

บทที่  
3

# มารูจัก Weka กันเถอะ !!

ในบทนี้ผมจะมาแนะนำถึงซอฟต์แวร์ open source ตัวหนึ่งที่ใช้ในการวิเคราะห์ข้อมูลด้วยเทคนิคการทำเหมืองข้อมูล ซอฟต์แวร์นี้ชื่อว่า “Weka” (อ่านว่า เวก้า) โดย weka นั้นได้รวบรวมเทคนิคในการวิเคราะห์ข้อมูลหลายๆ เทคนิคเข้าไว้ด้วยกัน แถมยังสามารถวิเคราะห์ข้อมูลเหล่านี้ได้ง่ายๆ ผ่านทางหน้าจอ GUI (Graphic User Interface) ของ Weka อีกด้วย ซึ่งนับว่าทำให้ผู้ใช้สามารถทำงานได้อย่างสะดวกสบายมากขึ้นอีกเยอะเลยครับ

## 3.1 ติดตั้งซอฟต์แวร์ Weka

คำว่า Weka ย่อมาจากคำว่า *Waikato Environment for Knowledge Analysis* พัฒนาขึ้นโดยใช้ภาษา java ครับ เห็นชื่อเต็มๆ แล้วอาจจะงงกับคำว่า Waikato ว่าแปลว่าอะไร คำนี้เป็นชื่อของมหาวิทยาลัยหนึ่งที่ตั้งอยู่ในประเทศนิวซีแลนด์ ครับ เพราะผู้พัฒนา Weka นี้เป็นอาจารย์ในมหาวิทยาลัยแห่งนี้ ท่านผู้อ่านสามารถดาวน์โหลดโปรแกรม Weka มาใช้ฟรีๆ ได้จาก <http://www.cs.waikato.ac.nz/ml/weka/> (ดังรูปที่ 3-1) ซึ่งในเว็บไซต์นั้นสามารถเลือกดาวน์โหลดซอฟต์แวร์ไปติดตั้งได้ทั้งบนระบบปฏิบัติการ Windows, Linux และ Mac OS เรียกว่ามีบริการให้อย่างครบถ้วนเลยทีเดียวครับ



รูปที่ 3-1 เว็บไซต์อย่างเป็นทางการของซอฟต์แวร์ Weka

ในส่วนของดาวน์โหลดโปรแกรมถ้าเลือกใช้นระบบปฏิบัติการ Windows จะสามารถเลือกดาวน์โหลดเวอร์ชัน ที่มี Java Virtual Machine มาให้หรือว่าจะเลือกเวอร์ชันที่มีเฉพาะ Weka เพียงอย่างเดียวก็ได้ ซึ่งในแบบหลังนี้ เครื่องของท่านผู้อ่านจะต้องมีการติดตั้งโปรแกรม Java Virtual Machine ไว้ก่อนแล้ว แต่สำหรับระบบปฏิบัติการอื่นๆ นั้น เครื่องของท่านผู้อ่านจะต้องติดตั้ง Java Virtual Machine ไว้เสียก่อน

หลังจากดาวน์โหลดเรียบร้อยแล้วก็ทำการติดตั้งตามปกติ ซึ่งในที่นี้ขอแสดงการติดตั้งโปรแกรมสำหรับระบบปฏิบัติการ Mac OS (ซึ่งสามารถใช้ได้กับระบบปฏิบัติการลินุกซ์ (Linux)) เป็นตัวอย่างนะครับ

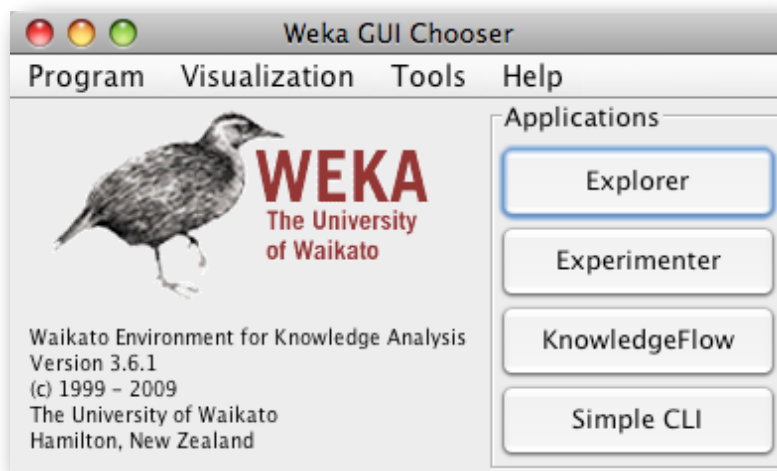
หลังจากดาวน์โหลดไฟล์เรียบร้อยแล้วจะได้ไฟล์ที่ชื่อ `weka-3-6-1.zip` จากนั้นให้ทำการแตกไฟล์ที่อยู่ใน zip ออกมาด้วยคำสั่ง

```
unzip weka-3-6-1.zip
```

