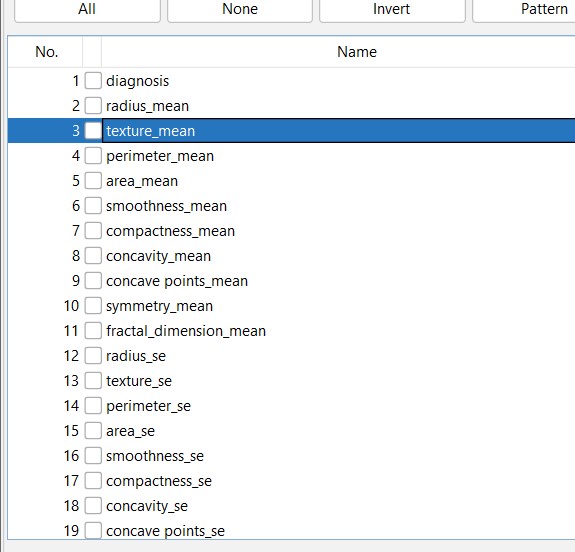
**Task 1:**

* **Cancer Data**

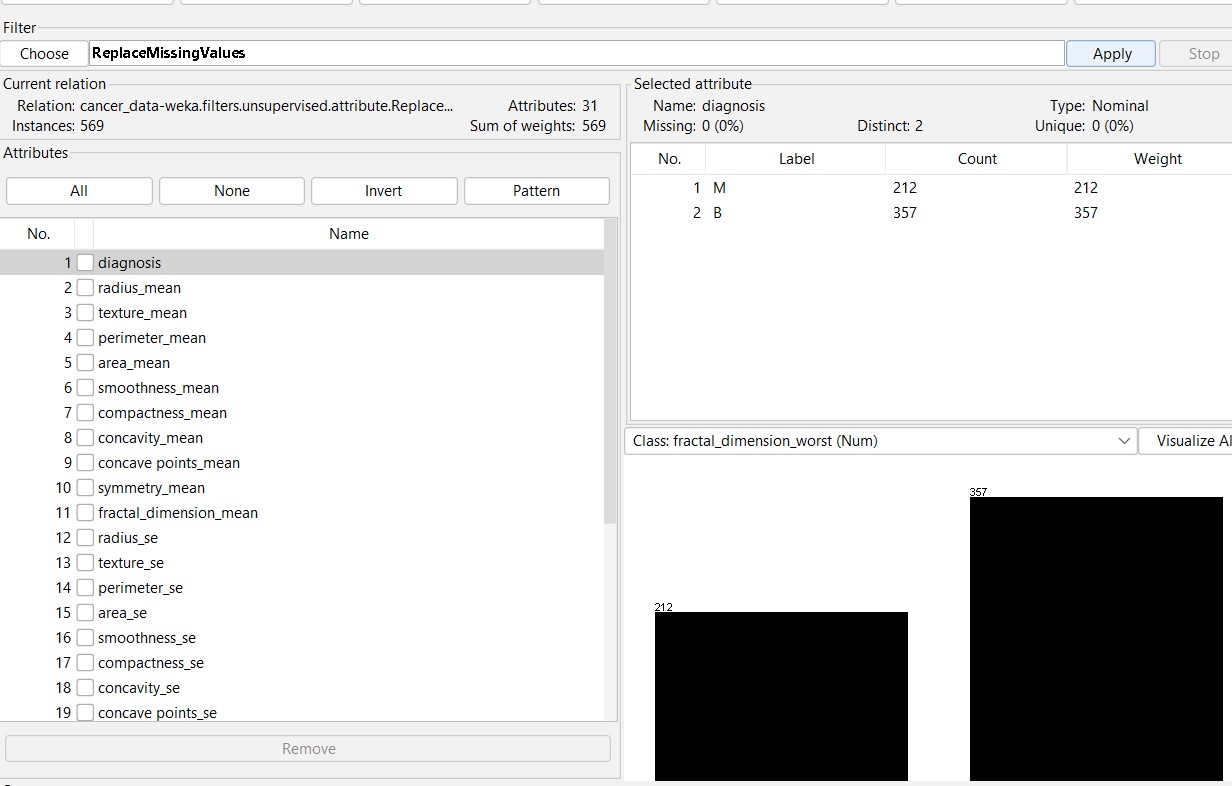
**Presprocessing :**

1. Dealing with the redundant feature



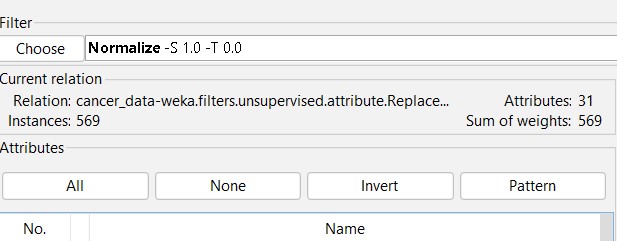
Here we have removed the ID attribute (Column as it was a redundant feature).

