**Aleksey Kramer**

**Homework 2 – Data Science 450**

**Table of Content**

[**Decision Model Accuracy** 2](#_Toc433906823)

[**Parameters Used to Generate Model** 3](#_Toc433906824)

[**Decision Tree** 4](#_Toc433906825)

[**ROC Curve** 5](#_Toc433906826)

[**Using Random Forest** 6](#_Toc433906827)

# **Decision Model Accuracy**

**=== Run information ===**

**Scheme:weka.classifiers.trees.J48 -C 0.25 -M 2**

**Relation: Bank Data**

**Instances: 600**

**Attributes: 11**

**age**

**sex**

**region**

**income**

**married**

**children**

**car**

**save\_act**

**current\_act**

**mortgage**

**pep**

**Test mode:10-fold cross-validation**

**Correctly Classified Instances** 537 89.5 %

**Incorrectly Classified Instances** 63 10.5 %

**TP Rate FP Rate Precision Recall F-Measure ROC Area Class**

0.854 0.071 0.911 0.854 0.881 0.877 YES

0.929 0.146 0.883 0.929 0.906 0.877 NO

**Weighted Avg.** 0.895 0.112 0.896 0.895 0.895 0.877

=== Confusion Matrix ===

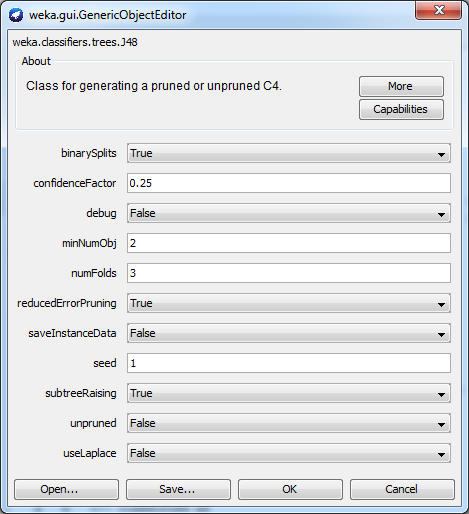
a b <-- **classified as**

234 40 | a = YES

23 303 | b = NO

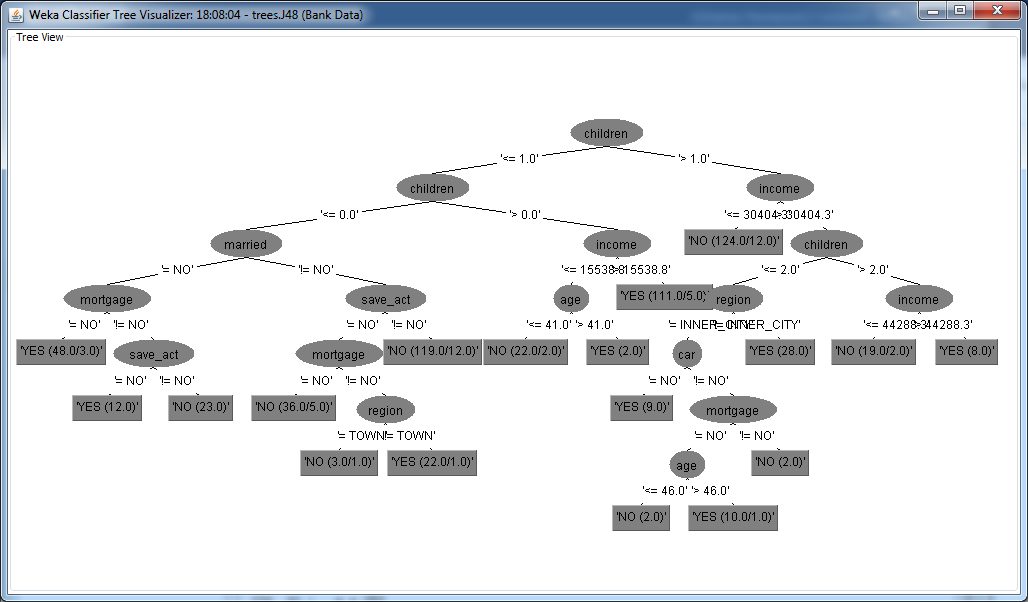
The results exhibit high recall of 89.5% and high precision of 89.6%. This appears to be a good classification model.

