

บทที่ 3

วิธีการดำเนินงานวิจัย

วิธีการดำเนินงานวิจัยของระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ ผู้วิจัยได้กำหนดขั้นตอนของการดำเนินงาน ดังนี้

- 3.1 การศึกษาข้อมูลที่เกี่ยวข้อง
- 3.2 รวบรวม คัดเลือก และจัดเตรียมข้อมูล
- 3.3 ทดสอบหาอัลกอริทึมและสร้างโมเดล
- 3.4 นำอัลกอริทึมที่ได้มาสร้างโมเดลและพัฒนาระบบ
- 3.5 การทดสอบความถูกต้องและความน่าเชื่อถือของโมเดล
- 3.6 สถิติที่ใช้ในการวิจัย

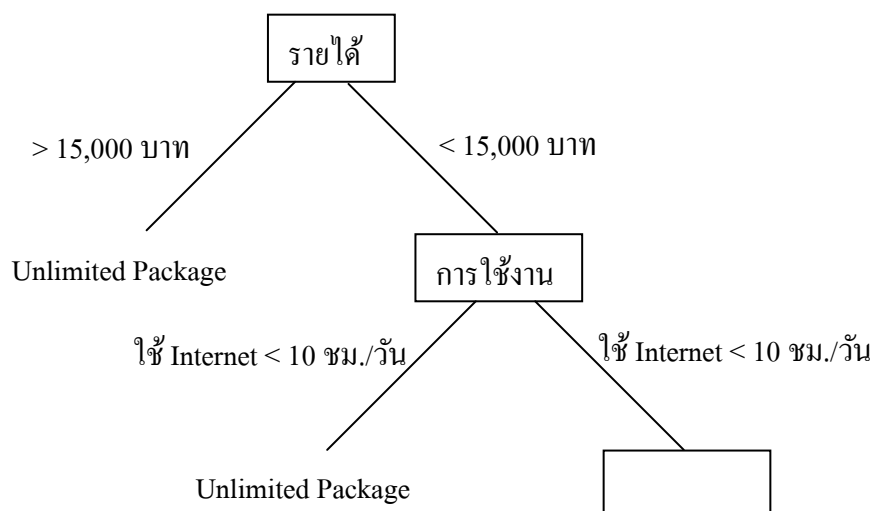
3.1 การศึกษาข้อมูลที่เกี่ยวข้อง

ในการทำวิจัยเรื่อง ระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ ผู้วิจัยต้องศึกษาค้นคว้าข้อมูลและความรู้ที่จะนำมาใช้ในงานวิจัย ดังนี้

3.1.1 ศึกษา และทบทวนวรรณกรรมหรือผลงานวิจัยที่เกี่ยวข้อง (Review Literature) เพื่อเป็นแนวทางในการดำเนินงานวิจัย

3.1.2 ศึกษาขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ต้นไม้ตัดสินใจ และอัลกอริทึมของเทคนิคต้นไม้ตัดสินใจแต่ละตัว

3.1.2.1 ศึกษาขั้นตอนการวิเคราะห์ข้อมูลโดยใช้ต้นไม้ตัดสินใจ เริ่มจากการศึกษาโครงสร้างของต้นไม้ตัดสินใจ ที่ประกอบด้วย ใบ (Leaf) กิ่ง (Branch) และราก (Root) ศึกษาถึงการเรียนรู้ของ Decision Tree ซึ่งเป็นการเรียนรู้จากการคาดคะเนเหตุการณ์ต่างๆ ที่เกิดขึ้น อาศัยเงื่อนไขเป็นตัวช่วยในการตัดสินใจว่าเมื่อมีเหตุการณ์ใดเกิดขึ้น จะแสดงผลออกมาอย่างไร ถูกนำเสนอด้วยรูปแบบการตัดสินใจ (if / then / else) ขั้นตอนต่อมาคือการสร้างแบบจำลองต้นไม้ตัดสินใจ โดยจะเริ่มพิจารณาที่โหนด “ราก” ก่อน การหาโหนดรากจะใช้วิธีคำนวณหาค่า Information gain ข้อมูลกลุ่มใดมีค่า Information gain มากที่สุดจะนำมาเป็นโหนดราก แล้วนำข้อมูลที่เหลือมาคำนวณ Information gain อีกครั้งเพื่อหาโหนดถัดไปที่เหมาะสม



ภาพที่ 3-1 ตัวอย่างต้นไม้ตัดสินใจที่มีข้อมูลรายได้เป็นโหนดราก

3.1.2.2 ศึกษาอัลกอริทึมของเทคนิคต้นไม้ตัดสินใจที่จะนำมาทดสอบ เพื่อหาอัลกอริทึมที่เหมาะสมในการสร้างโมเดลมากที่สุด จากการศึกษาผู้วิจัยได้เลือกอัลกอริทึมที่จะนำมาทดสอบ ดังนี้ ID3, C4.5 (J48)

3.1.2.3 ศึกษาถึงเครื่องมือ หรือซอฟต์แวร์ที่นำมาใช้ในการพัฒนาระบบ สร้างโมเดล และทดสอบอัลกอริทึม การวิจัยครั้งนี้ ผู้วิจัยได้เลือกใช้เครื่องมือ คือซอฟต์แวร์โปรแกรม Weka ซึ่งเป็นโปรแกรมที่สามารถนำมาทดสอบอัลกอริทึมของต้นไม้ตัดสินใจได้ เนื่องจากมีอัลกอริทึมที่ได้ระบุไว้ให้เลือกใช้โปรแกรมครบตามที่กำหนด และสามารถแสดงค่าความแม่นยำของอัลกอริทึมจาก Confusion Matrix ในโปรแกรม Weka ได้อีกด้วย นอกจากนี้ยังเลือกใช้โปรแกรมในการพัฒนาระบบ คือ โปรแกรม Microsoft Visual Studio 2008 ซึ่งใช้ภาษา ASP.NET ด้วย C# ในการสร้างหน้าเว็บ และสร้างโมดูลที่เกี่ยวข้องกับระบบงานทั้งหมด และใช้ร่วมกับโปรแกรม Microsoft SQL Server 2005 ในการจัดการและออกแบบฐานข้อมูลของระบบ โดยแสดงผลลัพธ์ผ่านทางโปรแกรมเว็บเบราว์เซอร์

3.2 รวบรวม คัดเลือก และจัดเตรียมข้อมูล

ทำการจัดเตรียมข้อมูลเพื่อใช้ในการสร้างโมเดล และทดสอบอัลกอริทึม มีรายละเอียด ดังนี้

3.2.1 กำหนดกลุ่มประชากร คือ กลุ่มบุคคลทั้งเพศชายและเพศหญิงที่ใช้บริการอินเทอร์เน็ตมือถือในเขตกรุงเทพมหานคร จำนวน 7,322 คน (สถิติจำนวนผู้ใช้อินเทอร์เน็ต จำแนกตามอุปกรณ์/

เครื่องมือเทคโนโลยีสารสนเทศและการสื่อสารในการเข้าถึงอินเทอร์เน็ต ภาค และเขตการปกครอง พ.ศ.2552 จากสำนักงานสถิติแห่งชาติ)

3.2.2 กำหนดกลุ่มตัวอย่าง คือ กลุ่มคนที่ใช้บริการอินเทอร์เน็ตมือถือในกรุงเทพมหานคร จำนวน 1,500 คน กำหนดขนาดของกลุ่มตัวอย่างโดยวิธีใช้เกณฑ์สัดส่วน วิธีนี้ใช้ในกรณีที่ผู้วิจัยทราบจำนวนหน่วยทั้งหมดที่มีอยู่ในประชากร หรือทราบขนาดของประชากรแล้ว จึงคำนวณหาจำนวนหน่วยตัวอย่างที่จะนำมาใช้เป็นกลุ่มตัวอย่างในงานวิจัย โดยมีเกณฑ์ในคำนวณ ดังนี้

3.2.2.1 ประชากรมีขนาดเป็นหลักพัน จะใช้หน่วยตัวอย่าง จำนวน 10 – 20 % เป็นกลุ่มตัวอย่าง

3.2.2.2 ประชากรมีขนาดเป็นหลักพัน จะใช้หน่วยตัวอย่าง จำนวน 5 – 10 % เป็นกลุ่มตัวอย่าง (เกียรติสุตา, 2549)

โดยงานวิจัยนี้มีประชากร จำนวน 7,322 คน มีขนาดเป็นหลักพัน หน่วยตัวอย่างที่ใช้จึงเท่ากับ 10-20% จึงทำให้ได้กลุ่มตัวอย่างในงานวิจัยนี้เท่ากับจำนวน 1,500 คน

3.2.2 จัดทำเครื่องมือในการเก็บรวบรวมข้อมูล โดยเครื่องมือที่ใช้คือ แบบสอบถามพฤติกรรม การใช้อินเทอร์เน็ตมือถือ ภายในแบบสอบถามแบ่งออกเป็น 2 ส่วน ดังนี้

