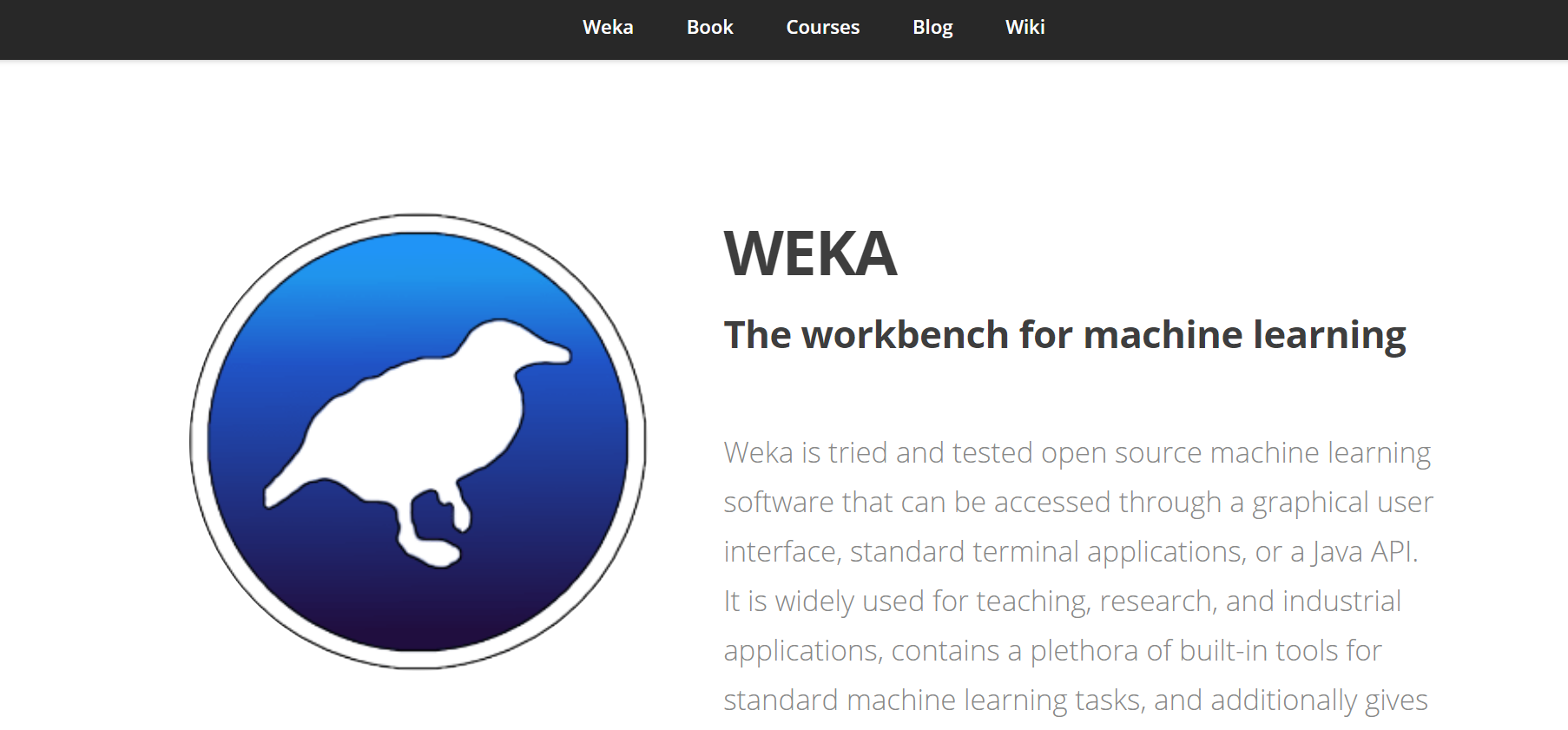
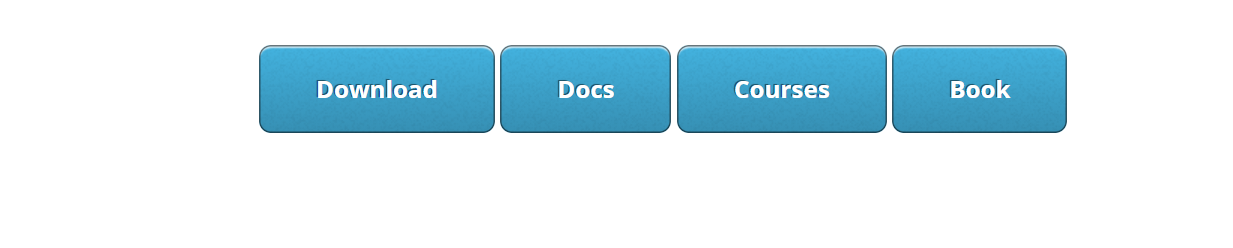
**WEKA INSTALLATION GUIDE & EXERCISES**

1. **Go to** [**https://www.cs.waikato.ac.nz/ml/weka/**](https://www.cs.waikato.ac.nz/ml/weka/)



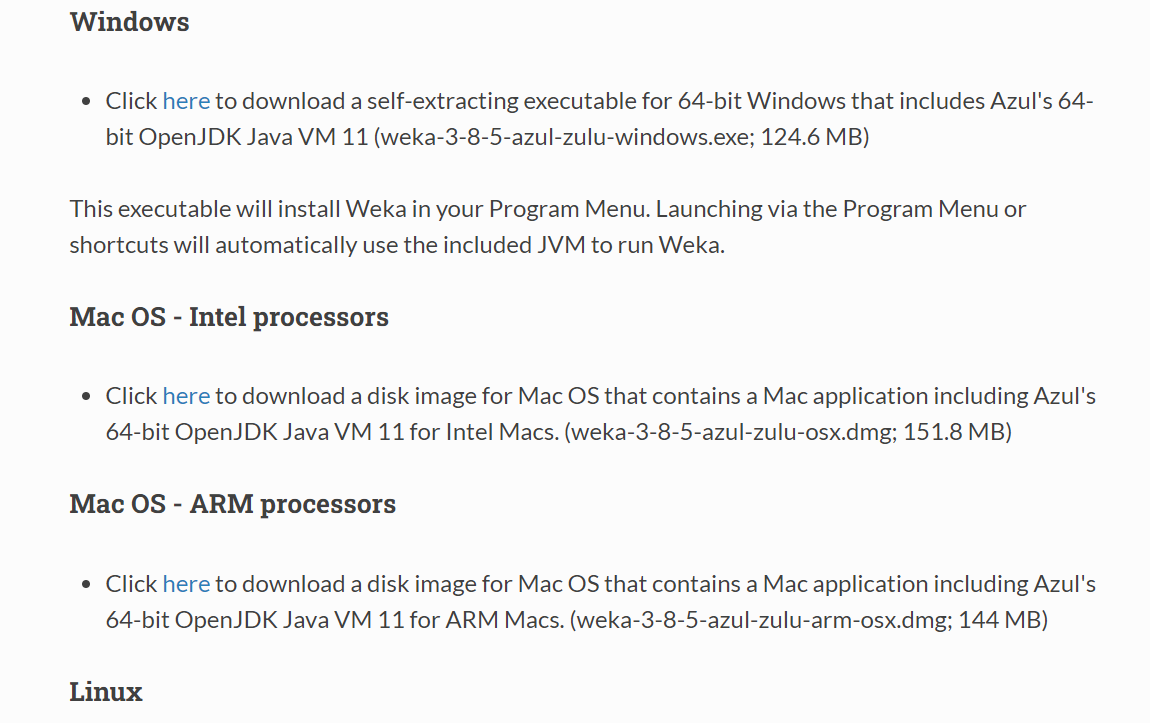
1. **Scroll down and click on Download**



