CADV Question 2 代码说明：

在问题二中，我们使用了gradient boosting classifier算法作为我们的核心算法，将用户当月和前一个月数据整合成当月数据，并通过两组差异较大，但预测结果均较好的模型进行整合得到最终的模型。在测试模型效果阶段用用户2-6月份信息进行训练，模型预测9月份未来三个月的换机情况，最终选出最佳搭配的变量组合，再用1-9月份进行训练，来预测16年1-3月份用户是否换机，其中我们用到的两组features包括：

1、第一套变量

|  |  |
| --- | --- |
| 变量编号 | 变量代码 |
| 1 | 未来3个月各牌子用户流失率（16年1，2，3月用15代替） |
| 2 | 用户前N个月平均电话通话数 |
| 3 | 用户该月使用手机牌子登机 |
| 4 | 用户使用手机价格（中关村上爬虫获得） |
| 5 | 用户前N个月更换电话平均数 |
| 6 | 用户当月APRU |
| 7 | 用户当月性别 |
| 8 | 用户当月网段 |
| 9 | 用户当月年龄 |
| 10 | 用户当月流量 |
| 11 | 用户当月短信 |
| 12 | 前两月是否换手机 |
| 13 | 前1月是否换手机 |
| 14 | 当月手机百度搜索结果数 |
| 15 | 该牌子用户中使用总数 |

2、第一套变量

|  |  |
| --- | --- |
| 变量编号 | 变量代码 |
| 1 | 手机牌子当月360搜索指数（360index爬虫获得） |
| 2 | 用户该月使用手机牌子登机 |
| 3 | 用户使用手机价格（中关村上爬虫获得） |
| 4 | 用户前N个月更换电话平均数 |
| 5 | 用户当月APRU |
| 6 | 用户当月性别 |
| 7 | 用户当月网段 |
| 8 | 用户当月年龄 |
| 9 | 用户当月流量 |
| 10 | 用户当月短信 |
| 11 | 前两月是否换手机 |
| 12 | 前1月是否换手机 |
| 13 | 当月手机百度搜索结果数 |
| 14 | 该牌子用户中使用总数 |