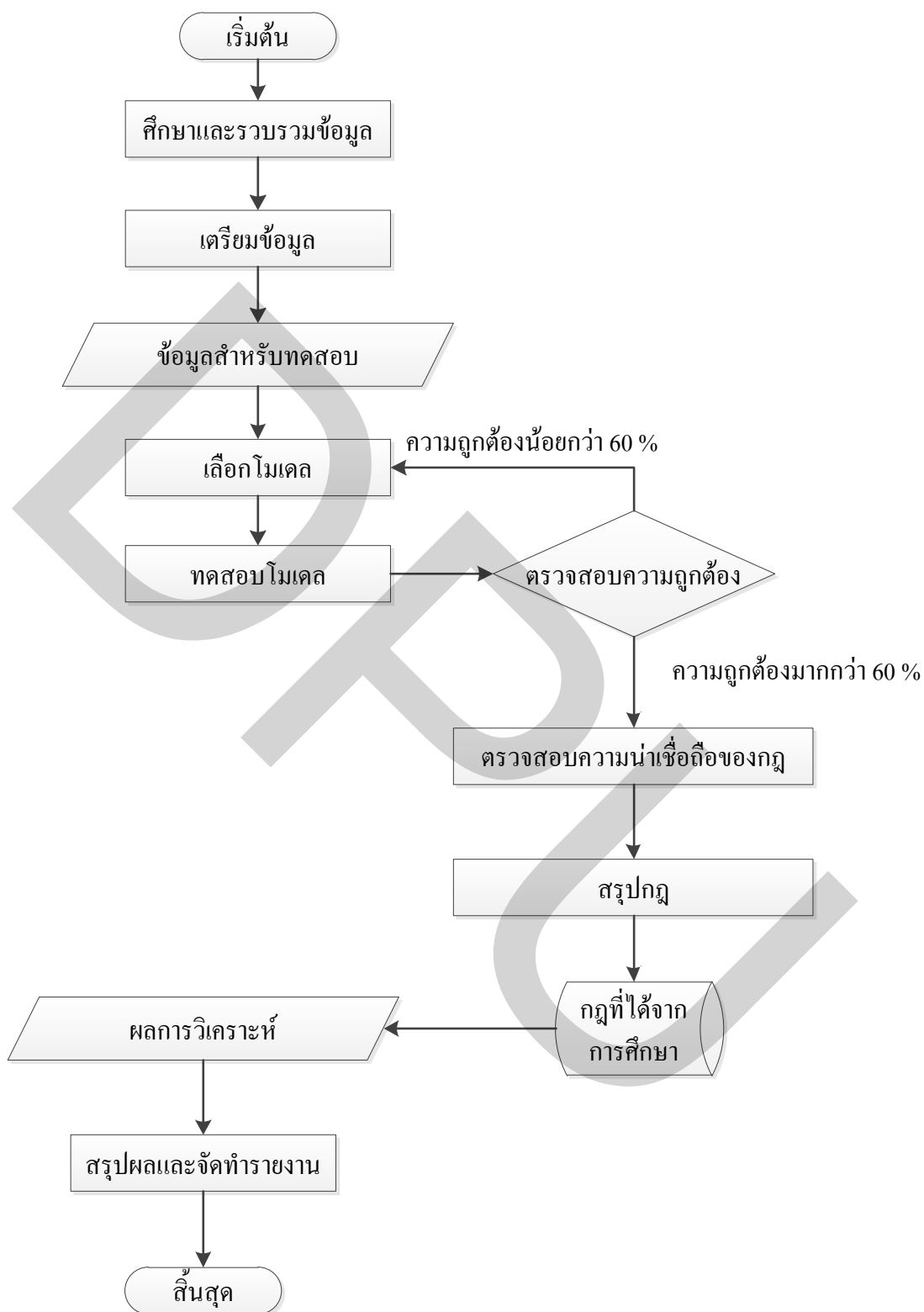
ลำดับ	ชื่อฟิลด์	ประเภท	ขนาด	คำอธิบาย	คีย์
1	id	int	100	ลำดับของบุคคลที่จัดเรียงภายในฐานข้อมูล	PK
2	fullname	Varchar	200	ชื่อ-นามสกุล	
3	username	Varchar	200	ชื่อ log in เข้าสู่ระบบ	
4	email	Varchar	200	ข้อมูลอีเมล	
5	tel	Varchar	45	หมายเลขโทรศัพท์	
6	password	Varchar	200	รหัสผ่านเข้าสู่ระบบ	
7	User_level	tinyint	4	สิทธิของการเข้าใช้งานในระบบ (99=สิทธิสูงสุด,1=สิทธิรองลงมา,2=สิทธิการใช้งานทั่วไป)	
8	date	date	-	วันที่จัดการข้อมูลบุคคล	

3.4 ระยะเวลาในการดำเนินการศึกษาค้นคว้า

ระยะเวลาในการดำเนินการศึกษาค้นคว้าสามารถแสดงขั้นตอนในการดำเนินการศึกษาค้นคว้า สามารถสรุปได้ดังตารางที่ 3.1

ตารางที่ 3.3 แสดงระยะเวลาในการดำเนินการศึกษาค้นคว้า

ที่	ขั้นตอน /ระยะเวลาในการดำเนินงาน(เดือน)	ต.ค. 57	พ.ย. 57	ธ.ค. 57	ม.ค. 58	ก.พ. 58	มี.ค. 58	เม.ย. 58	พ.ค. 58	มิ.ย. 58
1.	การศึกษาและรวบรวมข้อมูล	← →								
2.	การเตรียมข้อมูล		← →							
3.	การกำหนดคุณลักษณะ (Attribute) และค่าที่เป็นไปได้ของคุณลักษณะ				← →					
4.	การแปลงข้อมูลสำหรับการประมวลผล					← →				
5.	การศึกษา การทดสอบ และเลือกเทคนิคที่เหมาะสมนำมาใช้งาน						← →			
6.	การวิเคราะห์ผลและการสรุปผลการทดสอบ						← →			
7.	การพัฒนาระบบสนับสนุนการตัดสินใจช่วยคำนวณคติความให้กับผู้พิพากษา							← →		
8.	สรุปผลการศึกษาและข้อเสนอแนะ							← →		
9.	ดำเนินการเรียบเรียงงานค้นคว้าอิสระ				← →					



ภาพที่ 3.4 แผนผังแสดงขั้นตอนการดำเนินการศึกษา วิเคราะห์ข้อมูลในโปรแกรม Weka

บทที่ 4

ผลการดำเนินการศึกษาค้นคว้า

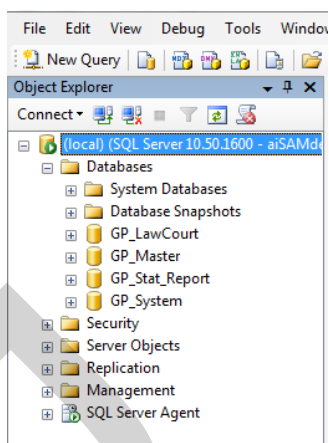
หลังจากที่ได้ศึกษาวิธีการต่างๆ เพื่อนำมาประยุกต์ใช้ในงานการศึกษาค้นคว้าอิสระแล้ว นั้น สามารถอธิบายผลการดำเนินการศึกษาออกเป็นส่วนย่อยได้ 5 ส่วน ดังรายละเอียดนี้

1. การศึกษาและรวบรวมข้อมูล
2. การเตรียมข้อมูล
3. การศึกษา การทดสอบ และเลือกเทคนิคที่เหมาะสม
4. การวิเคราะห์และสรุปผล
5. ขั้นตอนการพัฒนาระบบ

ผลการดำเนินการศึกษาค้นคว้าดังกล่าวสามารถสรุปขั้นตอนหรือกระบวนการทำงาน แสดงได้ดังรายละเอียดต่อไปนี้

4.1 การศึกษาและรวบรวมข้อมูล

การศึกษาโครงสร้างของฐานข้อมูลจากระบบการจัดการฐานข้อมูล Microsoft SQL server 2005 ซึ่งประกอบด้วยฐานข้อมูลจำนวน 4 ฐานข้อมูลซึ่งประกอบไปด้วย GP_LawCourt , GP_Master, GP_Start_Report, GP_System ดังแสดงในภาพที่ 4.1 ส่วนข้อมูลทางคดีที่จัดเก็บภายในระบบบริหารจัดการฐานข้อมูล ชื่อฐานข้อมูล GP_LawCourt จะอยู่ในตาราง thCommand มีจำนวนแอททริบิวต์ 140 แอททริบิวต์ที่จัดเก็บข้อมูลตั้งแต่เดือนตุลาคม 2552-มีนาคม 2557 มีจำนวน 127,055 แถว แต่ในการศึกษารั้งนี้จะใช้วิธีการจำลองข้อมูลให้คล้ายกับฐานข้อมูลจริง เนื่องจากข้อมูลทางคดีนั้นมีความสำคัญ และเป็นข้อมูลที่เป็นความลับ



ภาพที่ 4.1 แสดงรายละเอียดของฐานข้อมูล

4.2 การเตรียมข้อมูล(Data Preparation)

หลังจากที่เราเตรียมข้อมูลคุณลักษณะนั้น พบว่าข้อมูลที่ได้มานั้นเป็นข้อมูลที่ยังไม่สมบูรณ์ที่จะสามารถนำไปใช้งานผ่านกระบวนการศึกษาด้วยเทคนิคต่างๆของเหมืองข้อมูล จึงต้องมีการบริหารจัดการกับข้อมูลที่มีอยู่ โดยเรามีวิธีการจัดเตรียมข้อมูลเบื้องต้น ดังต่อไปนี้

4.2.1 การโอนย้ายข้อมูล (Data Transfer)

ในการโอนย้ายข้อมูลเป็นการโอนข้อมูลเดิมที่ถูกจัดเก็บอยู่ภายในระบบจัดการฐานข้อมูล Microsoft SQL server 2005 ทำให้ไม่สามารถนำข้อมูลที่ได้มาดำเนินการศึกษา จึงได้ดำเนินการ โอนย้ายข้อมูลจากระบบจัดการฐานข้อมูล Microsoft SQL server 2005 ให้มาอยู่ในรูปแบบไฟล์ Microsoft Excel และบันทึกข้อมูลให้อยู่ในรูปแบบไฟล์ .CSV เพื่อให้สามารถนำไปใช้งานได้และสะดวกต่อการนำเข้าไปโปรแกรม Weka ตามภาพที่ 4.2

BlackRecGen	BlackNo	BlackRunNo	PenalNo	RedRunNo	CodeLawCourt	RedNo	Type	TypeCourt	DocumentCode	DocNo	Budget	ReceiveDate	SendDate	Status	Recomm
1	01900000277	๑๗1752/2551	1752/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
2	01900000278	๑๗1753/2550	1753/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
3	01900000279	๑๗1753/2551	1753/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
4	01900000281	๑๗1754/2551	1754/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
5	01900000282	๑๗1755/2550	1755/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
6	01900000283	๑๗1755/2551	1755/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
7	01900000284	๑๗1756/2550	1756/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
8	01900000286	๑๗1757/2550	1757/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
9	01900000287	๑๗1757/2551	1757/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
10	01900000289	๑๗1758/2551	1758/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
11	01900000290	๑๗1759/2550	1759/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
12	01900000291	๑๗1759/2551	1759/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
13	01900000292	๑๗176/2546	176/2546	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
14	01900000293	๑๗176/2547	176/2547	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
15	01900000295	๑๗176/2549	176/2549	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
16	01900000296	๑๗176/2550	176/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
17	01900000298	๑๗1760/2550	1760/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
18	01900000300	๑๗1761/2550	1761/2550	NULL	NULL	019	NULL	152	0	5	NULL	NULL			0
19	01900000302	๑๗1762/2550	1762/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
20	01900000304	๑๗1763/2550	1763/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
21	01900000306	๑๗1764/2550	1764/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
22	01900000307	๑๗1764/2551	1764/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
23	01900000308	๑๗1765/2550	1765/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
24	01900000309	๑๗1765/2551	1765/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
25	01900000311	๑๗1766/2551	1766/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0

