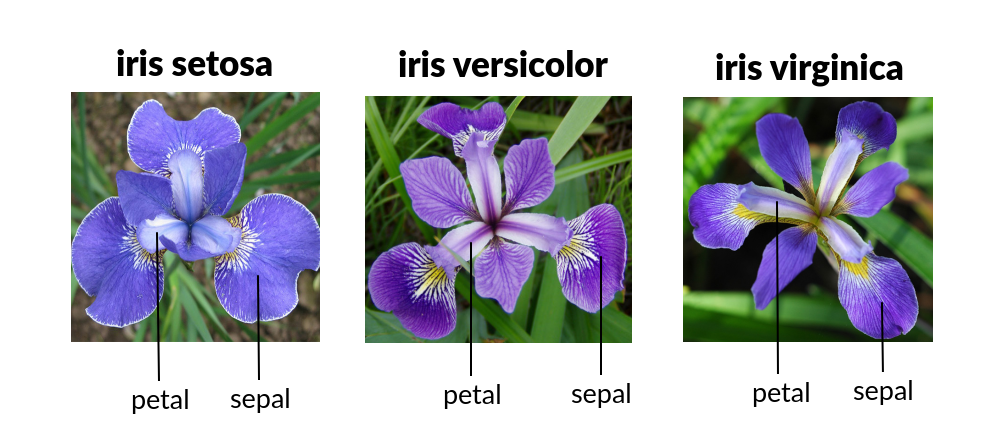
**วิเคราะห์ฐานข้อมูล iris ด้วยโปรแกรม Weka**

Iris dataset เป็นฐานข้อมูลเกี่ยวกับดอกได้ iris กว่า 150 ดอก โดยรายละเอียดจะมี ความยาวกลีบเลี้ยง (Sepal length), ความกว้างกลีบเลี้ยง (Sepal width), ความยาวกลีบดอก (petal length) และ ความกว้างกลีบดอก (Petal width) หน่วยวัดเป็นหน่วยเซนติเมตร (cm) และ คลาสสายพันธุ์ (class) ของดอก iris แบ่งเป็น 3 กลุ่ม คือ Setosa, Versicolor และ Virginiga.



ภาพจาก https://kongruksiamza.medium.com

ตัวอย่างชุดข้อมูล

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| sepal\_length | sepal\_width | petal\_length | petal\_width | class |
| 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa |
| 4.9 | 3 | 1.4 | 0.2 | Iris-setosa |
| 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa |
| 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa |
| 5 | 3.6 | 1.4 | 0.2 | Iris-setosa |

Attribute/ features = 5 (sepal length, sepal width, petal length, petal width, class)

Class Lebel to predict = 3 (Iris Setosa, Iris Versicolor, Iris Virginica)

Instances = 150 records (แต่ละ class จะมี 50 records) และไม่มี missing value

Graphical user interface, application, Word

Description automatically generated

โปรแกรม weka เป็นซอฟต์แวร์ที่ใช้ทำ machine learning, pattern recognition หรือ data mining เป็นต้น เริ่มการทำงานโดยกดที่ปุ่ม Explorer ในหน้าต่างหลัก

A screenshot of a computer

Description automatically generated

ใช้เปิดไฟล์ที่โปรแกรมรองรับ เช่น \*.csv , \*.arff เป็นต้น สามารถลองเปิดไฟล์ที่ตัวโปรแกรมแถมมาได้ที่ C:\Program Files\Weka-3-8-5\data

A screenshot of a computer

Description automatically generated

หน้าต่างแทบ Preprocess

แทบ Class ใช้แสดงการกระจายตัวของฟีเจอร์ที่เลือกจากแทบ Attribute โดยแกนนอนแทนค่าของฟีเจอร์จากน้อยไปมาก แกนตั้งแทนค่าแทนจำนวน สีแทนชั้นข้อมูล (class)

แทบ Selected attribute ใช้แสดงใช้แสดงค่าสถิติของฟีเจอร์ที่เลือกจาก แทบ Attribute

แทบ เปิดไฟล์

แทบ Attribute ใช้แสดงฟีเจอร์ของไฟล์ที่เปิด

A picture containing graphical user interface

Description automatically generated.

หน้าต่างแทบ Classify โดย Classify เป็นส่วนที่ใช้ในการวิเคราะห์ข้อมูลด้วยวิธีการจำแนกข้อมูล (classification) หรือทำนายข้อมูล (prediction) ซึ่งมีวิธีการต่างๆ ให้เลือกมากมาย

แสดงผลการประเมินเมื่อกดปุ่ม Start ในแต่ละครั้ง

แทบ Test options มีไว้สำหรับเลือกวิธีประเมินตัวจำแนก โดย **Use Training set** คือ ใช้เซตข้อมูลทั้งหมดเป็นทั้ง train set และ test set , **Supplied test set** คือเลือก test set ต่างจาก train set , **Cross-validation** คือ , **Percentage Split** คือแบ่ง train/test set ตามเปอร์เซ็นต์ จากนั้นกดปุ่ม Start เพื่อเริ่มการประเมิน

กด Choose เพื่อเลือกตัวจำแนกที่ต้องการใช้

Graphical user interface, text, application

Description automatically generated

หน้าต่างแทบ Visualize สำหรับการแสดงผล การกระจายของแบบ จากแต่ละชั้นข้อมูล

การเลือกคู่ของฟีเจอร์ที่ต้องการ ให้แสดงผลใน 2 แกน

สีสัญลักษณ์ที่ระบุว่า แบบในสีใดอยู่ในชั้นข้อมูลไหน

หน้าต่าง Visualizing สำหรับการแสดงผล การกระจายของแบบ จากแต่ละชั้นข้อมูล