1. Dealing with the missing values



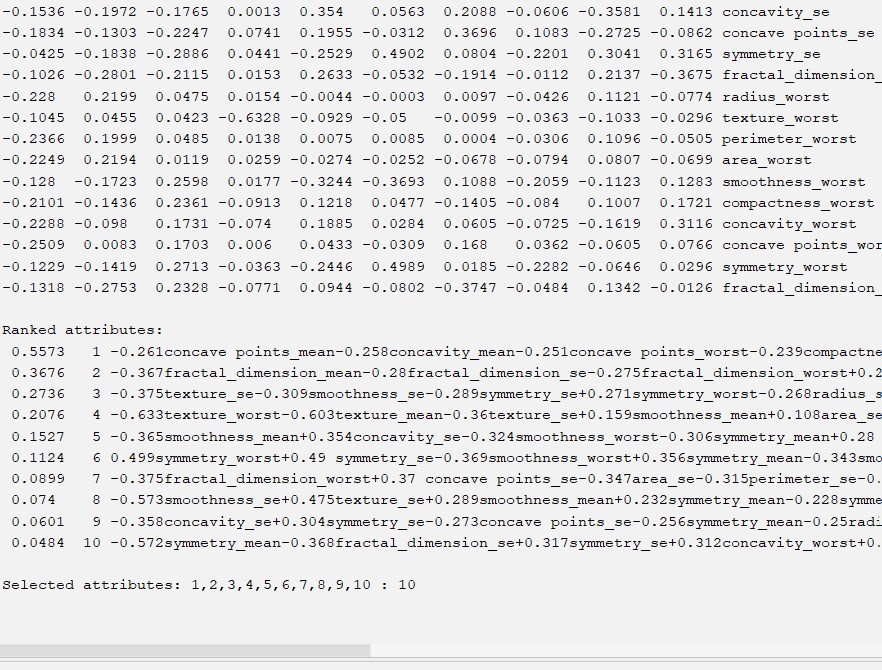
Here we have dealt with the missing values.

1. Normalization:



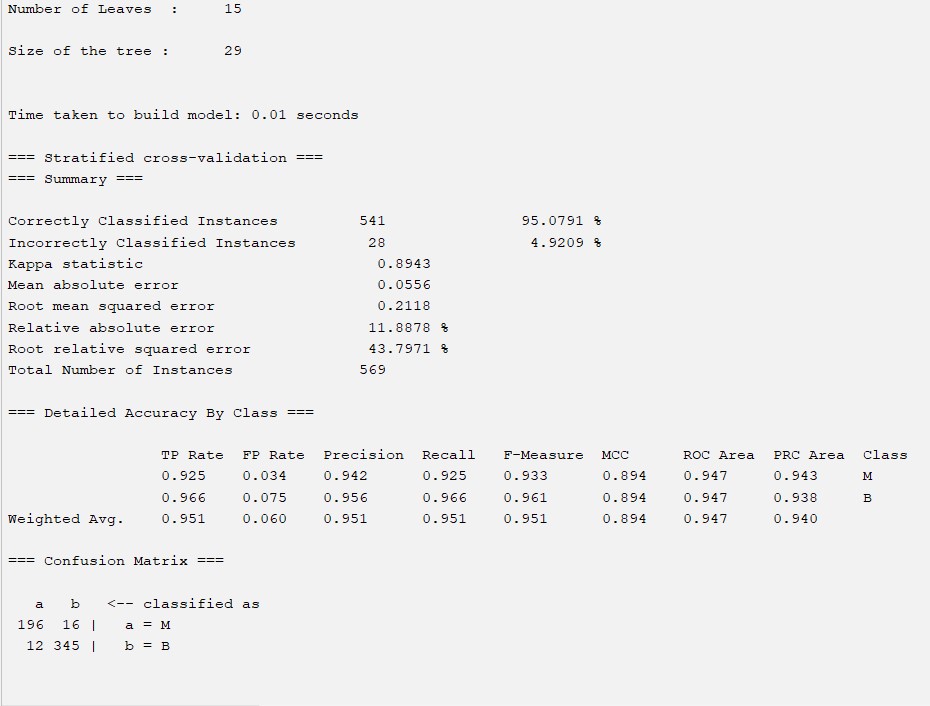
Here we have normalized the complete table. Now our dataset is completely preprocessed.