**There are 2 versions of WEKA that you will see on the website once you scroll down:**

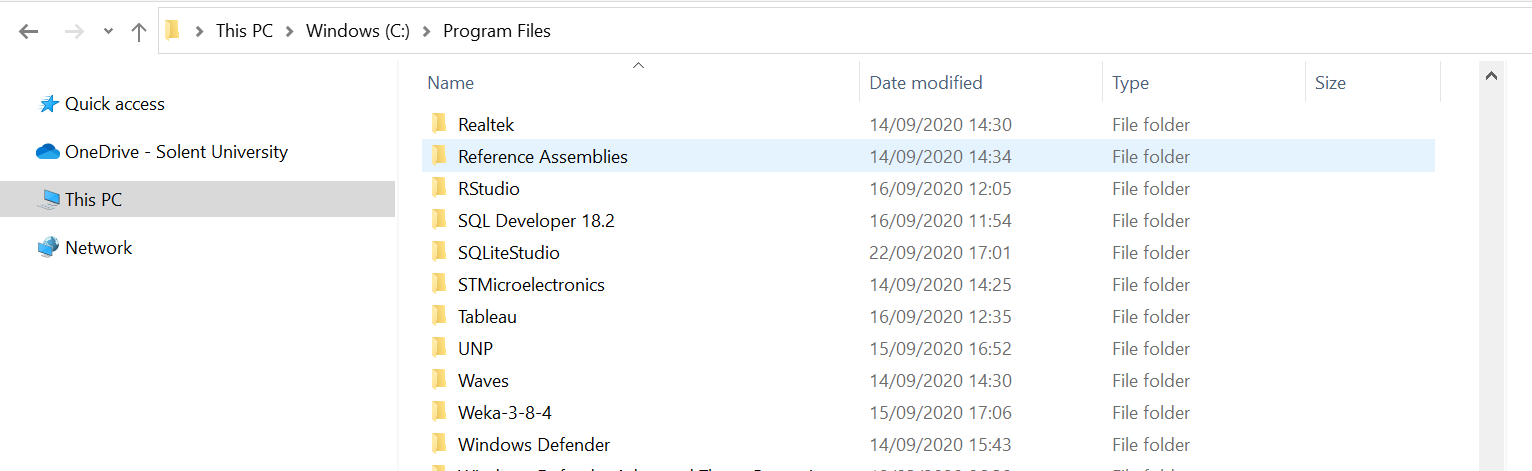
1. **Stable Version**
2. **Developer Version**

**Make sure you download the Stable version. It is available on both Windows and Mac. It is completely free to download. Weka 3.8 is the latest stable version and Weka 3.9 is the development version.**

