

在本次大数据预测项目中，我使用了线性回归算法进行机器学习，将 2013 年学生的各个方面有用的特征提取出来，进行 2014 年学生成绩排名的预测。

groupAndAnalyzeSubmissions.py 这个文件把 2013 年数据中的 submission 按照学生 id 来分组，然后对每一组进行分析，并提取出特征。提取出的特征有：学生得到满分时的那次提交的时间、学生第一次提交作业的时间、学生第一次提交作业的分数、特定的某一题学生提交的总次数。

cal2013examGrades.py 这个文件把 2013 年所有学生 exam 的成绩算出来，用做训练。

calIndividualCharacters.py 这个文件为每个同学提取一次特征，特征有：完成一道题的平均提交次数、平均提交的分数、平均第一次提交的时间、平均满分时提交的时间。

combineFeaturesAndTraining.py 这个文件将所有同学的特征与 exam 成绩结合用做样本训练。

applyAndGenerateResults.py 这个文件把由 2013 年的数据训练出的模型应用在 2014 年的数据上进行成绩预测，并生成预测结果。