และจะได้ไดเรกทอรีที่ชื่อ `weka-3-6-1` ออกมา หลังจากนั้นให้เปลี่ยนไปยังไดเรกทอรี `weka-3-6-1` แล้วใช้คำสั่ง

```
java -jar weka.jar
```

เพื่อทำการเรียกซอฟต์แวร์ Weka ให้ทำงาน ซึ่งหน้าต่าง Weka GUI Chooser จะปรากฏออกมาให้เห็นดังรูปที่ 3-2



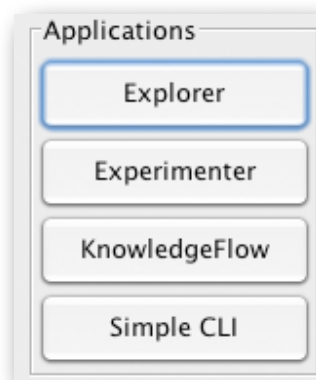
รูปที่ 3-2 หน้าต่าง Weka GUI Chooser

ส่วนการเรียกโปรแกรมใน Windows นั้นสามารถทำได้ 2 แบบ คือ

- เรียกผ่าน Start > Programs > Weka 3.4.12 > Weka หลังจากนั้นจะปรากฏหน้าต่างของ Weka GUI Chooser ขึ้นมา
- เรียกผ่าน command line โดยการเข้าไปที่ไดเรกทอรีที่เก็บ Weka อยู่หลังจากนั้น ใช้ คำสั่ง
- java -jar weka.jar (เหมือนกับระบบปฏิบัติการ ลินุกซ์ และ Mac OSX) หลังจากนั้นจะปรากฏหน้าต่างของ Weka GUI Chooser ขึ้นมาดังในรูปที่ 3-2

### 3.2 เริ่มต้นใช้งาน Weka

เมื่อนำหน้าต่าง Weka GUI Chooser ปรากฏขึ้นมาแล้วหลายคนอาจจะงงว่าแล้วจะใช้ Weka อย่างไร จะสังเกตเห็นว่าที่หน้าต่าง Weka GUI Chooser นี้จะมีเมนูหลักๆ ให้เลือก 4 อันคือ (รูปที่ 3-3)



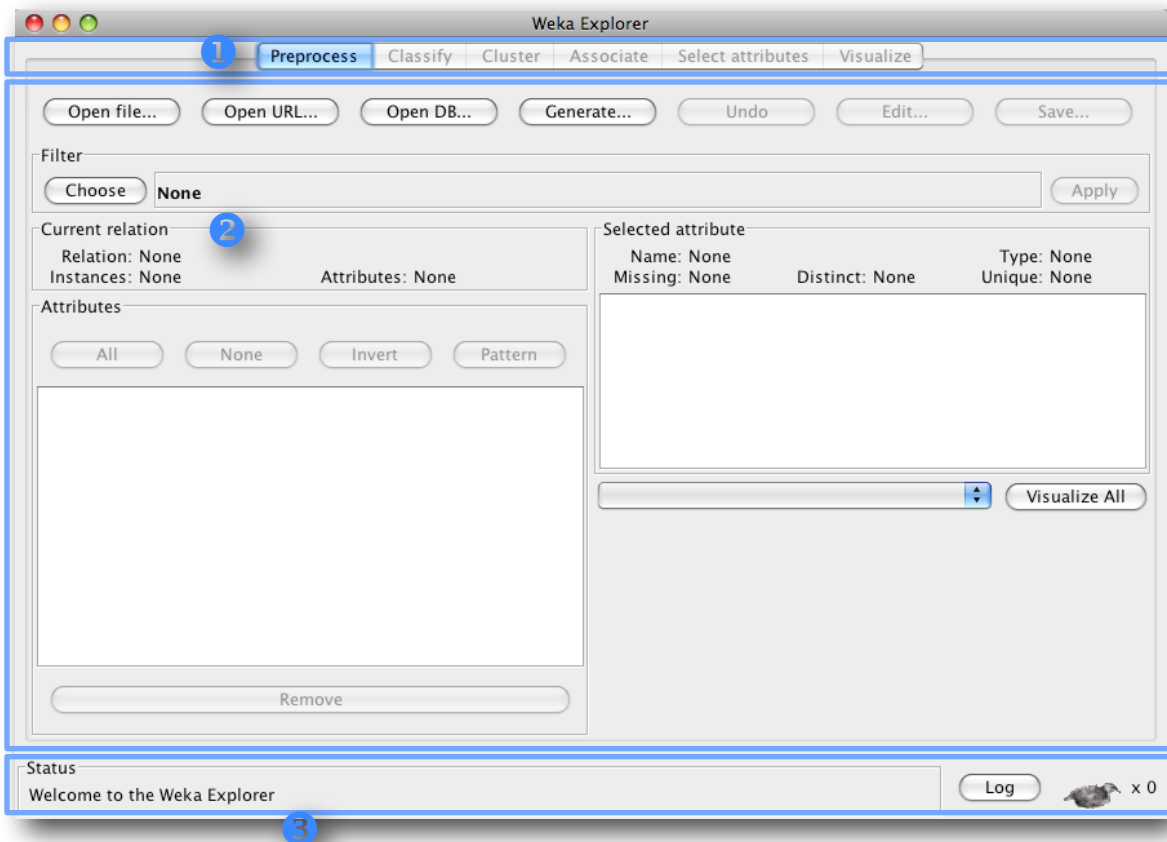
รูปที่ 3-3 เมนูหลักในหน้าต่าง Weka GUI Chooser