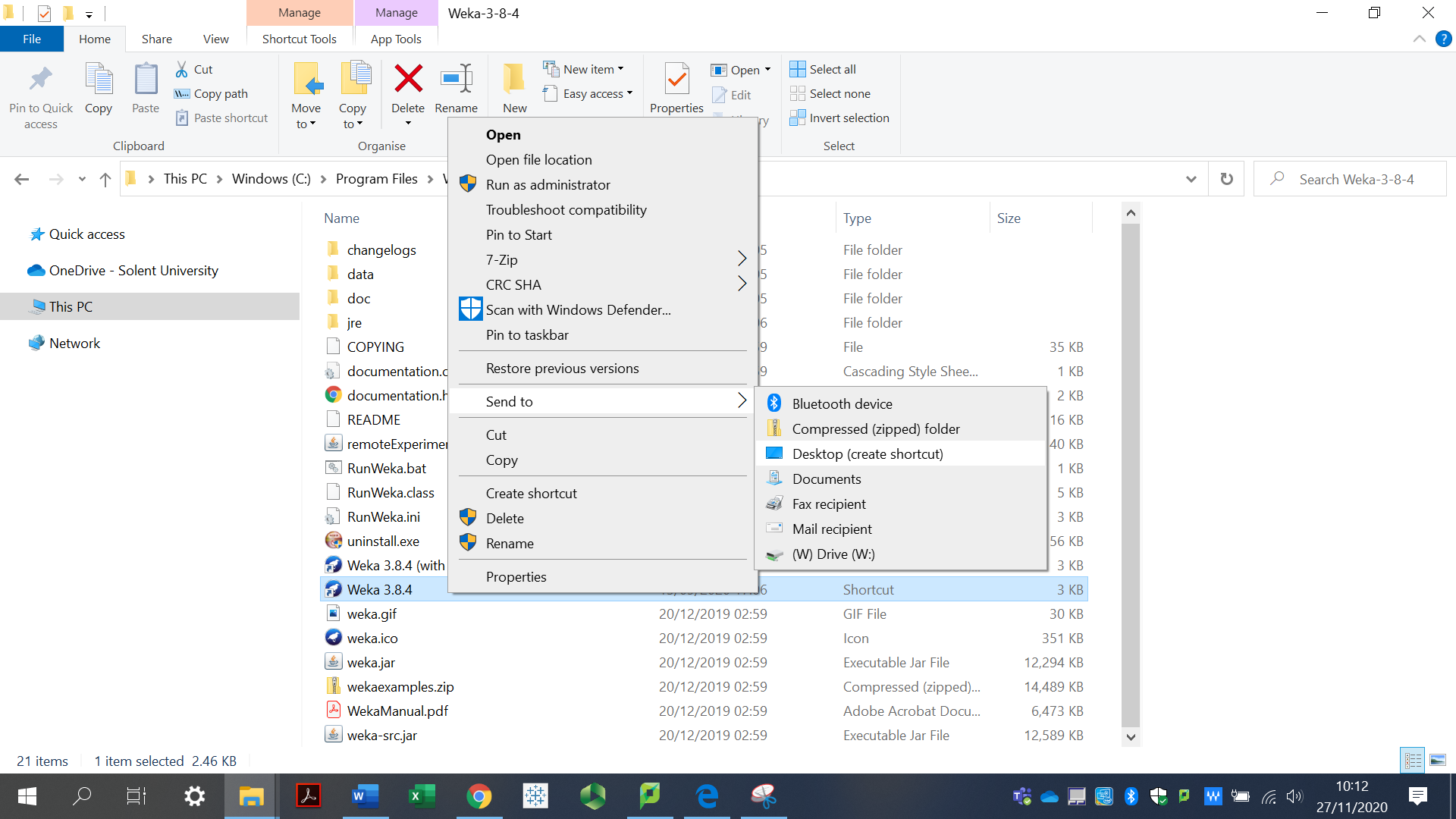


**Click on Next until the installation is complete. Make sure you know where your downloaded file is located.**

**You should see the WEKA Folder here:**

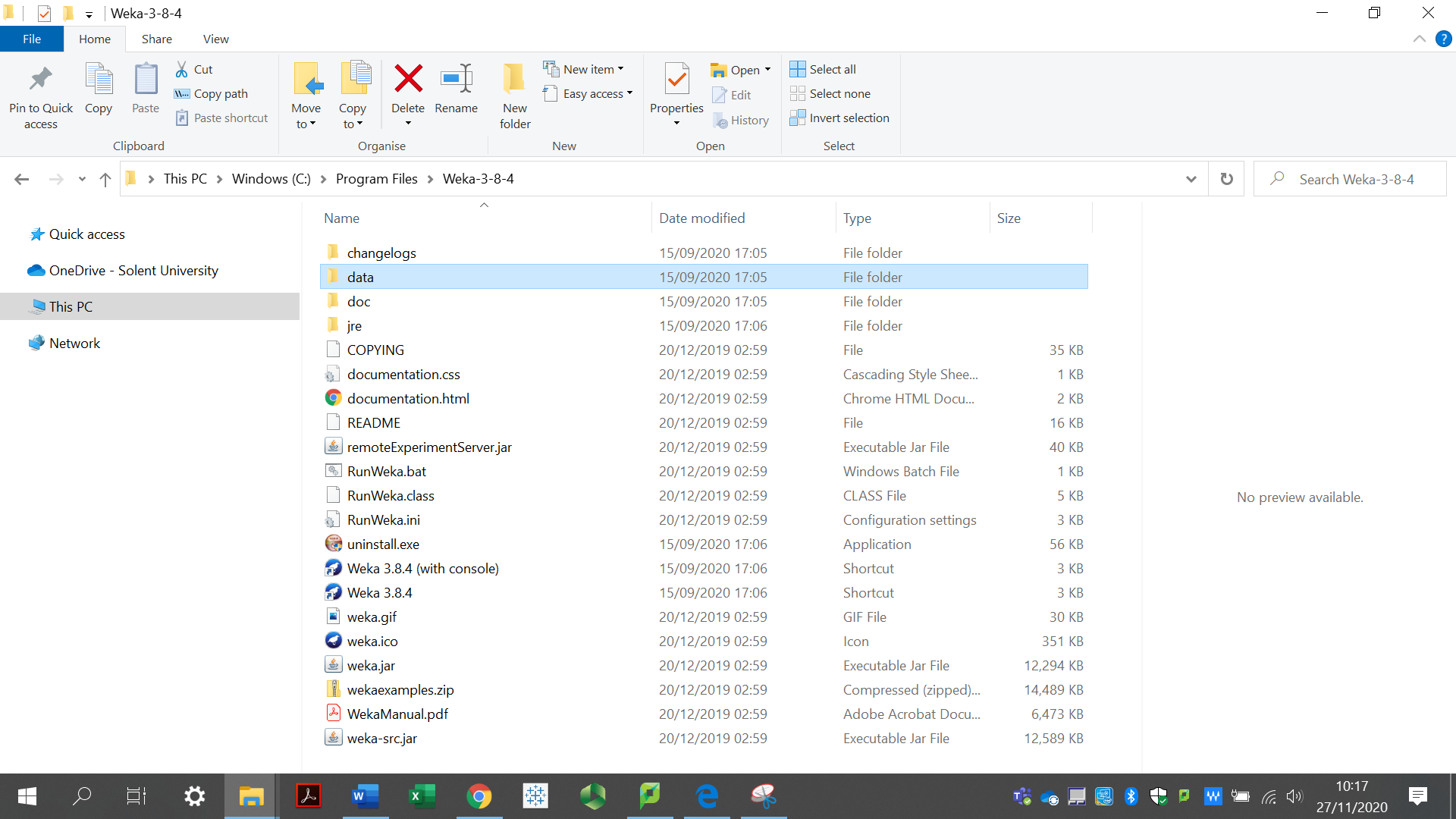


**You can open this folder and select Weka 3.8.5, right click> Send to> Desktop and create a shortcut to your Desktop.**



In the same folder, you will also see a **data folder**, you can copy and paste this folder to your preferred location where it is more convenient for you to access this. Once you have copied this data folder to another location, rename this folder to **Weka Dataset**.

I have also provided you with the dataset on SOL. You can download the folder and save it.



**This part is only for Mac Users, if you are not a Mac user, go to the next section to continue.**

**How to open apps from unidentified developers on Mac in macOS (if you are having issues)**

For Mac users, if you are having issues, you might need to go to System Preferences> Security & Privacy>App Downloaded from Appstore> Anywhere

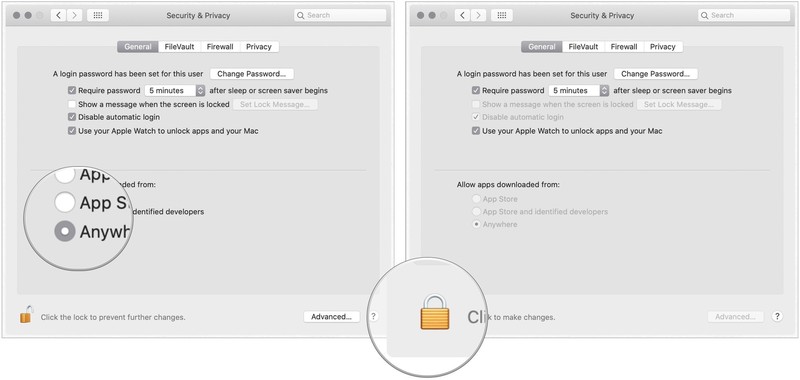
Changing your settings

Now, it's time to allow your Mac to open any app.

1. Click on **System Preferences** on your Mac Dock.
2. Choose **Security & Privacy**.
3. Tap the **lock** at the bottom left of the screen.

