Modeling **Meiling Liu**
May 5, 2019

After feature engineering, I would like to try several models on the processed datasets. The work flow is as follows:

1. split training data as training and validation datasets;
2. run processed data with random forest, xgboost and neural network algorithms.
3. Other ensemble algorithm could be used to improve the accuracy, such as stacking.
4. **control overfitting**
    - learning curve: use validation set to prevent overfitting;
    - for neural network: use drop-out which will remove co-adaption;
    - for xgboost: 1) tunning regularization parameters such as lambda and alpha; 2) learning rate; 3) column subsampling, which is commonly used in random forest.