README

There will be four parts in this project report excluding the Introduction Part.

**First Part: Exploratory Data Analysis (EDA)**

In this part, we analyze the data set in an exploratory way to understand basic properties of the data. We try to show the simple relationship between features and income variable visually so that we could simplify and regroup the data set in a cleaning way.

**Second Part: Analysis and Methods (model building)**

In this part, we analysis the data by applying different models. Keep in mind that our goal is to predict whether a person could make over 50K per a year based on the other variables. We used five different models (Classification Tree, Bagged Tree, Random Forest, Neural Networks, Support Vector Machine) to see which one could give us the best result based on the highest test accuracy rates and AUC values.

**Third Part: Model Selection**

In this part, we select the best model based on the results from the previous two parts and discuss why we select that method.

**Fourth Part: Conclusion**

In this part, we make overall conclusion of the whole project.