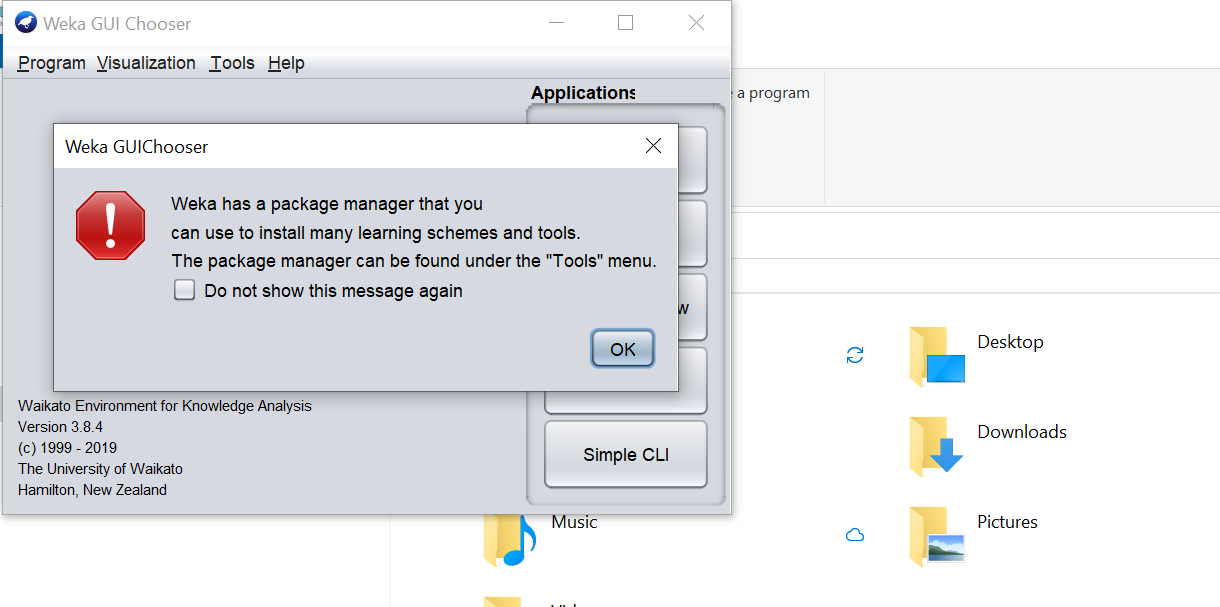


1. Enter your **password** to unlock Security and Privacy.
2. Choose the **Anywhere** under Allow apps downloaded from. Prior to making the change, this option wasn't available.
3. Click the **unlocked lock** to keep the change.

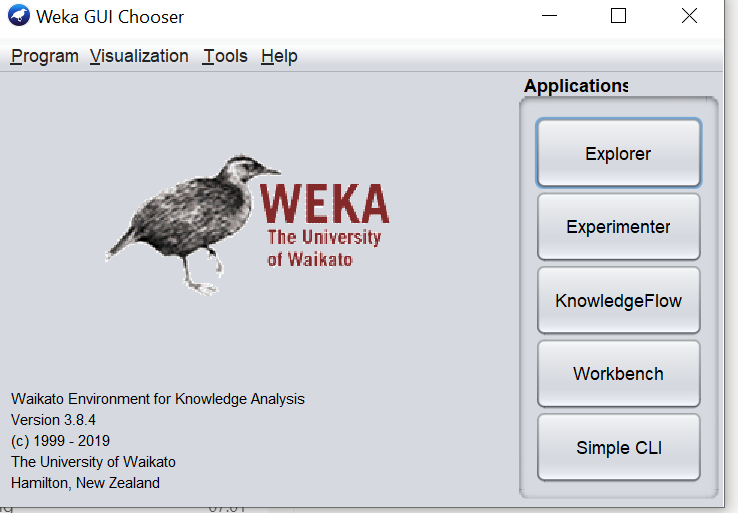
[](https://www.imore.com/sites/imore.com/files/styles/xlarge/public/field/image/2019/10/change-macos-security-any-app.jpg?itok=fj5bQK-2)

This should allow you to install WEKA. You might want to change this back to its original settings after you download and install WEKA successfully for security reasons.