ตารางที่ 3-1 รายละเอียดของเมนูต่างๆ ในซอฟต์แวร์ Weka

<i>Explorer</i>	<i>Explorer</i> เป็นส่วนที่ผู้ใช้สามารถใช้ฟังก์ชันการทำงานต่างๆ ของ Weka ผ่านทางหน้าจอ GUI ซึ่งเป็นส่วนที่เหมาะสมสำหรับผู้เริ่มต้นใช้งาน Weka เพราะผู้ใช้จะสามารถเรียกฟังก์ชันการทำงานต่างๆ ได้เพียงแค่คลิกและเปลี่ยนแปลงพารามิเตอร์ในหน้าฟอร์มเท่านั้น ต้องยอมรับว่าส่วนนี้เป็นจุดขายอีกอย่างของ Weka ที่ช่วยให้ผู้ใช้สามารถวิเคราะห์ข้อมูลได้อย่างสะดวกมากยิ่งขึ้น
<i>Experimenter</i>	<i>Experimenter</i> เป็น GUI อีกส่วนหนึ่งที่ยอมให้ผู้ใช้สามารถทดลองเปลี่ยนแปลงเทคนิคในการวิเคราะห์ข้อมูลและค่าพารามิเตอร์ต่างๆ ที่เกี่ยวข้องได้หลากหลายรูปแบบ จนได้ผลลัพธ์ที่น่าพอใจ ซึ่งการใช้ <i>Explorer</i> นั้นไม่สะดวกเนื่องจากต้องมาทำการเปลี่ยนแปลงค่าต่างๆ เอง
<i>Knowledge Flow</i>	<i>Knowledge Flow</i> เป็น GUI อีกส่วนหนึ่งที่ยอมให้ผู้ใช้สามารถนำเทคนิคต่างๆ ของ Weka มาเรียงต่อกันเพื่อช่วยให้การวิเคราะห์ข้อมูลทำงานได้ตามที่ต้องการ

<p><i>Simple CLI</i></p>	<p><i>Simple CLI</i> เป็นส่วนที่ผู้ใช้สามารถเรียกฟังก์ชันการทำงานของ Weka มาใช้ผ่านทาง command line ได้ ซึ่งการเรียกใช้ฟังก์ชันผ่านทาง command line นี้จะช่วยให้ผู้ใช้เข้าใจการเรียกฟังก์ชันต่างๆ เบื้องหลังหน้าจอ GUI ของ Weka ได้ และสามารถนำไปประยุกต์ใช้ในการเขียนโปรแกรมเพื่อเรียก Weka ในการใช้งานได้อีกด้วย</p>
--------------------------	--

ในบทที่ 3 นี้จะแนะนำการใช้งานส่วน Explorer เป็นหลักนะครับ เมื่อคลิกที่ปุ่ม Explorer จะปรากฏหน้าจอ *Weka Explorer* ขึ้นมา ดังในรูปที่ 3-4 ในหน้าต่าง *Weka Explorer* นี้สามารถแบ่งพื้นที่ได้เป็น 3 ส่วนใหญ่ๆ คือ