ส่วนที่ 1 ข้อมูลพื้นฐานของผู้ตอบแบบสอบถาม

ส่วนที่ 2 พฤติกรรมการใช้อินเทอร์เน็ตมือถือ

เพื่อนำข้อมูลที่ได้จากแบบสอบถามดังกล่าวมาใช้ในการสร้างโมเดลเพื่อให้ระบบเรียนรู้

3.2.3 การเก็บรวบรวมข้อมูล ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลโดยใช้แบบสอบถาม จำนวน 1,500 ชุด ตามจำนวนของกลุ่มตัวอย่างที่ได้ระบุไว้ ซึ่งผู้วิจัยและผู้ช่วยวิจัย ไปเก็บข้อมูลเอง โดยแจกแบบสอบถามให้กลุ่มตัวอย่างที่ใช้แพคเกจอินเทอร์เน็ตมือถือ โดยเก็บรวบรวมข้อมูลจากประชากรในกรุงเทพมหานครให้ครบตามจำนวนทั้งหมด และเก็บข้อมูลจากสถานที่ต่างๆ ดังนี้

3.2.3.1 รถไฟฟ้า บีทีเอส สถานีหมอชิต

3.2.3.2 รถไฟฟ้า บีทีเอส สถานีอนุสาวรีย์ชัยสมรภูมิ

3.2.3.3 รถไฟฟ้า บีทีเอส สถานีสยาม

3.2.3.4 ห้างสรรพสินค้าเดอะมอลล์งามวงศ์วาน

3.2.3.5 ห้างสรรพสินค้าเดอะมอลล์บางกะปิ

3.2.3.6 ห้างสรรพสินค้าเซลทรัลปิ่นเกล้า

3.2.3.7 ห้างสรรพสินค้าเซลทรัลบางนา

3.2.3.8 ศูนย์การค้าฟิวเจอร์พาร์ครังสิต

3.2.4 จัดเตรียมข้อมูล เมื่อรวบรวมแบบสอบถามทั้งหมดจากกลุ่มตัวอย่างแล้ว ผู้วิจัยต้องทำการจัดเตรียมข้อมูล ดังนี้

3.2.4.1 ตรวจสอบความสมบูรณ์ของข้อมูล ว่าข้อมูลมีความครบถ้วนสมบูรณ์หรือไม่

3.2.4.2 นำข้อมูลที่ได้จากแบบสอบถามจำนวน 1,500 ชุดมาจัดเก็บลงในฐานข้อมูล เพื่อเตรียมเข้าโปรแกรมสำหรับสร้างโมเดลและทดสอบหาอัลกอริทึมที่เหมาะสมต่อไป

3.3 ทดสอบหาอัลกอริทึมและสร้างโมเดล

หลังจากขั้นตอนการรวบรวมและจัดเตรียมข้อมูลเสร็จสิ้น จะได้ข้อมูลซึ่งอยู่ในรูปแบบฐานข้อมูล จากนั้นจะเข้าสู่ขั้นตอนการทดสอบหาอัลกอริทึม โดยมีวิธีการ ดังนี้

3.3.1 ผู้วิจัยได้ทำการเลือกอัลกอริทึมของต้นไม้ตัดสินใจไว้ ซึ่งได้แก่ ID3, C4.5 (J48) เพื่อนำมาทดสอบ

3.3.2 สร้างแบบจำลอง (Model Building) นำข้อมูลในฐานข้อมูลมาสร้างการแบ่งประเภท โดยใช้เทคนิคต้นไม้ตัดสินใจ มีขั้นตอนดังนี้

3.3.2.1 ข้อมูลนำเข้า (Input) เป็นข้อมูลจากฐานข้อมูลที่เก็บรวบรวมจากแบบสอบถามเพื่อนำไปสร้างแบบจำลอง โดยจะทำการกำหนดข้อมูลที่จะนำมาสร้างแบบจำลอง และจะแสดงข้อมูลที่นำมาสร้างแบบจำลองดังตารางต่อไปนี้

ตารางที่ 3-1 ชื่อข้อมูล ชนิดข้อมูล และรายละเอียดของข้อมูลที่จะนำมาสร้างแบบจำลอง

ลำดับ	ชื่อข้อมูล	ชนิดข้อมูล	รายละเอียด
1	Sex	String	เพศ
2	Age	String	อายุ
3	Occupation	String	อาชีพ
4	Income	String	รายได้เฉลี่ยต่อเดือน
5	Chat	String	ปริมาณการสนทนาผ่านอินเทอร์เน็ต
6	SocialNetwork	String	ปริมาณการใช้งาน Social Network
7	Youtube	String	ปริมาณการใช้งาน Youtube
8	Website	String	ปริมาณการใช้งาน Website
9	Email	String	ปริมาณการใช้งาน Email
10	UploadPic	String	ปริมาณการใช้งาน Upload รูปภาพ
11	DownloadPic	String	ปริมาณการใช้งาน Download รูปภาพ

ตารางที่ 3-1 (ต่อ) ชื่อข้อมูล ชนิดข้อมูล และรายละเอียดของข้อมูลที่นำมาสร้างแบบจำลอง

ลำดับ	ชื่อข้อมูล	ชนิดข้อมูล	รายละเอียด
12	UploadMusic	String	ปริมาณการใช้งาน Upload เพลง
13	DownloadMusic	String	ปริมาณการใช้งาน Download เพลง
14	GameOnline	String	ปริมาณการใช้งาน Game Online
15	Package	String	ประเภทแพคเกจอินเทอร์เน็ตที่ผู้ตอบแบบสอบถามเลือกใช้

ตารางที่ 3-2 การแบ่งประเภทเพศ (Sex)

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Sex 1	เพศหญิง
2	Sex 2	เพศชาย

ตารางที่ 3-3 การแบ่งประเภทอายุ (Age)

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Age 1	อายุระหว่าง 15-25 ปี
2	Age 2	อายุระหว่าง 26-35 ปี
3	Age 3	อายุมากกว่า 35 ปี

ตารางที่ 3-4 การแบ่งประเภทอาชีพ (Occupation)

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Occupation 1	นักเรียน / นักศึกษา
2	Occupation 2	ค้าขาย / ธุรกิจส่วนตัว
3	Occupation 3	ลูกจ้างเอกชน / พนักงานบริษัท
4	Occupation 4	ข้าราชการ / รัฐวิสาหกิจ

ตารางที่ 3-5 การแบ่งประเภทรายได้เฉลี่ยต่อเดือน (Income)

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Income 1	รายได้น้อยกว่า 10,000 บาท
2	Income 2	รายได้ระหว่าง 10,000-20,000 บาท
3	Income 3	รายได้ระหว่าง 20,000-30,000 บาท
4	Income 3	รายได้ 30,000 บาท ขึ้นไป

ตารางที่ 3-6 การแบ่งประเภทปริมาณการสนทนาผ่านอินเทอร์เน็ต (Chat)

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Chat 1	น้อยกว่า 3 ชั่วโมงต่อวัน
2	Chat 2	3-10 ชั่วโมงต่อวัน
3	Chat 3	มากกว่า 10 ชั่วโมงต่อวัน

ตารางที่ 3-7 การแบ่งประเภทปริมาณการใช้งาน Social Network

ลำดับ	รหัสข้อมูล	รายละเอียด
1	SocialNetwork 1	น้อยกว่า 3 ชั่วโมงต่อวัน
2	SocialNetwork 2	3-10 ชั่วโมงต่อวัน
3	SocialNetwork 3	มากกว่า 10 ชั่วโมงต่อวัน

ตารางที่ 3-8 การแบ่งประเภทปริมาณการใช้งาน Youtube

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Youtube 1	น้อยกว่า 3 ชั่วโมงต่อวัน
2	Youtube 2	3-10 ชั่วโมงต่อวัน
3	Youtube 3	มากกว่า 10 ชั่วโมงต่อวัน

ตารางที่ 3-9 การแบ่งประเภทปริมาณการใช้งาน Website

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Website 1	น้อยกว่า 3 ชั่วโมงต่อวัน
2	Website 2	3-10 ชั่วโมงต่อวัน
3	Website 3	มากกว่า 10 ชั่วโมงต่อวัน

ตารางที่ 3-10 การแบ่งประเภทปริมาณการใช้งาน E-mail

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Email 1	น้อยกว่า 10 ฉบับต่อวัน
2	Email 2	10-20 ฉบับต่อวัน
3	Email 3	มากกว่า 20 ฉบับต่อวัน

ตารางที่ 3-11 การแบ่งประเภทปริมาณการใช้งาน Upload รูปภาพ

ลำดับ	รหัสข้อมูล	รายละเอียด
1	UploadPic 1	น้อยกว่า 5 รูปต่อวัน
2	UploadPic 2	5-10 รูปต่อวัน
3	UploadPic 3	มากกว่า 10 รูปต่อวัน

ตารางที่ 3-12 การแบ่งประเภทปริมาณการใช้งาน Download รูปภาพ

ลำดับ	รหัสข้อมูล	รายละเอียด
1	DownloadPic 1	น้อยกว่า 5 รูปต่อวัน
2	DownloadPic 2	5-10 รูปต่อวัน
3	DownloadPic 3	มากกว่า 10 รูปต่อวัน

ตารางที่ 3-13 การแบ่งประเภทปริมาณการใช้งาน Upload เพลง

ลำดับ	รหัสข้อมูล	รายละเอียด
1	UploadMusic 1	น้อยกว่า 5 เพลงต่อวัน
2	UploadMusic 2	5-10 เพลงต่อวัน
3	UploadMusic 3	มากกว่า 10 เพลงต่อวัน

ตารางที่ 3-14 การแบ่งประเภทปริมาณการใช้งาน Download เพลง

ลำดับ	รหัสข้อมูล	รายละเอียด
1	DownloadMusic 1	น้อยกว่า 5 เพลงต่อวัน
2	DownloadMusic 2	5-10 เพลงต่อวัน
3	DownloadMusic 3	มากกว่า 10 เพลงต่อวัน

