**Case study on supervised learning algorithm using Weka tool.**

**Weka** is a free open-source software with a range of built-in machine learning algorithms that you can access through a graphical user interface! “

**WEKA** stands for **Waikato Environment for Knowledge Analysis** and was developed at the University of Waikato, New Zealand.

Weka has multiple built-in functions for implementing a wide range of machine learning algorithms from linear regression to neural network. This allows you to deploy the most complex of algorithms on your dataset at just a click of a button! Not only this, Weka gives support for accessing some of the most common machine learning library algorithms of Python and R!

With Weka you can preprocess the data, classify the data, cluster the data and even visualize the data! This you can do on different formats of data files like ARFF, CSV, C4.5, and JSON. Weka even allows you to add filters to your dataset through which you can normalize your data, standardize it, interchange features between nominal and numeric values, and what not!

# What is a Decision Tree?

A Decision Tree is a categorization approach made up of three parts: the root node, the branch (edge or link), and the leaf node. The root node contains the test condition for various attributes, the branch node represents all possible outcomes in the test, and the leaf nodes contain the label of the category to which it corresponds. The root node is located at the beginning of the tree, often known as the tree’s top. Classification and regression challenges can be helped by decision trees.

Classification and Regression Trees are newer terms for decision trees (CART). They work by constructing a tree to analyze a piece of data, beginning at the root and progressing to the leaves (roots) until a forecast can be produced. The method of developing a decision tree involves greedily selecting the optimal split point to make predictions and then repeating the process until the tree reaches a fixed depth.  
Following the construction of the tree, it is pruned to increase the model’s capacity to generalize to new data.

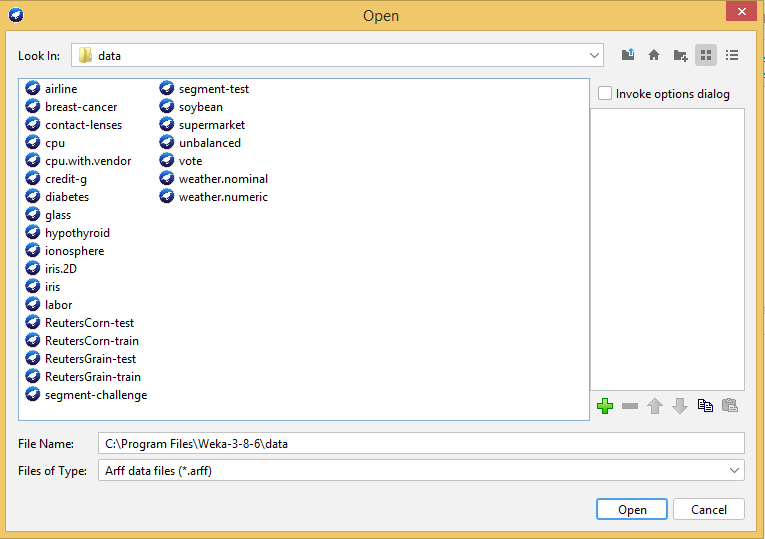
# What is the J48 Classifier?

C4.5 generates an algorithm for generating a decision tree (an extension of ID3). It is often referred to as a statistical classifier. A database is required for decision tree categorization. The vote dataset described in the preceding [article](https://medium.com/@cse.190840131011/how-to-implement-data-preprocessing-in-weka-d4a739f00927)will be used.

Follow the steps below to load the dataset and perform classification in Weka:

1) Open WEKA explorer.

2) Select vote. arff file from the “Open file” under the preprocess tab option.



3) Go to the “Classify” tab for classifying the unclassified data. Click on the “Choose” button. From this, select “trees -> J48”. Let us also have a quick look at other options in the Choose button:

· Bayes: It is a density estimation for numerical attributes.

