

Outline for final project

Group 1

2024-11-30

Outline for the project

Some reminders

1. due 10 th
2. Linear regression for multi variable
3. multilinearity
4. selecting the important features

Step 1: reading and processing the data:

collecting and reading some relative papers about this topic and the use of machine learning

cleaning the data and doing transformation on the original data set dummy variables for string type normalization on most variables care about the percentage variable (whether to /100)

due: 1st' night Dec.

Step 2: EDA(graphs and describing the data)

plot the original data using box-plot, hist-gram, and so on create the descriptive table(maybe put in the appendix)

due: 3rd' night Dec

Second meeting on Dec 3rd

Third meeting on Dec 5th evening/afternoon

Step 3: feature selection

(choose both language Python and R)

using PCA or Lasso method to filter important predictors trying to use clustering to group the original data set

fitting models and test

due: 6th' 9:30am to 2:30pm

Step 5: paper writing

more than 15 ~ 20 or more, not including the reference latex may be useful

due: 8th Dec night

Step 6: submitting the paper

due : 8th Dec