



**刘家明** 具有数据挖掘经验，熟悉基本机器学习算法，熟悉hadoop生态圈  
男 | 23岁 | 汉族 | 广州市 | 1年工作经验 | 15625122990 | 290671981@qq.com | 暨南大学 |  
中国电信综合平台中心

## 自我评价

- 编程基础: 熟练使用python，熟悉java和scala。
- 熟悉常用机器学习算法: 有随机森林、逻辑回归、决策树的开发经验，同时熟悉协同过滤、基于模型等基本推荐算法。
- 机器学习开发包熟悉程度: 使用scikit-learn和spark mllib进行开发。
- 熟悉hadoop生态圈: 曾搭建过公司大数据cdh测试集群，了解hbase、hdfs和hadoop, 熟练使用hive和spark。
- 学习能力: 本科学历，在校成绩良好，获得了国家奖学金和两次获得学校二等奖学金，有数学建模大赛获奖经历。
- 其他熟悉工具: 熟悉linux、shell、sql等常用辅助工具,以及python机器学习辅助包numpy和pandas
- 个人地址: <https://github.com/liujm7/>

## 教育背景

- 2012.9-2016.7 暨南大学（本科）数学与应用数学

## 工作经验

- 2016.7-至今 中国电信综合平台开发运营中心(21cn) 大数据分析  
主要职责:使用spark、scikit-learn,构建浅层机器学习模型,参与模型部署落地,支撑运营(189邮箱、天翼云、流量800)提数。  
参与流量需求预测项目:根据过去的用户流量使用量来预测下一个用户是否有流量需求。  
参与流量宝异常用户分析项目:根据用户过去8天的行为进行异常评分，寻找用户是否为可疑用户。  
使用cdh来进行大数据测试集群的搭建。
- 2015.12-2016.4 美商天睿有限公司广州分公司(Teradata) 数据挖掘实习生  
参与广东建设银行高校营销项目，在广州高校学生识别项目中负责抽取数据，数据处理，数据统计，以及进行数据建模。  
目标是识别广东高校附近ATM机的交易的学生客户。工具主要是SAS 和R

## 项目经验

- 2017.5-2017.6 流量宝异常用户分析 大数据处理/挖掘
  - 项目描述: 基于2016年10月-2017年3月的流量宝异常行为日志以及流量宝历史冻结用户清单，根据最近7天的行为特征，自动化搜索出异常用户
  - 项目效果: 异常用户分析的准确率超过80%,召回率超过70%。
  - 开发工具: hive、spark-core和spark-mllib、shell、python、hbase
  - 工作职责:
    - 使用hive进行特征构建，实现了46个特征的构建，使用python，scikit-learn进行离线分析，寻找重要特征。
    - 使用spark-mllib进行模型的训练、预测并进行效果评估，使用随机森林回归及随机森林分类模型进行异常评分监测。
    - 将异常评分的数据导入到hbase。
    - 使用shell启动定时任务来实现构建特征、模型评分和数据入库的自动化

2016.11-2017.7	基于大数据流量推荐	大数据处理/挖掘
<ul style="list-style-type: none"><li>项目描述:基于2016年2-6月的客户账单数据以及客户查询记录，预测用户未来是否会用超流量，辅助运营部门进行流量影响</li><li>项目效果:预测账单用户在下一个流量需求的准确率90%，召回率86%</li><li>开发工具: hive、spark、python、scikit-learn</li><li>工作职责: 使用spark进行账单解析，使用规则提取出有效的账单数据；使用hive，进行构建特征，将两份数据进行加工合并；使用python、scikit-learn进行模型的训练、预测、调优并选择随机森林作为最后预测模型；用python完成模型的流程化及自动化</li></ul>		
2015.12-2016.4	广东高校识别项目	数据挖掘实习生
<ul style="list-style-type: none"><li>项目描述:根据ATM交易行为数据以及客户信息数据，使用数据挖掘的方法识别广东高校的学生客户</li><li>项目效果:识别学生客户的准确率达到80%,召回率达到78% 开发工具：sql、sas以及R语言</li><li>工作职责:使用sas及sql进行抽取数据、数据处理、样本标签提取；数据建模, 通过逻辑回归模型识别广东高校附近ATM机交易的学生客户</li></ul>		

荣誉奖项

2015.10获得国家奖学金  
2014.10获得广东省数学建模大赛省级三等奖  
2013.11和2014.11两次获得暨南大学二等奖学金

技能特长

数据挖掘	<div><div></div></div>	熟练	python	<div><div></div></div>	熟练
hive/sql	<div><div></div></div>	精通	spark	<div><div></div></div>	良好
java/scala	<div><div></div></div>	良好			



感谢您的阅读，扫一扫查看我的手机简历