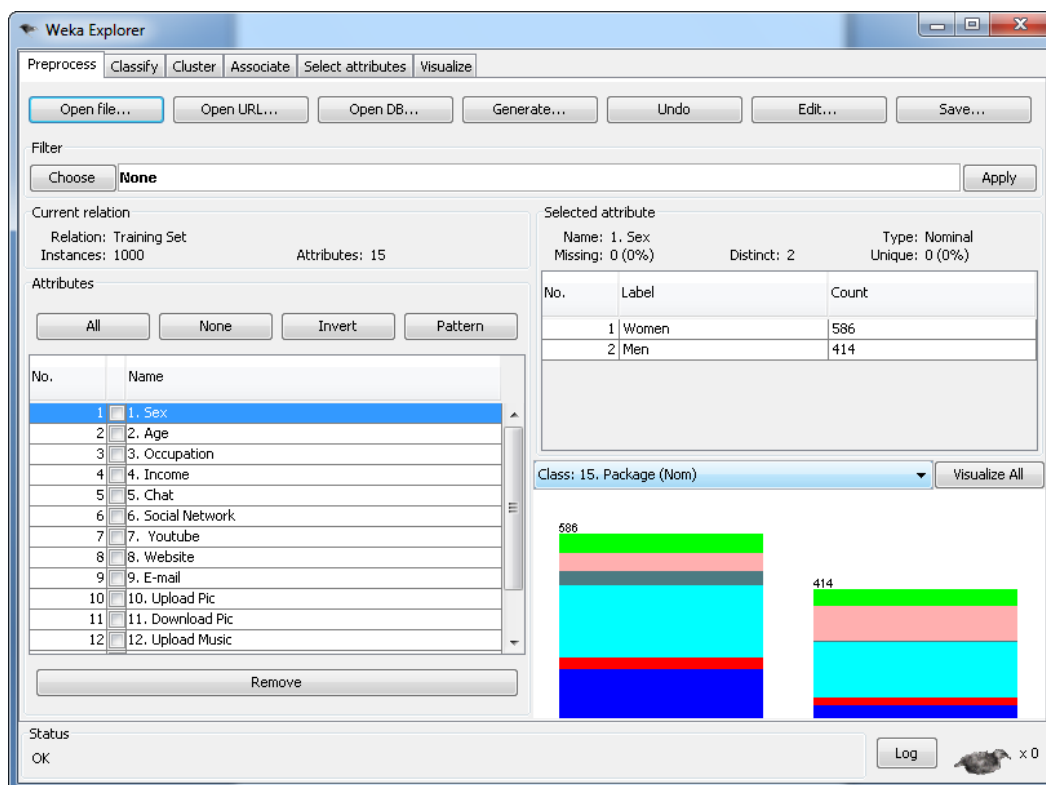
ตารางที่ 3-15 การแบ่งประเภทปริมาณการใช้งาน Game Online

ลำดับ	รหัสข้อมูล	รายละเอียด
1	GameOnline 1	น้อยกว่า 3 ชั่วโมงต่อวัน
2	GameOnline 2	3-10 ชั่วโมงต่อวัน
3	GameOnline 3	มากกว่า 10 ชั่วโมงต่อวัน

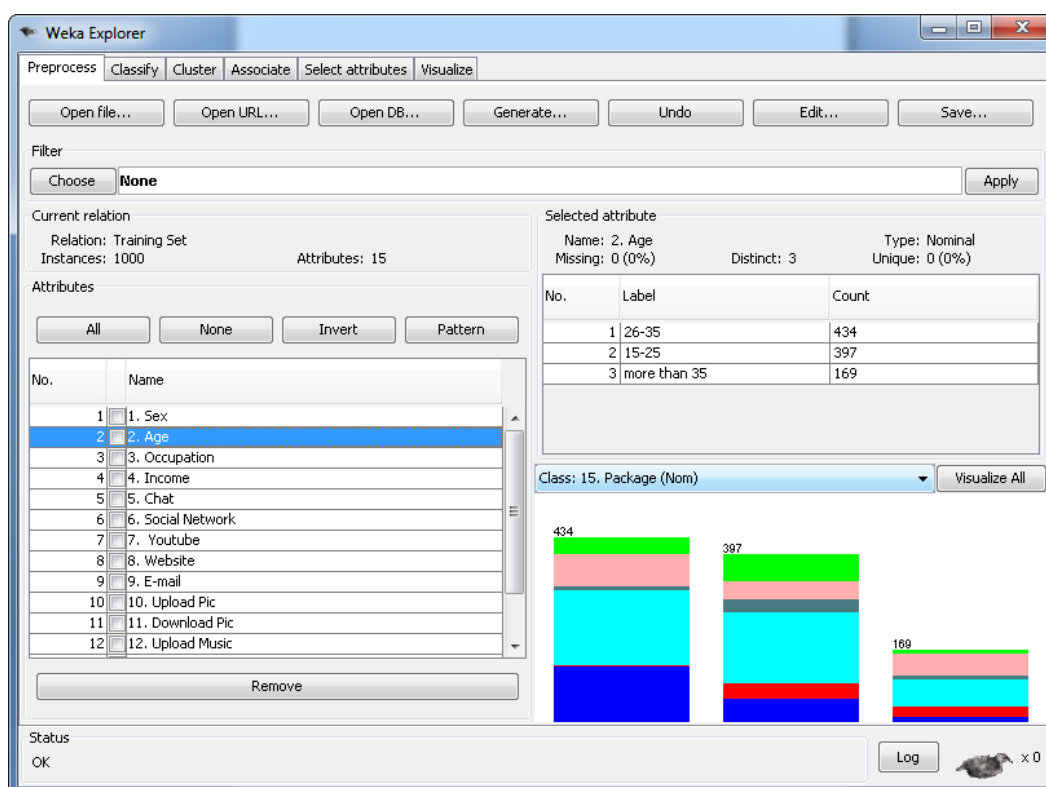
ตารางที่ 3-16 ประเภทแพคเกจอินเทอร์เน็ตที่ผู้ตอบแบบสอบถามเลือกใช้ (Package)

ลำดับ	รหัสข้อมูล	รายละเอียด
1	Package 1	แพคเกจอินเทอร์เน็ตแบบไม่จำกัด แบบรายเดือน
2	Package 2	แพคเกจอินเทอร์เน็ตแบบไม่จำกัด แบบรายสัปดาห์
3	Package 3	แพคเกจอินเทอร์เน็ตแบบไม่จำกัด แบบรายวัน
4	Package 4	แพคเกจอินเทอร์เน็ตแบบคิดตามปริมาณการใช้งาน (เป็นเมกะไบต์/Mb)
5	Package 5	แพคเกจอินเทอร์เน็ตแบบคิดตามระยะเวลาใช้งาน (ชั่วโมง) แบบรายเดือน
6	Package 6	แพคเกจอินเทอร์เน็ตแบบคิดตามระยะเวลาใช้งาน (ชั่วโมง) แบบรายสัปดาห์

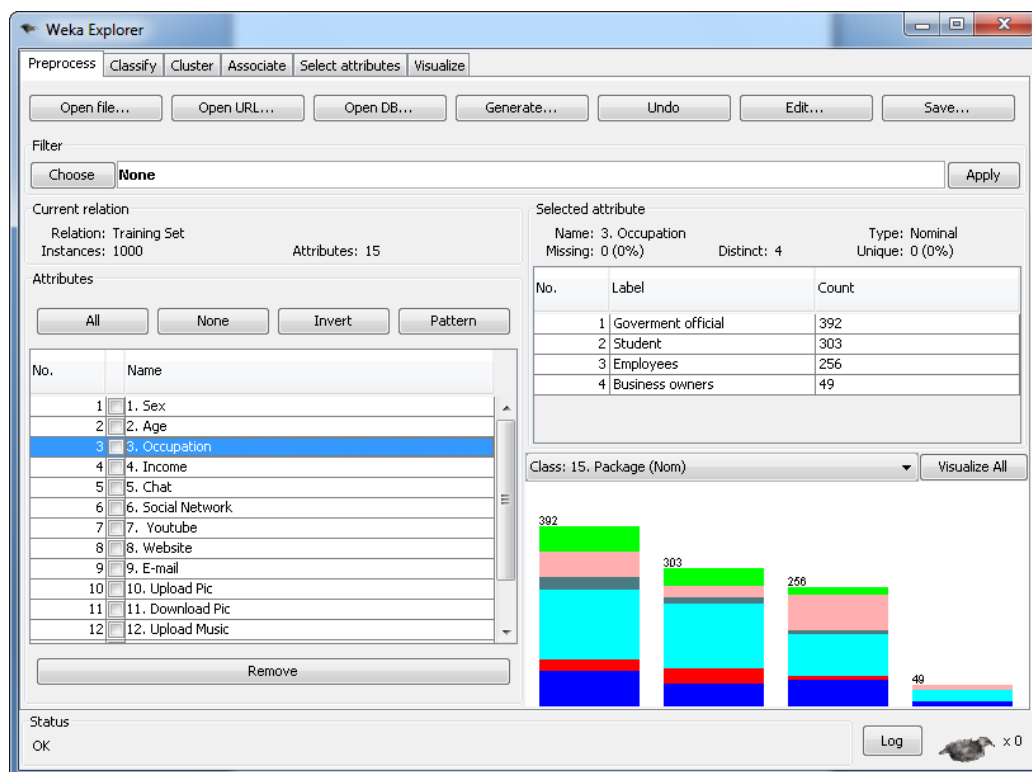
จากข้อมูลในตารางที่จะนำมาสร้างแบบจำลอง เมื่อนำเข้าโปรแกรม Weka แสดงข้อมูลได้ดังรูปภาพต่อไปนี้



