

个人简历

基本资料

姓 名： 朱军 毕业院校： 湖北工业大学
联系电话： 13662215641 学 历： 统招一本
微 信： 2973114353 工作经验： 3 年
电子邮件： haha_zhuge@163.com 求职意向： 数据分析师



职业技能

1. 熟练掌握 python 编程；
2. 熟悉 sql 语句，能够熟练操作 mysql、sql server、redis、mongodb 等常见数据库；
3. 能熟练使用 numpy、pandas、matplotlib、jieba、wordcloud 等工具进行数据处理及可视化操作；
4. 掌握线性回归、逻辑回归、朴素贝叶斯、决策树&随机森林、支持向量机、主成分分析等模型及分类算法，熟悉 scikitlearn 库；
5. 熟练使用 excel、powerpoint 办公软件；
6. 熟悉 scrapy 爬虫框架，会使用 urllib、requests、bs4、selenium 等常见的库，能够独立编写脚本爬取网页数据；
7. 熟练掌握 linux 各种常见命令；
8. 英语六级，阅读相关技术文档无压力

工作经历及项目经验

2017.03-2018.12 深圳南方华通信息技术有限公司 数据分析师

项目一：自动化风险评估模型

1. 项目概述：信贷公司在筛选目标用户时，需要综合考虑用户信用情况及还款能力做整体评估，通过构建好的最优模型来给申请人和潜在用户进行评分，判断是否符合放款条件，以及放款后用户出现违约和逾期的可能性，从而在

规避风险的同时保证公司的利润最大化。

2 项目内容

- 2.1 数据获取及预处理：数据获取的渠道主要有征信中心、银行、房管局、小额贷款平台等。从获取到的数据中提取出基本属性、偿还能力、信用往来、财产状况等初始变量，通过 describe 函数对异常和缺失情况初步了解，然后进行填充或删除；从 scikitlearn 中导入 train_test_split 对数据集进行拆分；
- 2.2 探索性分析：对已有的数据在尽量少的先验假设下进行尝试性分析，作出直方、散点图、箱形图，初步了解各变量的分布情况看是否符合统计分析假设；
- 2.3 变量选择：先进行变量分箱处理，计算各个分箱的 woe 值，剔除没有价值的标通过 seaborn 工具分析变量之间的相关性，计算各个变量的 iv，确定变量的预测能力；
- 2.4 模型评价：通过 woe 转换将 logistic 回归模型转化为标准评分卡模式，然后用开始预留的 test 数据，同时调用 sklearn.metrics 计算 roc、auc 评估模型的拟合能力；

项目二：用户价值 k-means 分析

1. 项目概述：通过客户公司提供的数据，对不同用户类别进行特征分析，比较不同用户的价值，对不同价值的用户提供个性化服务，制定相应的营销策略
2. 项目内容：
 - 2.1 RFM 模型指标的确定：初步确定了消费时间间隔(R)、入会时间长度(L)、消费频率(F)、观测窗口结束时累计的飞行里程(M)、折扣系数的均值(C)作为影响用户价值的指标；
 - 2.2 数据预处理：通过客户提供的近七万条用户的详细记录，抽取出会员 id、年龄、级别、观测窗口结束时的积分、消费次数、平均折扣率等 44 个属性作为初始特征，使用 numpy、pandas 进行数据清洗，包括缺失值检查、空值剔除、数据筛选等步骤；
 - 2.3 数据变换：将 L、R、F、M、C 作为新的特征并进行标准化处理然后构建特征矩阵，这些特征的值是从上面处理后的数据中提取到的；

2.4 模型构建：调用 scikitlearn 库中的 k-means 算法训练模型，进行值计算、数组化、数聚合并、导出数据表等步骤构建模型；

2.5 聚类分析：分析各个客户群的数量和分布情况，定义用户级别对重要用户应优先投放资源，重要挽留用户应通过营销手段延长用户生命周期；

2015.10-2017.01 武汉众信科技有限公司 数据分析师

项目：商品销量预测模型

1. 项目概述：对京东、淘宝上销量超过 5 万件的爆款女士内衣的销售情况进行统计分析、找出影响内衣销量的关键变量、构建 logistic 回归模型，分析用户的行为偏好，对现有商品的销量做预测，从而为后续经营提供指导。

2. 项目内容：

2.1 数据获取：搭建 scrapy 框架，构建 user-agent 应对反爬，导入 requests、urllib2、bs4 等库，通过 css、xpath 定位到目标节点，将爬取到的数据保存到 mongodb 中；

2.2 数据预处理：使用数据分析工具对缺失值进行填充或删除，构建特征矩阵，选取的特征包括尺码、颜色、款式、词条比重等；

2.3 探索性分析：使用 matplotlib 工具作出直方图、散点图、饼状图看各个变量对销量的影响比重，同时看各个变量的分布情况；

2.4 特征提取：通过去平均值、计算协方差矩阵、求特征向量等步骤提取特征

2.5 模型构建：选 logistic 回归模型，使用 gridsearch 和 cross-validation 对参数调优；

自我评价

喜欢交友、喜欢户外活动；

对互联网行业有着浓厚的兴趣、希望以后可以成为技术大牛；

做事踏实肯钻研，学习能力强