BlackRecGen	BlackNo	BlackRunNo	PenalNo	RedRunNo	CodeLawCourt	RedNo	Type	TypeCourt	DocumentCode	DocNo	Budget	ReceiveDate	SendDate	Status	Recomm
1	01900000277	๑๗1752/2551	1752/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
2	01900000278	๑๗1753/2550	1753/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
3	01900000279	๑๗1753/2551	1753/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
4	01900000281	๑๗1754/2551	1754/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
5	01900000282	๑๗1755/2550	1755/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
6	01900000283	๑๗1755/2551	1755/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
7	01900000284	๑๗1756/2550	1756/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
8	01900000286	๑๗1757/2550	1757/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
9	01900000287	๑๗1757/2551	1757/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
10	01900000289	๑๗1758/2551	1758/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
11	01900000290	๑๗1759/2550	1759/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
12	01900000291	๑๗1759/2551	1759/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
13	01900000292	๑๗176/2546	176/2546	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
14	01900000293	๑๗176/2547	176/2547	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
15	01900000295	๑๗176/2549	176/2549	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
16	01900000296	๑๗176/2550	176/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
17	01900000298	๑๗1760/2550	1760/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
18	01900000300	๑๗1761/2550	1761/2550	NULL	NULL	019	NULL	152	0	5	NULL	NULL			0
19	01900000302	๑๗1762/2550	1762/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
20	01900000304	๑๗1763/2550	1763/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
21	01900000306	๑๗1764/2550	1764/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
22	01900000307	๑๗1764/2551	1764/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
23	01900000308	๑๗1765/2550	1765/2550	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0
24	01900000309	๑๗1765/2551	1765/2551	NULL	NULL	019	NULL	151	0	4	NULL	NULL			0

ภาพที่ 4.2 แสดงการโอนย้ายข้อมูลจาก ระบบจัดการฐานข้อมูลมาอยู่ในรูปแบบไฟล์ Microsoft Excel

4.2.2 การลดขนาดข้อมูล (Data Reduction)

เนื่องจากข้อมูลที่เราได้มีปริมาณมากและไม่ได้ได้รับความสะดวกต่อการศึกษา เพื่อให้ได้ข้อมูลที่มีความสมบูรณ์ ถูกต้องและครบถ้วน จึงต้องดำเนินการลดขนาดของข้อมูลลงเพื่อให้สะดวกต่อการนำข้อมูลที่ได้ไปดำเนินการวิเคราะห์ในโปรแกรม Weka และได้ลดขนาดข้อมูลอยู่ 3 ส่วนด้วยกันคือ ส่วนที่ 1 ลดขนาดของจำนวนแอตทริบิวต์ จะได้จำนวนแอตทริบิวต์ ดังตารางที่ 4.1 ส่วนที่ 2 ดำเนินการลดจำนวนข้อมูลแอตทริบิวต์ที่มีข้อมูลไม่ครบถ้วน พบว่ามีข้อมูลบางส่วนที่มีข้อมูลไม่ครบถ้วน ดังนั้น จึงดำเนินการตัดข้อมูลที่ไม่ครบถ้วนออกไปคงเหลือข้อมูลที่ครบถ้วนมาใช้ในการศึกษาด้วยเทคนิคของเหมืองข้อมูลในลำดับต่อไป ส่วนที่ 3 ลดความหลากหลายของ

จำนวนแอตทริบิวต์ ได้แก่ แอตทริบิวต์ Difficulty case แต่เดิมได้จัดแบ่งออกเป็น 3 คุณลักษณะ คือ Easy General และ Difficulty คงเหลือเพียง 2 คุณลักษณะดังตารางที่ 4.1

ตารางที่ 4.1 แสดงคุณลักษณะ (Attribute) ที่ใช้ดำเนินการศึกษา

รายละเอียดของค่าในแต่ละคุณลักษณะ (Attribute)	
คุณลักษณะที่ 1 Reasoning	แบ่งข้อมูลออกเป็น 3 คุณลักษณะ
	Special case management คดีจัดการพิเศษ
	General litigation คดีสามัญ
	Special case คดีพิเศษ
คุณลักษณะที่ 2 Have Relate	แบ่งข้อมูลออกเป็น 2 คุณลักษณะ
	Relate มีคดีเกี่ยวเนื่อง
	Not relate ไม่มีคดีเกี่ยวเนื่อง
คุณลักษณะที่ 3 Consider	แบ่งข้อมูลออกเป็น 6 คุณลักษณะ
	Justice ตัดสิน/พิพากษา
	Sent to justice สมควรส่งให้ภาคตรวจ
	Lawyer requested ขอแรงทนาย
	Psychologists requested มีหนังสือขอนักจิตวิทยา
	High incarceration rate อัตราโทษสูง
	Interpreter requested มีหนังสือขอล่าม
คุณลักษณะที่ 4 Time	แบ่งข้อมูลออกเป็น 3 คุณลักษณะ
	Morning เช้า
	Afternoon บ่าย
	All day ทั้งวัน

ตารางที่ 4.1 (ต่อ)

คุณลักษณะที่ 5 Document	แบ่งข้อมูลออกเป็น 2 คุณลักษณะ
	Complete เอกสารครบ
	Not Complete เอกสารไม่ครบ
คุณลักษณะที่ 6 Gender	แบ่งข้อมูลออกเป็น 2 คุณลักษณะ
	Male ชาย
	Female หญิง
คุณลักษณะที่ 7	แบ่งข้อมูลออกเป็น 2 คุณลักษณะ
	Easy คดีง่าย คดีทั่วไป
	Difficulty คดียาก

4.2.3 การทำความสะอาดข้อมูล (Data Cleaning)

การทำความสะอาดข้อมูลนี้เป็นการแก้ไขค่าของข้อมูลที่มีความหมายเหมือนกัน ให้มีค่าเป็นมาตรฐานเดียวกัน แสดงได้ดังตารางที่ 4.2

ตารางที่ 4.2 แสดงการแปลงค่าของข้อมูล

ที่	สถานะภาพเดิม	สถานะภาพใหม่
1	08.30	Morning
2	09.00	
3	09.30-11.30	
4	13.00	Afternoon
5	13.30	
6	13.00-16.00	
7	08.30-16.00	All day
8	09.00-16.00	

4.3 การศึกษา ทดสอบ และเลือกเทคนิคที่เหมาะสม

เทคนิคการจำแนกกลุ่มข้อมูลนั้นมีหลากหลายเทคนิคด้วยกัน แต่ละเทคนิคก็ให้ผลลัพธ์ในรูปแบบที่แตกต่างกัน ในการศึกษาครั้งนี้ต้องการนำโมเดลที่ได้ไปดำเนินการพัฒนาระบบสนับสนุนการตัดสินใจของผู้บริหารในการตัดสินใจการแบ่งกลุ่มประเภทคดี เพื่อรองรับการจ่ายสำนวนให้กับผู้พิพากษาในการพิจารณาตีความของศาลจังหวัดปทุมธานี ดังนั้น เทคนิคการจำแนกกลุ่มข้อมูลที่น่ามาดำเนินการศึกษาในครั้งนี้จึงเป็นเทคนิคที่ให้ผลลัพธ์ที่เป็นกฎทั้งหมด โดยรายละเอียดการทำงานของแต่ละเทคนิคได้อธิบายไว้ในบทที่ 2 และเทคนิคที่ใช้ศึกษาได้แก่เทคนิค

- 1) เทคนิคการจำแนกกลุ่มแบบต้นไม้ตัดสินใจ (Decisions tree) หรือ J48 ในโปรแกรม Weka
- 2) เทคนิคการจำแนกกลุ่มด้วยวิธีการเรียนรู้แบบอย่างง่าย (Naïve Bayes) ในโปรแกรม Weka

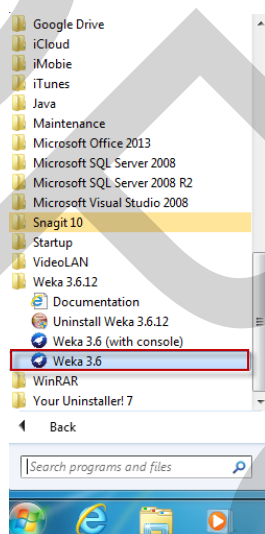
การศึกษานี้ผู้ศึกษาได้ทดสอบโมเดลโดยใช้โปรแกรมประยุกต์ Weka เวอร์ชัน 3.6 ข้อมูลที่ใช้ในการทดสอบมีจำนวน 1,280 แถว ในการทดสอบได้ใช้การทดสอบแบบ 10-fold cross validation ซึ่งได้แก่ การแบ่งข้อมูลออกเป็นจำนวน 10 กลุ่ม โดยในแต่ละรอบของการทดสอบจะนำ

กลุ่มข้อมูลจำนวน 9 กลุ่มเป็นกรณีศึกษา (Training Set) และกลุ่มข้อมูลที่เหลือเป็นกลุ่มทดสอบ (Testing Set) โดยทำการทดสอบซ้ำๆกันจำนวน 10 รอบ เพื่อทำการเปลี่ยนกลุ่มสำหรับการทดสอบให้ครบทุกกลุ่ม

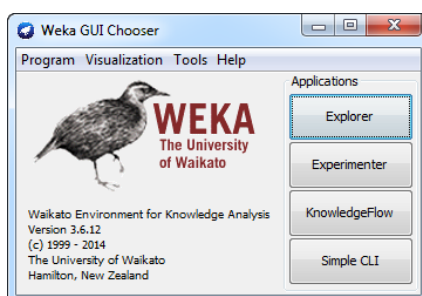
4.3.1 การทดสอบข้อมูล

หลังจากที่ได้ เตรียมข้อมูลสำหรับการทดสอบที่มีความสมบูรณ์ ครบถ้วน และดำเนินการลดความหลากหลายของข้อมูลแล้ว จึงนำข้อมูลดังกล่าว มาทดสอบกับโปรแกรม Weka เวอร์ชัน 3.6 ซึ่งมีขั้นตอนการทำงานดังนี้

ขั้นตอนที่ 1. คลิก (Click) เปิดโปรแกรม Weka เวอร์ชัน 3.6 ขึ้นมา ตามภาพที่ 4.3 และแสดงหน้าต่างโปรแกรม ตามภาพที่ 4.4

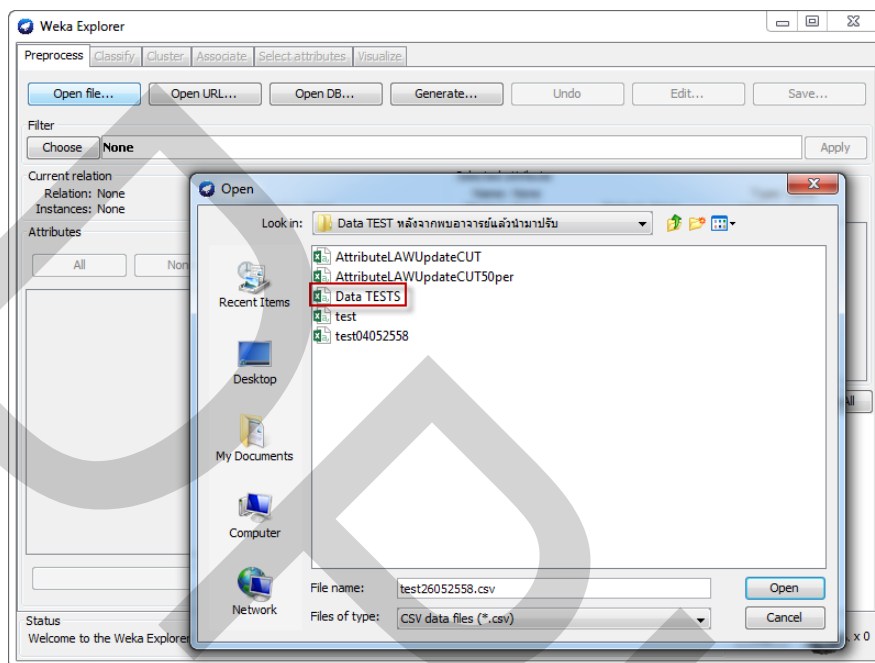


ภาพที่ 4.3 แสดงการเปิดโปรแกรม Weka เวอร์ชัน 3.6

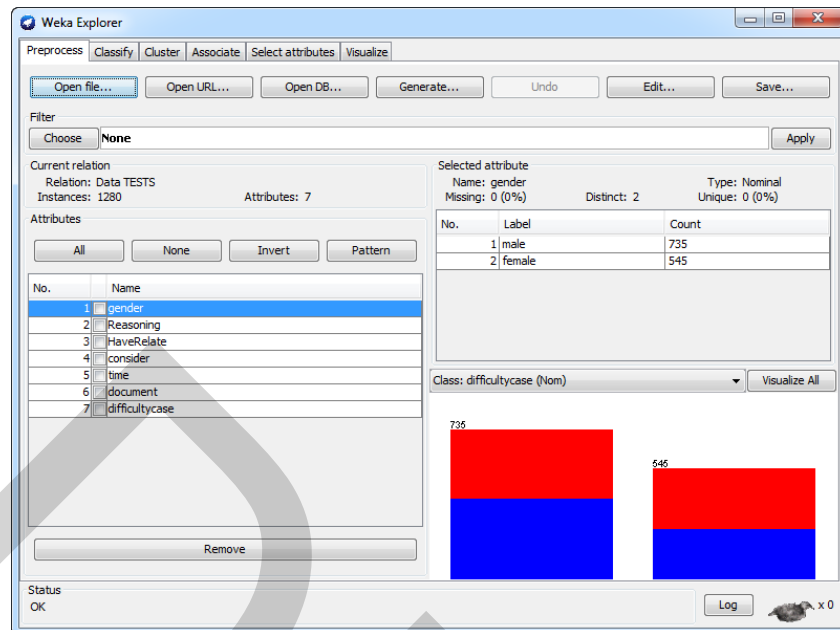


ภาพที่ 4.4 แสดงการเข้าหน้าจอโปรแกรม Weka เวอร์ชัน 3.6

ขั้นตอนที่ 2. นำข้อมูลที่ได้จัดเตรียมไว้ โดยเลือกที่ Applications >> Explorer >> Open file เลือกไฟล์ข้อมูลที่ต้องการนำมาทดสอบ ตามภาพที่ 4.5 และหลังจากนั้น โปรแกรมจะแสดงผลหน้าจอข้อมูล ตามภาพที่ 4.6 ซึ่งหน้าจอนี้จะทำการวิเคราะห์หาจำนวนข้อมูล