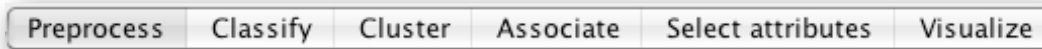


รูปที่ 3-4 เมนูหลักในหน้าต่าง Weka Explorer

- ❶ ส่วนบนสุดจะเป็นแท็บ (tab) ซึ่งมีด้วยกันทั้งหมด 6 แท็บวางเรียงกันอยู่ทางด้านบน ซึ่งแท็บต่างๆ เหล่านี้จะเป็นเมนูให้ผู้ใช้สามารถใช้งานเทคนิคต่างๆ ของ Weka ได้นั่นเอง
- ❷ ส่วนที่อยู่ตรงกลางซึ่งจะเปลี่ยนไปตามการกดแท็บต่างๆ ส่วนนี้เป็นส่วนของการเลือก option ต่างๆ ในการวิเคราะห์ข้อมูล และส่วนการแสดงผลลัพธ์หลังจากทำการวิเคราะห์ข้อมูลเสร็จแล้ว ส่วนนี้บางครั้งจะเรียกว่า Workspace

③ ส่วนที่อยู่ด้านล่างสุด จะเป็นส่วนที่บอกสถานะ (status) ของการทำงานในแต่ละขั้นตอน และปุ่มทางขวามือจะเป็นปุ่มสำหรับดู Log ไฟล์ เมื่อคลิกจะแสดงหน้าจอของ log ขึ้นมา

ในบทนี้จะอธิบายถึงภาพรวมของ Weka Explorer ก่อน แล้วจึงลงรายละเอียดในการทำงานแต่ละส่วนในบทถัดๆ ไป ดังนั้น ส่วนนี้ผมจะขอแนะนำเมนูที่อยู่ใน แท็บ ทั้ง 6 แท็บนี้เสียก่อน ดังในรูปที่ 3-5 นะครับ



รูปที่ 3-5 แท็บเมนูหลักในหน้าต่าง Weka Explorer

ตารางที่ 3-2 รายละเอียดของแท็บเมนูต่างๆ ในหน้าต่าง Weka Explorer

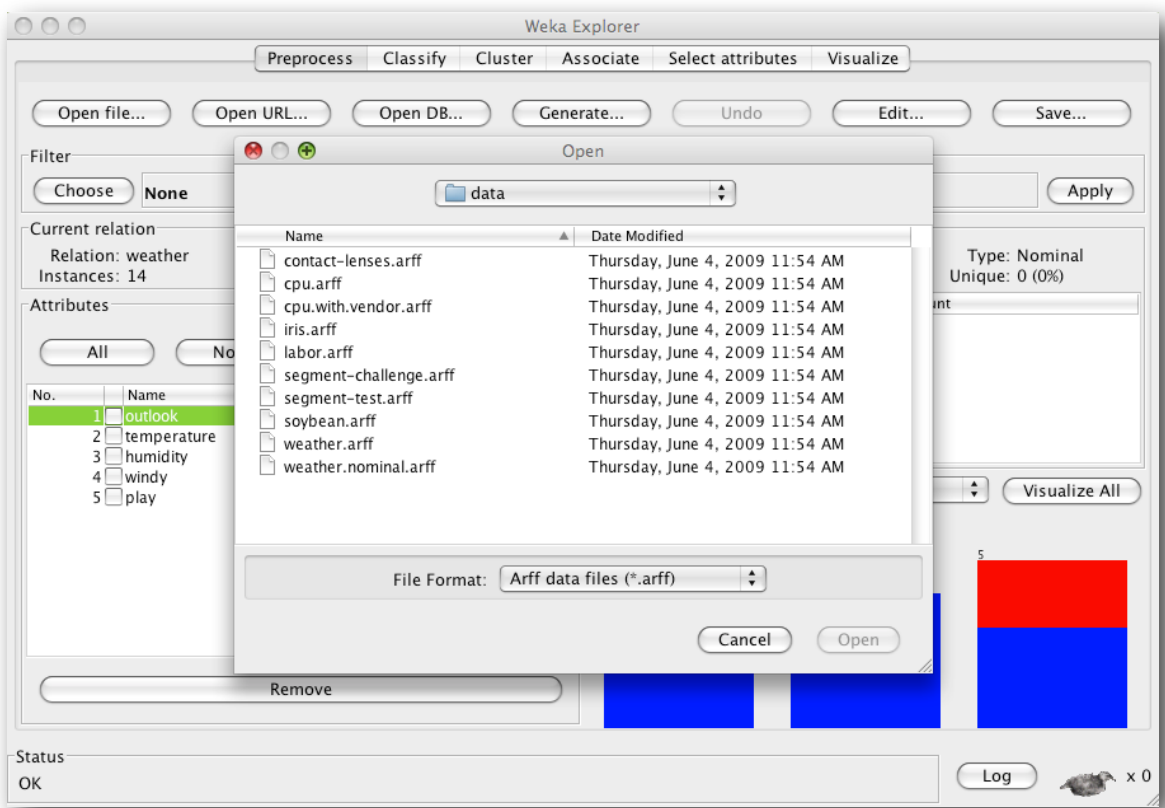
<i>Preprocess</i>	เป็นส่วนที่ใช้ในการเลือกไฟล์ข้อมูลสำหรับเป็นอินพุท (input) เพื่อทำการวิเคราะห์ด้วยเทคนิคต่างๆ ต่อไป นอกจากนั้นยังมีวิธีการต่างๆ ที่ช่วยในการจัดการข้อมูลต่างๆ เบื้องต้น (รายละเอียดจะกล่าวถึงในบทที่ 4)
<i>Classify</i>	เป็นส่วนที่ใช้ในการวิเคราะห์ข้อมูลด้วยวิธีการจำแนกข้อมูล (classification) หรือทำนายข้อมูล (prediction) ซึ่งจะมีวิธีการต่างๆ ให้เลือกมากมาย (รายละเอียดจะกล่าวถึงในบทที่ 5)
<i>Cluster</i>	เป็นส่วนที่ใช้ในการวิเคราะห์ข้อมูลด้วยวิธีการจัดกลุ่มข้อมูล (clustering) โดยจะจัดกลุ่มข้อมูลที่มีลักษณะคล้ายๆ กันหรือมีความสัมพันธ์กันเข้าไว้ด้วยกัน (รายละเอียดจะกล่าวถึงในบทที่ 6)
<i>Associate</i>	เป็นส่วนที่ใช้ในการวิเคราะห์ข้อมูลด้วยวิธีการหาความสัมพันธ์ของข้อมูลที่เก็บอยู่ในฐานข้อมูล (association rule) เพื่อดูว่าข้อมูลแบบไหนที่มีโอกาสเกิดร่วมกันบ่อยๆ (รายละเอียดจะกล่าวถึงในบทที่ 7)
<i>Select attributes</i>	เป็นส่วนที่คล้ายๆ กับส่วน Preprocess แต่จะเน้นที่การหาว่าตัวแปรไหนที่สำคัญและไม่สำคัญในชุดข้อมูลบ้าง ซึ่งตัวแปรที่ไม่สำคัญนี้จะถูกกำจัดทิ้งไปก่อนที่จะวิเคราะห์ข้อมูลด้วยวิธีต่างๆ
<i>Visualize</i>	เป็นส่วนของการ plot จุดข้อมูลในรูปแบบ 2 มิติ