Now Apply PCA

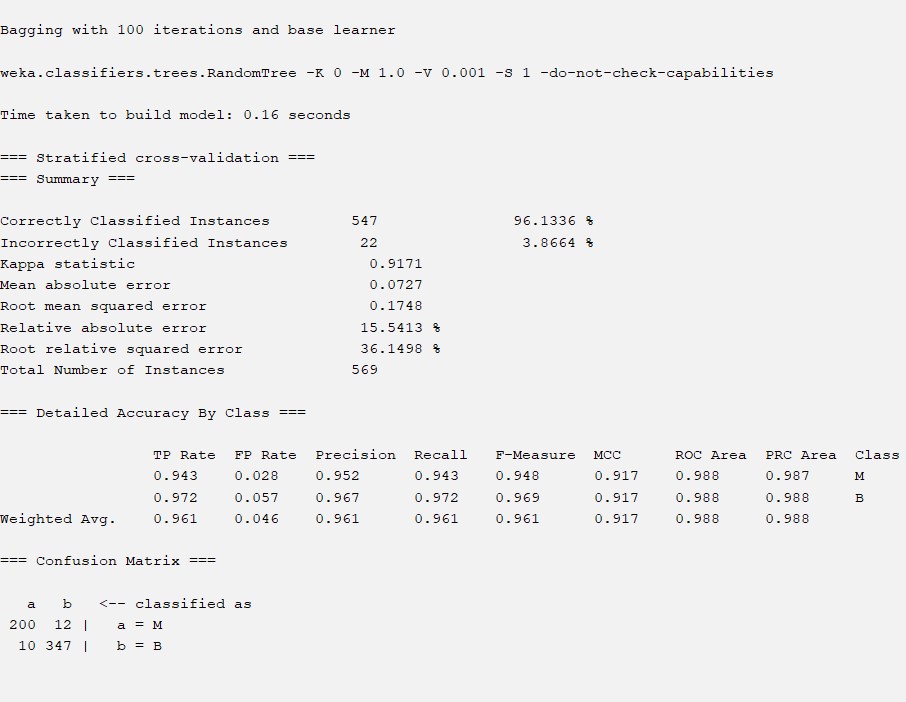


Now Apply all the Classifications

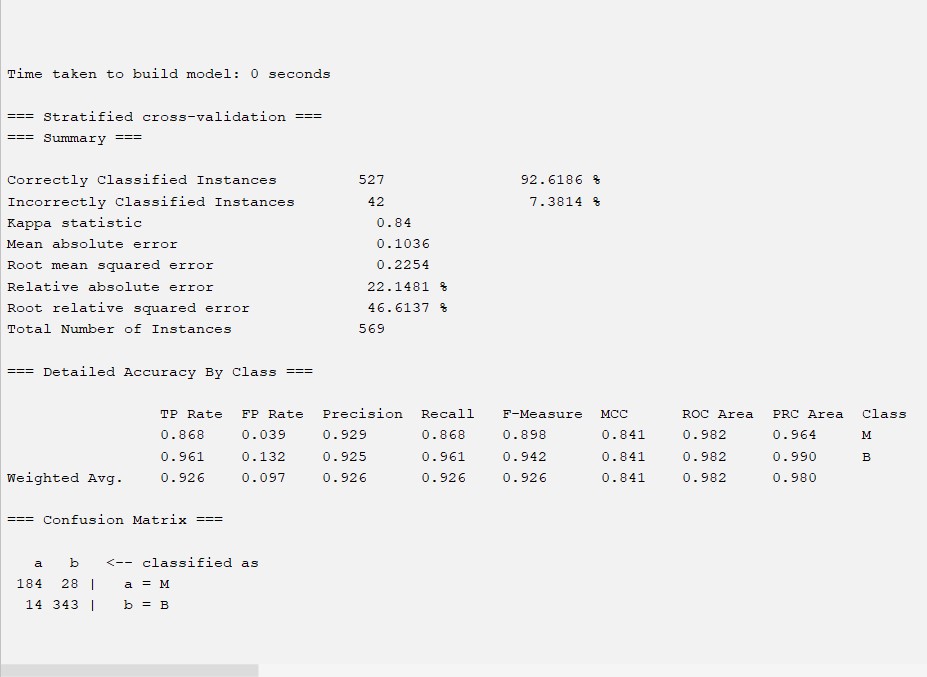
1. Decision trees



1. Random Forest



