Problem description:

Good job - selecting and describing the dataset you have chosen.

The description could use more details to explain why the dataset is interesting from an ML perspective (e.g., reasons could be number of features, type of features, number of data points, the correlation between features, the non-trivial relationship between features and output, class imbalance, or number of classes).

It is expected that you choose two distinct datasets and explain your thoughts on why they are suitable for comparing and analyzing various algorithms.

It’s expected that you explain the performance metric used.

Algorithms:

- It is expected that you plot learning and model complexity curves for each dataset. There is not much analysis as the description is lacking the required details to make a connection with the problem. It looks like plots/analysis is missing for all the algorithms. Why did you get the results you did? How do you compare and contrast the different algorithms? What sort of changes might you make to each of those algorithms to improve performance?

While it’s nice that you tried out the experiments, the analysis depth needs to be improved much more. It might help to think/write down

- The objective for running the experiment. – Why scaling?

- What input is ideal, what output is expected, how to measure the performance? - How to scale? How scaling improve the performance? How to measure the improvement?

- What algorithm is appropriate, what are free parameters and why?

- What is unknown and needs to be tested (e.g., the effect of hyperparameters on the performance)? – Why tune those hyperparameters? How they affect the performance?

- What results are observed and what do they tell us about the objective? e.g., how does performance vary with different hyperparameter values? – Whether the results are as expected and why?

- Have we found the best algorithm / parameters? What is best (if at all)? Is it going to work equally well on the unseen data? Why / why not?