**Once you start WEKA, you might have an error like this:**

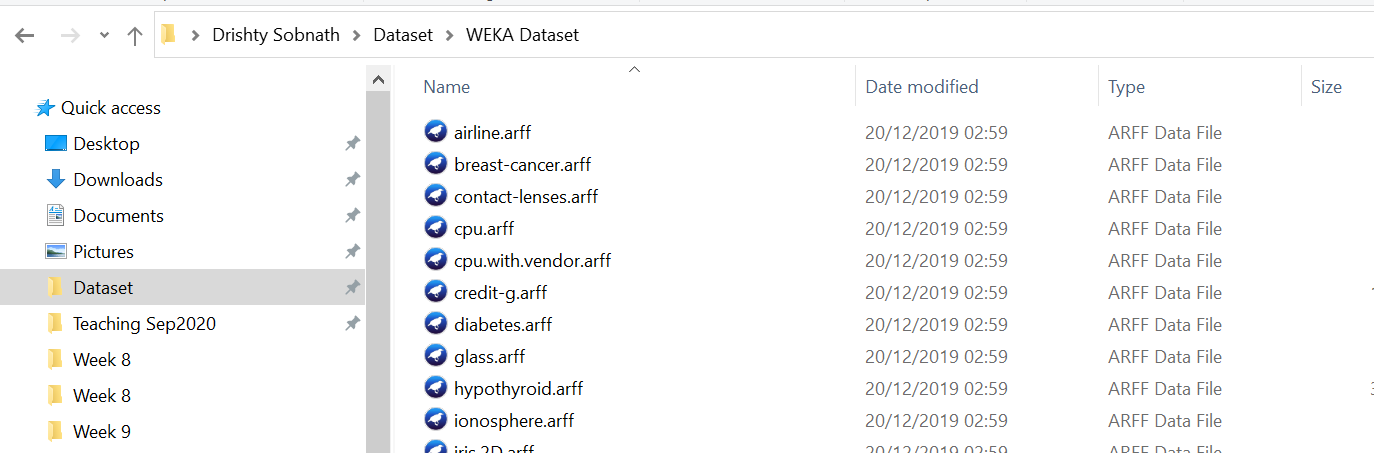


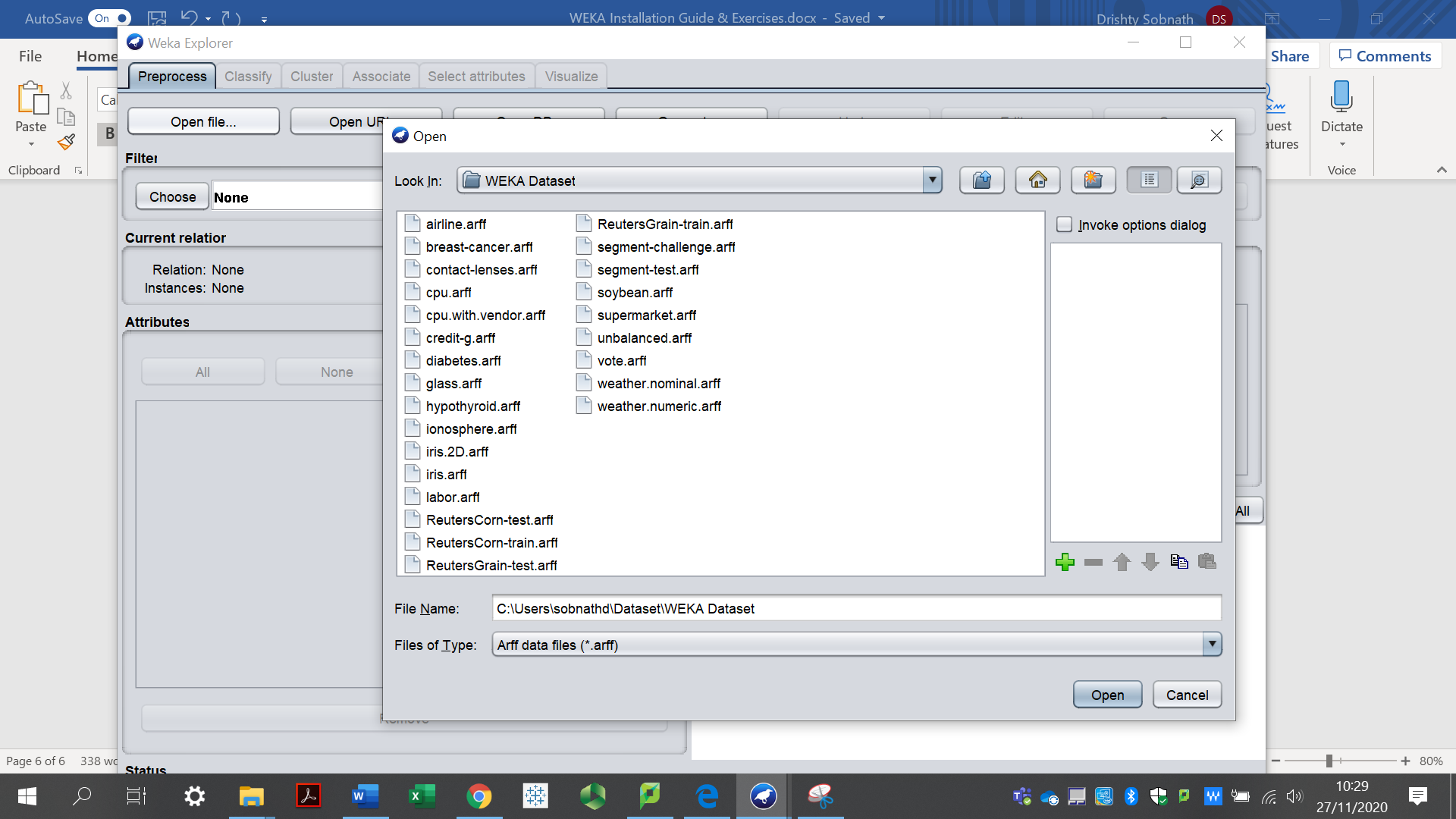
Click on OK and do not worry about this. You should see:



In the Explorer there are several panels: Pre-process; Classify; Clustering; Association rules; Attribute selection; and Visualization. We will be using the Explorer One. Click on Explorer, Under Pre-process> Go to Open File

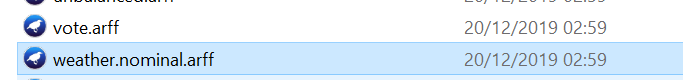
Make sure you know where you have saved your dataset.





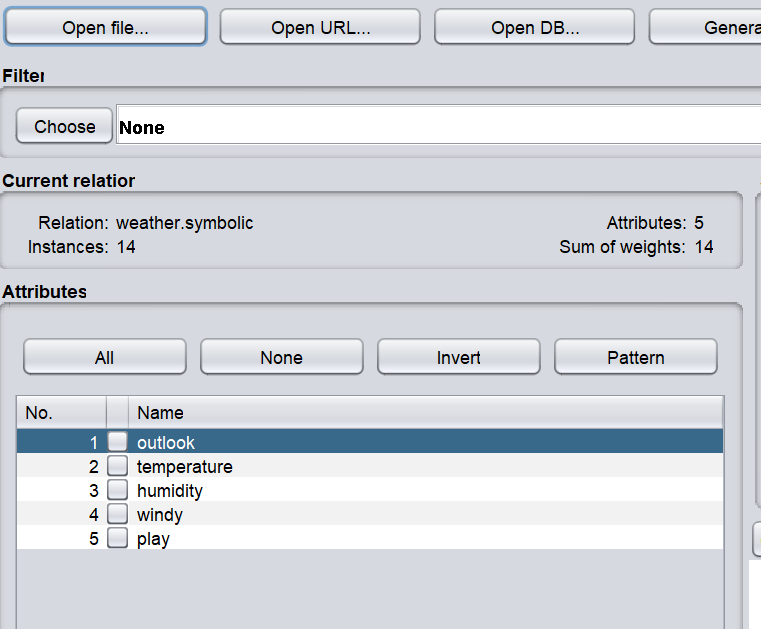
WEKA provides you with some datasets already and most of them are related to health. ARFF is the data format that WEKA supports.

You will find two weather datasets, nominal and numeric, open the weather.nominal.arff file.

