

黄 轩

年 龄：27 岁

性 别：男

居住地：广东深圳南山区

联系电话：(+86) 18520641475

邮 箱：huangxuan1517@163.com

求职意向：数据分析师

婚姻状况：未婚

籍 贯：广东韶关

微信：18520641475

QQ：526737867



教育背景

2010.9-2014.6 广东外语外贸大学 南国商学院 金融学 本科（统招）学士学位

主修课程：微积分、线性代数、概率论、西方经济学、货币银行学、公司财务分析、证券投资分析等

专业技能

- 1.熟练掌握 Python 语言，有良好的代码阅读与代码编写能力。
- 2.熟练使用 Sql、Python、SPSS 等常用数据统计分析工具，熟练掌握 Mysql、Mongodb、Redis 等关系型、非关系型、内存型数据库的使用。
- 3.熟练掌握 Numpy、Pandas、Matplotlib、Seaborn 和 Jupyter 等数据科学工具库和数据分析工具的使用，能够根据具体业务场景，对数据进行清洗、整理、分析和可视化呈现。
- 4.熟练使用 excel，熟练运用 vlookup、lookup 等常用数学统计函数以及数据透视表工具。
- 5.熟悉机器学习常用算法，如线性回归(LR、SGD)、逻辑回归(Logistic)、朴素贝叶斯(Bayes)、支持向量机(SVM)、决策树(DecisionTree)、随机森林(RandomForest)、k 近邻(Knn)和 k-means 等，能够独立运用 scikit-learn 对数据模型进行参数调优和分析预测。
- 6.熟悉 jieba 和 wordcloud 的使用，能够根据需求进行文本挖掘。
- 7.熟悉 scrapy 爬虫框架和基本库的使用，熟悉正则表达式 Re、Xpath、CSS、BS4 等网页信息抽取技术。
- 8.熟悉简单的前端开发技术 html、css 和 js，以及 Django、Bootstrap，JQuery Mobile 等前后端 web 框架。
- 9.熟练运用 linux 下的 ubuntu、redhat 操作系统，掌握常用的命令操作，能编写简单的 shell 脚本语言，具备良好的编码习惯。
- 10.具备良好的英文文档阅读能力，良好的团队沟通和协作能力。

工作经历

2016.5-至今 深圳市中合电子商务有限公司 数据分析师

工作描述

- 1.利用自有数据和可利用的外部数据，挖掘用户行为特征，为产品、运营提供参考依据。

- 2.关注数据变化，及时发现其背后的原因，整理成相关报告，提出业务发展及改进的建议。
- 3.负责数据库数据的收集和整理，根据需求梳理数据，并用合理的方式分析和展示数据。
- 4.熟悉各业务规则，与各个部门进行有效的沟通和协作，形成各类业务分析报表，配合其他部门完成数据统计和分析的工作。
- 5.结合公司现状，不定期进行数据分析报告的撰写和汇报，推动产品和业务的改善。

2014.6-2016.4 深圳金融电子结算中心——深圳黄金投资有限公司 数据分析专员

工作描述

- 1.根据公司整体经营发展需要和风控管理目标，开展分析及预测工作。
- 2.建设并完善风险预警监控体系，管理日常报表体系，撰写专题分析报告。
- 3.参与营销数据分析项目，协助分析主管进行数据提取、清洗、整合及后期的分析工作。
- 4.每日监控运营业务数据,通过数据发现异常波动情况进行问题定位和深入分析。
- 5.定期撰写业务数据分析报告,定期向公司领导、业务相关主管汇报。

项目经验

项目一【网贷平台违约率预测】

项目描述

预测某 p2p 公司的客户申请借款后的违约率，细分客户并采取有针对性的预防策略，对数据进行分析、筛选、判断，提供策略进行风险预警，更好地识别和评估风险，本人将贷款状态分成完成和违约两种，把该项目定义为一个二元分类问题。

项目流程

- 1.按照项目需求导入数据样本，简单查看、理解数据后对数据做描述性统计，利用 Numpy 和 Pandas 对数据进行清洗，并对样本的缺失值、异常值进行处理。
- 2.对清洗后的数据做探索性分析，通过数据可视化来加强对数据的理解，使用 Matplotlib 对数据的变量特征绘制直方图，观察数据的分布情况。
- 3.为保证算法模型有更好的训练效果，对数据特征进行选定，采用过滤法（方差选择法）把一些低方差特征剔除，采集常用变量特征如还款期限、每月还款额、月薪、年收入、工作年限、历史还款情况、借款与收入比例、信用等级等，同时将获取的原始数据转化为可用作模型开发的格式化数据。
- 4.将获取的数据拆分为训练集和测试集，分别使用决策树（DecisionTree）、随机森林（RandomForest）对训练集的数据进行训练，然后对测试集数据进行预测，运用网格搜索、交叉验证搜索其他最佳参数。
- 5.根据之前训练得到的最佳参数拟合模型，预测结果输出，并比较预测结果与真实还款情况。对算法模型进行评估，利用 scikit-learn 中 metrics 提供的 ROC 曲线和 AUC 值来评估模型的拟合能力，并形成模型评估报告。