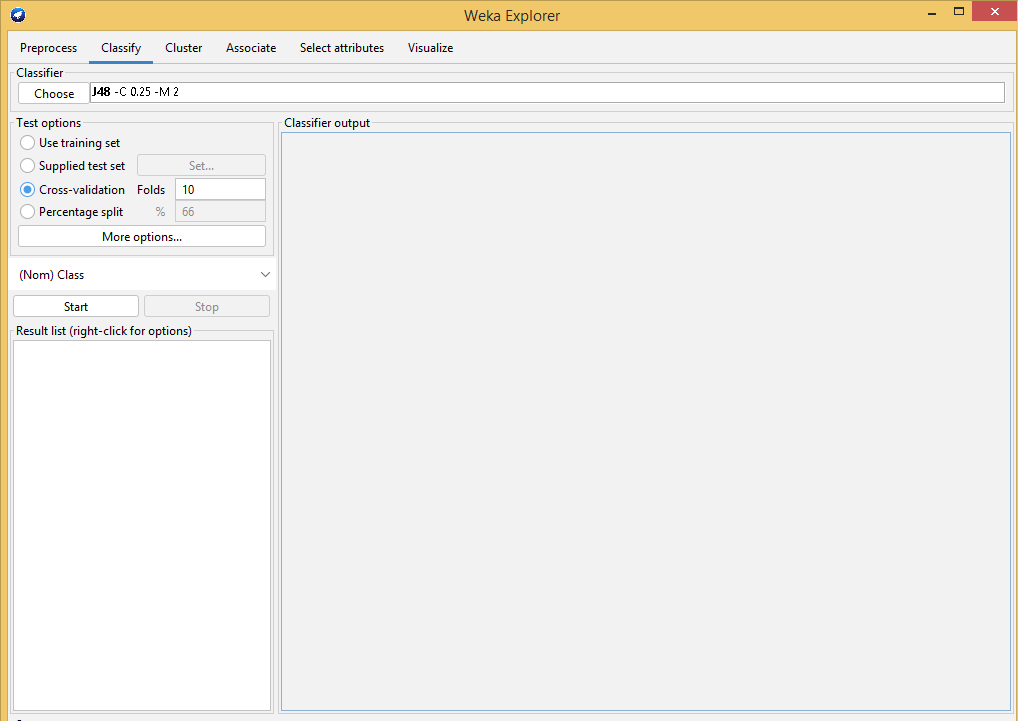
· Meta: It is a multi-response linear regression.

· Functions: It is logistic regression.

· Lazy: It sets the blend entropy automatically.

· Rule: It is a rule learner.

· Trees: Trees classify the data.



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.

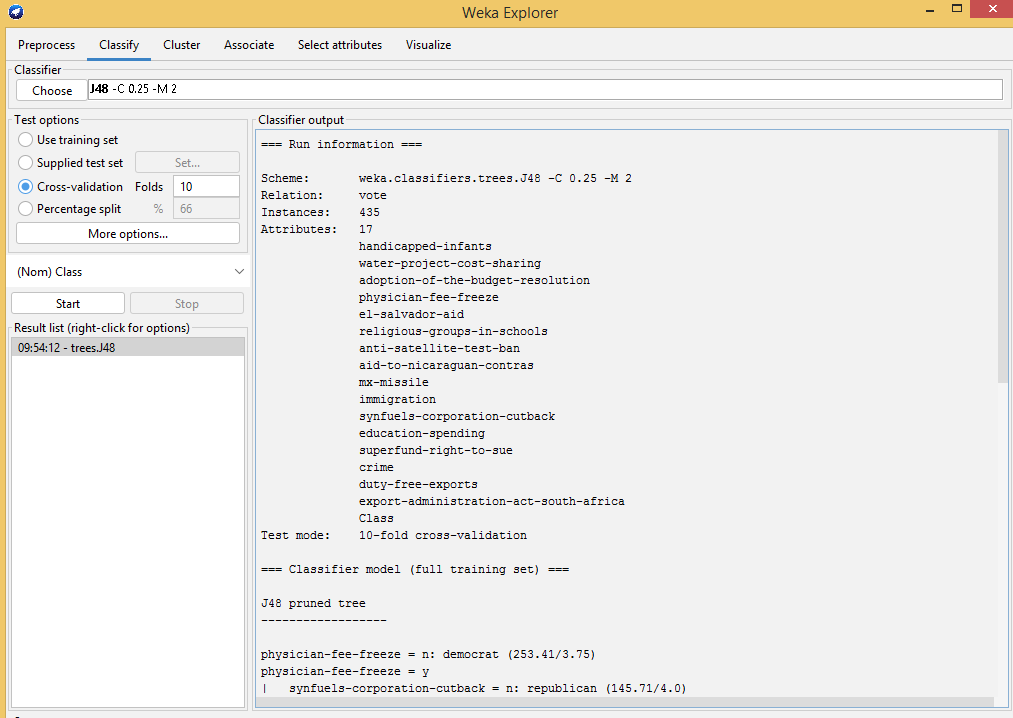
· Instances: Number of data rows in the dataset.