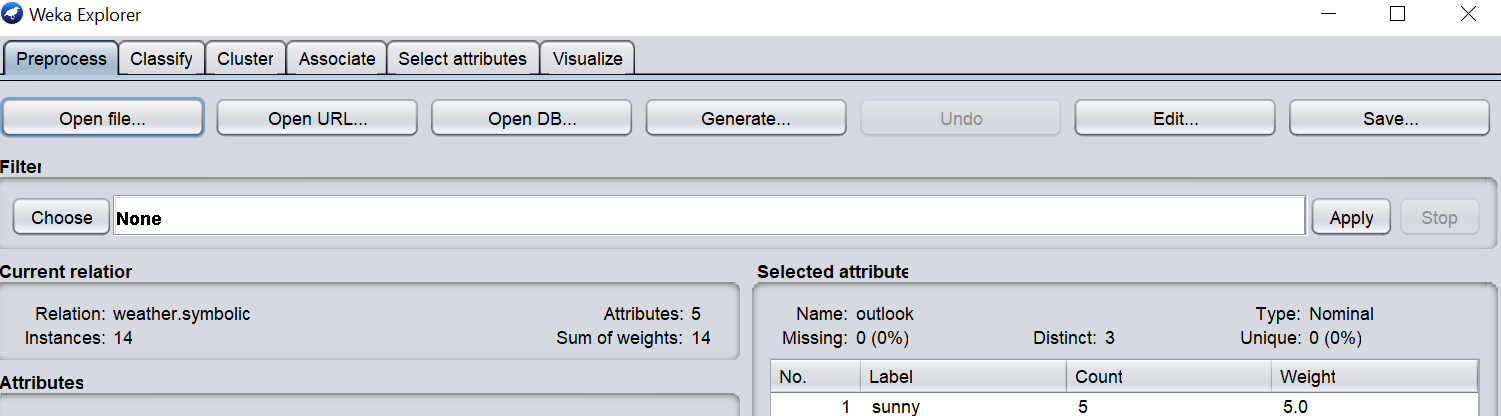


**Exploring the data**

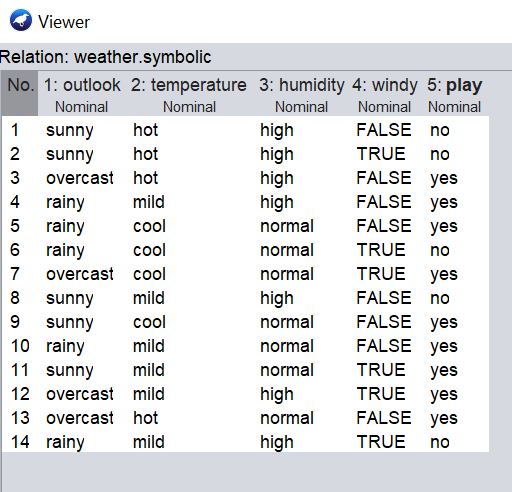
You will see that there are 14 instances, meaning there are 14 rows and 5 instances, meaning there are 5 columns.



To see the dataset, click on Edit at the top



You will see this:

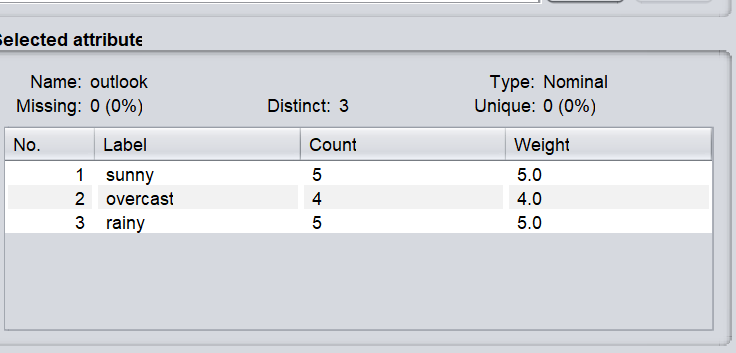


This dataset consists of 2 weeks of data regarding the weather, from Day 1 to Day 14. The columns are

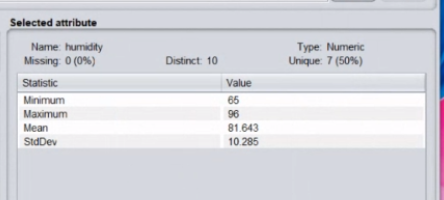
* Outlook
* Temperature
* Humidity
* Windy
* Play (This is the column/class that deals with the prediction, based on the weather, we are trying to see whether it is a good day to play any outdoor games or not.

This is a classification problem.

The dataset contains 5 counts of Sunny days, 4 counts of overcast days and 5 counts of rain days.