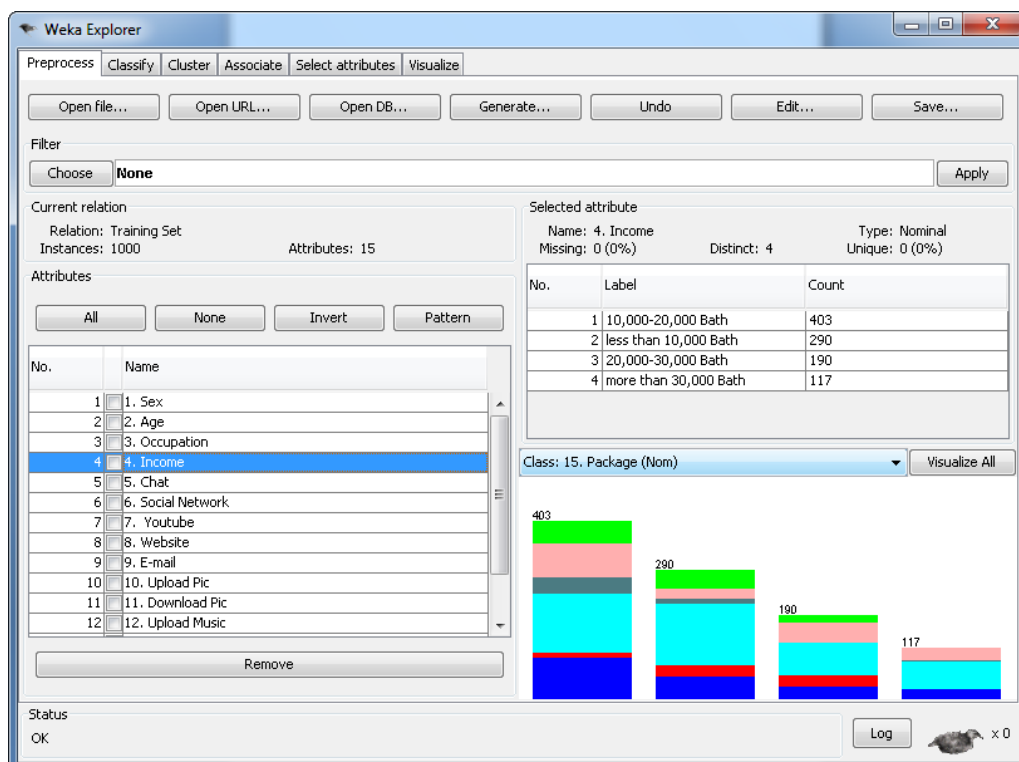
ภาพที่ 3-2 แสดงข้อมูลเพศ ใน Weka



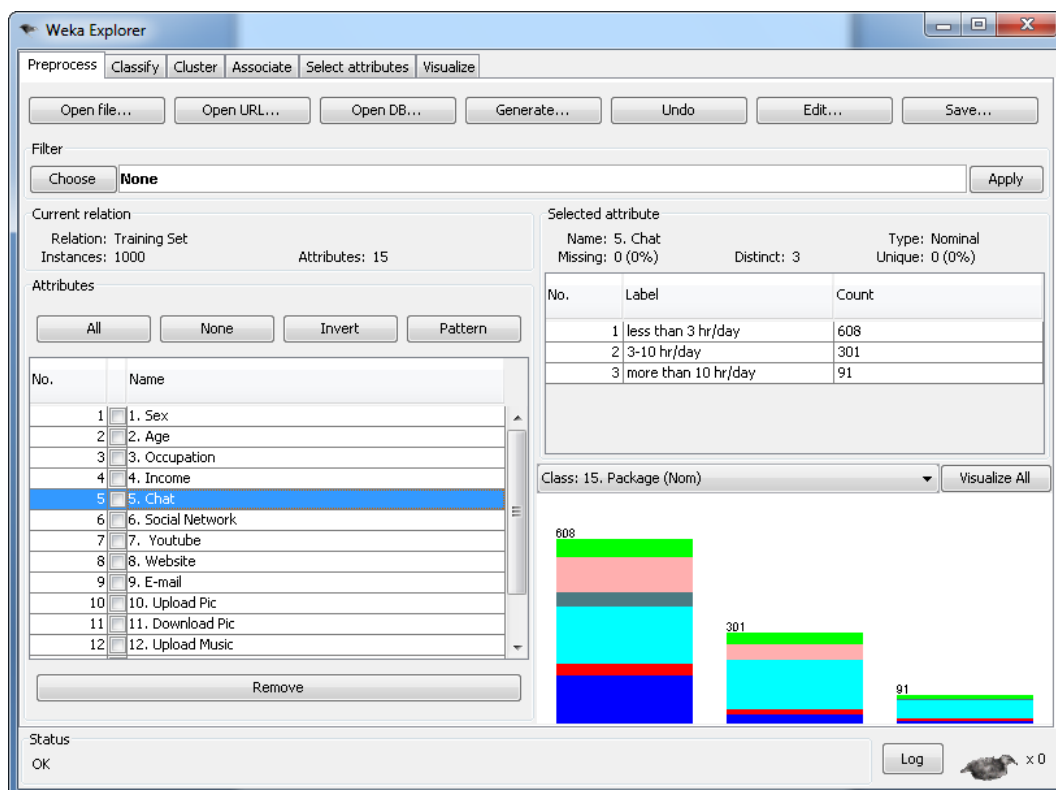
ภาพที่ 3-3 แสดงข้อมูลอายุ ใน Weka



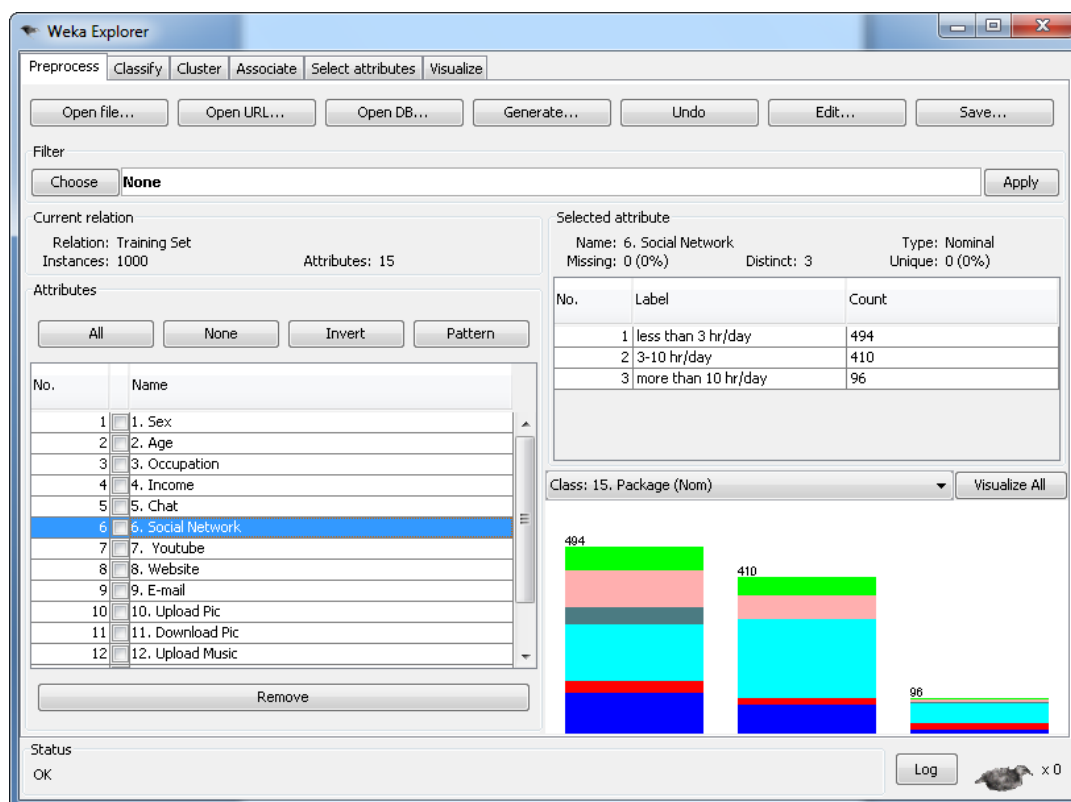
ภาพที่ 3-4 แสดงข้อมูลอาชีพ ใน Weka



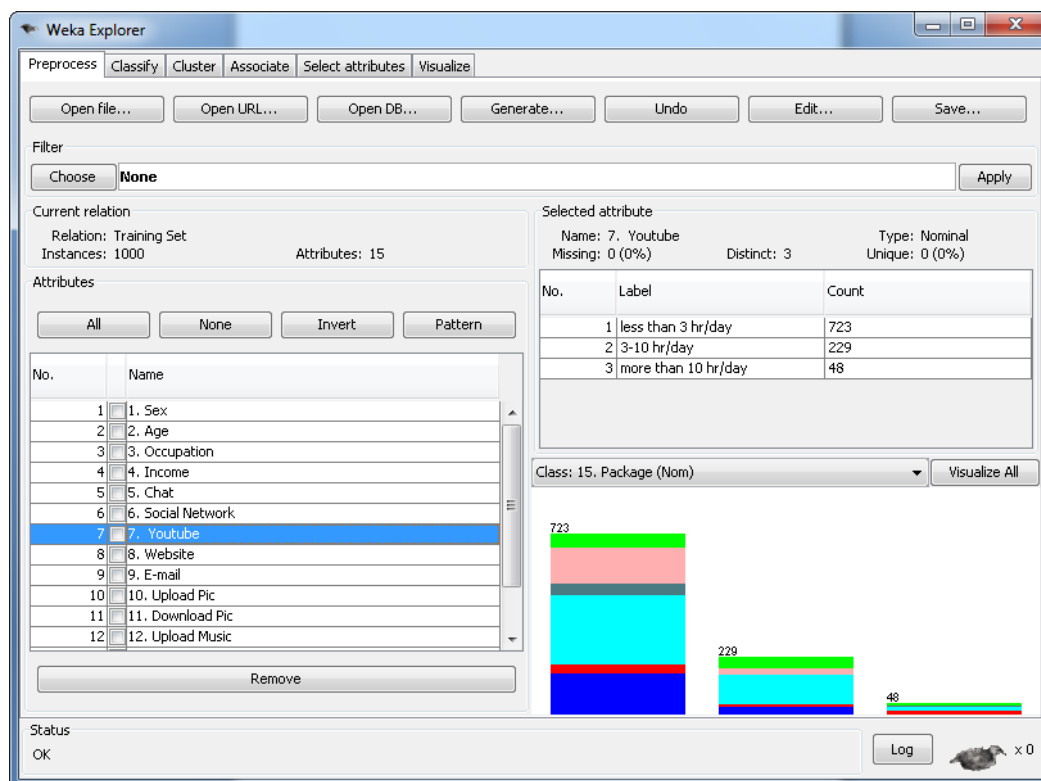
ภาพที่ 3-5 แสดงข้อมูลรายได้เฉลี่ยต่อเดือน ใน Weka