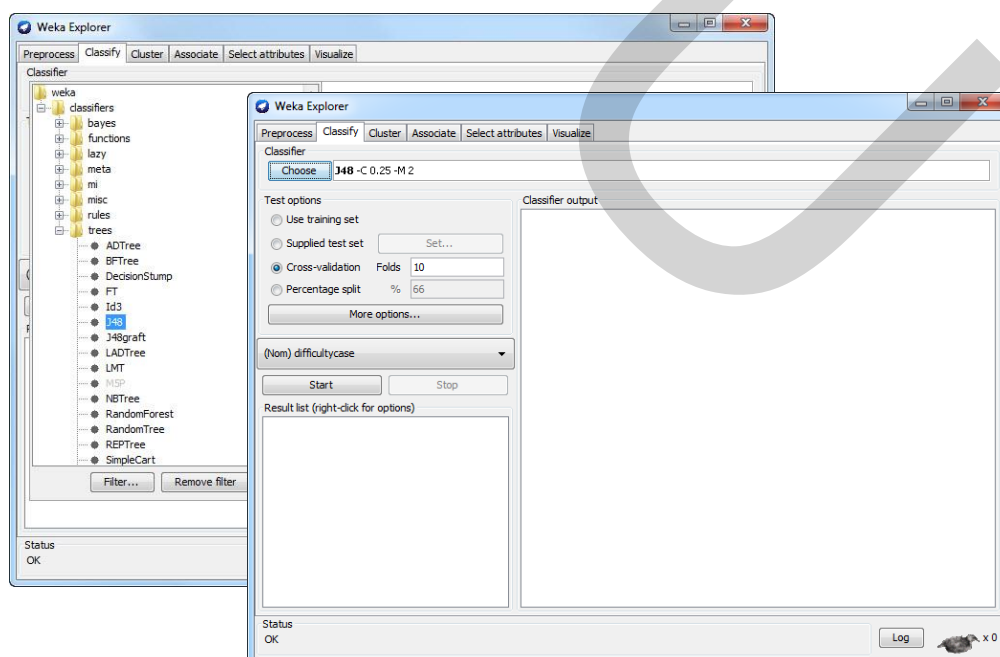


ภาพที่ 4.5 แสดงการนำเข้าข้อมูลใน โปรแกรม Weka เวอร์ชัน 3.6

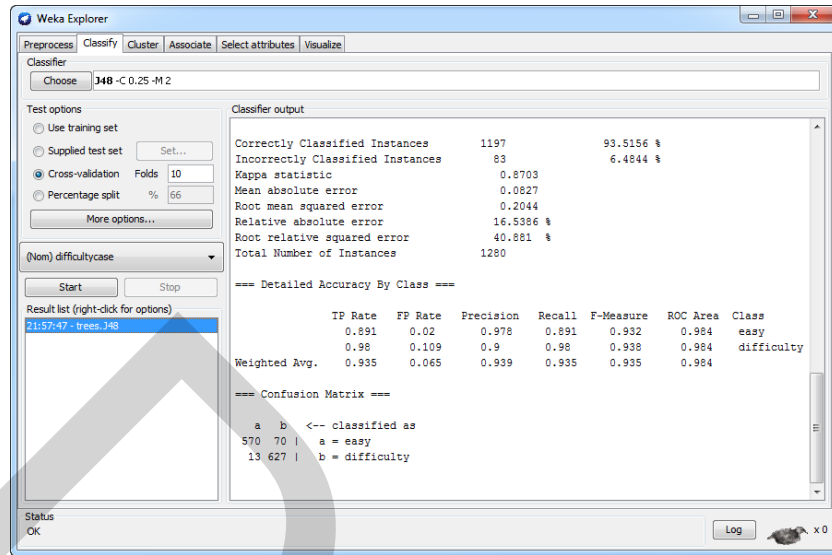


ภาพที่ 4.6 หน้าจอแสดงข้อมูลที่นำเข้าไปในโปรแกรม Weka เวอร์ชัน 3.6

ขั้นตอนที่ 3. ดำเนินการเลือกเทคนิคที่ต้องการ ได้แก่ เทคนิคการจัดกลุ่มข้อมูลโดยเลือก Classification >> Choose >> tree >> เลือกเทคนิคที่ต้องการ ในที่นี้ผู้ศึกษาเลือกใช้เทคนิค J48 จากนั้นให้กดปุ่ม Start ตามภาพที่ 4.7 จะแสดงผลลัพธ์ที่ได้ตามภาพที่ 4.8

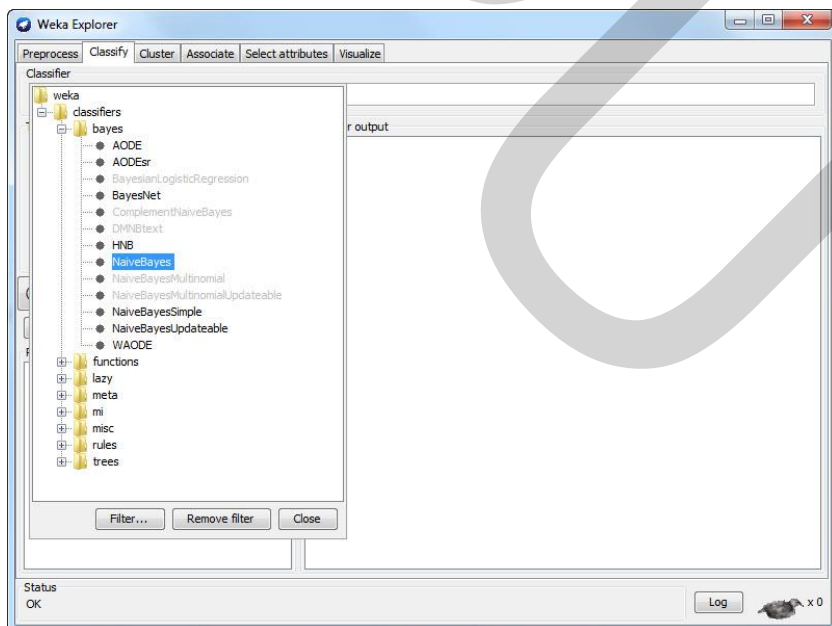


ภาพที่ 4.7 การเลือกเทคนิคที่ใช้ในการศึกษาในการจำแนกข้อมูลแบบ Decision Tree: J48

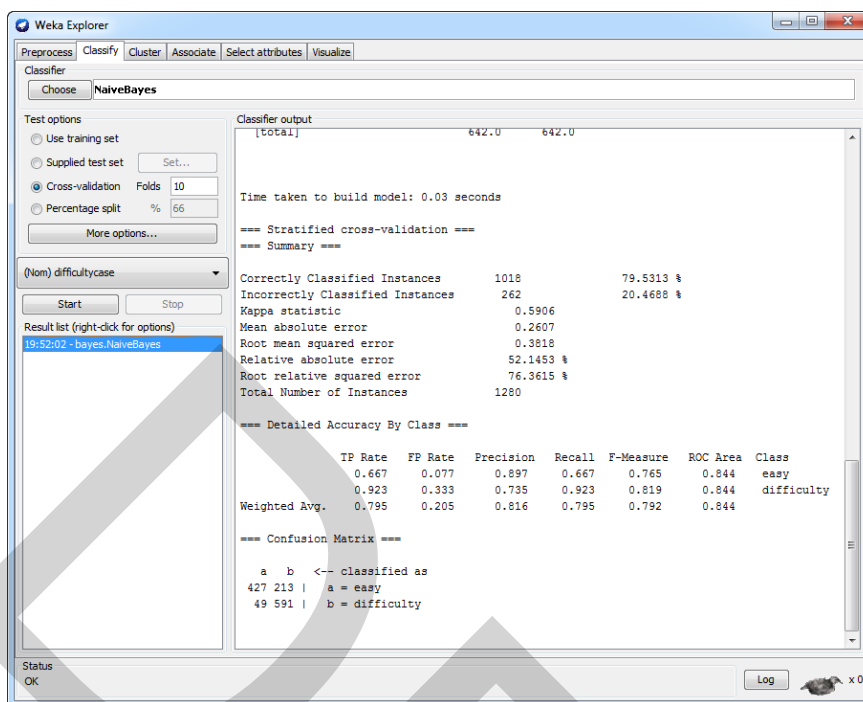


ภาพที่ 4.8 หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Decision Tree: J48

ขั้นตอนที่ 4. ดำเนินการเลือกเทคนิคที่ต้องการ ได้แก่ เทคนิคการจัดกลุ่มข้อมูลโดยเลือก Classification >> Choose >> Bayes >> Naïve Bayes จากนั้นให้กดปุ่ม Start ตามภาพที่ 4.9 จะแสดงผลลัพธ์ที่ได้ตามภาพที่ 4.10



ภาพที่ 4.9 การเลือกเทคนิคที่ใช้ในการศึกษาในการจำแนกข้อมูลแบบ Naive Bayes



ภาพที่ 4.10 หน้าจอผลลัพธ์ของโมเดลการจำแนกข้อมูลแบบ Naive Bayes

4.4 การวิเคราะห์และสรุปผล

4.4.1 ผลการจำแนกกลุ่ม แบบ Decision Tree: J48

จากผลการทดลองพบว่าเทคนิค Decision Tree: J48 ให้ผลลัพธ์การจำแนกประเภทดี ง่าย (Easy) และคดียาก (Difficulty) มีความถูกต้องสูงถึง 93.52 % และได้ถูกจำนวน 15 กฎ ดังภาพที่ 4.11 แสดงผลลัพธ์ที่เป็นกฎในลักษณะของต้นไม้การตัดสินใจที่มีกิ่งแตกออกมา

=== Classifier model (full training set) ===

J48 pruned tree

```

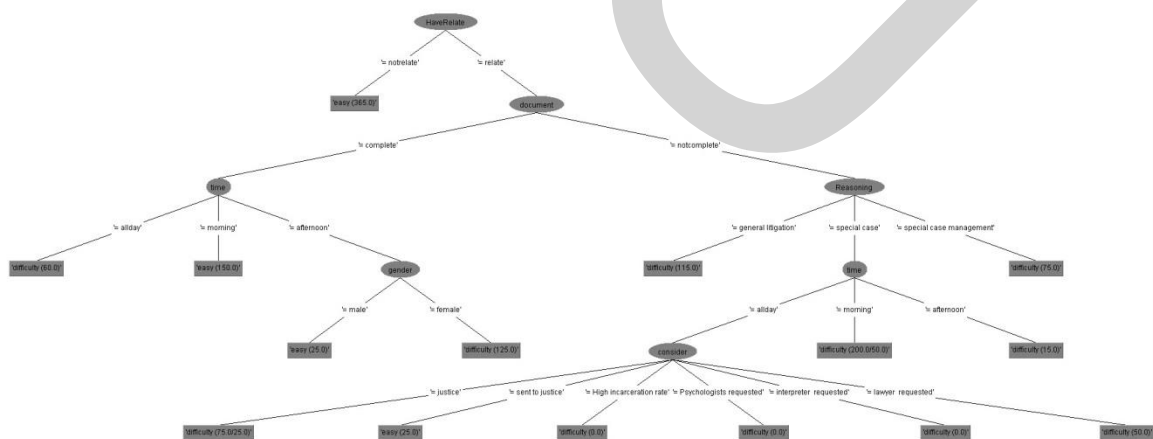
-----

HaveRelate = notrelate: easy (365.0)
HaveRelate = relate
  | document = complete
  | | time = allday: difficulty (60.0)
  | | time = morning: easy (150.0)
  | | time = afternoon
  | | | gender = male: easy (25.0)
  | | | gender = female: difficulty (125.0)
  | | document = notcomplete
  | | Reasoning = general litigation: difficulty (115.0)
  | | Reasoning = special case
  | | | time = allday
  | | | | consider = justice: difficulty (75.0/25.0)
  | | | | consider = sent to justice: easy (25.0)
  | | | | consider = High incarceration rate: difficulty (0.0)
  | | | | consider = Psychologists requested: difficulty (0.0)
  | | | | consider = Interpreter requested: difficulty (0.0)
  | | | | consider = lawyer requested: difficulty (50.0)
  | | | | time = morning: difficulty (200.0/50.0)
  | | | | time = afternoon: difficulty (15.0)
  | | | Reasoning = special case management: difficulty (75.0)

Number of Leaves :    15
Size of the tree :    22