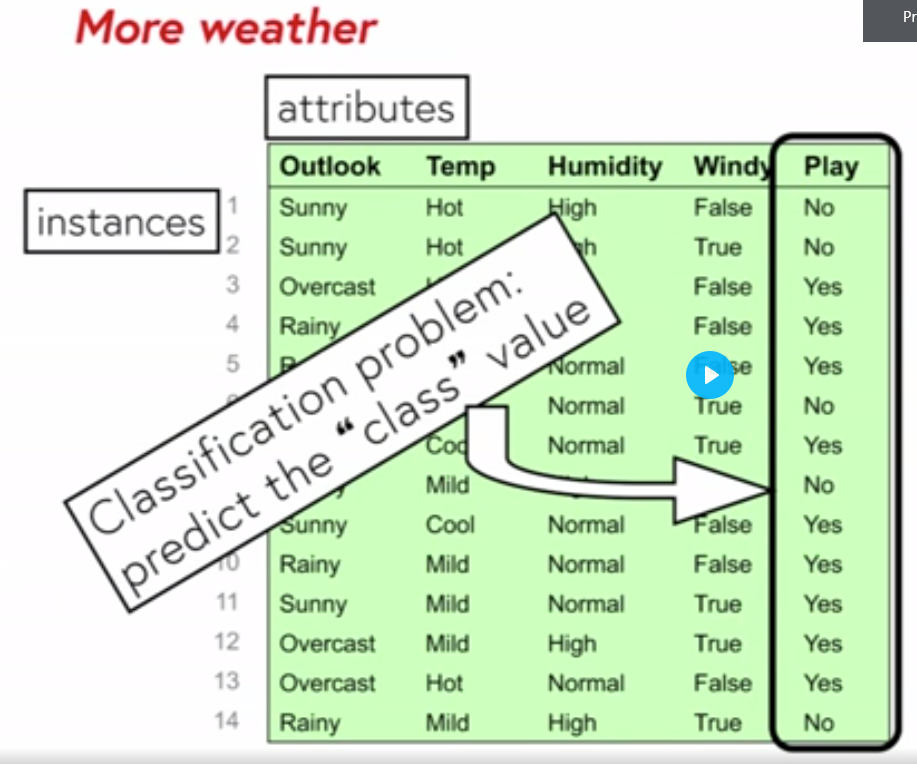


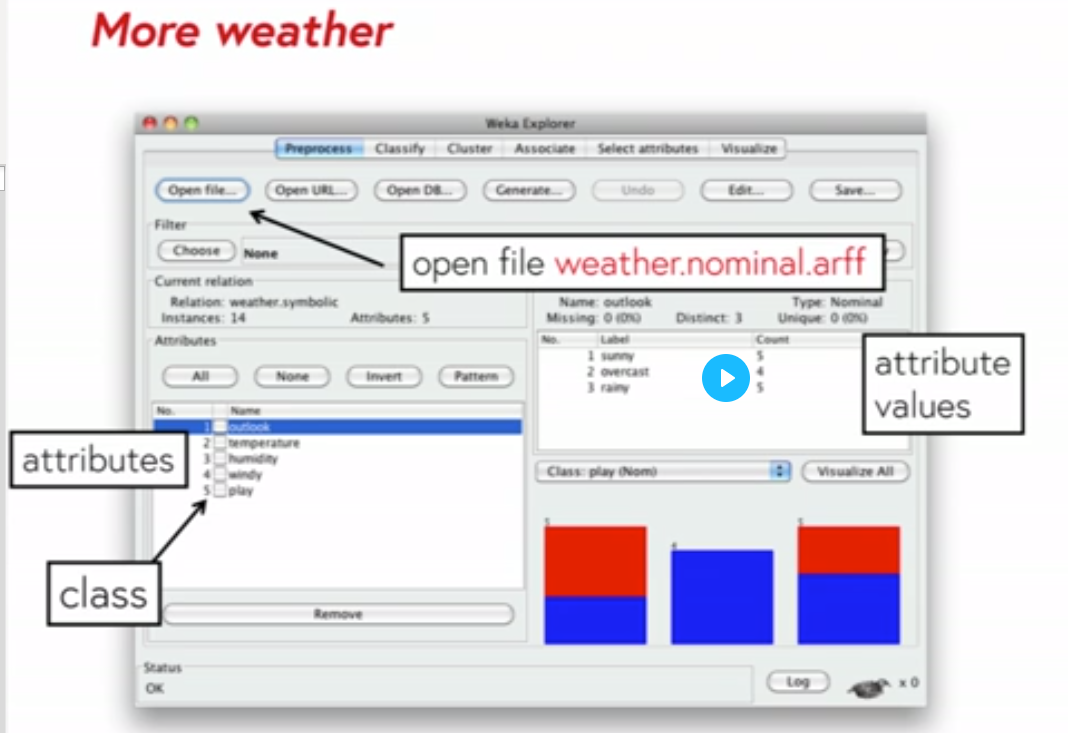
If your dataset contains numeric data, once you click on the column, you will see the mean and standard deviation. You can check this by opening the weather.numeric.arff provided from the data folder from WEKA.

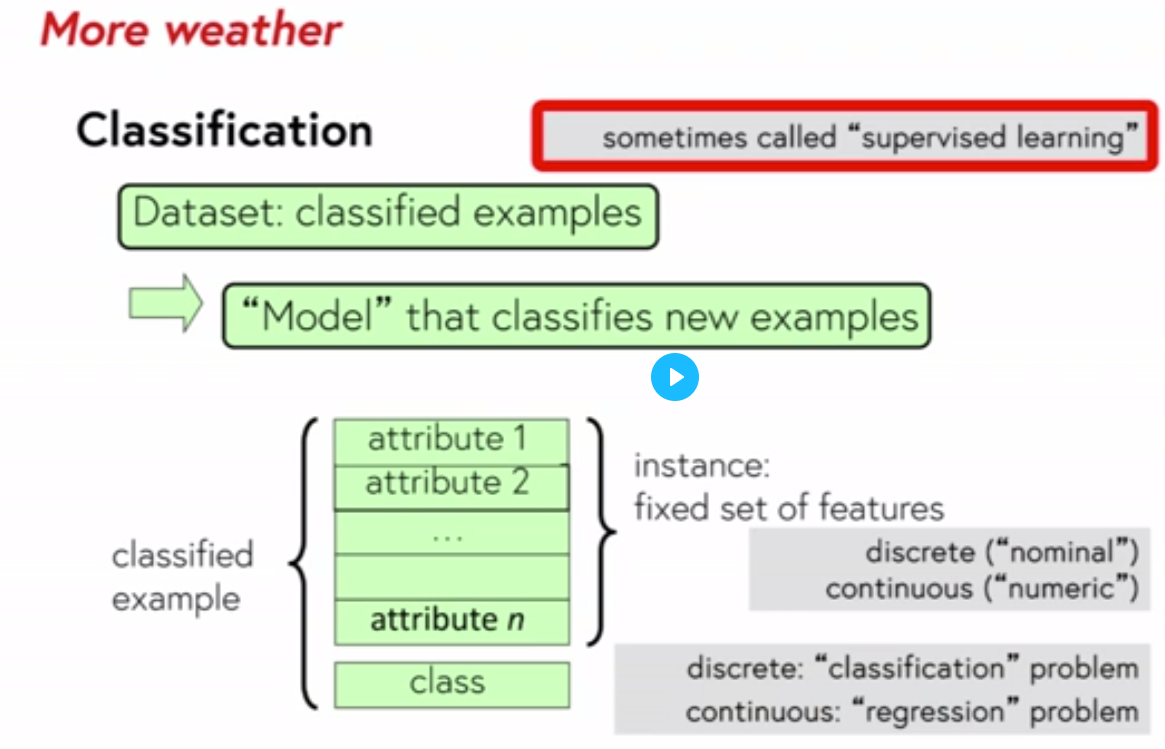


**Let’s get back to our weather.nominal.arff dataset**

It is a sample dataset present in the direct of WEKA. This dataset predicts if the weather is suitable for playing cricket. The dataset has 5 attributes and 14 instances. The class label “play” classifies the output as “yes’ or “no”.