ตอนนี้ท่านผู้อ่านก็คงพอรู้จัก Weka Explorer กันสักเล็กน้อยแล้วนะครับ ต่อไปผมจะขอยกตัวอย่างการใช้งาน Weka Explorer เพื่อให้ท่านผู้อ่านเข้าใจการใช้งานมากยิ่งขึ้นอีกสักนิดนะครับ

**ตัวอย่างที่ 3-1 : มาแข่งขันกีฬากันเถอะ**

ตอนนี้อากาศก็เริ่มเปลี่ยนแปลงแล้วนะครับ อยู่ดีๆ อากาศก็หนาวขึ้นมาผิดปกติ ดังนั้นในการแข่งขันกีฬาประเภทต่างๆ เรื่องของดินฟ้าอากาศถือเป็นเรื่องสำคัญเรื่องหนึ่งที่ต้องคำนึงถึง เพราะถ้าวันไหนฝนตกและลมแรงมาก ๆ ก็คงจะแข่งไม่สนุกแน่ๆ จะเป็นการดีกว่าไหมครับถ้าท่านผู้อ่านรู้ว่าวันนี้ควรจะแข่งกีฬาหรือไม่

ที่ผมเกริ่นมานี้เพื่อจะยกตัวอย่าง (ขออย่าว่าเป็นข้อมูลตัวอย่างจริงๆ นะครับ) การทำนายข้อมูล (Data Prediction) โดยใช้ Weka ครับ การทำนายข้อมูลล่วงหน้าจะต้องมีข้อมูลย้อนหลังที่น่าเชื่อถือได้เก็บไว้เสียก่อน (ข้อมูลเหล่านี้จะเรียกว่า training data) แล้วจึงนำข้อมูลที่เก็บไว้เหล่านี้มาทำการวิเคราะห์และสร้างโมเดล (model) เพื่อเก็บไว้สำหรับใช้ทำนายเมื่อมีข้อมูลใหม่เข้ามา (ข้อมูลใหม่ที่เข้ามาเรียกว่า testing data) โดยในตัวอย่างนี้ผมจะขอเสนอการทำนายข้อมูลด้วยวิธี Decision Tree ซึ่งเป็นวิธีที่เข้าใจค่อนข้างง่าย เริ่มแรกหลังจากที่เราเปิดหน้าต่าง Weka Explorer ขึ้นมาแล้ว