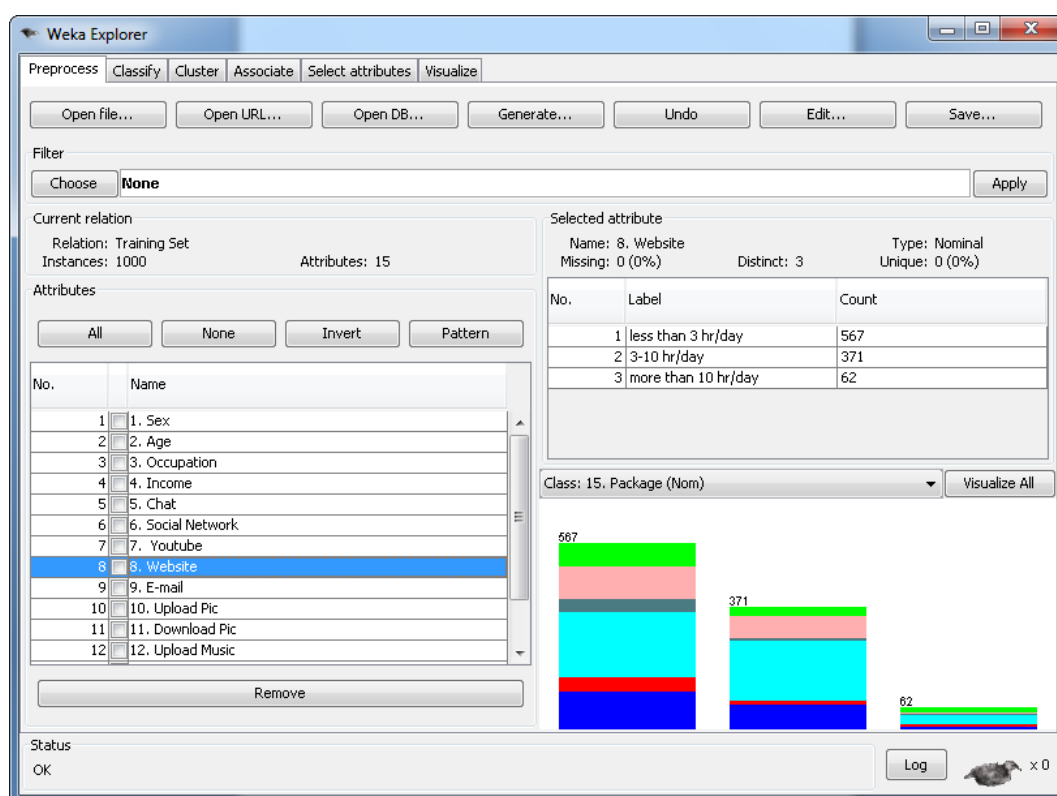
ภาพที่ 3-6 แสดงข้อมูลปริมาณการใช้โปรแกรมสนทนาผ่านอินเทอร์เน็ตบนมือถือ ใน Weka



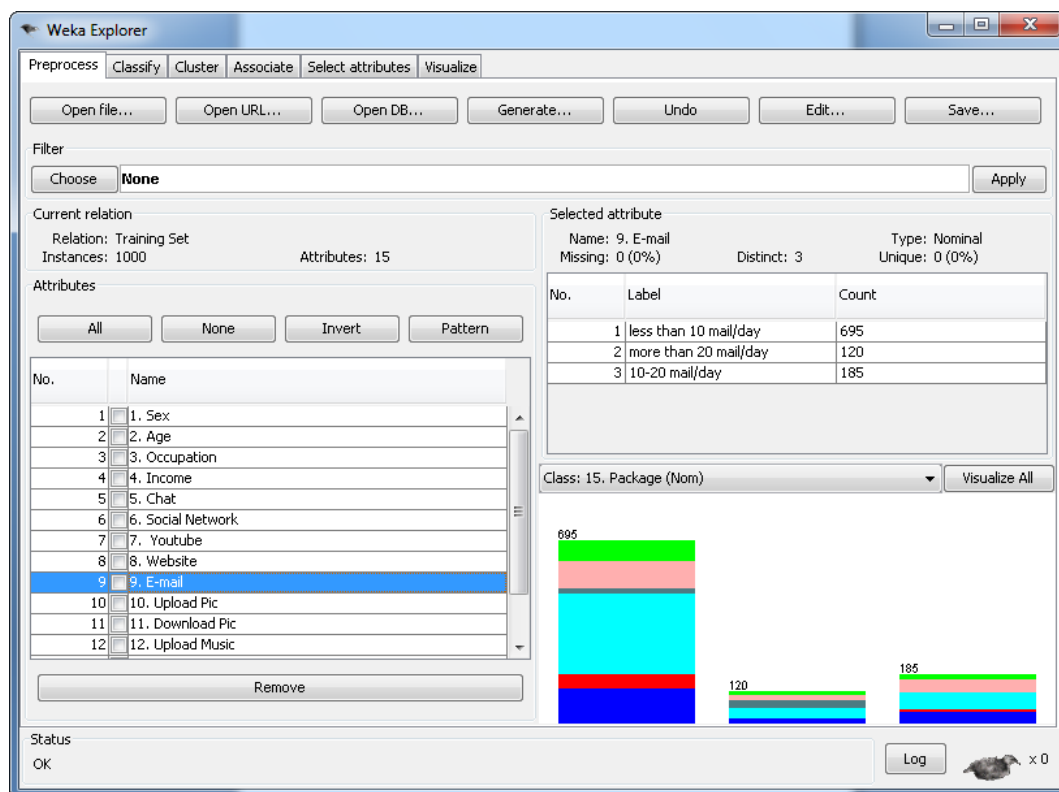
ภาพที่ 3-7 แสดงข้อมูลปริมาณการใช้งาน Social Network ใน Weka



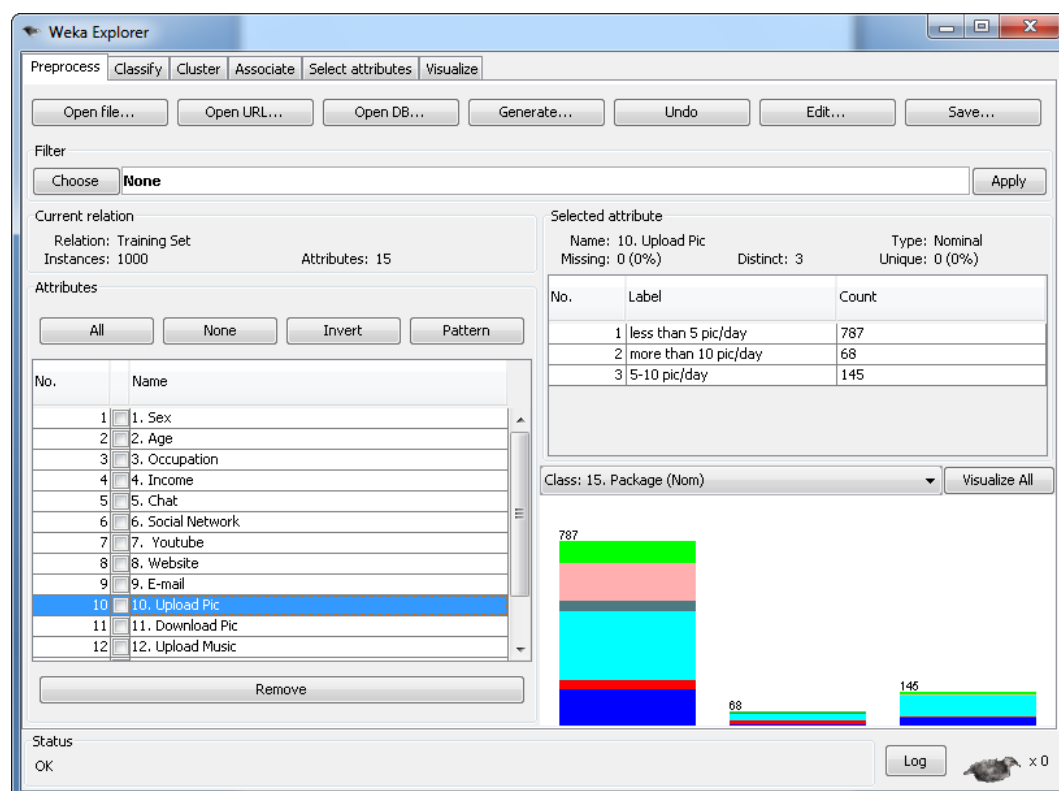
ภาพที่ 3-8 แสดงข้อมูลปริมาณการใช้งาน Youtube ใน Weka