```

ภาพที่ 4.11 ผลลัพธ์จากการจำแนกกลุ่มแบบ Decision Tree: J48



ภาพที่ 4.12 แสดงแผนภาพการจำแนกกลุ่มแบบ Decision Tree: J48

4.4.2 ผลการจำแนกกลุ่ม แบบ Naïve Bayes

จากผลการทดลองพบว่าเทคนิคของ Naïve Bayes ให้ผลลัพธ์การจำแนกคดีย่าง (Easy) และคดียาก (Difficulty) มีความถูกต้องเพียง 79.53 % ซึ่งน้อยกว่าเทคนิคของ Decision Tree: J48 แสดงได้ดังภาพที่ 4.11 ซึ่งวิธีการแบบนี้จะไม่มีกฎออกมาให้ศึกษาทำให้ยากต่อการนำไปใช้งาน ประกอบกับผลลัพธ์ที่ได้จากการทำนายให้ผลลัพธ์ต่ำกว่า Decision Tree: J48 ถึง 13.99%

```

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      1018          79.5313 %
Incorrectly Classified Instances    262           20.4688 %
Kappa statistic                     0.5906
Mean absolute error                 0.2607
Root mean squared error             0.3818
Relative absolute error             52.1453 %
Root relative squared error         76.3615 %
Total Number of Instances          1280

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  ROC Area  Class
          0.667    0.077    0.897     0.667   0.765     0.844    easy
          0.923    0.333    0.735     0.923   0.819     0.844    difficulty
Weighted Avg.   0.795    0.205    0.816     0.795   0.792     0.844

=== Confusion Matrix ===

  a  b  <-- classified as
427 213 | a = easy
 49 591 | b = difficulty

```

ภาพที่ 4.13 ผลลัพธ์จากการจำแนกกลุ่มแบบ Naïve Bayes

ดังนั้น จะใช้เทคนิคของการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้ในการศึกษา เนื่องจากให้ผลลัพธ์ในการทำนายที่สูงประกอบกับมีกฎที่สามารถทำนายได้จำนวน 15 กฎ ซึ่งสามารถนำไปใช้ในการแบ่งกลุ่มได้ตามเงื่อนไขได้ชัดเจนที่สุดและสามารถนำกฎที่ได้สามารถนำไปวิเคราะห์กฎต่อไปได้ (ผลการรันฉบับเต็มอยู่ที่ภาคผนวก) จากภาพที่ 4.11 และภาพที่ 4.12 โดยสามารถจำแนกกฎที่ได้ออกเป็น 4 กลุ่ม ดังต่อไปนี้

กลุ่มที่ 1. กฎที่จำแนกกลุ่มประเภทคดีย่าง (Easy) ได้ถูกต้อง 100 เปอร์เซ็นต์ มีจำนวน 4 กฎ ที่อยู่ในกลุ่มที่มีความน่าเชื่อถือมากที่สุด เพราะสามารถจำแนกกลุ่มตัวอย่างได้ถูกต้องทั้งหมด

IF Have Relate = Not relate THEN Difficulty case= Easy

หมายความว่า ถ้าคดีความไม่มีประเภทคดีที่เกี่ยวข้องกัน ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่ง่าย

IF Have Relate = Relate AND Document = Complete AND Time = Morning THEN Difficulty case = Easy

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารในการประกอบพิจารณามีความครบถ้วนสมบูรณ์และเวลาในการพิจารณาคดีอยู่ในช่วงเช้า ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่ง่าย

IF Have Relate = Relate AND Document = Complete AND Time = Afternoon AND Gender = Male THEN Difficulty case = Easy

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารในการประกอบพิจารณามีความครบถ้วนสมบูรณ์และเวลาในการพิจารณาคดีอยู่ในช่วงบ่ายและคดีส่วนใหญ่เกิดขึ้นกับเพศชาย ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่ง่าย

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Sent to justice THEN Difficulty case = Easy

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารในการประกอบพิจารณาไม่ครบถ้วน และประเภทคดีส่วนใหญ่เป็นคดีจัดการพิเศษและเวลาในการพิจารณาคดีได้พิจารณาคดีตลอดทั้งวันและคดีส่วนใหญ่มีผลการพิจารณานั้นพบว่าส่งเอกสารให้ภาคเป็นคนตรวจ ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่ง่าย

กลุ่มที่ 2 กฎที่จำแนกกลุ่มประเภทคดียาก (Difficulty) 1 ได้ถูกตั้ง 100 เปอร์เซ็นต์ มีจำนวน 6 กฎ ที่อยู่ในกลุ่มที่มีความน่าเชื่อถือมากที่สุด เพราะสามารถจำแนกกลุ่มตัวอย่างได้ถูกต้องทั้งหมด

IF Have Relate = Relate AND Document = Complete AND Time = All day THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาครบถ้วนสมบูรณ์และเวลาในการพิจารณามีการพิจารณาทั้งวัน ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Complete AND Time = Afternoon AND Gender = Female THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาครบถ้วนสมบูรณ์และเวลาในการพิจารณามีการพิจารณาอยู่ในช่วงบ่ายและคดีส่วนใหญ่เกิดขึ้นกับแพทหญิง ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = General litigation THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีสามัญ ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case management THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีจัดการพิเศษ ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = Afternoon THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่อยู่ในช่วงบ่าย ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Lawyer requested THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่มีการพิจารณาคดีทั้งวันและผลการพิจารณานั้นได้จัดทำหนังสือขอแรงทนายความ ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

กลุ่มที่ 3 กฎที่จำแนกกลุ่มประเภทคดียาก (Difficulty) ได้ถูกตั้งน้อยกว่า 100 เเปอร์เซ็นต์ มีจำนวน 2 กฎ คือ

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = Morning THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่อยู่ในช่วงเช้า ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Justice THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่มีการพิจารณาคดีทั้งวันและผลการพิจารณานั้นได้มีการตัดสินหรือมีคำพิพากษา ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

กลุ่มที่ 4 กฎที่จำแนกกลุ่มประเภทคดียาก (Difficulty) ที่ไม่มีตัวอย่างสนับสนุน มีจำนวน 3 กฎ คือ

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = High incarceration rate THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่มีการพิจารณาคดีทั้งวันและผลการพิจารณานั้นมีอัตราโทษสูง ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Psychologists requested THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่มีการพิจารณาคดีทั้งวันและผลการพิจารณานั้นได้มีหนังสือขอนักจิตวิทยา ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Interpreter requested THEN Difficulty case = Difficulty

หมายความว่า ถ้าคดีความมีประเภทคดีที่มีเกี่ยวเนื่องกันและเอกสารประกอบการพิจารณาไม่ครบถ้วนและประเภทคดีส่วนใหญ่เป็นคดีพิเศษและเวลาในการพิจารณาคดีส่วนใหญ่มีการพิจารณาคดีทั้งวันและผลการพิจารณานั้นได้มีหนังสือขอถาม ผลการพิจารณานั้นพบว่าประเภทคดีความนั้นจะเป็นประเภทคดีที่มีความยุ่งยาก ซับซ้อน

4.5 ขั้นตอนการพัฒนาระบบ

หลังจากที่เราได้เลือกเทคนิคของการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้งานในขั้นตอนต่อมาจะเข้าสู่ในส่วนของการพัฒนา ระบบ โปรแกรมตรวจสอบข้อมูลทางคดีของศาลจังหวัดปทุมธานี โดยเป็นการนำกฎที่ได้จากการทำเหมืองข้อมูลประเภทการจำแนกกลุ่มแบบ Decision Tree: J48 มาใช้เป็นเงื่อนไขในการพัฒนา โปรแกรมตรวจสอบข้อมูลทางคดี ของศาลจังหวัดปทุมธานี เพื่อเป็นเครื่องมือช่วยในการตัดสินใจของผู้บริหารในการแบ่งกลุ่มประเภทคดีง่ายหรือประเภทคดียาก โดยใช้ภาษา PHP ในการพัฒนา และ ระบบฐานข้อมูลที่ใช้คือ My SQL และใช้ phpMyAdmin เป็นเครื่องมือสำหรับใช้จัดการฐานข้อมูล

4.5.1 ออกแบบระบบฐานข้อมูล

ในส่วนของ การออกแบบระบบฐานข้อมูลจะประกอบไปด้วยจำนวนตารางอยู่จำนวน 2 ตารางคือ data_case และ ตาราง useraccount ที่ได้กล่าวไว้ในตารางที่ 3.1 และตารางที่ 3.2 ไปนั้น สามารถสรุปความสัมพันธ์ของตารางได้ดังนี้

Data_case

<u>id</u>	gender	reasoning	Have_rate	time	document	consider	<u>Id_useraccount</u>
-----------	--------	-----------	-----------	------	----------	----------	-----------------------

PK

FK

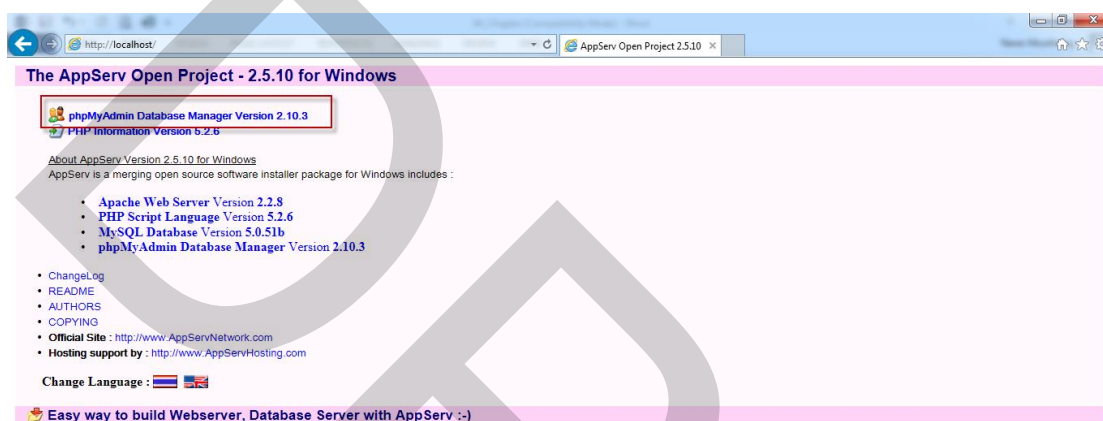
useraccount

<u>id</u>	fullname	username	email	tel	password	User_level	date
-----------	----------	----------	-------	-----	----------	------------	------

PK

หลังจากที่ได้ทำการออกแบบตารางความสัมพันธ์เป็นที่เรียบร้อยแล้ว หลังจากนั้นจะแสดงวิธีการติดตั้งโปรแกรมจัดการฐานข้อมูลหรือดำเนินการสร้างฐานข้อมูลที่ได้ออกแบบไว้ โดยสามารถแสดงขั้นตอนการทำงานได้ดังต่อไปนี้

ขั้นตอนที่ 1 ให้เราเปิด Browser เช่น IE Chrome Firefox ขึ้นมาแล้วพิมพ์ 127.0.0.1 หรือพิมพ์คำว่า http://localhost/ ที่ช่อง URL จากนั้นให้คลิกที่ข้อความ phpMyAdmin Database Manager ตามภาพที่ 4.14



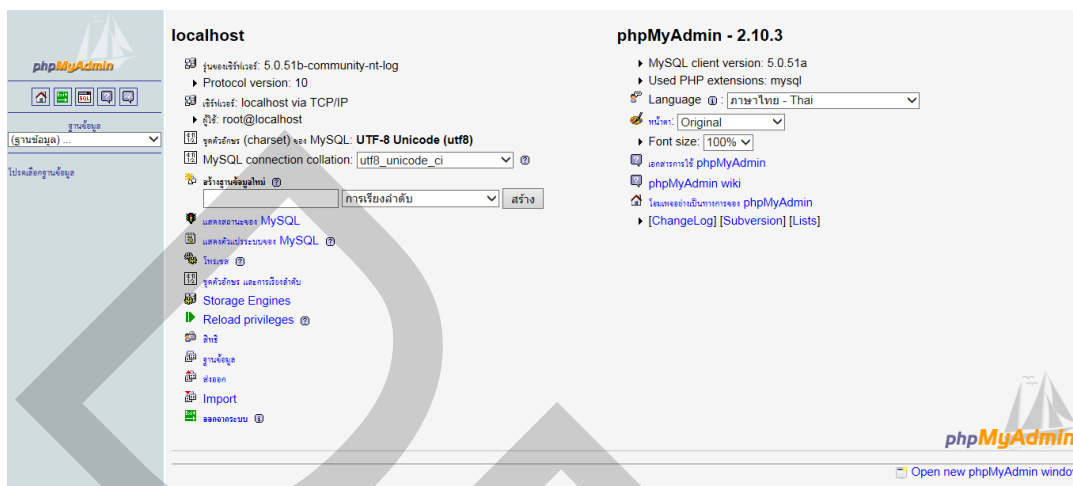
ภาพที่ 4.14 แสดงขั้นตอนการใช้งาน phpMyAdmin Database Manager