项目二【线上贵金属饰品营销】

项目描述

深度挖掘潜在客户，对存量客户分层，通过算法构建客户购买产品的预测模型，输出高潜客户和目标产品的匹配结果，为精准营销提供高质量的目标群体，同时为运营人员的活动策划提供决策依据。

项目流程

- 1.数据获取 :包括存量客户和潜在客户的个人信息数据。存量客户是指已经开展相关投资交易业务的客户,潜在客户是指已经在公司完成开户的客户。数据从后台系统直接调取,预先做好简单的分类。
- 2.数据预处理:主要工作包括数据清洗、缺失值处理、异常值处理,将获取的原始数据转化为可用作模型开发的格式化数据。
- 3.数据分析 :获取样本总体的大概情况,查看数据各特征的分布,以及特征之间是否存在相关关系等冗余。使用到的探索性数据分析方法有:散点图、直方图、箱线图,使用 matplotlib、pandas、seaborn 等工具可视化呈现。
- 4.数据筛选:基于 scikit-learn 提供的方法,对数据进行特征选择和主成分分析,尽可能降低原数据的维数(复杂度),损失少量信息,删除所有低方差特征,从而筛选出最有效的特征。
- 5.模型开发和评估:将数据拆分为测试集和训练集,调用 Logistic(逻辑回归)模型训练数据,在测试集对训练出来的算法模型进行评估,利用 scikit-learn 中 metrics 提供的 ROC 曲线和 AUC 值来评估模型的拟合能力,AUC 值为 0.87,模型预测效果可以达到预期。

项目三【客户反馈咨询分类】

项目描述

收集用户对客服人员反馈咨询的信息,分析客户对产品的需求和关注点,为内容运营人员运营微信公众号提供内容采集来源,把更好的内容呈现给用户。

项目流程

- 1.数据采集:从后台系统导出客户反馈咨询的相关文本信息,利用 numpy,pandas 等数据科学库进行数据清洗。
- 2.数据预处理:调用 jieba 库,使用精确模式对投诉咨询信息进行分词处理,接着导入中文停用词库,过滤掉一些低频词、停用词以及空格符号,然后将数据集进行 bunch 化。
- 3.特征工程 使用 TF-IDF 算法和 n-gram 对过滤后的反馈咨询词进行词频统计 将数据集矩阵转换成 tfidf 权重矩阵。调用 train_test_split 方法对数据集进行拆分(训练集取 75%)
- 4.训练模型:使用朴素贝叶斯算法对训练集进行训练,在测试集进行数据预测,对模型的拉普拉斯平滑参数进行微调。
- 5.评估:调用 classification_report 评分函数和交叉验证法对模型分类的准确率进行评估,精度达到 97% 以上。

项目四【电商网站女性口红市场数据分析】

项目描述

根据项目需求爬取天猫、京东、当当等电商网站的销售数据,如品牌、售价、成交量、评论、客户端等等,分析中国女性对口红选择偏好和分析品牌销售量的趋势。

项目流程

- 1.根据项目需求文档,从天猫、京东、当当等电商网站爬取相关信息。
- 2.分析目标网站相关网页 URL 结构和 Response 内容,针对有反爬虫机制的网站,使用伪造请求头、设置代理 IP 等手段进行 request 请求,获取完整有效的相应信息。
- 3.构建 Scrapy 爬虫框架,利用正则表达式 Re、Xpath、CSS、BS4 等网页信息抽取技术提取有效信息,转化为 json 数据。
- 4.将提取出来的数据保存为 CSV 文件存入 MongoDB 数据库。

5.导入 Matplotlib 工具库进行数据可视化处理，自定义函数绘制各类数据分布图，如：品牌销售量分布图、色号喜好分布图、售价分布图等，分析不同品牌的销售变化趋势。

6.调用 jieba 库分词以及中文停用词库过滤评论数据中的停用词，然后自定义函数对过滤后的评论进行词频统计，绘制 wordcloud 词云图，反映女性对口红功效的偏好。

自我评价

- 1.对数字敏感，有较强的数据洞察力，学习能力强；
- 2.具备数据统计分析能力和良好的文字功底；
- 3.为人沉稳谨慎,学习能力强,善于自我总结；
- 4.工作态度认真负责、严谨，具有团队协作精神。

附加信息

已获得基金从业资格、证券从业资格、大学英语六级。

本人平时热爱运动，闲暇时间喜欢泡健身房、打篮球和射箭。