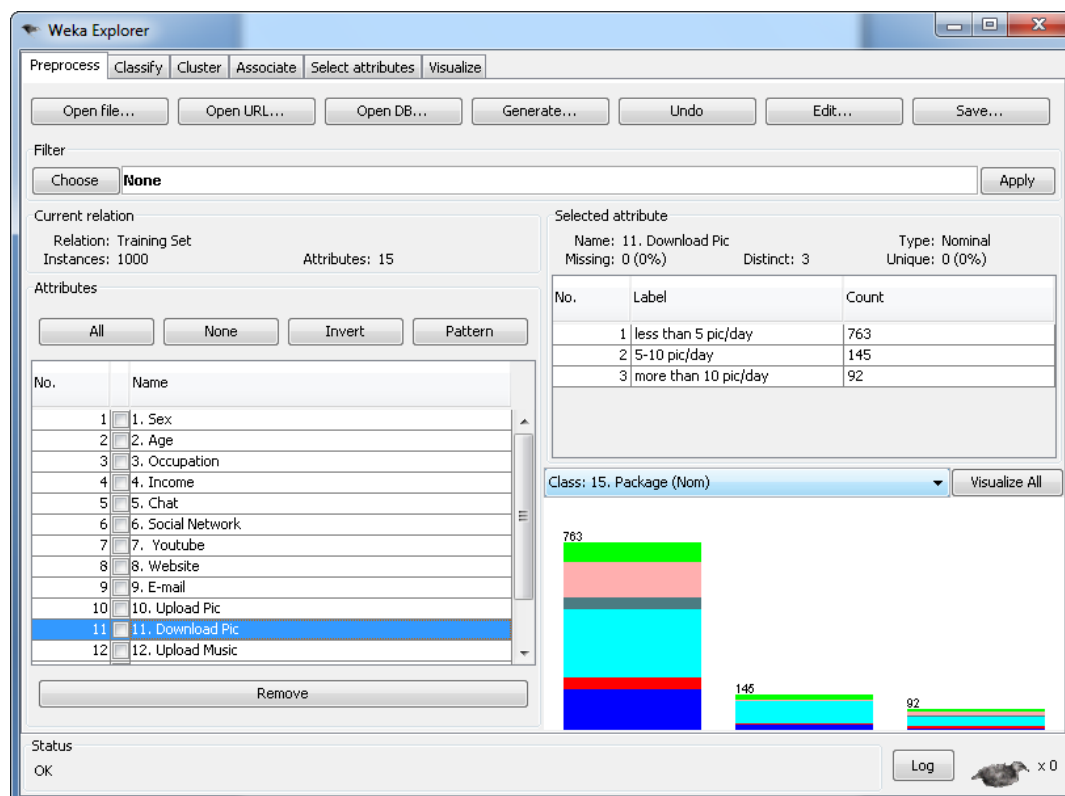
ภาพที่ 3-9 แสดงข้อมูลปริมาณการใช้งาน Website ใน Weka



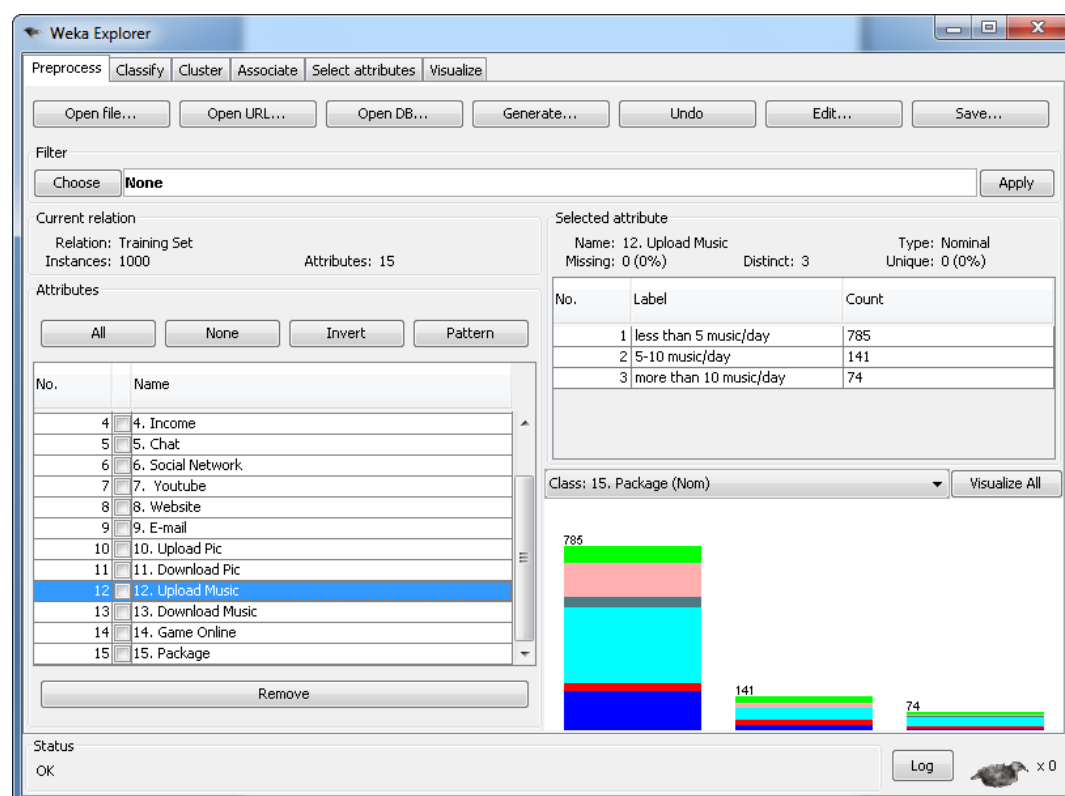
ภาพที่ 3-10 แสดงข้อมูลปริมาณการใช้งาน E-mail ใน Weka



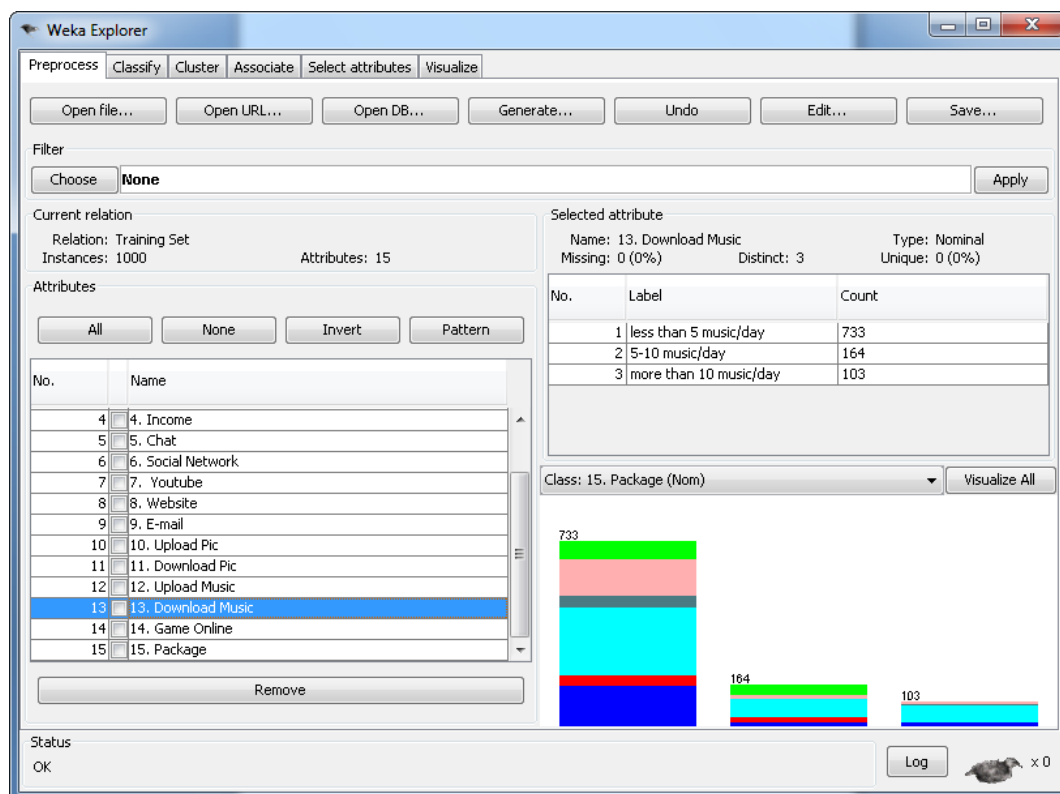
ภาพที่ 3-11 แสดงข้อมูลปริมาณการใช้งาน Upload รูปภาพ ใน Weka



