84%

When running

Mine freq before using apri

Pre-mined patterns were included so iniial sisdn’t have to .

Rules generated by the program.

Uses rules to classify tumors.

Sort into bins 3 for each point: sm med lg

Tumor has a list of bins it goes into

Transaction list is composed of tumors: only has bins the values go in

Mine the trans list for patterns freq. 15 support count.

Use those rules for Bayesian rule generation

Any tumor higher thana 55 or higher to be benign it is added to rules list.

If a tumor fits a rule it has a 55% chance of being correct

Did not calc an optimal prob

Did not gen rules for malig. Defaulted to malignant. Some maligs fit rules for benign

Reason for b84%

Benchmark is faster due to pre-mined patterns and we have to mine itself.

12 rules acc = 84%

Our project implements a Rules-based Classification system using a Naïve Bayesian Classifier. Our project uses this to classify cancer tumors as either malignant or benign. The report we borrowed data from performs this in a similar way, the only difference being that the patterns were pre-mined, and therefore it was not necessary for them to generate their own rule sets. Our project, on the other hand, our project must first mine for frequent patterns using Apriori then use those patterns to generate a rule set implementing the Naïve Bayes Classifier.

In order to perform this our project places the tumors into a list of bins; small, medium, and large. Each tumor has a list of bins it goes into. This list of bins for each tumor is then put into a transaction list where each tumor is represented by its unique list of bins. This transaction list is then mined for frequent patterns using a support count of 0.15. After the transaction list has been mined for a list of rules those rules are used to perform the Bayesian Rules Generation. To perform this, we take the odds a tumor is benign and the odds a tumor fits into a given rule to determine the odds that a tumor that fits within the rules it qualifies is also benign. Any tumor with greater than a 55% chance of being benign is labeled as such.

Overall our project generates 12 rules and has an 84% accuracy rate. Our code runs at a slower pace than the study we used as a benchmark, but this is due to having to use to use Apriori in our rules generation scheme, while the benchmark study used pre-generated rules. Our project also did not calculate an optimal probability that a tumor must reach in order to be classified as benign. We also did not generate rules to define malignant tumors, and instead had tumors default to malignant, only becoming benign if they breached the threshold. Two improvements that would provide an increase in accuracy would be to define an optimal probability for benignness and generate rules for malignant tumors, as well.