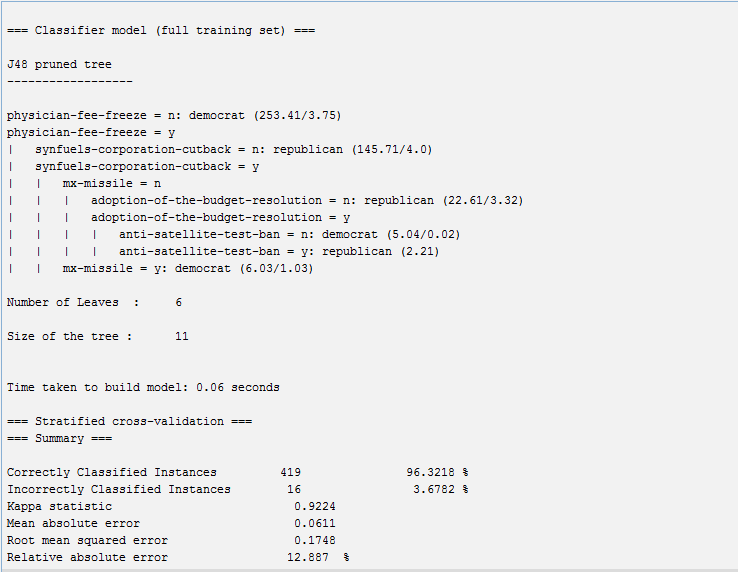
· Attributes: The dataset has 5 attributes.

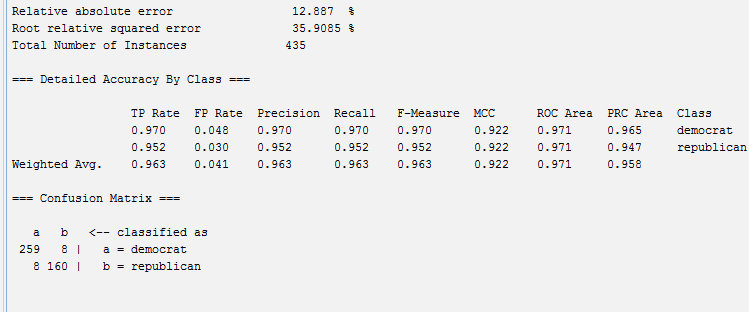
· The number of leaves and the size of the tree describes the decision tree.

· Time is taken to build the model: Time for the output.

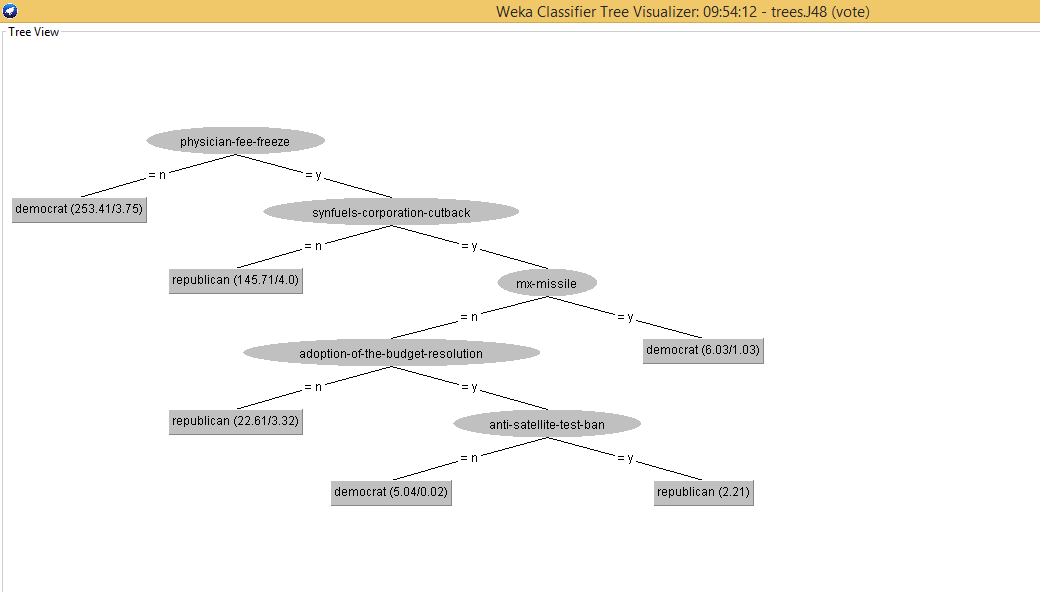
· Full classification of the J48 pruned with the attributes and number of instances.







