**What did you use for showing the features in chart?**

I use python and its libraries matplotlib and seaborn, actually it’s not a single chart it’s a multiple charts then I combine then together and make a single picture.

Explain the modeling part.

First we choose the data in weak from preprocess tab, when in classify tab we choose our model

Then I set the testing options, first I select the testing options cross-validation and set the folds

Equals to 10 and then click on start so model has been start training and testing then then it

Shows the result. This process is same for DT and RF

**Why did you choose DT and RF?**

**Decision tree**

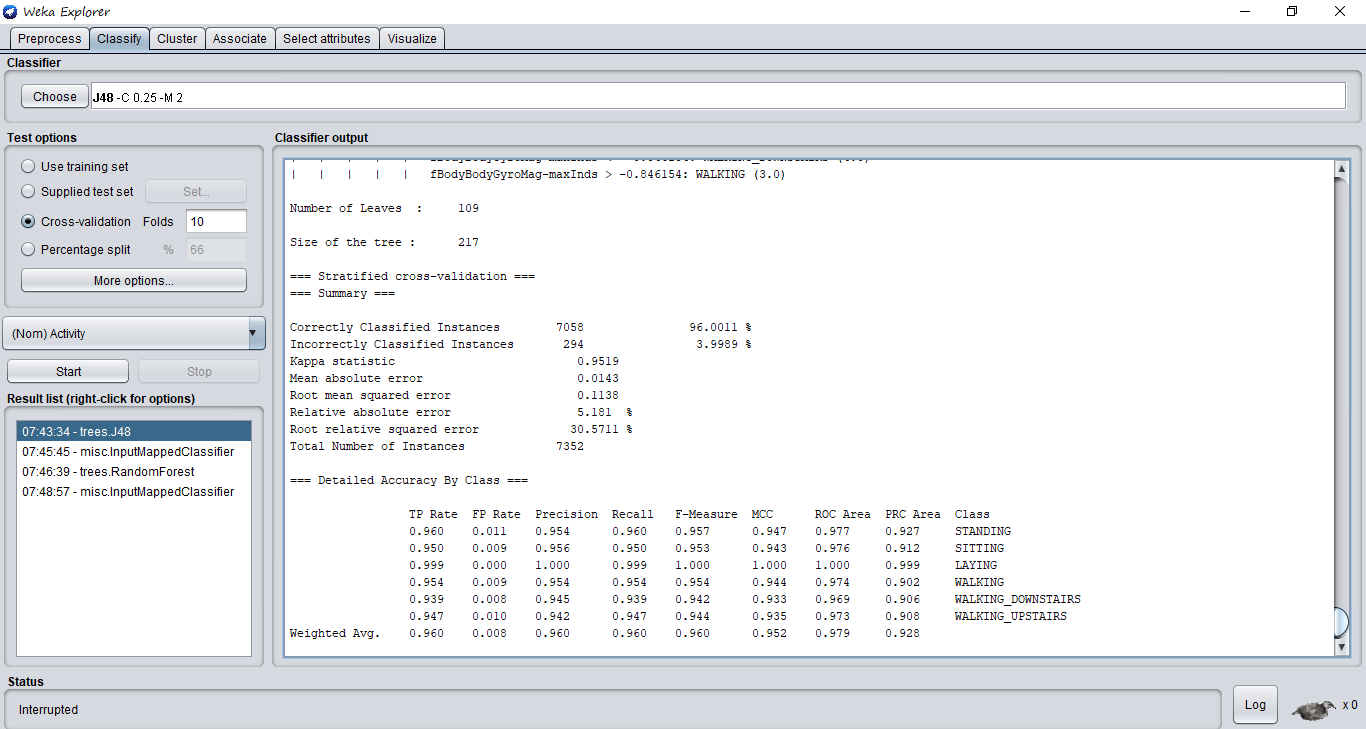
Decision trees gives an effective method for Decision Making and classification because they are clearly lay out the problem so that all options can be challenged and seen. DT Allow us to analyze fully the possible consequences of a decision. Decision tree is also one of the fastest way to identify most significant variables and relation between two or more variables, by using decision trees, we can create new features that has better power to predict target variable.

**Random Forest**

Random forest algorithm is one of the best algorithm for classification and as well as regression. RF provides higher accuracy. Random forest classifier can handle the missing values and maintain the accuracy of a large proportion of data. If we use trees in RF, it won't allow overfitting trees in the model. Under the hood it is an ensemble of randomized decision trees.

So when I use decision tree algorithms as our first classifier model then it is good idea to use Random forest because doing like this we can see what difference in accuracy is and time we use single decision tree and multiple DT known as RF.

