**Aleksey Kramer**

**Homework 2 – Data Science 450**

**Table of Content**

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

[**Parameters Used to Generate Model** 2](#_Toc433906824)

[**Decision Tree** 3](#_Toc433906825)

[**ROC Curve** 3](#_Toc433906826)

[**Using Random Forest** 4](#_Toc433906827)

# **Decision Model Accuracy**

**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

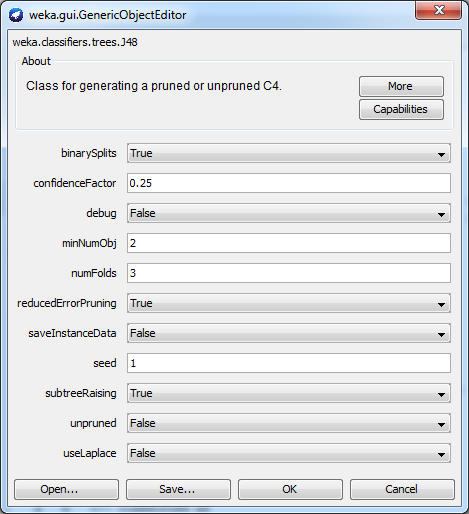
=== Confusion Matrix ===

a b <-- **classified as**

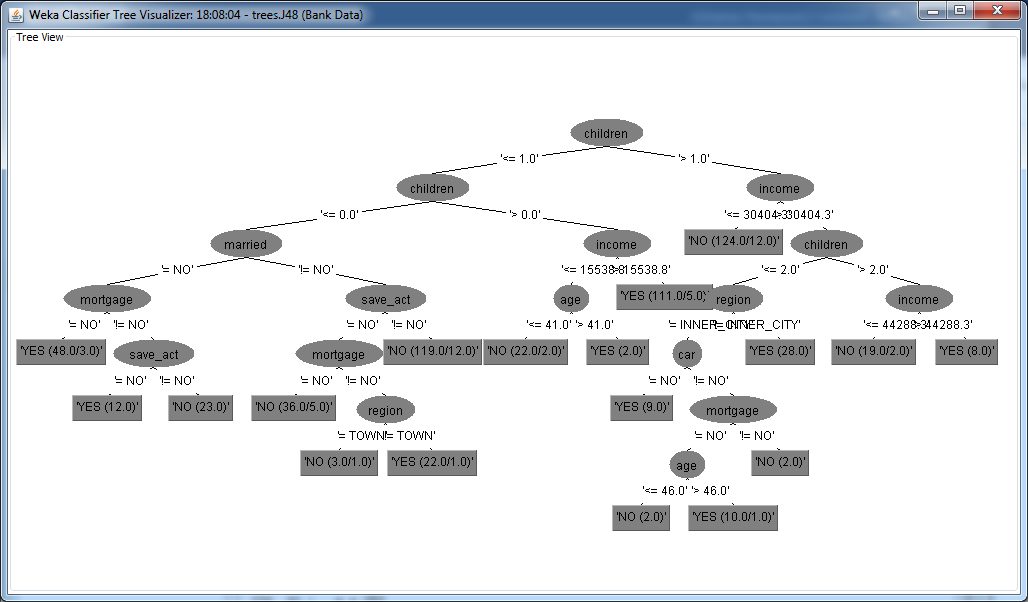
234 40 | a = YES

23 303 | b = NO

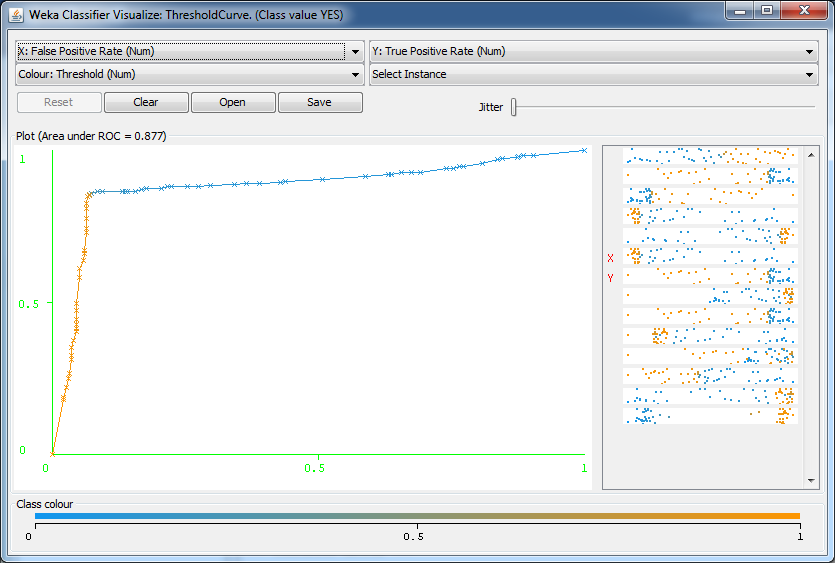
# **Parameters Used to Generate Model**



# **Decision Tree**



# **ROC Curve**



# **Using Random Forest**

I tried to use 1000 trees, but it seems that the best result I got was 88.8333% accuracy of classification.

