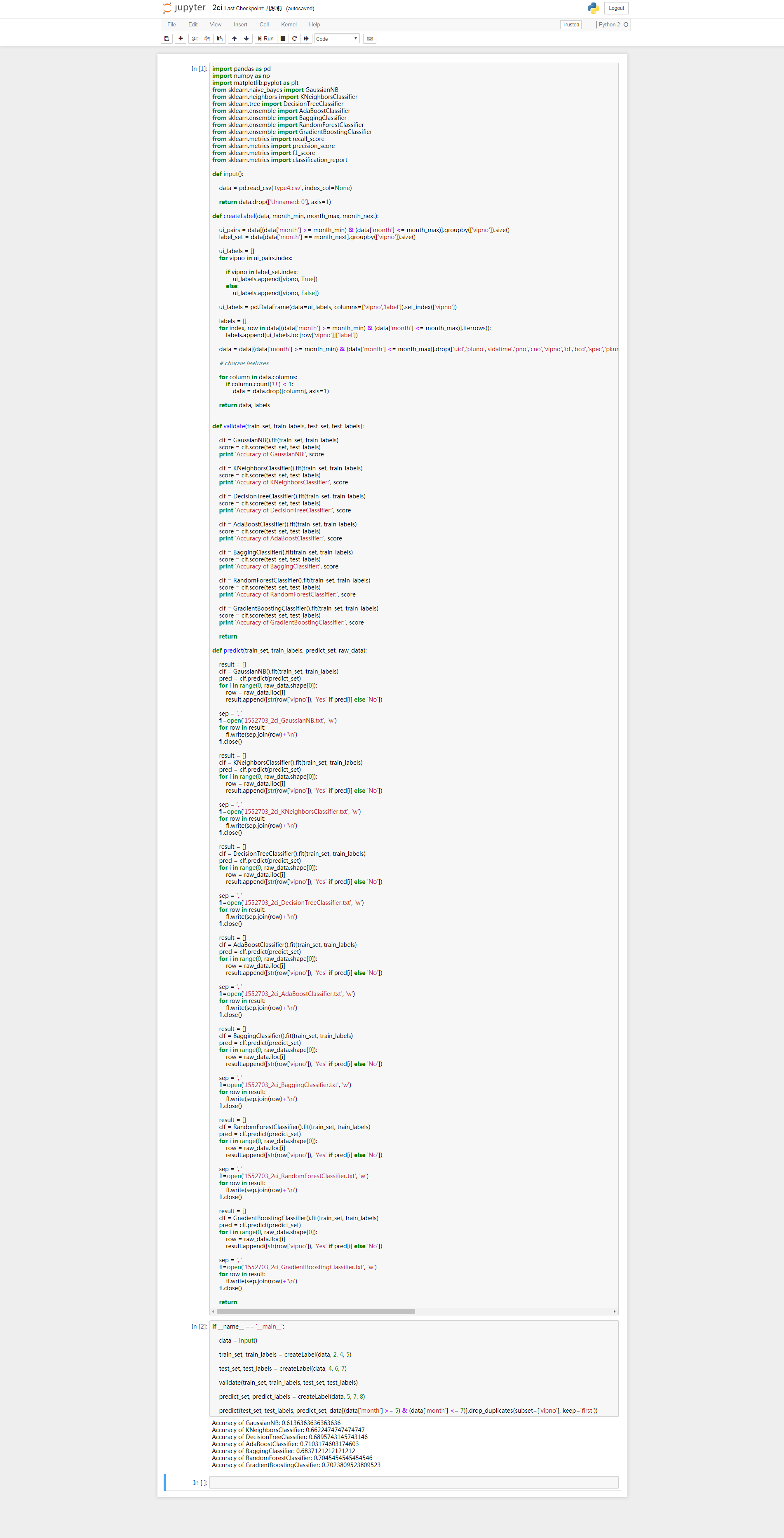
## 代码运行

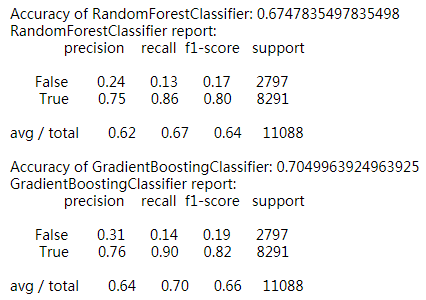
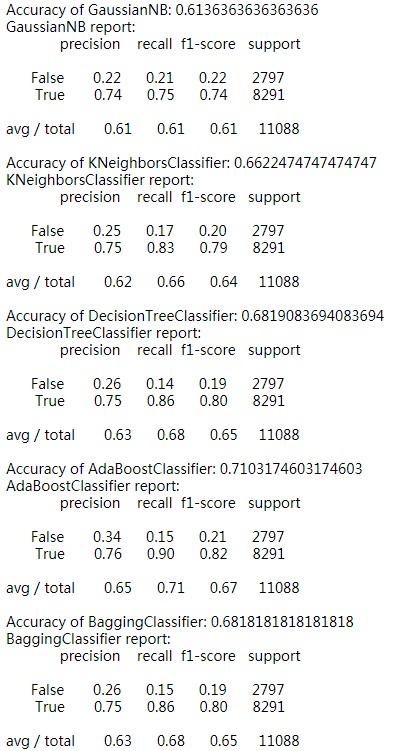
[点击查看原图](ci.png)

运行所需type4.csv文件可在a的特征结果中找到



### 讨论分析

1. 本题预测用户是否会在第k+1个月存在购买可能，依然使用1a中的七个分类器
2. 七个分类器的特点分析见上一小题
3. 本小题中并没有出现b小题中的问题，即使用全数据的分类效果现在好于使用该用户的第一条数据，分析如下：
   1. 本次预测中，进行的是二分类预测，不存在多标签问题中差异过小的问题，训练数据一定程度上来说越多越好。
   2. 一个用户一条数据的情况下，训练集数据将会过少，会存在b中不考虑的随机性问题，一条记录并不能代表足够的内容。
4. 本题中仅使用与用户（vipno）有关的特征，约为272列，在特征集命名中表现为含有“U”的列。
5. 预测模型评估如下：
   1. 准确率在0.61-0.71之间，相比于b结果较为一般
   2. 原因分析如下：
      1. 标签和特征联系较弱，训练集中所用特征和用户是否那个月会购买商品的联系远弱于和用户那个月是否会购买某样东西的联系。准确率理应有所下降。



### 性能比较

比较各分类器训练并预测8月数据所耗费的时间。