**รูปที่ 3-6** เลือกไฟล์อินพุต

① ให้ทำการเลือกไฟล์อินพุต (input) โดยการคลิกที่ปุ่ม **Open file...** แล้วเลือกไฟล์ที่ชื่อ weather.arff ซึ่งอยู่ในไดเรกทอรี Weka-3-6/data ดังในรูปที่ 3-6 ซึ่งมีข้อมูลทั้งหมด 14 แถวและมี 5 แอตทริบิวต์ (attribute) ขอให้สังเกตสัญลักษณ์ว่า ภายในไฟล์จะมีข้อความที่เขียนว่า @relation, @attribute และ @data อยู่ด้วย ดังรูปที่ 3-7

```

weather.arff
@relation weather

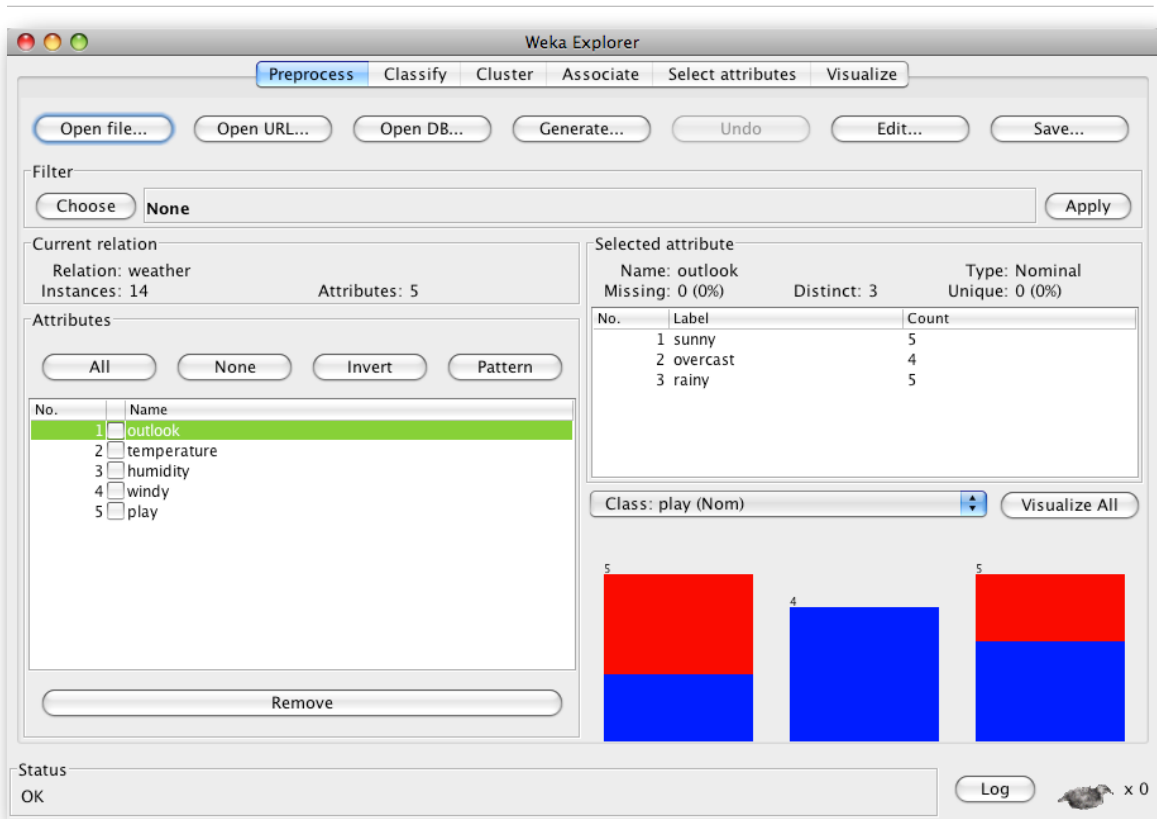
@attribute outlook {sunny, overcast, rainy}
@attribute temperature real
@attribute humidity real
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
    
```

รูปที่ 3-7 ไฟล์อินพุต

- @relation จะใช้ในการบอกชื่อของข้อมูลชุดนี้
- @attribute จะใช้ในการบอกว่าในข้อมูลชุดนี้มีแอตทริบิวต์ชื่ออะไรบ้าง และในแต่ละแอตทริบิวต์นั้นมีข้อมูลเป็นแบบไหน เช่น เป็นจำนวนจริง (real) หรือเป็นเซตของข้อมูล ({sunny, overcast, rainy}) และแอตทริบิวต์สุดท้ายจะเป็นสิ่งที่เราต้องการหาคำตอบ ซึ่งในที่นี้คือ ควรจะเล่นกีฬาหรือไม่
- @data ถัดลงมาจากบรรทัดนี้จะเป็นข้อมูลที่ใช้ในการวิเคราะห์

