1) Preprocessing Problem Data (preprocessingProblemData.py)

- The 'level_type' column is transformed to a numeric column and missing values have been replaced by the mean
- The missing values of 'point' column has been imputed by the mean of 'point' values that belong to the corresponding 'level_type'
- The 'tags' columns in converted into multiple columns denoting whether the problem belong to the particular domain or not.

2) Preprocessing User Data (preprocessingUserData.py)

- The 'rank' column is transformed to a numeric column and missing values have been replaced by the mean

3) Merging Data (mergingData.py)

- The data frames of problems and users were merged with respect to 'train_submissions.csv'

4) Prediction (analysis.y)

- Feature Scaling was applied to the features data frame.
- Test data is restructured so that the model can predict
- Random Forests Classifier was used to predict the 'attempts_range'.