ขั้นตอนที่ 2 เมื่อคลิกที่ PhpMyAdmin Database Manager จะมีหน้าต่างให้ใส่ ชื่อผู้ใช้ และ รหัสผ่าน ปรากฏขึ้นมาในช่องชื่อผู้ใช้ (Username) ให้ใส่คำว่า root ส่วนในช่องรหัสผ่าน (password) ให้ใส่รหัสผ่านที่เราสร้างไว้เมื่อตอนติดตั้ง Appserv ซึ่งก็คือ 07303126 จากนั้นกดปุ่มตกลง (OK)



ภาพที่ 4.15 แสดงการป้อน Username และ Password เพื่อเข้าสู่ระบบ

ขั้นตอนที่ 3 ถ้าล็อกอินเข้าสู่ระบบได้เป็นที่เรียบร้อย ระบบจะนำเข้าสู่หน้าต่างของโปรแกรม phpMyAdmin แสดงได้ดังรูปที่ 4.16



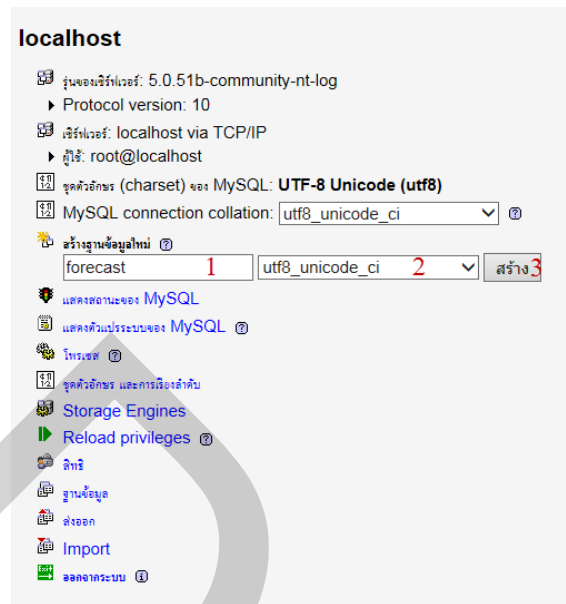
ภาพที่ 4.16 แสดงหน้าต่างของ phpMyAdmin

ขั้นตอนที่ 4 ให้ดำเนินการสร้างฐานข้อมูลชื่อ forecast ประกอบด้วย 2 ตาราง คือ ตาราง data_case และ ตาราง useraccount เพื่อเก็บรายละเอียดข้อมูลที่บันทึก ตัวอย่างของตารางที่จัดเก็บข้อมูลปรากฏอยู่ในตารางที่ 3.1 และตารางที่ 3.2 ของบทที่ 3

โดย หมายเลข 1 คือ ช่องสำหรับใส่ชื่อฐานข้อมูลที่เราจะสร้าง ในที่นี้จะสร้างฐานข้อมูลชื่อ Forecast ให้ดำเนินการพิมพ์ชื่อฐานข้อมูลลงไปในช่วงที่ 1

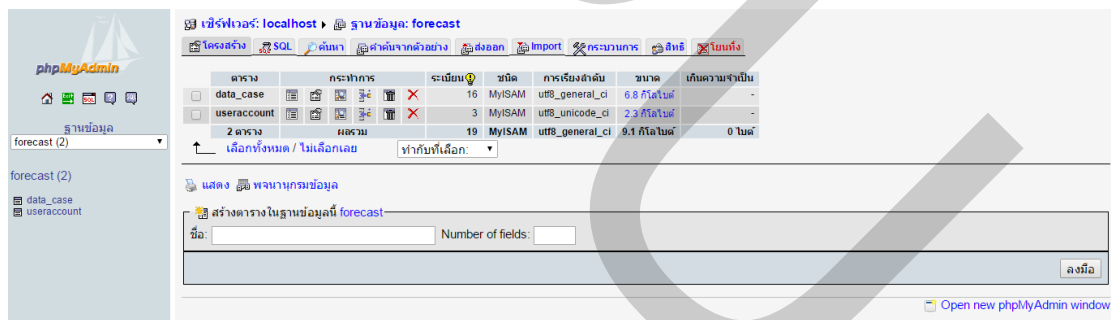
หมายเลข 2 คือ ช่องสำหรับเปลี่ยนรหัสอักขระ ให้คลิกเลือกเป็น utf8_unicode_ci

หมายเลข 3 คือ ปุ่ม สร้าง เพื่อสร้างฐานข้อมูล ตามภาพที่ 4.17



ภาพที่ 4.17 แสดงขั้นตอนการสร้างฐานข้อมูล

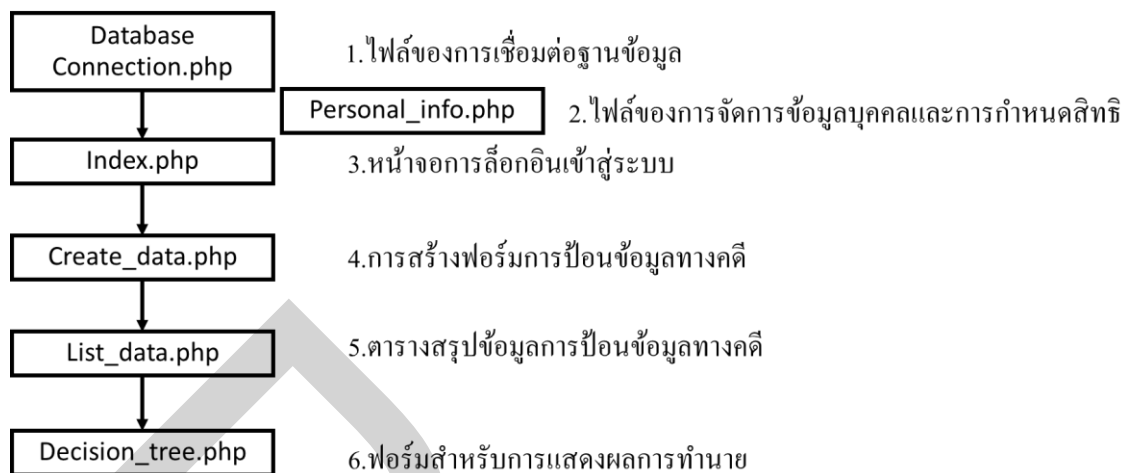
ขั้นตอนที่ 5 หลังจากที่เรสร้างฐานข้อมูลเสร็จในขั้นตอนที่ 4 ระบบจะแสดงผลการสร้างฐานข้อมูล ตามภาพที่ 4.18 จะประกอบไปด้วย 2 ตาราง คือตาราง data_case และ ตาราง useraccount



ภาพที่ 4.18 แสดงผลการสร้างฐานข้อมูล

4.5.2 การออกแบบหน้าจอ

ในการออกแบบหน้าจอโปรแกรมตรวจสอบข้อมูลทางคดี ของศาลจังหวัดปทุมธานี เพื่อดำเนินการบันทึกข้อมูลรายละเอียดต่างๆที่ใช้ในการวิเคราะห์ข้อมูลมีหน้าเว็บเพจที่เกี่ยวข้อง จำนวน 6 เว็บเพจ สามารถแสดงได้โดยแผนภาพ ดังรูปที่ 4.19




ภาพที่ 4.19 แสดง โครงสร้างหน้าเว็บเพจของโปรแกรมตรวจสอบข้อมูลทางคดี



ภาพที่ 4.20 หน้าจอแสดงการล็อกอินเข้าสู่ระบบ

โปรแกรมตรวจสอบข้อมูลทางคดีศาลจังหวัดปทุมธานี User Account manager1

หน้าแรก **บันทึกข้อมูลทางคดี** รายการข้อมูลทางคดี

 โปรแกรมตรวจสอบข้อมูลทางคดี ศาลจังหวัดปทุมธานี
Pathumthani Provincial Court

+ บันทึกข้อมูลทางคดี
กรุณากรอกข้อมูลในแบบฟอร์มให้ครบถ้วน

แบบฟอร์มบันทึกข้อมูลทางคดี

เพศ (Gender) ชาย(Male) หญิง(Female)

ประเภทคดี (Reasoning) คดีจัดการพิเศษ (Special case management)
 คดีสามัญ (General Litigation)
 คดีพิเศษ (Special Case)

ความเกี่ยวเนื่องของคดี (Have Relate) คดีมีความเกี่ยวเนื่อง(Relate)
 คดีไม่มีความเกี่ยวเนื่อง(Not relate)

เวลา(Time) เช้า(Morning)
 บ่าย(Afternoon)
 ทั้งวัน(All day)

ความครบถ้วนของเอกสาร (Document) เอกสารครบ (Complete)
 เอกสารไม่ครบ (NOT complete)

ผลการพิจารณา (Consider) นัดสืบ/พิพากษา (Justice)
 ส่งไปศาลตรวจ (Sent to justice)
 อัตราโทษสูง (High incarceration rate)
 มีหนังสือขอนักจิตวิทยา (Psychologists requested)
 มีหนังสือขอล่าม (Interpreter requested)
 ขอแรงทนาย (Lawyer requested)

[บันทึกข้อมูล](#)

ภาพที่ 4.21 หน้าจอบันทึกข้อมูลทางคดีที่ต้องการวิเคราะห์

โปรแกรมตรวจสอบข้อมูลทางคดีศาลจังหวัดปทุมธานี User Account manager1

หน้าแรก + รายการข้อมูลทางคดี

โปรแกรมตรวจสอบข้อมูลทางคดี ศาลจังหวัดปทุมธานี Pathumthani Provincial Court

รายการข้อมูลทางคดี

แสดงข้อมูลทางคดีในฐานข้อมูล

ID	เพศ(GENDER)	ประเภทคดี (REASONING)	ความเกี่ยวข้องของคดี (HAVE RELATE)	เวลา(TIME)	ความครบถ้วนของเอกสาร (DOCUMENT)	ผลการพิจารณา (CONSIDER)	ตรวจสอบข้อมูล
48	หญิง(Female)	คดีพิเศษ (Special Case)	คดีมีความเกี่ยวเนื่อง(Relate)	ทั้งวัน(All day)	เอกสารครบ (Complete)	อัตราโทษสูง (High incarceration rate)	Decisions tree:J48 ลบข้อมูล
47	ชาย(Male)	คดีจัดการพิเศษ (Special case management)	คดีมีความเกี่ยวเนื่อง(Relate)	เช้า(Morning)	เอกสารครบ (Complete)	อัตราโทษสูง (High incarceration rate)	Decisions tree:J48 ลบข้อมูล
46	ชาย(Male)	คดีจัดการพิเศษ (Special case management)	คดีมีความเกี่ยวเนื่อง(Relate)	เช้า(Morning)	เอกสารครบ (Complete)	ส่งไปศาลตรวจ (Sent to justice)	Decisions tree:J48 ลบข้อมูล
45	หญิง(Female)	คดีสามัญ (General Litigation)	คดีมีความเกี่ยวเนื่อง(Relate)	บ่าย(Afternoon)	เอกสารไม่ครบ (Not complete)	มีหนังสือขอล่าม (Interpreter requested)	Decisions tree:J48 ลบข้อมูล
44	ชาย(Male)	คดีจัดการพิเศษ (Special case management)	คดีไม่มีความเกี่ยวเนื่อง(Not relate)	บ่าย(Afternoon)	เอกสารครบ (Complete)	มีหนังสือขอจิตวิทยา (Psychologists requested)	Decisions tree:J48 ลบข้อมูล
43	หญิง(Female)	คดีพิเศษ (Special Case)	คดีไม่มีความเกี่ยวเนื่อง(Not relate)	ทั้งวัน(All day)	เอกสารไม่ครบ (Not complete)	ตัดสินพิพากษา (Justice)	Decisions tree:J48 ลบข้อมูล
42	ชาย(Male)	คดีพิเศษ (Special Case)	คดีมีความเกี่ยวเนื่อง(Relate)	ทั้งวัน(All day)	เอกสารไม่ครบ (Not complete)	ตัดสินพิพากษา (Justice)	Decisions tree:J48 ลบข้อมูล
41	หญิง(Female)	คดีจัดการพิเศษ (Special case management)	คดีมีความเกี่ยวเนื่อง(Relate)	เช้า(Morning)	เอกสารครบ (Complete)	ตัดสินพิพากษา (Justice)	Decisions tree:J48 ลบข้อมูล
40	ชาย(Male)	คดีพิเศษ (Special Case)	คดีมีความเกี่ยวเนื่อง(Relate)	ทั้งวัน(All day)	เอกสารไม่ครบ (Not complete)	อัตราโทษสูง (High incarceration rate)	Decisions tree:J48 ลบข้อมูล
39	หญิง(Female)	คดีสามัญ (General Litigation)	คดีมีความเกี่ยวเนื่อง(Relate)	บ่าย(Afternoon)	เอกสารไม่ครบ (Not complete)	ส่งไปศาลตรวจ (Sent to justice)	Decisions tree:J48 ลบข้อมูล