1. Naïve Bayes

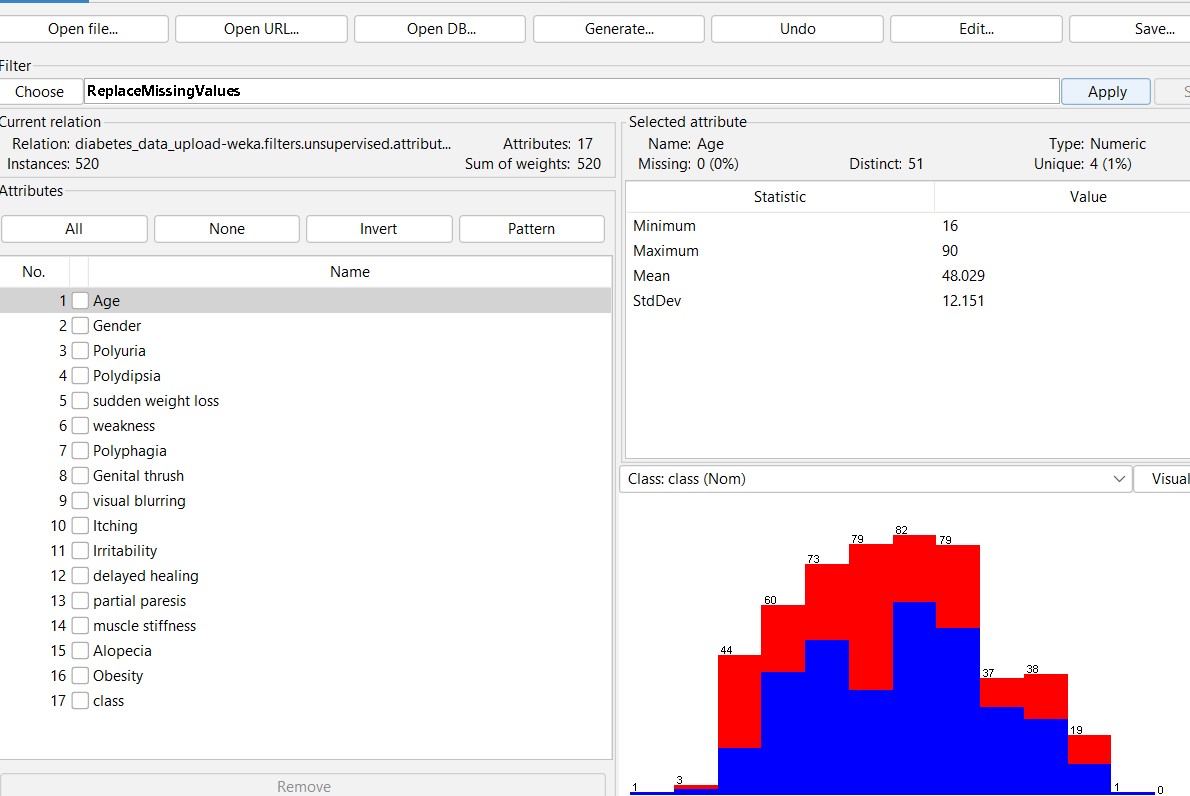


***Result:***

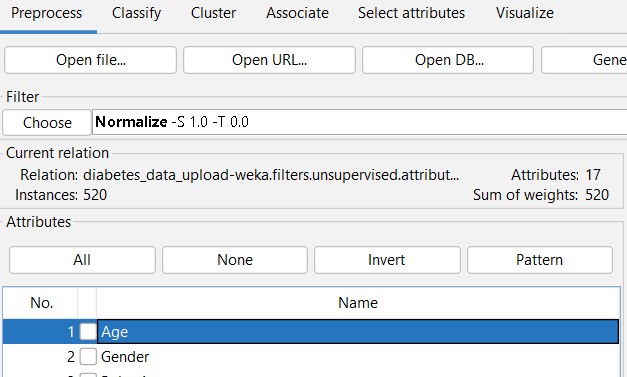
***The accuracies were improved of all the classifiers.***

