**Progress Report 5: Crossover Strategy, Tree Feature Extraction & Support Vector Regression**

This week I focused on finding a crossover strategy for our tree chromosomes, this allows us to combine trees from different clusters and mimic nature by “mating” them. I also spent some time extracting tree-based features in order to observe the relationship between or ByteBeat tracks and their underlying structure. Lastly, I attempted to used a Support Vector Regression to get a better prediction of ratings.

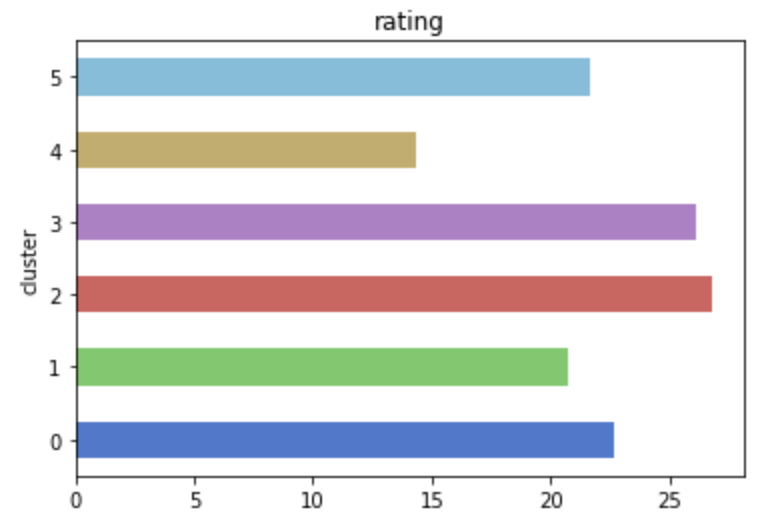
Starting with the crossover strategy, my approach is to select two trees from different sound-based clusters, and have them swap sub-trees. Sub-trees of lower heights have a higher probability of swapping while higher ones swap quite rarely. Additionally, the sub-trees are swapped at equal heights as shown in the figure below. The probability of crossover in out Genetic algorithm should be quite low, considering this is a significant mutation to the original tree and has the potential to break the original structure that made it musical.

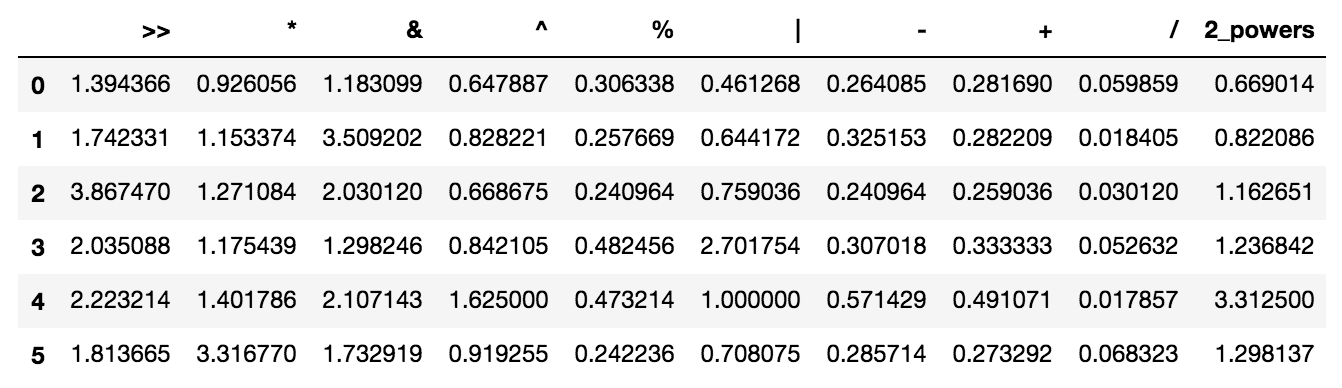


Once the trees can freely mutate and mate, I wanted to analyze these trees and how they relate to the music they generate. For this, I extracted the following tree-based features and ran a k-mean clustering algorithm based on them.

* Height: height of the tree.
* Leaves: number of leaves.
* t-count: number of ”t” leaves.
* Operators: number of operators.
* Count of each operator: 9 columns counting each operator.
* Powers of 2: number of leaf nodes that are powers of 2.
* Average Operand: Average value of operands

The clustering revealed some interesting patterns. The highest rated group contained many bit-shift and bit-wise “&” operators, which is in-line with Viznut’s research on the Sierpinski Harmonies. The lowest rated group had a surprisingly high-amount of powers of two, which runs contrary to Viznut’s hypothesis about these numbers.





The last bit of progress I made this week was trying out different models to try to predict a song’s rating based on its tree and spectral features. So far I have had little success in generating accurate predictions, but the best model so far has been the support-vector regression. I will meet with my Machine Learning TA to try to solve this issue.