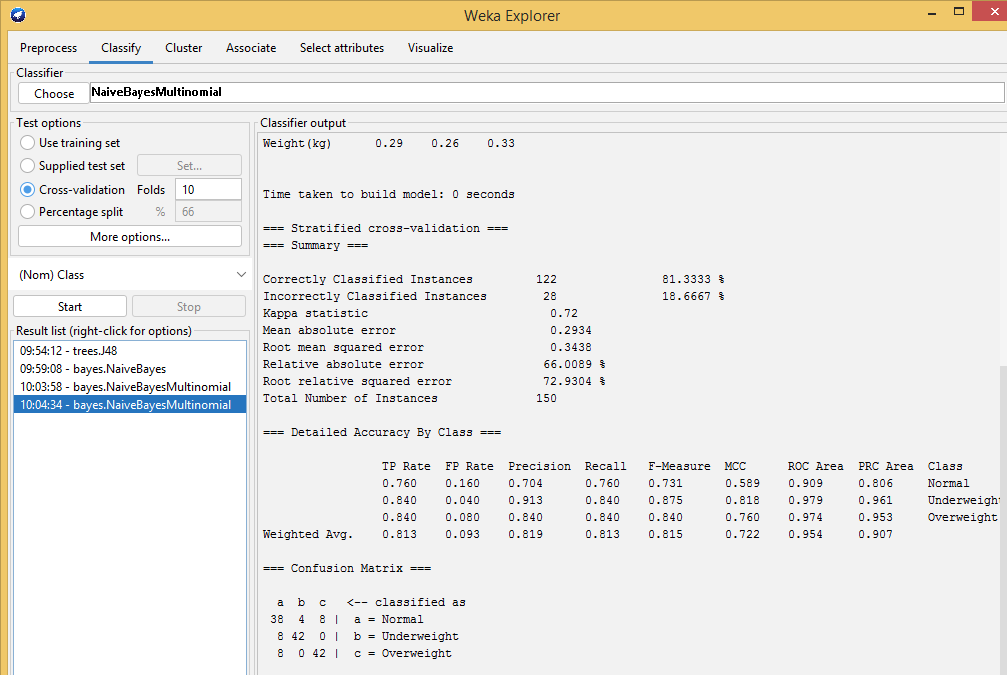
5) To visualize the tree, right-click on the result and select visualize the tree.

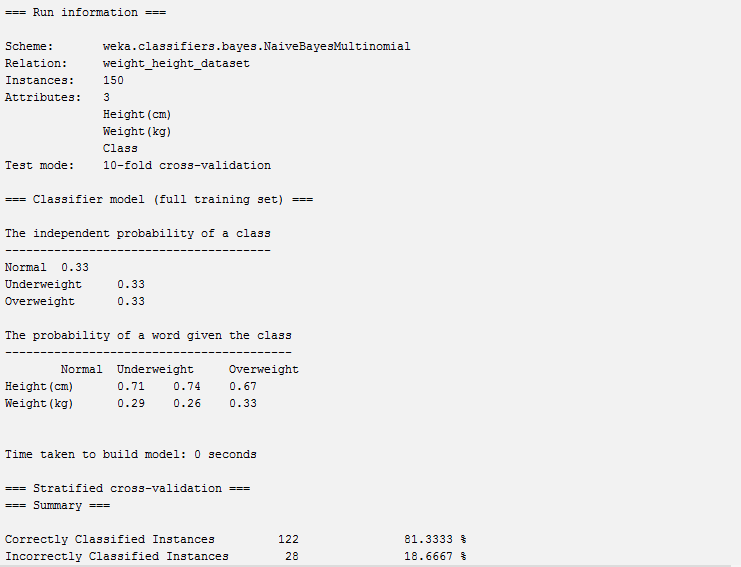


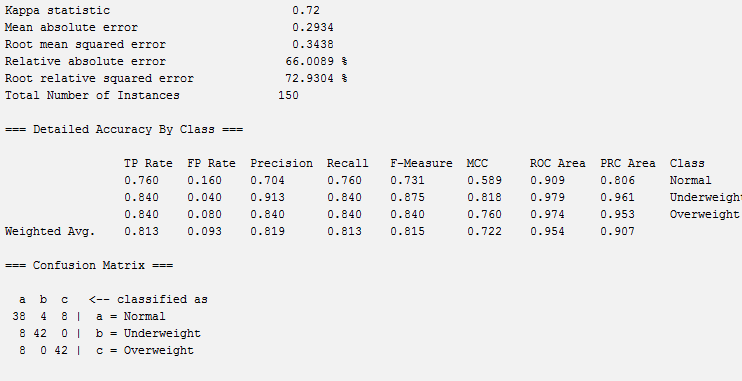
**NAIVE BAYES ALGORITHM**

**1.Select** height weight csv file from the “Open file” under the preprocess tab option.

2.Go to Classify tab and select Naivebayes Multinomial algorithm and click on start button.