* **Diabetes Data**

1. Replacing the null values

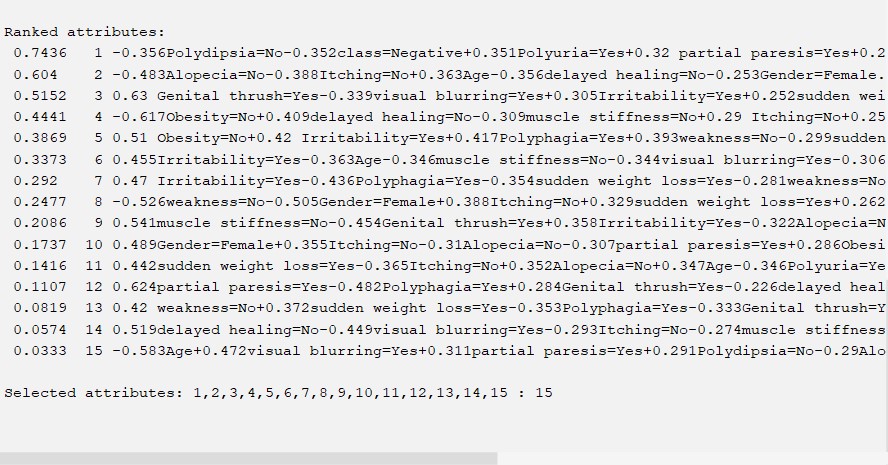


Normalization:



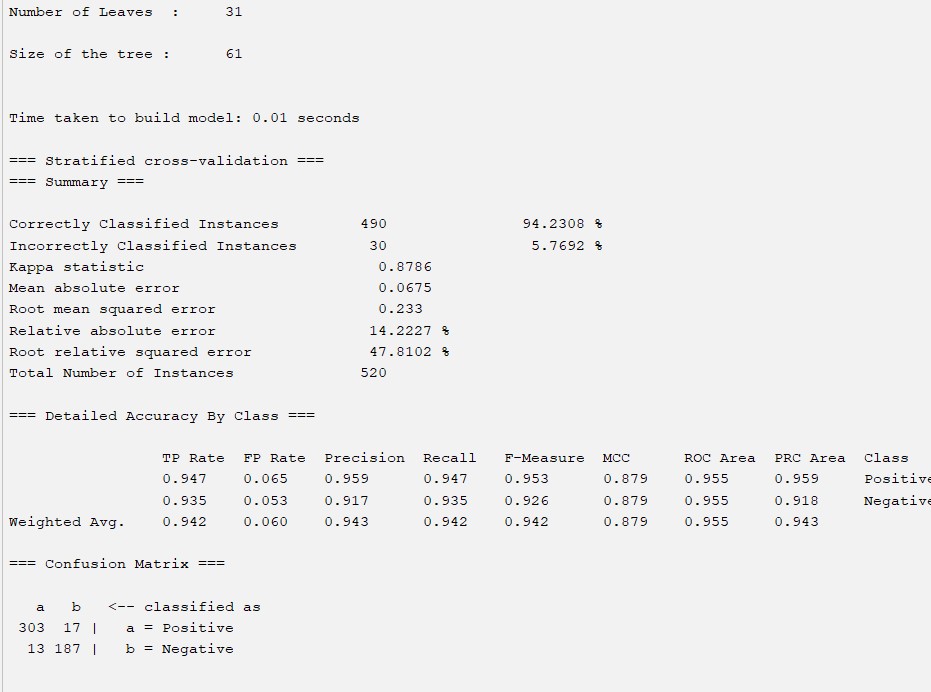
Now we will be applying the Classifications methods on the dataset

