This document will serve as the project write-up for the decision tree creation of Sean MacEachern and Zach Arnold at WPI (C Term 2015).

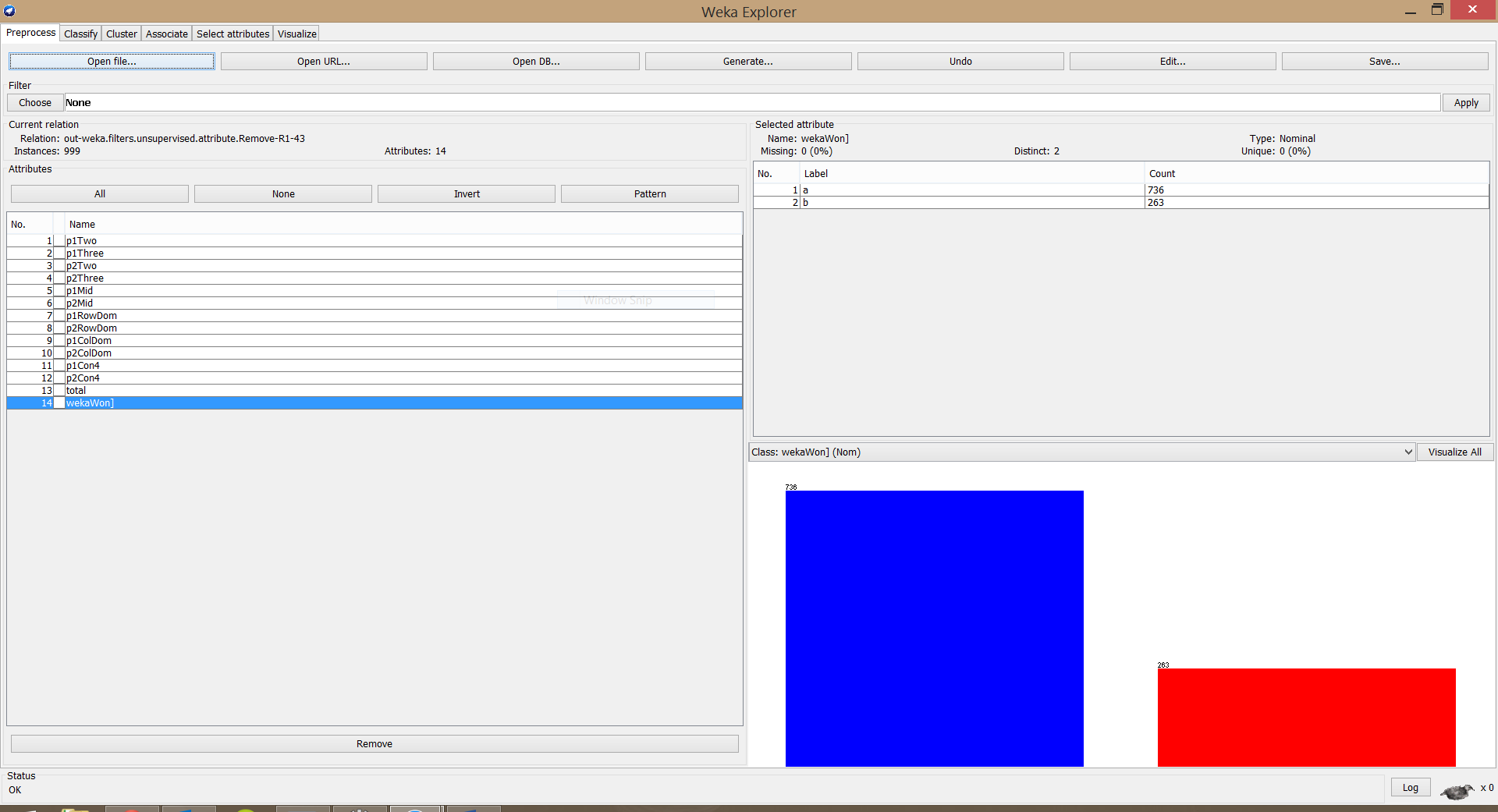
# Feature Selection:

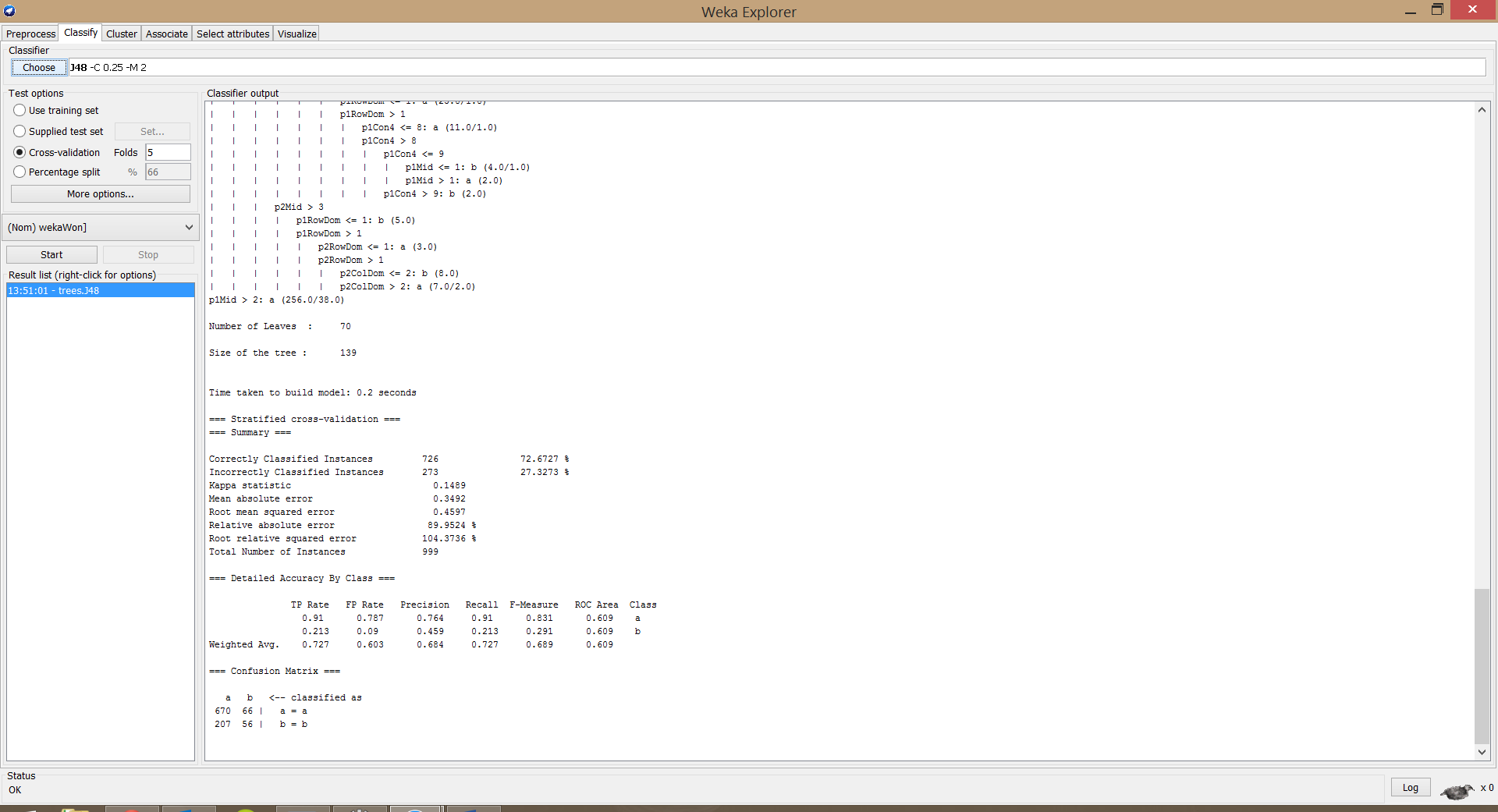
Through our own recollection of good heuristics from project 1 and research on the internet of good features for the solved game of connect 4, we have chosen the following features to classify each board state for WEKA.

* The total number of Connect 2’s each person has
  + Justification:
  + We thought this would be a good indicator of the “streak” heuristic which was our most successful. In essence, we were able to show through a thousand games that
* The total number of Connect 3’s each person has
  + Justification:
  + This one is quite easy to justify. Intuitively from playing the game, you need some series of connected two’s or three’s to win with a connect 4. This feature is basically guaranteeing a win as long as there’s not a blocking piece in the way. (We feel as though this will split the J48 tree nicely.)
* Which player has more columns and rows with pieces in them
  + Justification:
  + This is based off of the chess principle called: “zugzwang.” The formal definition of this strange German word is a situation where a player is forced to make a move when he would rather make no move at all. When one player dominates the board, there is more of an opportunity to win. (I have no idea if this will actually work.)
* Total number of potential Connect 4’s each person has
  + Justification:
  + This one is super easy to justify. If they are one token away from winning…they’re probably going to win.
* The total number of pieces on the board
  + Justification:
  + This is simply a measure of the completeness of the board state. How close to done are we? How mature is the game in this state? This feature is a measure of that.

# WEKA

* Mention complications with Numeric to Nominal
  + Add screenshot
* Show screenshots of actual data with real classification
* Kappa Statistic is better closer to 1
* 5-fold cross validation





|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Correctly Classified | Incorrectly Classified | Kappa Statistic | Mean Absolute Error | Relative Absolute Error |
| 72.6727% | 27.3273% | 0.1489 | 0.3492 | 89.9524% |
|  |  |  |  |  |