Total 11 Record : 2 Page : 1 [2] Next>>

โปรแกรมตรวจสอบข้อมูลทางคดีศาลจังหวัดปทุมธานี - Google Chrome

localhost/forecast/decision_tree.php?case_id=48

การตรวจสอบข้อมูลโดยวิธีต้นไม้ตัดสินใจ (Decisions tree:J48)

ผลการตรวจสอบ :

คดียาก(Difficulty)

โปรแกรมตรวจสอบข้อมูลทางคดีศาลจังหวัดปทุมธานี - Google Chrome

localhost/forecast/decision_tree.php?case_id=47

การตรวจสอบข้อมูลโดยวิธีต้นไม้ตัดสินใจ (Decisions tree:J48)

ผลการตรวจสอบ :

คดีง่าย(Easy)

ภาพที่ 4.22 หน้าจอแสดงผลการวิเคราะห์



โปรแกรมตรวจสอบข้อมูลทางคดี ศาลจังหวัดปทุมธานี
Pathumthani Provincial Court

+ เพิ่มผู้ใช้ใหม่

กรุณากรอกข้อมูลผู้ใช้ให้ครบถ้วนก่อนทำการลงทะเบียน

ชื่อ-สกุล:

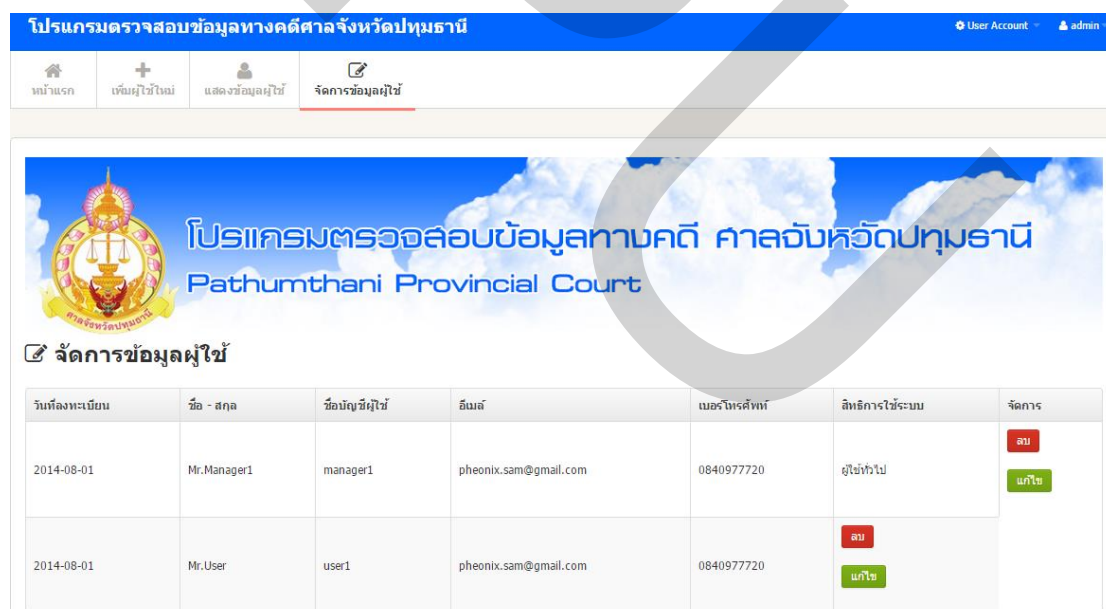
ชื่อปญญี่ผู้ใช้:

อีเมล:

เบอร์โทรศัพท์:

รหัสผ่าน:

ภาพที่ 4.23 หน้าจอของการจัดการเพิ่มข้อมูลบุคคล



โปรแกรมตรวจสอบข้อมูลทางคดีศาลจังหวัดปทุมธานี User Account admin

หน้าแรก + เพิ่มผู้ใช้ใหม่ แสดงข้อมูลผู้ใช้ จัดการข้อมูลผู้ใช้

จัดการข้อมูลผู้ใช้

วันที่ลงทะเบียน	ชื่อ - สกุล	ชื่อปญญี่ผู้ใช้	อีเมล	เบอร์โทรศัพท์	สิทธิ์การใช้ระบบ	จัดการ
2014-08-01	Mr.Manager1	manager1	pheonix.sam@gmail.com	0840977720	ผู้ใช้ทั่วไป	ลบ แก้ไข
2014-08-01	Mr.User	user1	pheonix.sam@gmail.com	0840977720	ลบ แก้ไข	

ภาพที่ 4.24 แสดงข้อมูลบุคคลผู้มีสิทธิใช้งานในระบบ

4.5.3 การเขียนโปรแกรม

เมื่อดำเนินการออกแบบหน้าจอและการแสดงผลที่ได้ เราจะใช้กฎที่ได้จากการทำเหมืองข้อมูลประเภทจำแนกกลุ่มแบบ Decision Tree: J48 มาเป็นเงื่อนไขในการเขียนโปรแกรม และจากกฎที่เราได้จำนวน 15 กฎนั้น จะทำการแปลงกฎที่ได้มาเป็น โค้ด (Code) ของตัวโปรแกรม โดยหนึ่งกฎที่ได้ คือหนึ่งเงื่อนไข ดังนั้นเราจะได้เงื่อนไขในการทำนายจำนวน 15 เงื่อนไข มาใช้ในการพัฒนาโปรแกรม และในส่วนนี้จะแสดงโค้ดโปรแกรมในส่วนของเงื่อนไขการทำนาย ดังตัวอย่างข้างล่างนี้พร้อมประกอบคำอธิบาย

ตัวอย่างเงื่อนไขในการเขียนโปรแกรมในส่วนของการทำนายผลลัพธ์

```

1 <?php
2
3 /* INCLUDE CONFIG DATABASE , HEADER */
4 include("config/database_connection.php");
5
6
7 /* PHP SCRIPT */
8
9 echo "<h1>การตรวจสอบข้อมูลโดยวิธีต้นไม้ตัดสินใจ<br>(Decisions tree:J48)</h1><br>";
10 echo "<h2>ผลการตรวจสอบ : </h2>";
11 $case_id = $_REQUEST['case_id'];
12 $sql = "select * from data_case where id = '$case_id'";
13 $data = mysql_fetch_array(mysql_query($sql));
14
15
16 /* 1.If Have Relate=not relate THEN Difficulty case= Easy*/
17 if ($data['have_relate'] == 'คดีไม่มีความเกี่ยวเนื่อง(Not relate)')
18 {
19 echo "<h2><font color=#33CC00>คดีง่าย(Easy)</font></h2>";
20 exit();
21 }else
22
23 /*2.If Have Relate = Relate AND Document = Complete AND Time = All day THEN Difficulty case = Difficulty*/
24 if ($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารครบ (Complete)' &&$data['time'] == 'ทั้งวัน(All day)')
25 {
26 echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
27 exit();
28 }else
29
30 /* 3. If Have Relate = Relate AND Document = Complete AND Time = Morning THEN Difficulty case = Easy*/
31 if ($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารครบ (Complete)' &&$data['time'] == 'เช้า(Morning)')
32 {
33 echo "<h2><font color=#33CC00>คดีง่าย(Easy)</font></h2>";
34 exit();
35 }else
36
37 /*4.If Have Relate = Relate AND Document = Complete AND Time = Afternoon AND Gender = Male THEN Difficulty case = Easy */
38 if ($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารครบ (Complete)' &&$data['time'] == 'บ่าย(Afternoon)'
&&$data['gender'] == 'ชาย(Male)')
39 {
40 echo "<h2><font color=#33CC00>คดีง่าย(Easy)</font></h2>";
41 exit();
42 }else
43
44 /* 5.If Have Relate = Relate AND Document = Complete AND Time = Afternoon AND Gender = Female THEN Difficulty case = Difficulty*/
45 if ($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารครบ (Complete)' &&$data['time'] == 'บ่าย(Afternoon)'
&&$data['gender'] == 'หญิง(Female)')
46 {
47 echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
48 exit();
49 }else
50
51 /* 6. If Have Relate = Relate AND Document = Not Complete AND Reasoning = General litigation THEN Difficulty case = Difficulty*/
52 if ($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีสามัญ
(General Litigation)')
53 {
54 echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
55 exit();
56 }else

```

```

57
58 /* 7. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case management THEN Difficulty case = Difficulty*/
59 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] ==
    'คดีจัดการพิเศษ (Special case management)')
60 {
61     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
62     exit();
63 }else
64
65 /* 8. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = Morning THEN Difficulty case = Difficulty*/
66 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'เช้า(Morning)')
67 {
68     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
69     exit();
70 }else
71
72 /* 9. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = Afternoon THEN Difficulty case = Difficulty*/
73 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'บ่าย(Afternoon)')
74 {
75     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
76     exit();
77 }else
78
79 /* 10. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Justice THEN Difficulty case =
    Difficulty*/
80 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'ทั้งวัน(All day)' &&$data['consider'] == 'ตัดสิน/พิพากษา (Justice)')
81 {
82     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
83     exit();
84 }else
85
86 /* 11. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Sent to justice THEN Difficulty case =
    Easy*/
87 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'ทั้งวัน(All day)' &&$data['consider'] == 'ส่งให้ภาคตรวจ (Sent to justice)')
88 {
89     echo "<h2><font color=#33CC00>คดีง่าย(Easy)</font></h2>";
90     exit();
91 }else
92
93 /* 12. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = High incarceration rate THEN
    Difficulty case = Difficulty*/
94 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'ทั้งวัน(All day)' &&$data['consider'] == 'อัตราโทษสูง (High incarceration rate)')
95 {
96     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
97     exit();
98 }else
99
100 /* 13. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Psychologists requested THEN
    Difficulty case = Difficulty*/
101 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'ทั้งวัน(All day)' &&$data['consider'] == 'มีหนังสือขอนักจิตวิทยา (Psychologists requested)')
102 {
103     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
104     exit();
105 }else
106
107 /* 14. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Interpreter requested THEN Difficulty
    case = Difficulty*/
108 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] == 'คดีพิเศษ
    (Special Case)'&&$data['time'] == 'ทั้งวัน(All day)' &&$data['consider'] == 'มีหนังสือขอล่าม (Interpreter requested)')
109 {
110     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
111     exit();
112 }else
113
114 /* 15. IF Have Relate = Relate AND Document = Not Complete AND Reasoning = Special case AND Time = All day AND Consider = Lawyer requested THEN Difficulty
    case = Difficulty*/
115 if($data['have_relate'] == 'คดีมีความเกี่ยวเนื่อง(Relate)' &&$data['document'] == 'เอกสารไม่ครบ (Not complete)' &&$data['reasoning'] ==
    'คดีพิเศษ(Special Case)' &&$data['time'] == 'ทั้งวัน(All day)' &&$data['consider'] == 'ขอเจรทนาย (Lawyer requested)')
116 {
117     echo "<h2><font color=#FF0000>คดียาก(Difficulty)</font></h2>";
118 }else
119 {
120     echo "<h2><font color=#33CC00>คดีง่าย(Easy)</font></h2>";
121 }
122 ?>

```

ภาพที่ 4.25 แสดงเงื่อนไขในการเขียนโปรแกรมในส่วนของการทำนายผลลัพธ์

คำอธิบายโค้ดโดยภาพรวม

บรรทัดที่ 1 เริ่มต้นการใช้งานภาษา PHP

บรรทัดที่ 4 คือ การเริ่มต้นในการเชื่อมต่อฐานข้อมูลที่ได้มีการสร้างไฟล์ไว้ที่ชื่อว่า database_connection.php

บรรทัดที่ 9-10 เป็นคำสั่ง echo หมายความว่า เป็นการแสดงผลข้อความทางจอภาพที่เราสามารถแทรกชื่อตัวแปรเข้าไปอยู่ในข้อความ ซึ่งเราต้องใช้เครื่องหมาย “ ” (Double code) ไว้

