# Data Mining

**Name – Surya Prakash Sekar**

**Email id – [sursekar@iu.edu](mailto:sursekar@iu.edu)**

## Decision Tree

## Observations

- When I use entropy and average the accuracy over 5 trials over random test and train splits on the Immunotherapy dataset, I can achieve an accuracy of approximately 81.4%.

- When I use Gini and average the accuracy over 5 trials over random test and train splits on the Immunotherapy dataset, I am able to achieve an accuracy of approximately 82.3%.

- When I use entropy and average the accuracy over 5 trials over random test and train splits on the Cryotherapy dataset, I am able to achieve an accuracy of approximately 84%.

- When I use Gini and average the accuracy over 5 trials over random test and train splits on the Cryotherapy dataset, I can achieve an accuracy of approximately 84.6%.

- There is a negligible difference whether I use Gini or info gain based upon entropy for splitting, but in my code, I have observed that the accuracy was a bit more when using Gini

- I noticed that entropy takes more computational time, it could be because it makes use of the logarithm functions.