Machine Learning Visualization
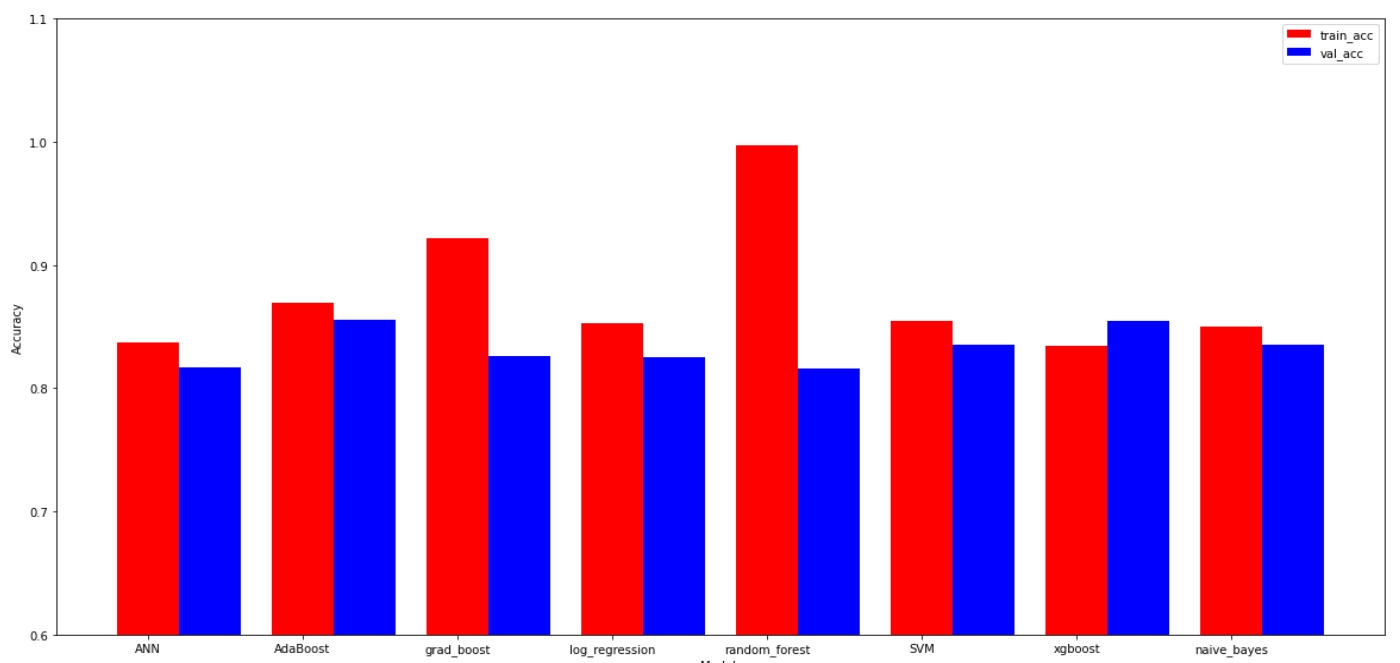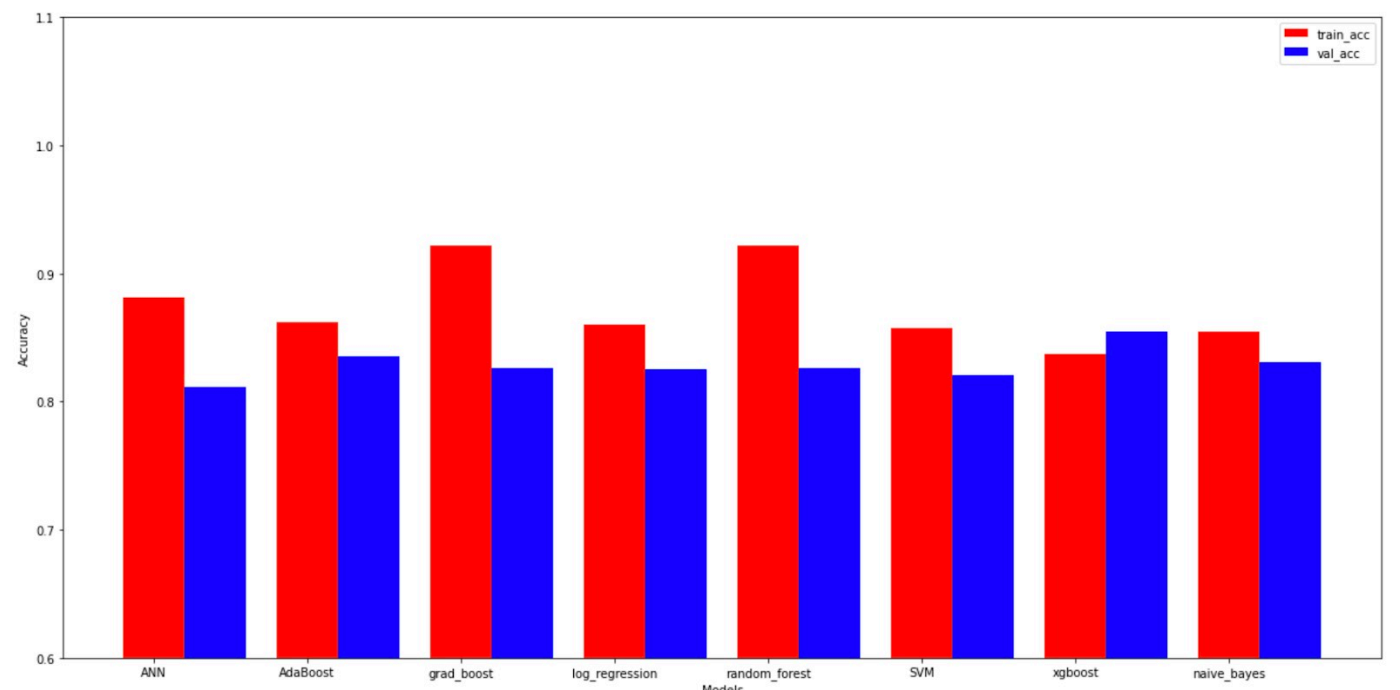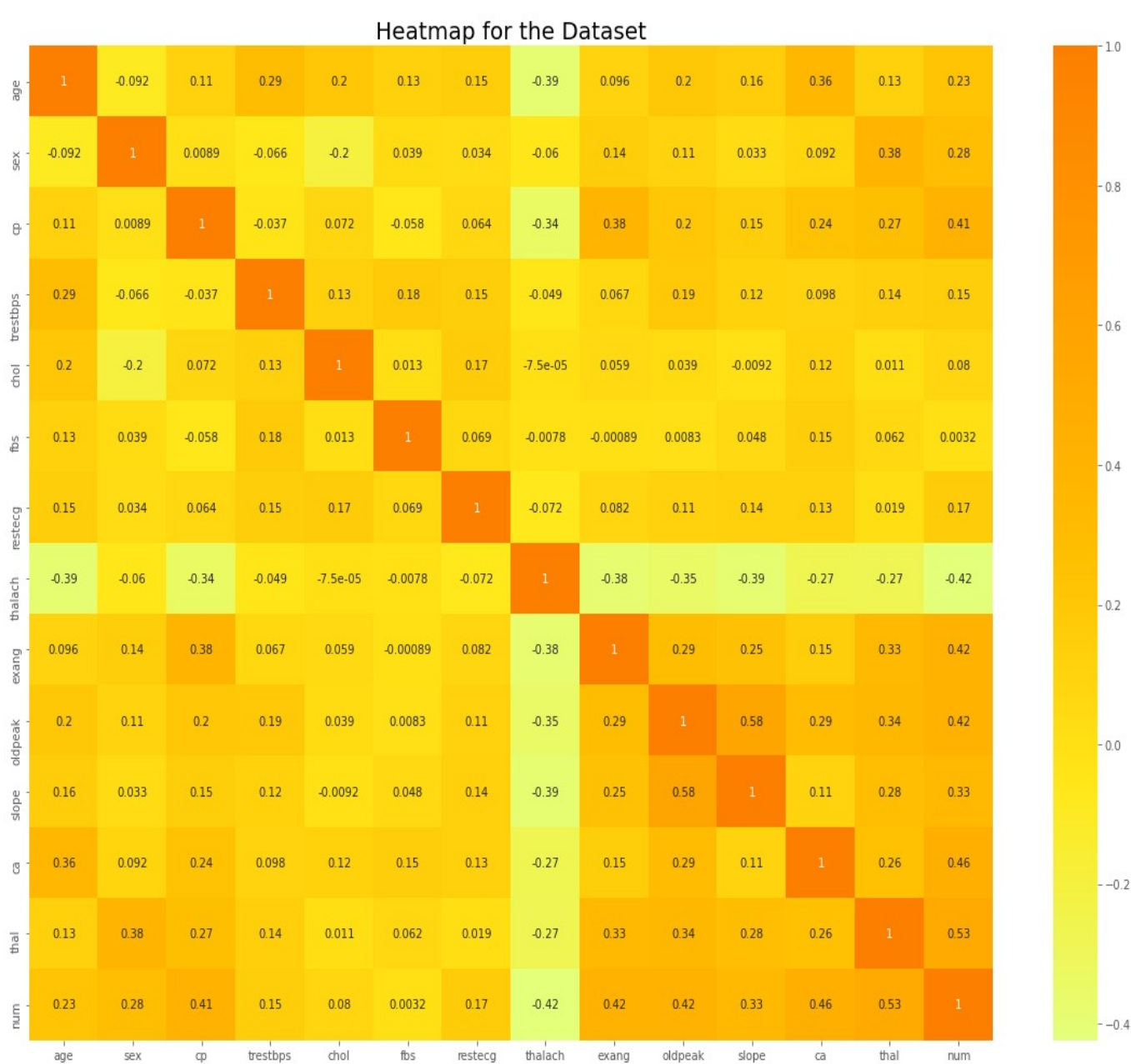## for Dimension Reduction based on Correlation

Student: Chen Feiyu   Supervisor: Dr Zheng Jianmin

Heatmap for the Dataset

Project Objectives:

The objective of the project is to investigate on the influence of using subset as input for machine learning to the accuracy of the corresponding model using different machine learning algorithm. The subset is selected by different methods based on visualization and other techniques.

- The research on different model training algorithm like ANN, SVM, XGBoost, AdaBoost, random forest.
- The processing of dataset using different type of visualization techniques. The result of the visualization enables us to make assumption on the importance of the features within the dataset, which can be tested by using model selection method during training.
- The implementation of different algorithms and different feature selection methodologies on the datasets.
- The testing and comparison of the performance of models. It is able to provide results and verify the assumption from the data visualization process.

Project Result:

From the experiment, it can be concluded that by using different feature selection methods and different training algorithms, the accuracy of the model using variance threshold and correlation threshold with threshold = 0.1 is better than the original model without using any dimension reduction and feature selection method.

As a result, using visualization tools and method to do data analytics can improve the performance of the machine learning model.