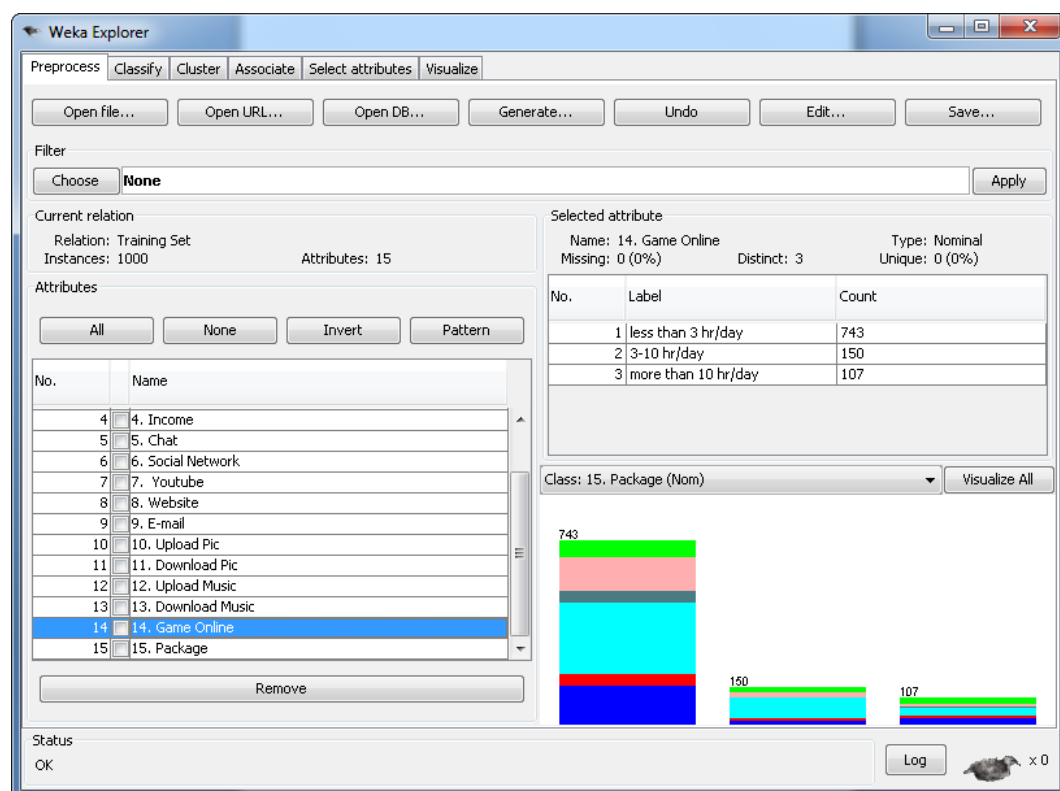
ภาพที่ 3-12 แสดงข้อมูลปริมาณการใช้งาน Download รูปภาพ ใน Weka



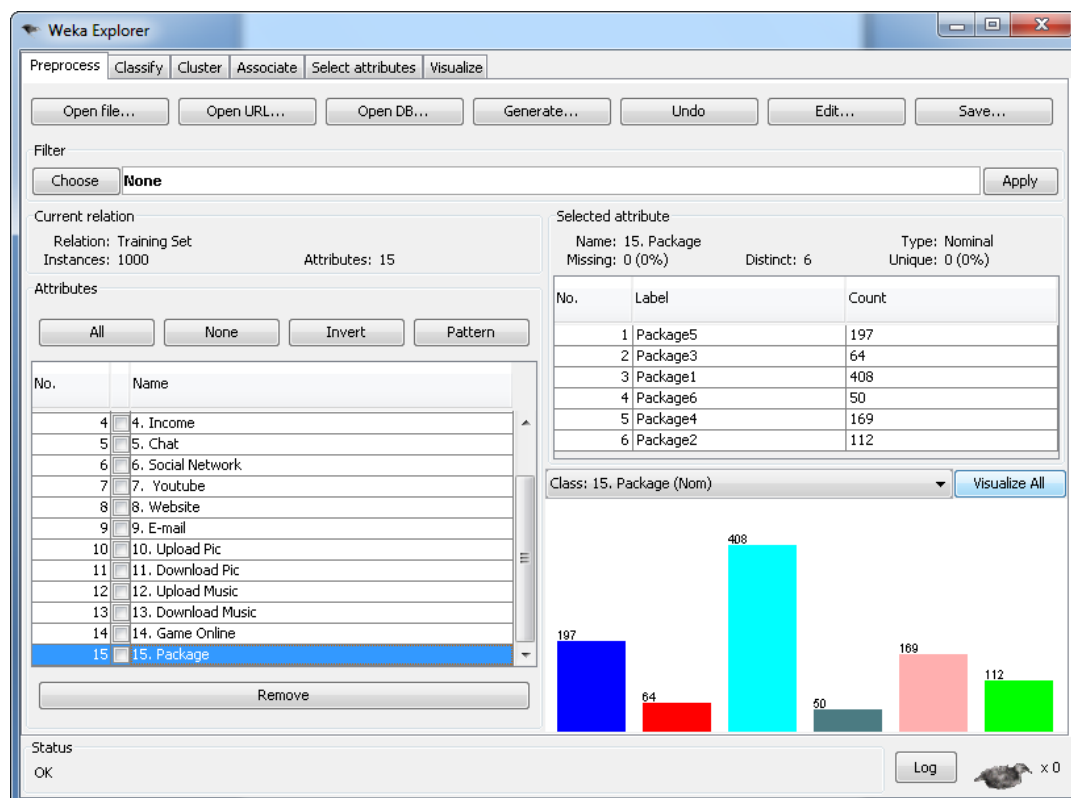
ภาพที่ 3-13 แสดงข้อมูลปริมาณการใช้งาน Upload เพลง ใน Weka



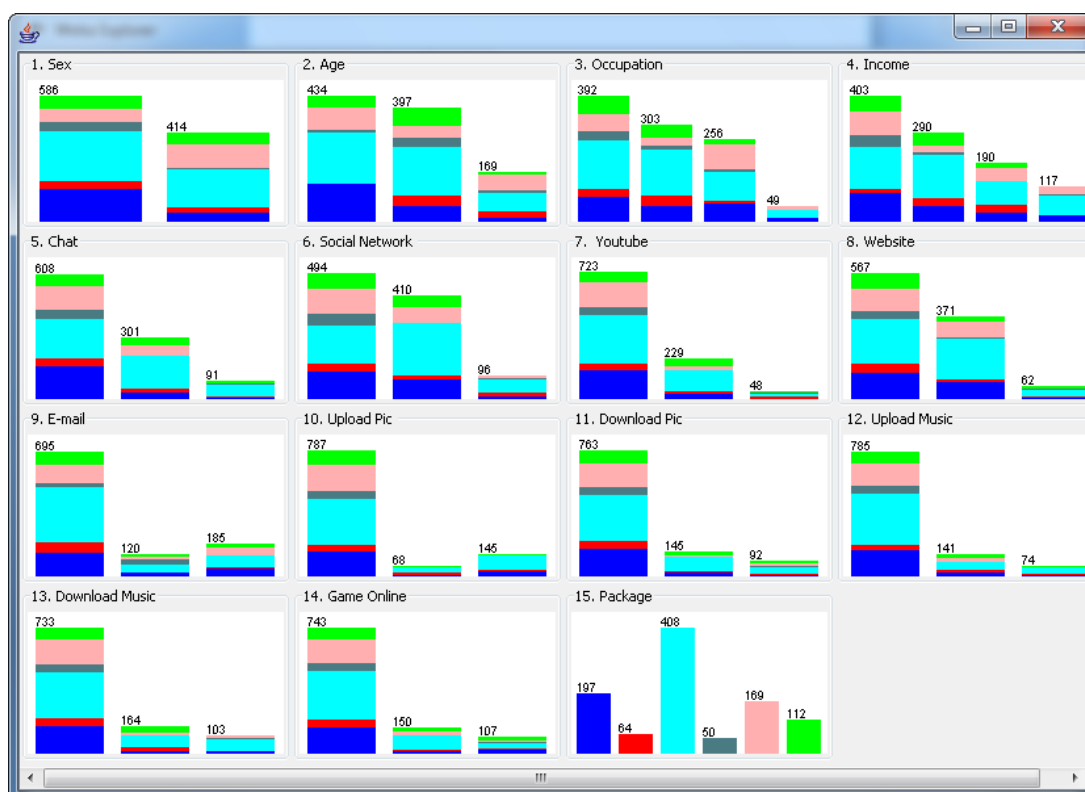
ภาพที่ 3-14 แสดงข้อมูลปริมาณการใช้งาน Download เพลง ใน Weka



