**个人简历**

姓 名：陈晓标 学 历：本科（统招）

性 别：男 电 话：18070157181

年 龄：26岁 邮 箱：18070157181 @163.com

微 信：qq/手机号 Q Q ：940890095

数据分析师

求职岗位：

教育背景

2012-09——2016-07 江西农业大学 信息与计算科学

主 修 ： 计算机科学基础、计算机图形学、算法与数据结构、数据库应用技术

技能描述

1.熟练掌握 SQL语言和数据库表体系，掌握关系型数据库 mysql、非关系型数据库 mongodb和内存

2.熟练掌握 Pandas、Numpy、Matplotlib、Tushare等数据分析库的使用，能独

立进行数据清洗、处理、分析与可视化操作

3.掌握 jieba使用，可以根据需求进行文本信息处理分析

4.熟悉爬虫基本库的使用，如 requests,urllib,bs4,xpath,selenium,正则表达

式等，能够独立进行爬虫项目的开发

5. 熟悉 Scrapy 爬虫框架及网页抓取原理，可以高效地进行抓取海量数据、解

析处理和入库等日常工作

6.熟悉朴素贝叶斯、决策树、线性回归、逻辑回归、聚类算法、PCA

7.熟悉Linux操作系统，掌握常用的命令操作

8.熟悉 excel可以对各项数据做各个维度的对比分析

9.了解Hadoop的基础架构和原理

工作经历

深圳中创易优网络科技有限公司

工作时间：2016.08-2019.01

在职岗位：数据分析师

工作描述：1.根据项目要求，清洗并整理相关数据，进行可视化的分析处理，制

定项目计划

2.了解项目相关知识，收集相关的行业数据

3.深度挖掘数据的特征关系，通过可视化操作将项目分析结果以文字

和图表的形式制作分析报告，提交给主管

4.跟踪项目的运营情况，根据反馈数据进行策略优化

5.总结项目经验，制作月、周工作报告

6.公司指派的其他任务

深圳蓝石投资管理有限公司

工作时间:2015.10-2016.6

工作岗位:数据分析助理

工作描述：1.对财经数据进行简单的数据清洗和处理

2.爬取财经网站的相关数据并保存入库

3.实时跟踪公司的交易数据，根据公司的投资策略，对数据变化进行简单的模型分析

4.整理交易数据，提交给上级

项目描述

**项目一.** 【利用朴素贝叶斯模型预测沪深300成分股涨跌】

项目描述：通过对股票的行业、市值、动量这三个维度的数据，利用伯努利分类器来预测下一日的涨跌幅情况，获取大概率会上涨的股票

细节描述：

1. 导入Tushare数据包，利用numpy、pandas，获取沪深A股所属行业信息、近三年每日市值、成交量、收盘价，涨跌幅,等数据并保存备份。
2. 读取数据，将市值数据进行对数化并读取交易量数据，排除停牌个股，然后通过pandas 的qcut()函数将数据分成5组。收盘价数据，将参数设为5日求出个股对应动量并进行相同的分组操作。将涨跌幅数据进行转化（0表示跌，1表示涨）。
3. 确定时间段，利用成交量数据日期索引制作交易日历。
4. 进行策略回测数据处理，回测方法是先通过i日的特征（行业、市值、动量）和i+1的涨跌幅数据（收益率大于零）训练分类器，再利用i+1日的特征预测i+2日的涨跌幅数据。首先需要将行业数据、市值数据、动量数据转化为哑变量，生成哑变量矩阵并合并得到训练矩阵，用相同原理生成预测哑变量矩阵。
5. 选择BernoulliNB进行模型训练，进行涨跌幅数据预测求与实际值平均数生成预测收益率数据并保存
6. 利用Matplotlib将处理后的数据进行可视化分析、评价策略效果
7. 构造函数，获取策略累计收益率、年化收益率、波动率、每日收益率、信息比率、盈亏比、最大回撤等指标
8. 通过各指标分析处理，得到策略累计收益率，最大回撤和波动率，胜率等，再进行可视化处理，生成项目分析报告提交上级

**项目二.** 【互联网行业岗位招聘需求分析】

项目描述：通过爬取前程无忧、智联、拉钩等招聘网站全国二线以上城市互联网相关岗位的招聘信息的爬取，如岗位、工作年限、学历、薪资、技能描述等，分析互联网行业岗位用人趋势

细节描述

1. 分析招聘网站的 URL结构和响应的内容， 创建 Scrapy项目和 Spider，在 Items里面

编写信息抓取的目标信息

2.在 Downloader Middlewares里面通过 selenium模拟浏览器发送网页请求，分页爬取每个

岗位对应的 目标信息

3.在spider里面解析爬取的信息，获取正则过滤后的目标信息

4. 在 pipelines里面设计需要存储的内容，将目标信息保存到 MySQL数据库中

5.从数据库中获取数据，对各网站的数据清洗并整理

6.将技能描述信息通过jieba中文分词库和TF-IDF方法统计词频，去除非技能词频的数据

7.从城市、岗位、薪资、工作年限、技能描述角度对数据进行排序，然后对各网站的数据进行均值话处理，利用Matplotlib工具对数据进行可视化处理。

1. 对数据进行分析，制作分析报告提交上级

**项目三.** 【股票客户投资行为分析】

项目描述：根据公司客户投资标的数据对客户进行投资行为分析，研究不同类型客户对各种投资品的兴趣度

1. 细节描述获取公司的客户提供的数据首先对无效数据进行清洗，对较少的缺失值进行均值填充。
2. 通过jieba库对用户特征描述进行分词处理，如风险偏好、专业性等，对特征值进行归一化
3. 将客户购买的投资品种进行整型的分类转化，并合并数据表
4. 进行模型评估，使用LR、RF、XGBOOST等进行训练及预测，对比结果，发现LR效果最好
5. 运用交叉验证进行参数调优
6. 对潜在客户数据进行相同处理，运用训练好的模型进行预测，通过Matplotlib进行可视化处理，并制作数据分析报告

自我评价

⯎ 对数字敏感，有较强的数据洞察力，学习能力强

⯎ 工作态度认真负责、严谨，具有团队协作精神

⯎ 具有良好的逻辑分析能力、沟通能力