**PCA:**

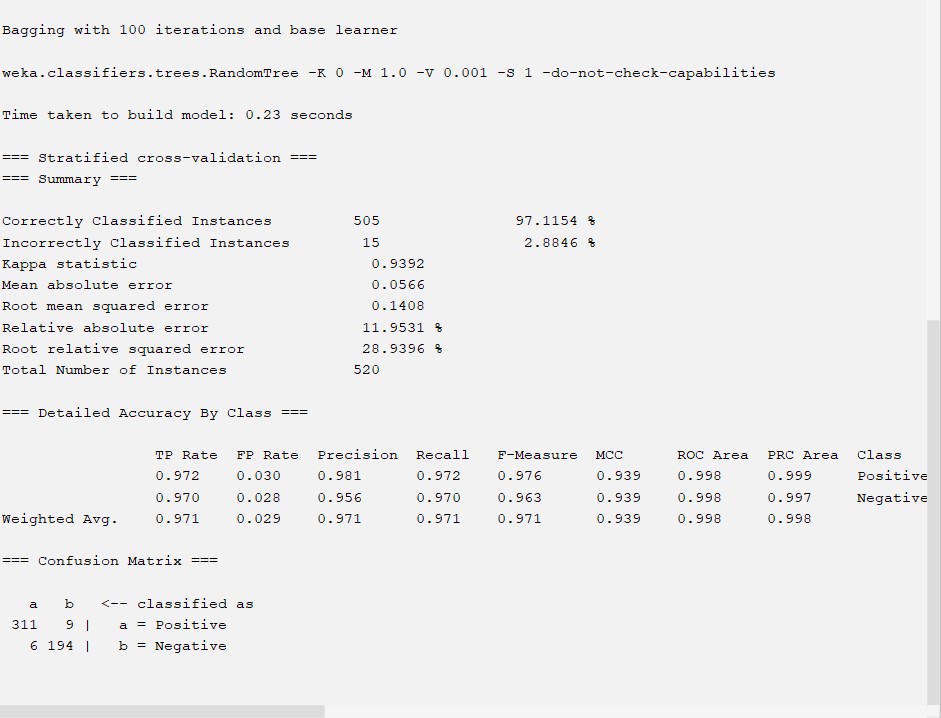
****

**Classification:**

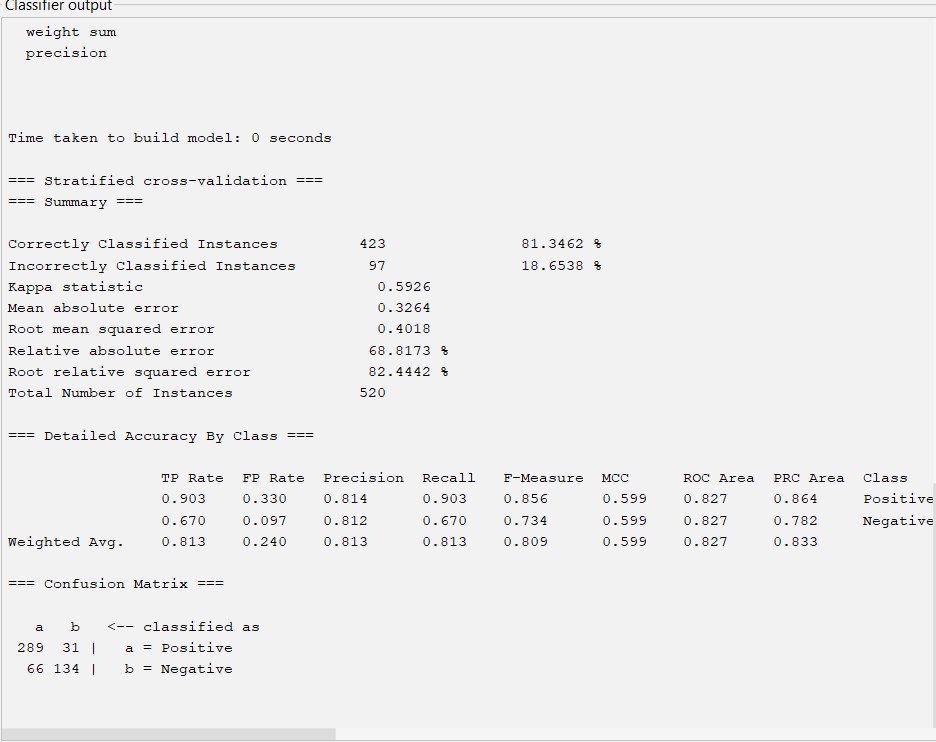
1. **Decision Tree**

****

1. **Random Forest**

****

1. **Naïve Bayes:**

****

**Result:**

Based on the above results, we can conclude that

1. Random Forest is best for classification in this case with 97.17% Accuracy
2. Decision tree is good for classification in this case with 94% Accuracy
3. Naïve Bayes is OK for classification in this case with 81 % Accuracy