With the drug consumption dataset, oversampling and under sampling of dataset lead to reduced accuracies. While this could be some what alleviated using parameter tuning, there wasn’t a significant difference in improvement. Oversampling also decreased the cost efficiency as MLP and other algorithms were more computationally expensive. Performing Parameter tuning for oversampled MLP data took hours to converge and there isn’t a significant improvement for that cost overhead. The accuracy values plateaued at 90.

With the labour negotiations dataset, Multiple columns had more than 50% of data missing while features like standby pay were almost 85% missing. While statistical methods could be done on < 50%, for other features they proved faulty. Use of other ML models to predict such data helped improve the accuracy of results. Feature selection was done by dropping values with least variance i.e.., if the values were similar for more than 85% of the data, they were dropped. For some models, parameter tuning using grid search didn’t yield great results. Thus, a Bayesian optimiser was used to find values for parameters in a range. Using them in grid search with other categorical parameters helped improve the accuracy to a great extent. The accuracy values plateaued at 95.

With the heart disease dataset, parameter tuning boosted the accuracies the most here for almost all algorithms. Feature selection was done by dropping values with least variance i.e.., if the values were similar for more than 85% of the data, they were dropped. After this parameter tuning was done similar to labour negotiations dataset using Bayesian optimiser and GridSearch. The accuracy plateaued at 84.

Finally, performing the Friedman test we realise that almost all algorithms performed similarly on given dataset. The accuracies plateaued and this led to reduced difference between the algorithms.   
Thus, the null hypothesis holds true and no significant difference was observed.