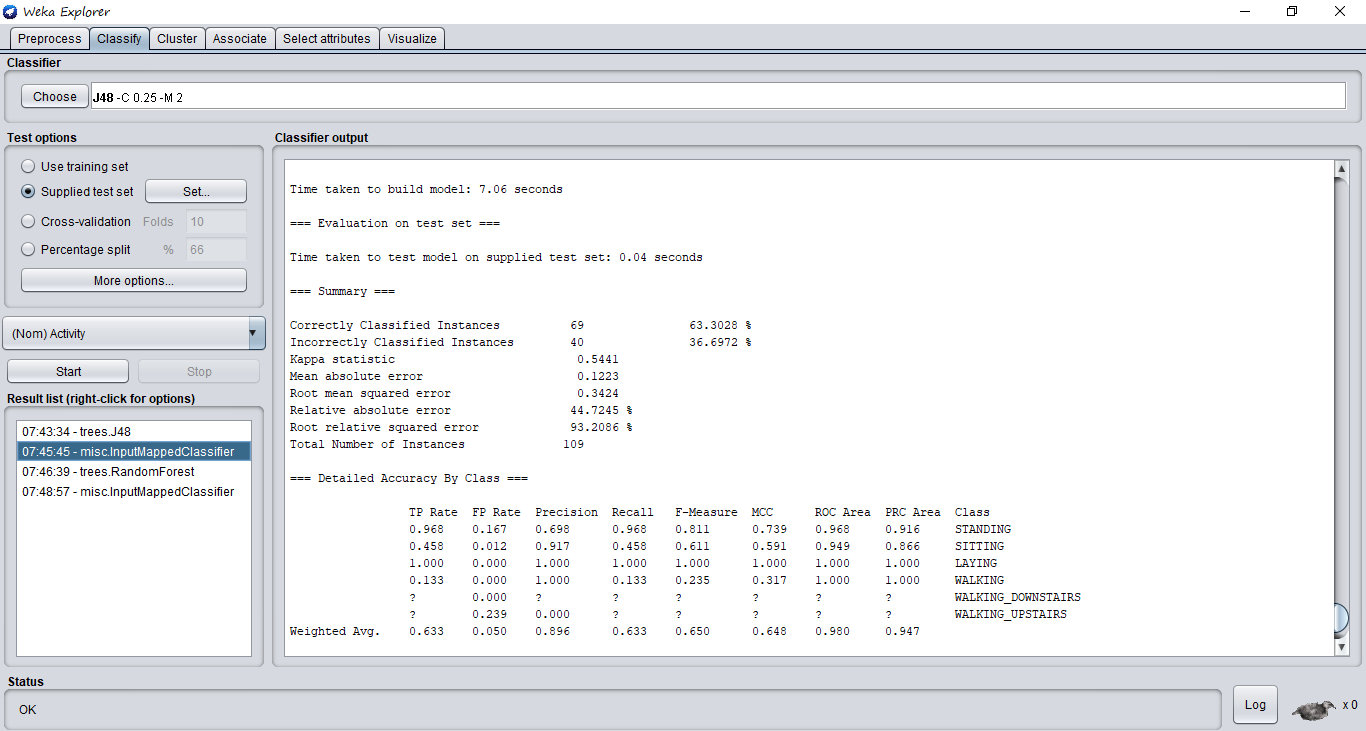
**Explain the visualization for weka photos**



In this picture we can clearly see that we have selected the algorithm j48-C 0.25-M 2

And we set testing option as cross-validation folds = 10, so as a result first row show that in our decision tree number of leaves are 109 leaves are those who have no child node. Then we have size of tree which tells the total number of nodes we have in our tree. We have correct results and incorrect results, we have number for both and as well as percentage. Then we have kappa statistic which a very good measure for multi class and imbalanced classification problems. For multi class problems, measures such as the accuracy, or precision/recall do not provide the complete picture of the performance of our classifier. So kappa value is 95%, then we have we have 4 types of errors which one have its one significance then we have total number of instance means samples. After that we have table in which each class accuracy has been shown using different methods like precision, recall, and ROC curve area by using this table we can see that how our model is predicting the results for each class, which class has been correctly predicted more accurately and which is not.

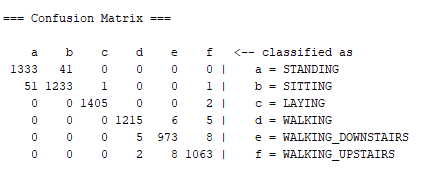
**Now I’m going to explain another picture.**



In this pic we are using DT but for testing we are not using cross validation we are using test data this time. So first we have model build and train time and then we have testing time then summary is same as shown in previous image I have same accuracy table in which multiple methods for calculating accuracy has been used each one have its own significance.

I have explain for DT then RF have same so we don’t need to explain again every thing.

**Explain the confusion matrix.**



Ok in confusion matrix the rows are corresponds to what model predicts and columns are representing the actual values, so the diagonal values are true positives which means those are the numbers of values which are correctly predicted by our model and other then diagonals are errors which means these are not correctly predicted by our model, we can see for the class a which is standing our model predict the 1333 samples belongs to class a and they are true positive means yes they are actually belongs to class a

Same as for class b, c and so on. If we see first row and 2nd column we have number 41 which means there are 41 samples which are actually belongs to class b which is sitting but our model predict it as a class a which is standing so this is the error and same for other off diagonals values.