**Project 3: Classification Algorithms**

Group #: 1 (you own group number, same as in Proj1)

Tianle Ma, 50096690, [tianlema@buffalo.edu](mailto:tianlema@buffalo.edu)

Other members

1. **Understanding of the algorithms**

**For each algorithm** (neural network, decision tree, naïve Bayes, random forest), use your own words to answer these questions (directly copy from slides or other sources will cause you losing points):

1. How it works: workflow of the algorithm and your implementation. (4 points)
2. When it works and when it doesn’t. State pros and cons. (4 points)

In addition,

1. For neural network, how do you handle categorical attributes? Any alternative ways? (2 points)
2. For decision tree and naïve Bayes, how do your handle both numerical and categorical attributes? Are there any other better ways? (2 points)
3. **Result analysis**
4. Neural network (8 points)

* Try a few of parameter settings (number of layers, number of variables in each layer) on both dataset 1 and dataset 2 (for dataset 2 you need to handle categorical features). These settings should be adjustable in your implementation.
* Adopt 10-fold Cross Validation to evaluate the performance on the provided datasets in terms of Accuracy, Precision, Recall, and F-1 measure. Show your results in tables and plots. Compare and analyze your results.

1. Decision tree (8 points)

* Implement your decision tree use two or more metrics (including Gini impurity and information gains).
* Try to change the way how to split a node. Adopt 10-fold Cross Validation to evaluate the performance on the provided datasets in terms of Accuracy, Precision, Recall, and F-1 measure. Show your results in tables and plots. Compare and analyze your results.

1. Naïve Bayes (6 points)

Implement naïve Bayes. Adopt 10-fold Cross Validation to evaluate the performance on the provided datasets in terms of Accuracy, Precision, Recall, and F-1 measure. Try your best to improve the performance. Show your results in tables and plots. Compare and analyze your results.

1. Random forest (6 points)

Implement random forest. Adopt 10-fold Cross Validation to evaluate the performance on the provided datasets in terms of Accuracy, Precision, Recall, and F-1 measure. Try your best to improve the performance. Show your results in tables and plots. Compare and analyze your results.

1. **Comparative Analysis** (4 points)

For dataset1 and dataset2, compare the performance among all four (or more) algorithms and analyze the results. Which works best? What are the possible reasons? Do you have any other findings?

1. **Classification strategies for dataset3**

(1) State your method in detail. Why do you choose this model? (10 points)

(2) How do you choose model parameter? (5 points)

1. **Other considerations** (10 points)

If you have done extra good work, depending on what it is, we may or may not consider to give you extra score. Please state it in your report **as well as during the demo**.