บรรทัดที่ 11 เป็นการตั้งชื่อตัวแปรแบบอาร์เรย์ เพื่อมารองรับค่า ในที่นี้จะรองรับค่าที่ส่งมาคือ case_id

บรรทัดที่ 12 เป็นการใช้คำสั่ง SELECT ข้อมูลจากฐานข้อมูลมาแสดงผล โดยมีการอ้างอิงข้อมูลจากหมายเลข case_id

บรรทัดที่ 13 คือการคืนค่า ค่าข้อมูล ของผลลัพธ์ในแถวที่ชื่ออยู่ และเก็บไว้ที่ array และเลื่อนไปตัวชี้ไปยังตำแหน่งถัดไปโดย array ที่ได้มี key คือคอลัมน์ที่หรือชื่อฟิลด์โดยจะคืนค่า false ถ้าเกิดความผิดพลาดขึ้นในที่นี้จะทำการคืนค่าไปยังคำสั่ง SQL ของแต่ละ case_id ที่คลิกผลการทำนาย

บรรทัดที่ 16-21 คือเงื่อนไขการทำงานกฎข้อที่ 1 ที่ว่าถ้าคดีไม่มีความเกี่ยวเนื่อง จะส่งผลลัพธ์การทำนายไปแสดงว่าเป็นประเภทคดีง่าย แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 23-28 คือเงื่อนไขของการทำงานกฎข้อที่ 2 ที่ว่าถ้าคดีมีความเกี่ยวเนื่อง และเอกสารประกอบการพิจารณาครบถ้วนสมบูรณ์ และเวลาในการพิจารณาคดีเป็นเวลาทั้งวัน จะส่งผลลัพธ์การทำนายไปแสดงว่าเป็นประเภทคดียาก แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 30-35 คือเงื่อนไขของการทำงานกฎข้อที่ 3 ที่ว่าถ้าคดีมีความเกี่ยวเนื่อง และเอกสารประกอบการพิจารณาครบถ้วนสมบูรณ์ และเวลาในการพิจารณาคดีเป็นเวลาเช้า จะส่งผลลัพธ์การทำนายไปแสดงว่าเป็นประเภทคดีง่าย แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 37-42 คือเงื่อนไขของการทำงานกฎข้อที่ 4 ที่ว่าถ้าคดีมีความเกี่ยวเนื่อง และเอกสารประกอบการพิจารณาครบถ้วนสมบูรณ์ และเวลาในการพิจารณาคดีเป็นเวลาบ่าย และคดีส่วนใหญ่เป็นเพศชาย จะส่งผลลัพธ์การทำนายไปแสดงว่าเป็นประเภทคดีง่าย แล้วดำเนินการออกจากเงื่อนไข

ส่วนใหญ่ทั้งวัน และผลการพิจารณาเบื้องต้นคือมีอัตราโทษสูง จะส่งผลลัพธ์การทำงานไปแสดงว่าเป็นประเภทดียาก แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 99-104 คือเงื่อนไขของการทำนายกฎข้อที่ 13 ที่ว่าถ้าถ้าคดีมีความเกี่ยวเนื่อง และเอกสารประกอบการพิจารณาไม่ครบ และประเภทคดีเป็นคดีพิเศษ และเวลาในการพิจารณาคดีส่วนใหญ่ทั้งวัน และผลการพิจารณาเบื้องต้นคือมีหนังสือขอนักจิตวิทยา จะส่งผลลัพธ์การทำงานไปแสดงว่าเป็นประเภทดียาก แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 106-111 คือเงื่อนไขของการทำนายกฎข้อที่ 14 ที่ว่าถ้าถ้าคดีมีความเกี่ยวเนื่อง และเอกสารประกอบการพิจารณาไม่ครบ และประเภทคดีเป็นคดีพิเศษ และเวลาในการพิจารณาคดีส่วนใหญ่ทั้งวัน และผลการพิจารณาเบื้องต้นคือมีหนังสือขอล่าม จะส่งผลลัพธ์การทำงานไปแสดงว่าเป็นประเภทดียาก แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 113-116 คือเงื่อนไขของการทำนายกฎข้อที่ 15 ที่ว่าถ้าถ้าคดีมีความเกี่ยวเนื่อง และเอกสารประกอบการพิจารณาไม่ครบ และประเภทคดีเป็นคดีพิเศษ และเวลาในการพิจารณาคดีส่วนใหญ่ทั้งวัน และผลการพิจารณาเบื้องต้นคือมีหนังสือขอแรงทนาย จะส่งผลลัพธ์การทำงานไปแสดงว่าเป็นประเภทดียาก แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 119 คือกรณีที่มีเงื่อนไขนอกเหนือจากเงื่อนไขที่ 1 ถึงเงื่อนไขที่ 15 จะส่งผลลัพธ์การทำงานไปแสดงว่าเป็นประเภทดีง่าย แล้วดำเนินการออกจากเงื่อนไข

บรรทัดที่ 122 คือปิดการใช้งาน PHP

บทที่ 5

สรุปผลการศึกษาค้นคว้า และข้อเสนอแนะ

5.1 สรุปผลการศึกษาค้นคว้า

จากการดำเนินการศึกษาวิจัย เรื่องการตรวจสอบข้อมูลทางคดี ของศาลจังหวัดปทุมธานี ด้วยเทคนิคของเหมืองข้อมูล มีวัตถุประสงค์เพื่อการศึกษาและค้นคว้ารูปแบบการเกิดความสัมพันธ์ของตัวแปรของข้อมูลทางคดีที่มีอยู่และนำความรู้ที่ได้จากการศึกษาค้นคว้านำมาสร้างโมเดลการทำนายผลลัพธ์ที่คาดว่าจะเกิดขึ้นภายในอนาคต นั้นมีขั้นตอนการทำงานหลักๆอยู่ 5 ส่วน ได้แก่

ส่วนที่ 1 ดำเนินการศึกษาและเก็บรวบรวมข้อมูลทางคดี โดยใช้วิธีการจำลองข้อมูลให้คล้ายกับฐานข้อมูลที่จัดเก็บจริง จำนวน 1,280 แถวให้อยู่ในรูปแบบไฟล์ Excel

ส่วนที่ 2 จัดเตรียมข้อมูลที่ได้นำไปทำตามมาตรฐานเดียวกัน โดยทำการโอนย้ายข้อมูล (Data Transfers) ทำการลดขนาดข้อมูล (Data Reduction) คัดเลือกข้อมูล และทำความสะอาดข้อมูล (Data Cleaning) เมื่อผ่านกระบวนการจัดเตรียมข้อมูล จะคัดเลือกเฉพาะข้อมูลที่สมบูรณ์และมีรูปแบบที่ถูกต้อง ครบถ้วนที่สุด เพื่อนำมาวิเคราะห์ในโปรแกรม Weka 3.6 ก่อนที่จะทำการวิเคราะห์ข้อมูลในโปรแกรมดังกล่าวต้องทำการแปลงไฟล์ให้อยู่ในรูปแบบ .CSV

ส่วนที่ 3 การทดสอบข้อมูล และเลือกเทคนิคที่เหมาะสม ซึ่งในการศึกษา ค้นคว้า วิจัยครั้งนี้ได้ทดสอบเทคนิคของเหมืองข้อมูล โดยใช้เทคนิคการจำแนกกลุ่ม จำนวน 2 เทคนิค คือ โมเดลการจำแนกกลุ่มแบบ Decision Tree:C4.5 หรือ โมเดล Classifier tree J48 และ โมเดลการจำแนกกลุ่มแบบ Bayes: Naïve Bayes หรือ โมเดล Classifier Naïve Bayes ในโปรแกรม Weka จากผลการทดสอบการจำแนกกลุ่มดังกล่าวพบว่า การจำแนกกลุ่มแบบ Decision Tree:C4.5 ให้ผลลัพธ์ที่ดีที่สุดและให้กฎออกมาที่สามารถนำไปวิเคราะห์กฎต่อไปได้ โดยผลลัพธ์มีความถูกต้องสูงถึง 93.52 เปอร์เซ็นต์ และได้กฎจำนวน 15 กฎ ส่วนผลลัพธ์ของโมเดลการจำแนกกลุ่มแบบ Bayes: Naïve Bayes สามารถให้ผลลัพธ์จากการทำนายสูงเช่นเดียวกัน แต่มีข้อด้วยตรงที่ไม่มีความสามารถศึกษาหรือนำไปต่อยอดได้

ส่วนที่ 4 การวิเคราะห์ข้อมูล จากผลการจำแนกกลุ่มข้อมูลแบบ Decision Tree: C4.5 เป็นการจับกลุ่มที่เหมาะสมที่สุดในการทดสอบ เนื่องจากมีกฎออกมาให้เราได้ดำเนินการศึกษาและนำไปต่อยอดได้ ประกอบกับมีเงื่อนไขที่เหมาะสมสามารถทำนายได้ถูกต้อง 100 เปอร์เซ็นต์ และมีบางกรณีที่มีเงื่อนไขไม่เหมาะสม รวมถึงมีกฎในการทำนายที่มีความถูกต้องน้อยกว่า 100

เปอร์เซ็นต์ มีจำนวน 15 กฎ ที่สามารถนำมาใช้เป็นเงื่อนไขในการพัฒนาโปรแกรมตรวจสอบข้อมูลทางคดีของศาลจังหวัดปทุมธานีต่อไป

ส่วนที่ 5 การพัฒนาโปรแกรมตรวจสอบข้อมูลทางคดีของศาลจังหวัดปทุมธานี เป็นการนำกฎที่ได้จากการทำเหมืองข้อมูลมาใช้เป็นเงื่อนไขในการเขียน โปรแกรม เพื่อเป็นเครื่องมือช่วยผู้บริหารช่วยการตัดสินใจการจ่ายสำนวนคดีให้กับผู้พิพากษาในการพิจารณารรถคดีความ โดยใช้โปรแกรม PHP และใช้ฐานข้อมูล My SQL ในการพัฒนาโปรแกรมโดยโปรแกรมที่พัฒนานั้นมีจำนวนเพจให้ใช้งานจำนวน 6 เพจ ประกอบไปด้วย ไฟล์ของการเชื่อมต่อฐานข้อมูล (Database Connection.php), ไฟล์ของการจัดการข้อมูลบุคคลและการกำหนดสิทธิ (Personal_info.php), หน้าจอการล็อกอินเข้าสู่ระบบ (Index.php), การสร้างฟอร์มการป้อนข้อมูลทางคดี (Create_data.php), ตารางสรุปข้อมูลการป้อนข้อมูลทางคดี (List_data.php) และฟอร์มสำหรับการแสดงผลการทำนาย (Decision_tree.php) โดยโปรแกรมที่ได้ช่วยอำนวยความสะดวกให้กับผู้บริหารได้เป็นอย่างดี

5.2 ปัญหาและแนวทางแก้ไข

1. ในการทำเหมืองข้อมูลครั้งนี้มีขั้นตอนในการเตรียมข้อมูลยังขาดข้อมูลบางส่วนที่มีความจำเป็นในการพิจารณา เช่น ปริมาณคดีที่อยู่ระหว่างการพิจารณาพิพากษาอรรถคดี หากมีการนำข้อมูลเหล่านี้มาศึกษา อาจจะทำให้ผลการศึกษาค้นคว้าครั้งนี้มีความน่าเชื่อถือมากขึ้น

2. ในการใช้เครื่องมือ (Tool) ต่างๆในการศึกษาค้นคว้า ได้ใช้ซอฟต์แวร์แบบโอเพนซอร์ส (Open source) ซึ่งเอกสารประกอบการใช้งานมีน้อย ทำให้ยากต่อการศึกษาวิธีการใช้งาน การวิเคราะห์ผลลัพธ์ จึงได้ดำเนินการหาเอกสารจากทางเว็บไซต์ที่เป็นภาษาอังกฤษ หรือบทความมาดำเนินการศึกษา หรืออ่านประกอบ

3. ในการทำเหมืองข้อมูลได้ใช้ข้อมูลทดสอบหรือข้อมูลตัวอย่างที่ไม่ได้แบ่งแยกประเภทที่ชัดเจนมาดำเนินการศึกษา ทำให้ยากต่อการวิเคราะห์ข้อมูล หรือได้ผลลัพธ์ที่ไม่เป็นที่น่าพอใจ จึงได้ดำเนินการแยกประเภทข้อมูลที่ชัดเจนออกมาดำเนินการทดสอบ ทำให้ผลลัพธ์มีความชัดเจนมากขึ้น

4. ในการพัฒนาโปรแกรมเกิดข้อผิดพลาดบ่อยครั้ง เนื่องจากไม่มีระยะเวลาในการศึกษามากพอ จึงได้สอบถามจากผู้มีประสบการณ์และหาหนังสือมาศึกษาเพิ่มเติม