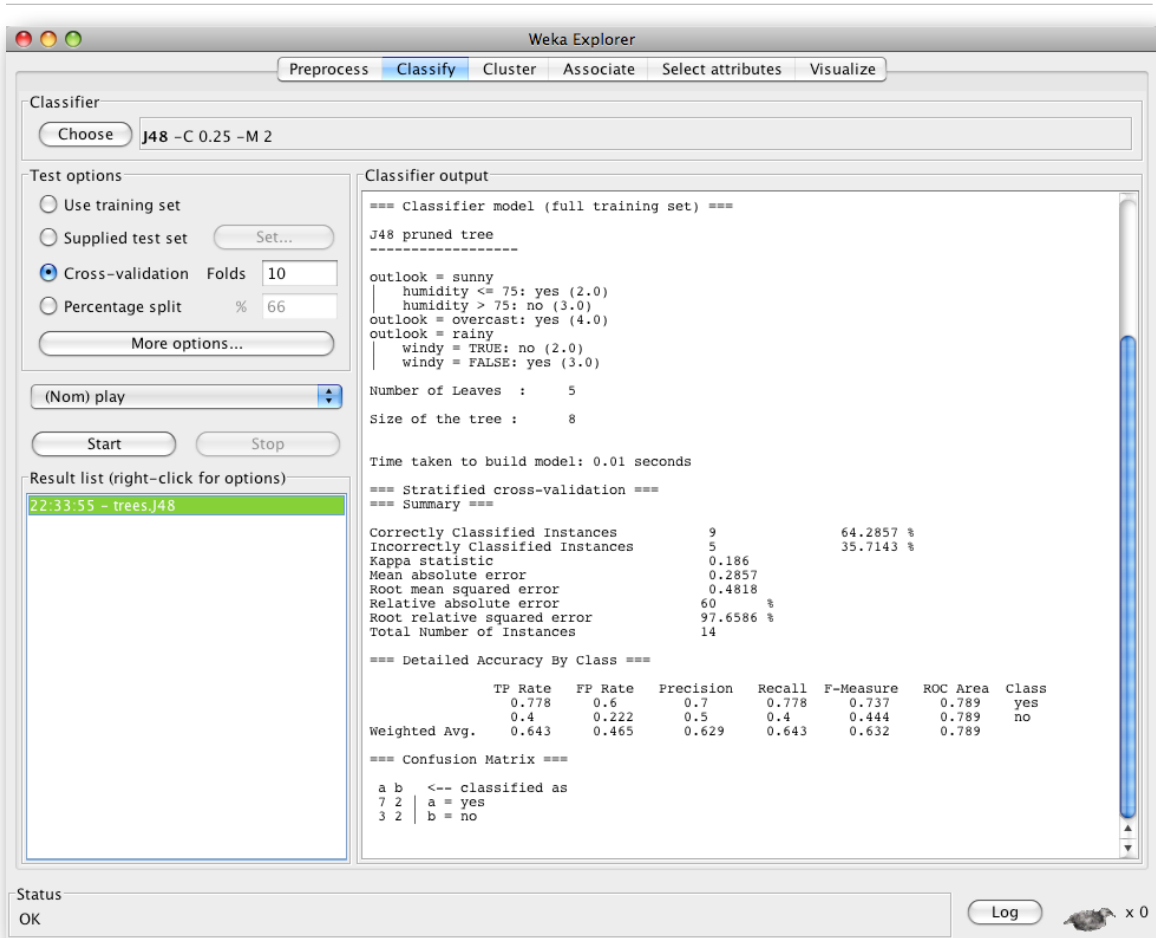
ไฟล์ข้อมูลที่ใช้เป็นอินพุตสำหรับ Weka จะต้องมีรูปแบบเช่นนี้ซึ่งทางผู้พัฒนาเรียกว่าไฟล์แบบ ARFF (Attribute-Relation File Format) ซึ่งในตัวอย่างนี้ผมจะใช้ข้อมูลนี้เป็นข้อมูลในการเรียนรู้ (training data) เพื่อสร้างโมเดลในการทำนายข้อมูลในอนาคต หลังจากเลือกไฟล์ที่ต้องการได้แล้วหน้าต่างในส่วน workspace ก็จะไปดังรูปที่ 3-8