ภาพที่ 3-15 แสดงข้อมูลปริมาณการใช้งาน Game Online ใน Weka

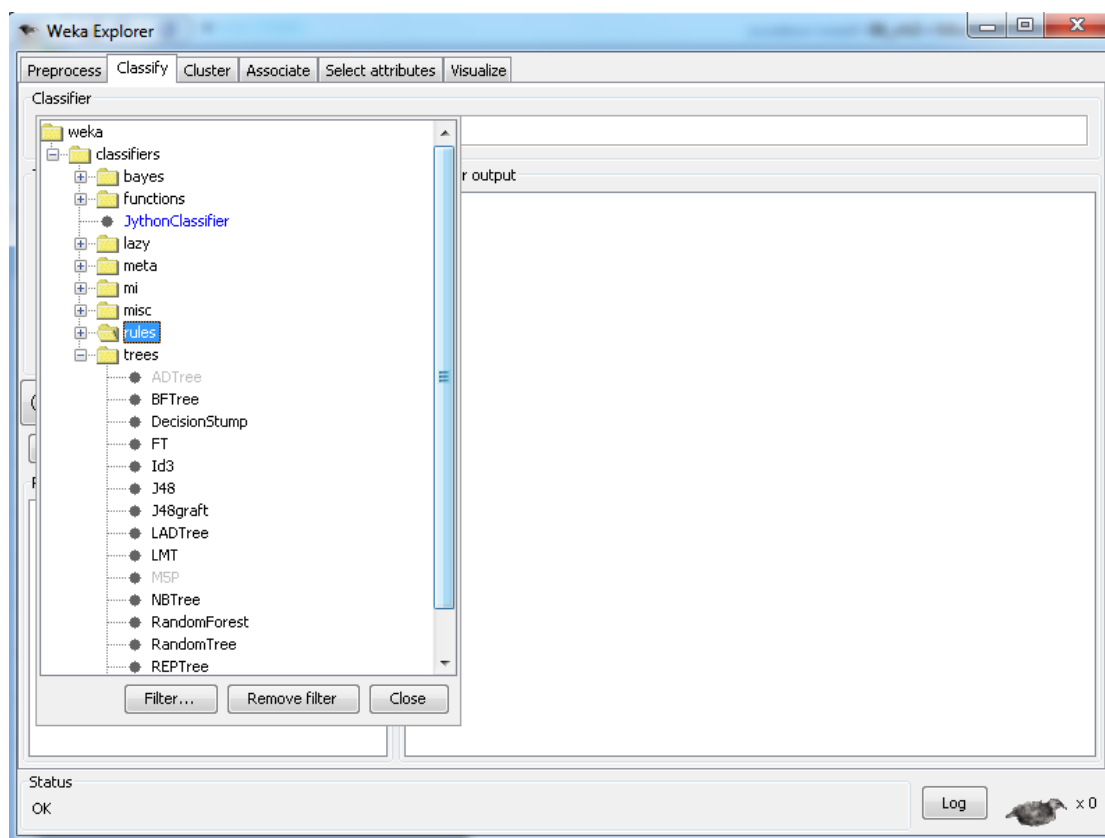


ภาพที่ 3-16 แสดงข้อมูลประเภทแพคเกจอินเทอร์เน็ตที่ผู้ตอบแบบสอบถามเลือกใช้ใน Weka



ภาพที่ 3-17 แสดงลักษณะข้อมูลทั้งหมดที่เตรียมพร้อมสำหรับวิเคราะห์ใน Weka

3.3.2.2 กระบวนการทำงาน (Process) จะใช้ข้อมูลที่นำเข้าที่ทราบค่าของ Class ที่ต้องการ มาสร้างการแบ่งประเภทโดยใช้อัลกอริธึมที่เลือกไว้ โดยทดสอบทีละอัลกอริธึม โดยใช้โปรแกรม Weka ในการทดสอบ ดังภาพที่ 3-18



ภาพที่ 3-18 แสดงอัลกอริธึมที่จะใช้ทดสอบใน Weka

3.2.1.3 ผลลัพธ์ (Output) เมื่อโปรแกรมสร้างรูปแบบหรือโมเดลแล้ว จะแสดงรูปของ Tree ที่ได้จากการประมวลผล

3.2.1.4 และโปรแกรมจะแสดงค่า Confusion Matrix ซึ่งเป็นค่าความแม่นยำของอัลกอริธึม

3.2.1.5 ทดสอบนำเข้าข้อมูล ประมวลผล และแสดงผลลัพธ์ ซึ่งมีค่า Confusion Matrix ให้ครบทุกอัลกอริธึม

3.3.3 นำค่า Confusion Matrix ของแต่ละอัลกอริธึมที่ทำการทดสอบมาเปรียบเทียบกัน เพื่อหาอัลกอริธึมที่มีค่าความแม่นยำมากที่สุด

3.3.4 นำอัลกอริธึมที่มีค่าความแม่นยำมากที่สุดไปใช้ในการสร้างโมเดลและพัฒนาระบบต่อไป

3.4 นำอัลกอริทึมที่ได้มาสร้างโมเดลและพัฒนาระบบ

จากการทดสอบจนได้อัลกอริทึมที่เหมาะสมที่สุดแล้ว ขั้นตอนต่อมาคือการพัฒนากระบวนงานสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ ซึ่งผู้วิจัยได้พัฒนาระบบโดยใช้เทคโนโลยี .NET Framework ในพัฒนาระบบโดยใช้โปรแกรม Microsoft Visual Studio 2008 ซึ่งใช้ภาษา ASP.NET ด้วย C# ในการสร้างหน้าเว็บ และสร้างโมดูลที่เกี่ยวข้องกับระบบงานทั้งหมด และใช้ร่วมกับโปรแกรม Microsoft SQL Server 2005 ในการจัดการและออกแบบฐานข้อมูลของระบบ และแสดงผลลัพธ์ผ่านทางโปรแกรมเว็บเบราว์เซอร์ โดยผู้วิจัยได้นำหลักการของอัลกอริทึมที่ทดสอบได้มาเขียนโปรแกรมจนกระทั่งได้ระบบที่ต้องการ จากนั้นนำข้อมูลที่เก็บจากแบบสอบถามมาเป็นข้อมูลที่ให้ระบบเรียนรู้ (Training Data) จำนวน 1,000 ชุด เพื่อทำการสร้างโมเดลของระบบขึ้น และทำการทดสอบความถูกต้องและความน่าเชื่อถือของโมเดลต่อไป

3.5 การทดสอบความถูกต้องและความน่าเชื่อถือของโมเดล

หลังจากพัฒนาระบบและสร้างโมเดลแล้ว ขั้นตอนต่อมาคือการทดสอบความถูกต้องและความน่าเชื่อถือของโมเดล โดยการนำข้อมูลจากแบบสอบถามเป็นข้อมูลกลุ่มทดสอบ (Testing Data) จำนวน 500 ชุด เพื่อทดสอบความถูกต้องและความน่าเชื่อถือของโมเดล โดยวัดจากค่า ความแม่นยำ Confusion Matrix

3.6 สถิติที่ใช้ในการวิจัย

ในการวิจัยเรื่องระบบสนับสนุนการตัดสินใจในการเลือกใช้แพคเกจอินเทอร์เน็ตมือถือ โดยใช้ต้นไม้ตัดสินใจ ผู้วิจัยเลือกใช้สถิติในการทดสอบหาค่าความแม่นยำของอัลกอริทึมและทดสอบความถูกต้องและความน่าเชื่อถือของโมเดลโดยใช้การประเมินค่าความแม่นยำ Confusion Matrix

Confusion Matrix คือ การประเมินผลลัพธ์การทำนาย (หรือผลลัพธ์จากโปรแกรม) เปรียบเทียบกับผลลัพธ์จริง

โดยที่

a, d คือ จำนวนข้อมูลที่มีในการทำนายถูก

b, c คือ จำนวนข้อมูลที่มีในการทำนายผิด

		<i>PREDICTED CLASS</i>	
		<i>Class=Yes</i>	<i>Class=No</i>
<i>ACTUAL CLASS</i>	<i>Class=Yes</i>	<i>a</i>	<i>b</i>
	<i>Class=No</i>	<i>c</i>	<i>d</i>

ภาพที่ 3-19 Confusion Matrix

แสดงเป็นสมการได้ดังนี้

Sensitivity or Recall คือ ค่าที่บอกว่าโปรแกรมทำนายได้ว่าจริง เป็นอัตราส่วนเท่าไรของจริงทั้งหมด คำนวณได้จากสมการที่ 3-1

$$Recall = \frac{a}{a+b} \quad (3-1)$$

Specificity คือ ค่าที่บอกว่าโปรแกรมทำนายได้ว่าไม่จริง เป็นอัตราส่วนเท่าไรของจริงทั้งหมด คำนวณได้จากสมการที่ 3-2

$$specificity = \frac{d}{d+c} \quad (3-2)$$

Precision คือ ค่าที่บอกว่าโปรแกรมทำนายว่าจริง ถูกต้องเท่าไร คำนวณได้จากสมการที่ 3-3

$$precision = \frac{a}{a+c} \quad (3-3)$$

Correctly Classified Instances คือ ค่าที่บอกว่ามีการทำนายข้อมูลถูกต้อง และมีค่าความแม่นยำเท่าไร ในการทำนายคำนวณได้จากสมการที่ 3-4

$$\text{Correctly Classified Instances} = \frac{(a+d)}{(a+b+c+d)} \quad (3-4)$$