5.3 ข้อเสนอแนะสำหรับการพัฒนาต่อไป

1. ในการทำเหมืองข้อมูล สิ่งที่สำคัญที่สุดในการวิเคราะห์ข้อมูลคือ ผู้ดำเนินการศึกษา จะต้องทราบ โครงสร้างของฐานข้อมูลเบื้องต้น เพื่อความถูกต้องในการคัดเลือกแอตทริบิวต์มาใช้งานเพื่อทำนายผลลัพธ์ที่เกิดขึ้นด้วยเทคนิคของเหมืองข้อมูล ซึ่งจะส่งผลให้ได้ผลลัพธ์ที่มีประสิทธิภาพสูงสุด

2. ในการทำเหมืองข้อมูลควรใช้ข้อมูลในการทดสอบที่หลากหลายและมีจำนวนมาก เพื่อให้ได้ผลลัพธ์ในการศึกษา วิเคราะห์ที่มีความน่าเชื่อถือมากยิ่งขึ้น

3. ควรดำเนินการใช้เทคนิคของเหมืองข้อมูลด้วยเทคนิคอื่นนอกเหนือจากการศึกษามาใช้ในการทดสอบกับข้อมูลที่มีอยู่เพื่อให้ได้ผลลัพธ์ที่ถูกต้อง ชัดเจน มากขึ้น

4. ในการพัฒนาต่อยอด โปรแกรมตรวจสอบข้อมูลทางคดีของศาลจังหวัดปทุมธานีนั้น เห็นควรมีการเพิ่มฐานข้อมูลในส่วนของผลลัพธ์จากการทำนาย เพราะจะทำให้การใช้งานจากผลลัพธ์การทำนายได้อย่างสะดวกและรวดเร็ว ทำให้ลดขั้นตอนในกระบวนการตรวจสอบข้อมูลทางคดีได้

5. เมื่อเราได้บันทึกผลลัพธ์ที่ได้จากการทำนายนั้นที่เราได้ดำเนินการแบ่งกลุ่มประเภทคดี เห็นควรศึกษาเทคนิคของเหมืองข้อมูลในส่วนของพยากรณ์ ในลำดับต่อไป เพื่อสามารถทำนายผลลัพธ์ของคดีความที่เราแบ่งกลุ่ม ว่าแต่ละประเภทคดีนั้นใช้เวลาในการพิจารณาคดีมากน้อยเพียงใด และจะเป็นประโยชน์ต่อการบริหารจัดการกับข้อมูลทางคดีได้

กรม
การ
การ

บรรณานุกรม

บรรณานุกรม

ภาษาไทย

กฤษณะ ไวยมัย, ชิดชนก สังกศิริ, และชนาวินท์ รักธรรมานนท์. (2544). การใช้เทคนิคดาต้าไมนนิ่ง เพื่อพัฒนาคุณภาพการศึกษานิสิตคณะวิศวกรรมศาสตร์, *The Nectec Thailand Jurnal.3(11)* , : 134-142.

กฤษณะ ไวยมัย,และ ชีระวัฒน์ พงษ์ศิริปริดา. (2544.) การใช้เทคนิค Association Rule Discovery เพื่อการจัดสรรกฎหมายในการพิจารณาคดีความ . *The Nectec Thailand Jurnal.3 (11)* , : 143-152.

ชนวัฒน์ ศรีสอาน. (2551). *ฐานข้อมูล คลังข้อมูล และเหมืองข้อมูล* .กรุงเทพฯ : มหาวิทยาลัยรังสิต.

ซัดชัย แก้วตา, และอัจฉรา มหาวิวัฒน์. (2553). *การวินิจฉัยคดีด้วยเทคนิคค้นไม้ตัดสินใจ*.

(วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). อุบลราชธานี: มหาวิทยาลัยอุบลราชธานี.

ทิพย์ธิดา วงศ์พิพันธ์. (2556). *การใช้เหมืองข้อมูลช่วยในการตัดสินใจการให้สินเชื่อ กรณีศึกษา:*

บริษัท กรุงไทยคาร์เร้นท์ แอนด์ ลีส จำกัด(มหาชน) (สารนิพนธ์ปริญญาโทบริหารธุรกิจ).

กรุงเทพฯ: มหาวิทยาลัยธุรกิจบัณฑิตย์.

ชนาคม ชัยสิทธิ์, และเนืองวงศ์ ทวยเจริญ. (2556). *การพัฒนาระบบสนับสนุนการตัดสินใจเลือกสาขา*

วิชาเอก ของ ผู้ที่ ต้องการ สมัคร เข้าศึกษา ต่อ ใน ระดับ ปริญญาตรี ของ

มหาวิทยาลัยสุโขทัยธรรมมาธิราชด้วยเหมืองข้อมูล. การประชุมวิชาการทางวิศวกรรมไฟฟ้า

ครั้งที่ 36 (EECON-36). มหาวิทยาลัยเกษตรศาสตร์.

ชีระวัฒน์ พงษ์ศิริปริดา. (2544). *การจัดสรรกฎหมายที่เหมาะสมให้กับคดีความโดยอัตโนมัติ*.

(วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ). กรุงเทพฯ : มหาวิทยาลัยเกษตรศาสตร์.

บัญชา ปะสีละเตสัง. (2553). *พัฒนาเว็บแอปพลิเคชันด้วย PHP ร่วมกับ MySQL และ Dreamweaver.*

กรุงเทพฯ: ซีเอ็ดดูเคชั่น .

- ปณิธิ แก้วสวัสดิ์. (2553). เหมืองข้อมูล: การค้นหาความรู้และการขุดค้นข้อมูล . คอมพิวเตอร์และ
เทคโนโลยีขั้นสูง . 11 , 51-60.
- สาขชล สันสมบูรณ์ทอง. (2558).การทำเหมืองข้อมูล *Data Mining*. กรุงเทพฯ: จามจุรีโปรดักท์.
- สิทธิโชค มุกดาสกุลภินาส.(2551). การวัดประสิทธิภาพของขั้นตอนวิธีตัวจำแนก *C4.5 ,ADTree*
และ *Bayes* ในการจำแนกข้อมูลการชุกซ่อนสิ่งเสพติดสำหรับไปรษณีย์ระหว่างประเทศ.
(วิทยานิพนธ์ปริญญาโทบริหารธุรกิจ).กรุงเทพฯ : มหาวิทยาลัยเกษตรศาสตร์.
- สิทธิธร เจริญรัตน์, และชฎารัตน์ พิพัฒน์นันท์. (2556). การศึกษาปัจจัยที่มีผลต่อรายได้ของประชากร
ในประเทศไทยโดยใช้เทคนิคเหมืองข้อมูล. วารสารวิชาการ มหาวิทยาลัยหอการค้าไทย.
33(1) , 131-151 .
- อนรรฆรงค์ คุณมณี. (2555). พัฒนา *Web App* แบบมืออาชีพด้วย *PHP+AJAX* และ *jQuery*.
นนทบุรี: อดิษฐ์ พรีเมียร์.
- เอกสิทธิ์ พัชรวงศ์ศักดิ์. (2556). คู่มือการใช้งาน *Weka Explorer* เบื้องต้น.กรุงเทพฯ: บริษัท เอเชีย
ดิจิทัลการพิมพ์ จำกัด.
- เอกสิทธิ์ พัชรวงศ์ศักดิ์. (2557). การวิเคราะห์ข้อมูลด้วยเทคนิคค้ำดำ ไม่นิ่ง เบื้องต้น. กรุงเทพฯ:
บริษัท เอเชีย ดิจิทัลการพิมพ์ จำกัด.

ภาษาต่างประเทศ

- Ian H. Witten, Frank Eibe, & Mark A. Hall. (2011). *Data mining : practical machine learning*
tools and techniques(3rd Edition). Morgan Kaufmann,San Francisco.
- Jiawei Han, Micheline Kamber, & Jian Pei . (2011). *DATA MINING: CONCEPTS AND*
TECHNIQUES (3RD EDITION) Morgan Kaufmann,San Francisco.

Pang-Ning Tan, Michael Steinbach, & Vipin Kumar. (2006). *Introduction to data mining*.

Addison-Wesley Longman Publishing Co. Inc. Boston Ma,USA.

T.Pongsiripreeda, & K. Waiyamai. (2000). *Using Data Mining technique to select the law and article for lawsuit(in thai language)*. The National Computer Science and Engineering Conference (NCSEC'2000). Thailand, Bangkok.November

U.Fayyad, G. Piatetsky-Shapiro, & P. Smyth. (1997). *From Data Mining to Knowledge Discovery in Database*. *AI MAGAZINE*, 1 , 39-45.

กรม
พาณิชย์
และ
อุตสาหกรรม

ภาคผนวก

ผลการรันโปรแกรมฉบับสมบูรณ์ classifiers.bayes.NaiveBayes

=== Run information ===

Scheme:weka.classifiers.bayes.NaiveBayes

Relation: test

Instances: 1280

Attributes: 7

gender

Reasoning

HaveRelate

consider

time

document

difficultycase

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute	Class	
	easy	difficulty
	(0.5)	(0.5)

=====

gender		
male	396.0	341.0
female	246.0	301.0
[total]	642.0	642.0

Reasoning		
general litigation	171.0	141.0
special case	216.0	426.0
special case management	256.0	76.0
[total]	643.0	643.0

HaveRelate		
notrelate	366.0	1.0
relate	276.0	641.0
[total]	642.0	642.0

consider		
justice	346.0	371.0
sent to justice	116.0	66.0
High incarceration rate	66.0	16.0
Psychologists requested	41.0	41.0
interpreter requested	26.0	56.0
lawyer requested	51.0	96.0
[total]	646.0	646.0

time

allday	146.0	206.0
morning	411.0	266.0
afternoon	86.0	171.0
[total]	643.0	643.0

document

complete	296.0	186.0
notcomplete	346.0	456.0
[total]	642.0	642.0

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	1018	79.5313 %
Incorrectly Classified Instances	262	20.4688 %
Kappa statistic	0.5906	
Mean absolute error	0.2607	
Root mean squared error	0.3818	
Relative absolute error	52.1453 %	
Root relative squared error	76.3615 %	
Total Number of Instances	1280	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.667	0.077	0.897	0.667	0.765	0.844	easy
	0.923	0.333	0.735	0.923	0.819	0.844	difficulty
Weighted Avg.		0.795	0.205	0.816	0.795	0.792	0.844

== Confusion Matrix ==

```

a   b <-- classified as
427 213 |   a = easy
49   591 |   b = difficulty

```

ผลการรันโปรแกรมฉบับสมบูรณ์ classifiers.trees.J48

=== Run information ===

Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: test04052558

Instances: 1280

Attributes: 7

gender

Reasoning

HaveRelate

consider

time

document

difficultycase

Test mode:10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

HaveRelate = notrelate: easy (365.0)

HaveRelate = relate

| document = complete

| | time = allday: difficulty (60.0)

| | time = morning: easy (150.0)

| | time = afternoon

| | | gender = male: easy (25.0)

| | | gender = female: difficulty (125.0)

| document = notcomplete

| | Reasoning = general litigation: difficulty (115.0)

| | Reasoning = special case

| | | time = allday

| | | | consider = justice: difficulty (75.0/25.0)

| | | | consider = sent to justice: easy (25.0)

| | | | consider = High incarceration rate: difficulty (0.0)

| | | | consider = Psychologists requested: difficulty (0.0)

| | | | consider = interpreter requested: difficulty (0.0)

| | | | consider = lawyer requested: difficulty (50.0)

| | | time = morning: difficulty (200.0/50.0)

| | | time = afternoon: difficulty (15.0)

| | Reasoning = special case management: difficulty (75.0)

Number of Leaves : 15

Size of the tree : 22

Time taken to build model: 0.13 seconds

==== Stratified cross-validation ====

==== Summary ====

Correctly Classified Instances	1197	93.5156 %
Incorrectly Classified Instances	83	6.4844 %
Kappa statistic	0.8703	
Mean absolute error	0.0827	
Root mean squared error	0.2044	
Relative absolute error	16.5386 %	
Root relative squared error	40.881 %	
Total Number of Instances	1280	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.891	0.02	0.978	0.891	0.932	0.984	easy
	0.98	0.109	0.9	0.98	0.938	0.984	difficulty
Weighted Avg.		0.935	0.065	0.939	0.935	0.935	0.984

=== Confusion Matrix ===

```
a    b <-- classified as
570  70 | a = easy
13   627| b = difficulty
```

ประวัติผู้เขียน

ชื่อ-นามสกุล

นายนเรศ หมูทอง

ประวัติการศึกษา

พ.ศ. 2552 วิทยาศาสตรบัณฑิต

สาขาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์

มหาวิทยาลัยเชียงใหม่

ตำแหน่งและสถานที่ทำงานปัจจุบัน

นักวิชาการคอมพิวเตอร์ปฏิบัติการ

ศาลจังหวัดปทุมธานี