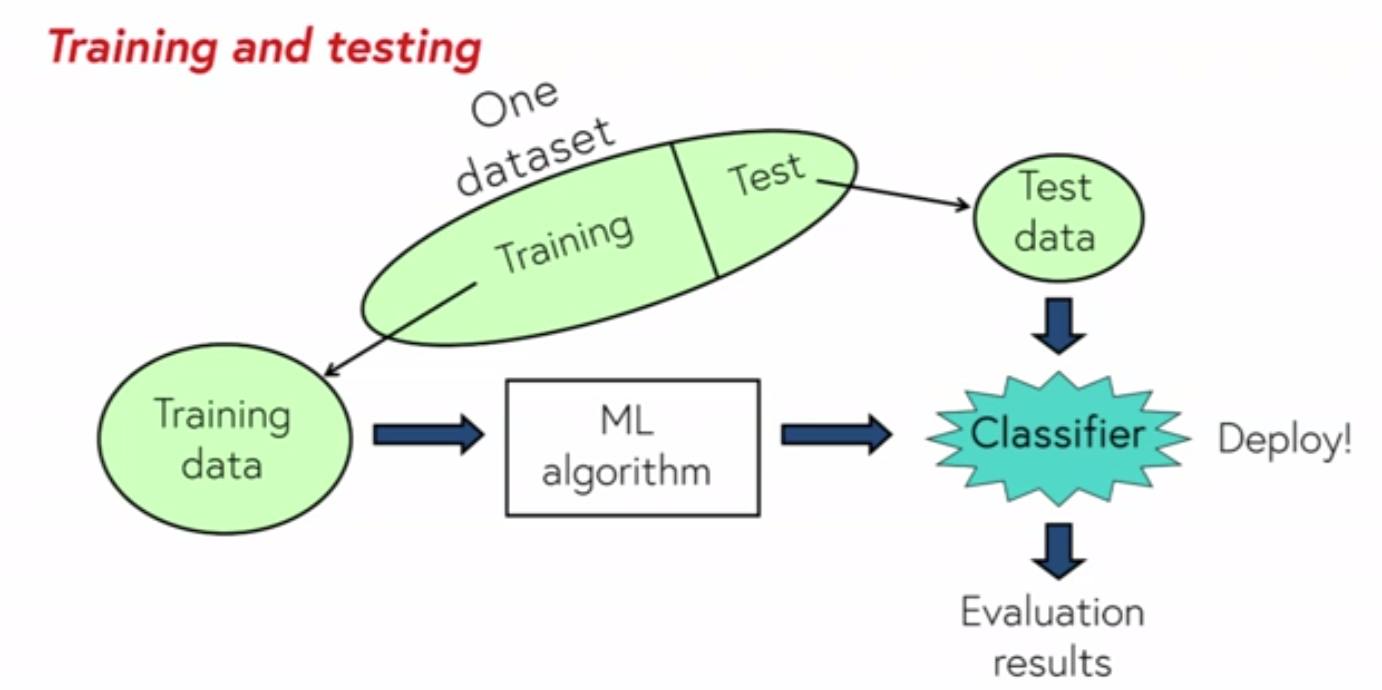






#### Training and testing

Evaluating what has been learned is an essential part of data mining. You should never evaluate on the training set! – the results will be overly optimistic. If you have a single dataset, hold some data back for testing.



### What Is Decision Tree

Decision Tree is the classification technique that consists of three components root node, branch (edge or link), and leaf node. Root represents the test condition for different attributes, the branch represents all possible outcomes that can be there in the test, and leaf nodes contain the label of the class to which it belongs. The root node is at the starting of the tree which is also called the top of the tree.

### J48 Classifier

It is an algorithm to generate a decision tree that is generated by C4.5 (an extension of ID3). It is also known as a statistical classifier. For decision tree classification, we need a database.

**Steps include:**

**#1)**Open WEKA explorer.

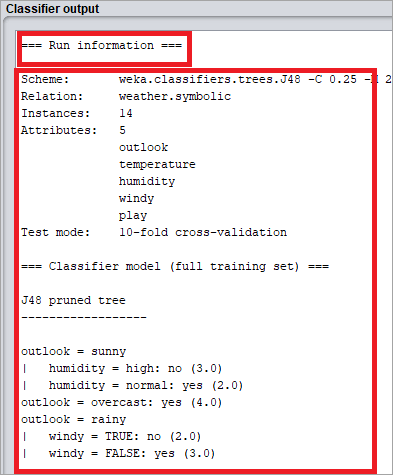
**#2)**Select weather.nominal.arff file from the “choose file” under the pre-process tab option.

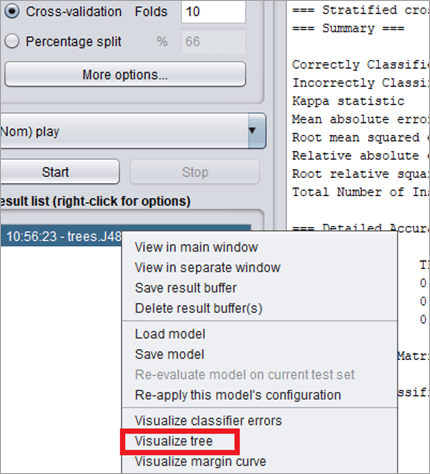
**#3)** Go to the “Classify” tab for classifying the unclassified data. Click on the “Choose” button. From this, select “trees -> J48”.

## Classify tab

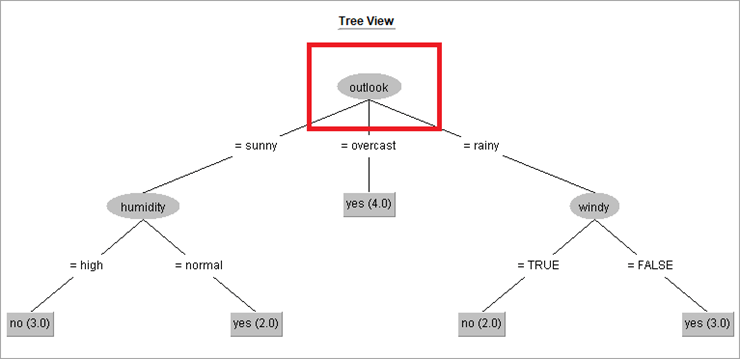
## **#4)**Click on Start Button. The classifier output will be seen on the Right-hand panel.**It shows the run information in the panel as:**

* **Scheme:** The classification algorithm used.
* **Instances:** Number of data rows in the dataset.
* **Attributes:** The dataset has 5 attributes.
* The number of leaves and the size of the tree describes the decision tree.
* **Time taken to build the model:** Time for the output.
* Full classification of the J48 pruned with the attributes and number of instances.





## **#5)**To visualize the tree, right-click on the result and select visualize the tree.



**Output**:

The output is in the form of a decision tree. The main attribute is “outlook”.

**If the outlook is sunny,** then the tree further analyses the humidity. If humidity is high, then class label play= “yes”.

**If the outlook is overcast,** the class label, play is “yes”. The number of instances which obey the classification is 4.

**If outlook is rainy,** further classification takes place to analyses the attribute “windy”. If windy=true, the play = “no”. The number of instances which obey the classification for outlook= windy and windy=true is 2.

## Conclusion

WEKA offers a wide range of sample datasets to apply machine learning algorithms. The users can perform machine learning tasks such as classification, regression, attribute selection, association on these sample datasets, and can also learn the tool using them.

WEKA explorer is used for performing several functions, starting from pre-processing. Pre-processing takes input as a .arff file, processes the input, and gives an output that can be used by other computer programs. In WEKA the output of pre-processing gives the attributes present in the dataset which can be further used for statistical analysis and comparison with class labels.

WEKA also offers many classification algorithms for decision tree. J48 is one of the popular classification algorithms which outputs a decision tree. Using the Classify tab the user can visualize the decision tree. If the decision tree is too populated, tree pruning can be applied from the Pre-process tab by removing the attributes which are not required and start the classification process again.

**Read more and try more examples here:**

# **How to Run Your First Classifier in Weka**

<https://machinelearningmastery.com/how-to-run-your-first-classifier-in-weka/>

<https://www.softwaretestinghelp.com/weka-datasets/>