รูปที่ 3-8 หน้าต่างหลังทำการโหลดไฟล์อินพุต

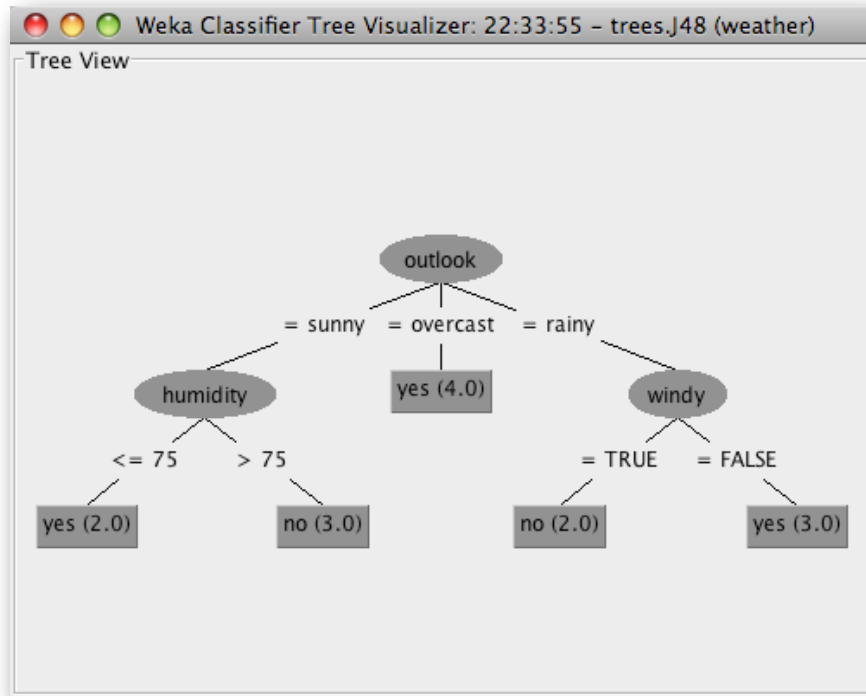
② คลิกที่แท็บ **Classify** หน้าตาของ workspace จะเปลี่ยนไปเป็นส่วนของแท็บ Classify แล้วคลิกที่ปุ่ม **Choose** จะมีลิสต์ (list) แสดงเทคนิคต่างๆ ของการ Classify ให้คลิกเลือกที่เมนู *trees* หลังจากนั้นให้เลือกที่เทคนิคการจำแนกข้อมูลด้วย J48 โดยคลิกที่เมนู *J48* หลังจากนั้นคลิกที่ปุ่ม **Start** Weka จะทำงานและแสดงผลการทำงานของงานดังในรูปที่ 3-9





รูปที่ 3-9 หน้าต่างผลลัพธ์การ Classify

3 ส่วนของ *Classifier output* ทางด้านขวาจะเปลี่ยนไป (รูปที่ 3-9) โดยจะแสดงผลการวิเคราะห์ในรูปแบบของ decision tree และข้อมูลทางสถิติอื่นๆ เช่น ค่าความถูกต้องในการทำนาย (Correctly Classified Instances) เป็นต้น ซึ่งข้อมูลที่เห็นนี้เป็นผลที่ได้จากการวิเคราะห์ข้อมูลด้วยเทคนิค J48 ซึ่งจะกล่าวถึงรายละเอียดต่อไปในบทที่ 4



รูปที่ 3-10 ต้นไม้ช่วยตัดสินใจ (decision tree)

④ ในช่อง **Result list** ทางด้านซ้ายเมื่อเราทำการคลิกขวาและเลือกเมนู **Visualize tree** จะปรากฏหน้าจอ **Tree View** ขึ้นมา ซึ่งต้นไม้ (tree) ที่แสดงนี้จะเรียกว่า ต้นไม้ช่วยตัดสินใจ หรือ decision tree ซึ่งเป็นต้นไม้ที่ใช้ในการสร้างกฎเพื่อนำไปทำนายข้อมูลใหม่ในอนาคต ตัวอย่างของกฎที่สร้างได้จากต้นไม้ช่วยตัดสินใจนี้ เช่น ถ้ามีฝนตก (rainy) และมีลมแรง (windy) ก็ไม่ควรจะเล่นกีฬา เป็นต้น นอกจากนี้ยังสามารถทำการ save model นี้ไว้ใช้ในการทำนายต่อไปได้โดยการคลิกขวาแล้วเลือกที่เมนู **Save model**

สุดท้ายนี้ผมหวังว่าท่านผู้อ่านคงได้เห็นภาพรวมการใช้งาน Weka กันบ้างแล้วซึ่งก็ไม่ได้ยากเท่าไรนัก แต่สิ่งที่ยากก็คือการนำแนวคิดเหล่านี้ไปประยุกต์ใช้กับสิ่งที่ท่านผู้อ่านต้องการวิเคราะห์