I’m going to be honest, I don’t know if this will get me the 5 points needed. But I felt like if I had more pages to write, I could’ve hit all of these points. It is a risk I’m willing to take though, because I feel like the biggest reason I missed points was due to a misconception as to what ‘analysis’ meant. I spent more time examining the model behavior and how that was affected by different parameters and you wanted to see us show what was shown to us in the lectures. I feel like my analysis into the parameters, how the models did on each dataset but running out of room for the cross dataset model analysis. I was focusing more on giving an actual analysis and not just making basic observations of ‘See, this parameter does better than the other parameter so it is better’. Note in the DT portion my analysis between the Gini and Entropy parameters.

Overall, one of the main reasons I felt I did poorly on this assignment was due to the page length. I had all of the portions mentioned below (learning curves, in depth examination into the datasets, class imbalance and how that effected the models, etc) but with all of this information I was close to 20 pages. I had to trim it down somehow, so I cut out the portions that I thought were basic observations and not analysis (looking at a learning curve, and seeing the class imbalance in the dataset). Due to the lack of a rubric I decided that as these portions were more general observations (and didn’t provide what I thought was an analysis) I removed them from the paper to make room for what I thought was a deeper analysis. I understood the analysis to be analyzing how the parameters affected the models performance, not just stating what has already been stated in the lectures.

I also read through a friends paper, who scored well above me, and I noticed that in his paper it was all general observations and didn’t quite “boil down to: why... WHY WHY WHY?”, but in my paper I attempted to not just notice a behavior but give what I believed to be they WHY behind the behavior, which is why I had to cut some of these portions out. I literally helped him with this analysis and when we got to the paper portion I told him to do the best analysis he could, looking into the parameters and behaviors of the models.

Even though I didn’t include the learning curves, I feel like my exploration into other aspects for each of the models should’ve made up for this. I felt like putting a graph for the learning curve and stating the obvious wouldn’t be good enough as in the description the analysis portion is weighted so heavily.

* Detailed description for both the problems from an ML perspective in terms of sample sizes, length of features, class imbalances, categorical / continuous variables, etc. is required. Insights into why these datasets might produce differing results when used with same ML algos is needed.
  + I had a whole page on this section but had to cut it out to meet the page requirements.
    - I talked about how due to a major class imbalance, the data had less than 10% of its values as 1 while the rest were all 0. I had graphs that showed what this does to a model and how drastically you can chance the performance based on the balance of the data.
* DT-Learning curves (validation scores v/s number of training samples) is missing here. Analysis on bias variance is required (eg. which parameters contributed to the bias/variance, how can these issues can be mitigated by tuning hyperparameters / cross validation, etc.)
  + I know one of the biases of a DT is that it prefers shorter trees, and that can clearly be seen in the graph, it has been demonstrated
  + I know the biases of all of the models including, DT shorter tress, ANN smaller networks and smaller weights, Boost more focused on stumps, KNN location matters, and SVM linearly separable data
* Very good job on comparing all the algos on both datasets along with the confusion matrices! A bit more insight into which features are more important than the other, how performance can be improved, comparisons through cross validation, occam's razor, etc. could have been done.
  + I would’ve loved to provide more information into feature importance, why one variable might affect the performance, cross validation across the different parameters but again, 12 pages is hard to fit all this information into!

Conclusion, well I believe I deserve more than a 47. Yes, there were no learning curves and I didn’t make those basic observations so obviously not a 90 either. Based on your feedback, and the analysis that was performed I believe I should fall somewhere within the 60-70 range. If you want to sit down and test my knowledge of these things, or ask the friend how I helped him, or subject me to any other type of test to prove that I’m not just saying these things and that I actually know them and so I focused on an analysis and not just the basics, I would happily submit.