****

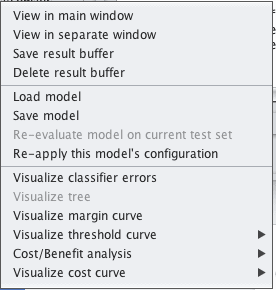
****

****

You can easily save a trained model to file in the Weka Explorer interface.

1. Right click on the result item for your model in the “Result list” on the “Classify” tab.

2. Click “Save model” from the right click menu.



Weka Save Model to File

3. Select a location and enter a filename such as “heightwt”, click the “Save button.

Your model is now saved to the file “heightwt.model”.

**Load a Finalized Model**

You can load saved Weka models from file.

The Weka Explorer interface makes this easy.

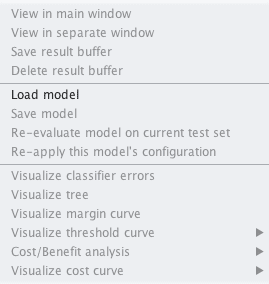
1. Open the Weka GUI Chooser.

2. Click the “Explorer” button to open the Weka Explorer interface.

3. Load any old dataset, it does not matter. We will not be using it, we just need to load a dataset to get access to the “Classify” tab. If you are unsure, load the *data/diabetes.arff* file again.

4. Click the “Classify” tab to open up the classifiers.

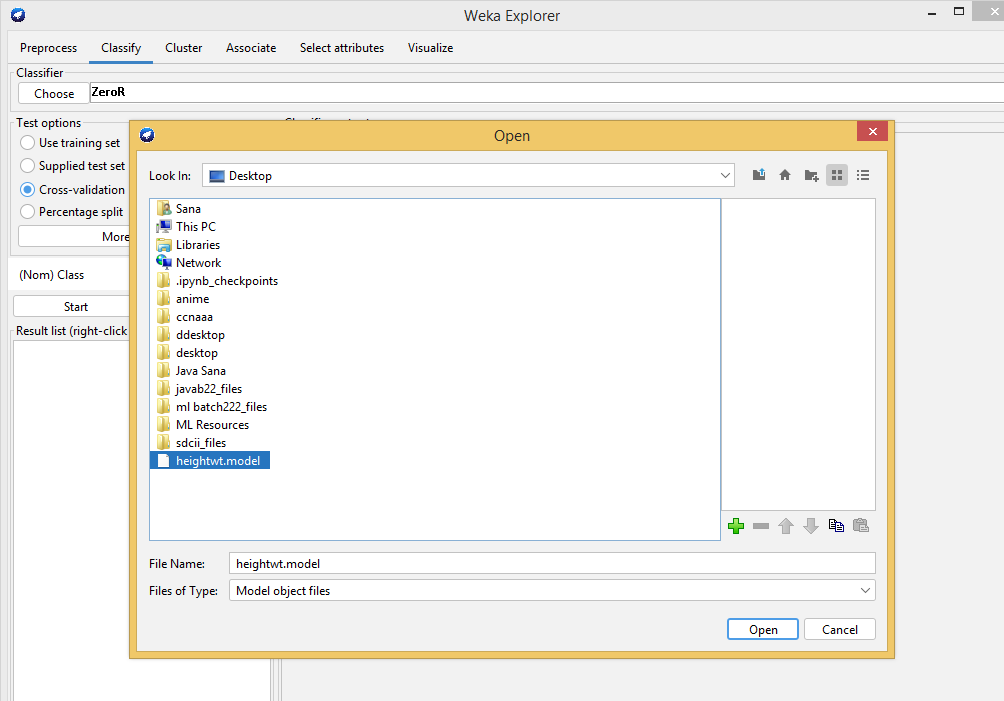
5. Right click on the “Result list” and click “Load model”, select the model saved in the previous section “logistic.model”.



Weka Load Model From File

The model will now be loaded into the explorer.

We can now use the loaded model to make predictions for new data.

****