# **Parameters Used to Generate Model**



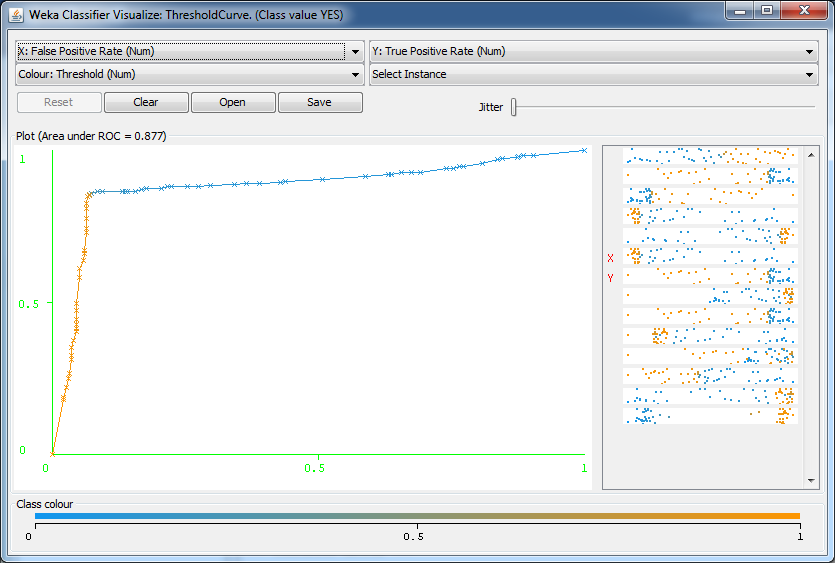
Displaying options selected for the J48 algorithm

# **Decision Tree**



The Classification tree above visually displays the output of the J48 algorithm allowing to track each decision based on the threshold selected by the j48 algorithm with numbers that obey classifications and number of exceptions.

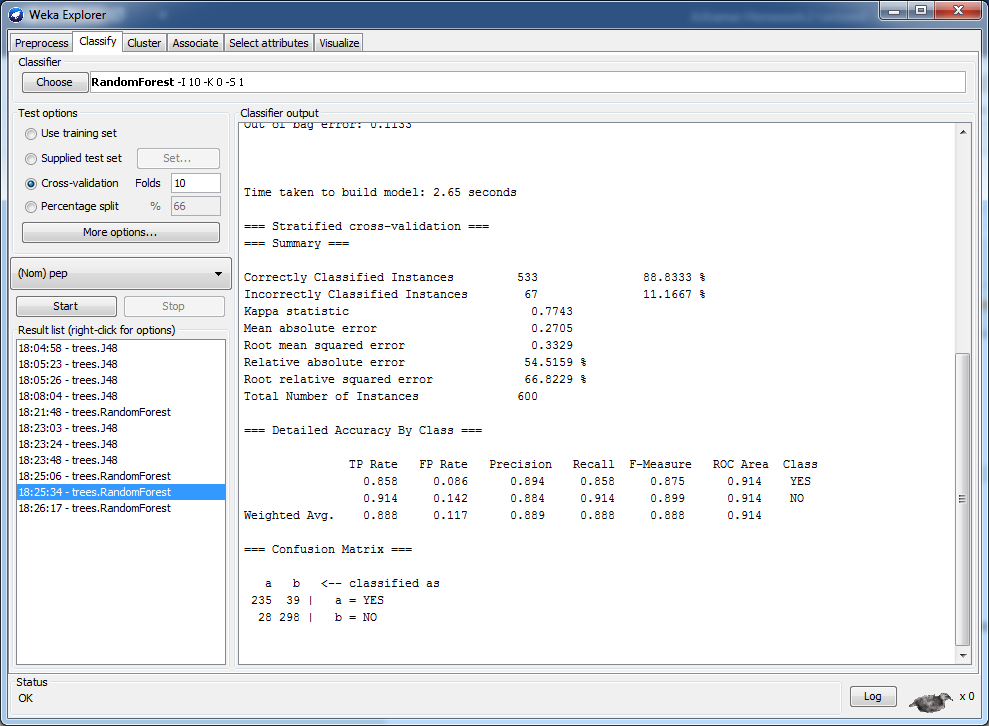
# **ROC Curve**



Visual representation of the TP vs. FP ROC curve showing 98% area under the curve. Pretty good.

# **Using Random Forest**

I tried to use 1000 trees, but it seems that the best result I got was 88.8333% accuracy of classification. Apparently J48 algorithm works better with minimal intervention producing better quality estimates.



The picture below shows the options selected for the Random Forest algorithm.

