

高帅

邮箱: leoshuaigao@163.com | 手机: 18811580546 | 微信号: Demon_First

博客网址: leoshuaigao.com

教育经历

墨尔本皇家理工大学, 数据科学, 硕士 墨尔本, 澳大利亚 | 2017 年 07 月-2019 年 07 月

河北工业大学城市学院, 测绘工程, 本科 天津, 中国 | 2012 年 09 月-2016 年 07 月

工作及实习经历

Coles, 数据分析部, 高级分析实习生 墨尔本, 澳大利亚 | 2019 年 03 月-2019 年 06 月

- Coles 是澳洲第二大零售商, 基于全澳州用户对于冷冻分类的购买行为及个人信息统计, 使用非监督机器学习算法进行用户信息群分;
- 负责数据处理, 从 SQL 提取数据, 清洗数据, 特征选择并进行统计分析, 实现数据可视化;
- 结果: 根据用户的 RFM(最近购买, 购买频率, 购买开支)对用户价值评分, 得到团队肯定, 后期将联动用户对折扣倾向、品质要求、购买数量价格等信息进行群分, 奠定广告投放依据;

IT MAN, 机器学习项目, 数据科学家实习生 墨尔本, 澳大利亚 | 2018 年 06 月-2019 年 02 月

- 独立负责北京工作市场的调研并且建立机器学习算法进行职位薪资预测;
- 负责抓取 51job 网职位数据, 进行统计分析和特征选择, 实现清洗和处理数据;
- 使用深度学习、随机森林、SVM 等机器学习模型, 预测职位薪资水平;
- 结果: 当用户给出职位信息的描述, 可以基本预测薪资水平, 预测精度达到 80%以上;

项目经验

云计算——云平台数据分析网页部署, 个人项目 墨尔本 | 2019 年 01 月-2019 年 02 月

- 描述: 使用谷歌云计算平台, 进行数据提取、清洗、可视化等处理并移入云平台, 使用多个第三方 API 达到用户可快速便捷得到更多感兴趣的统计信息;
- 独立完成项目中云平台部署, 网页设计, 解决在部署过程中各种平台不兼容问题;
- 利用据分布式 BigQuery 提取数据、清洗、储存、可视化等处理步骤移入云平台并使用 Python 编写代码;
- 亮点: 使用分布计算技术 (BigQuery) 缩减数据提取处理时间; 得到项目审查老师的表扬;

澳洲劳动力 (全职/兼职) 男女比例 shiny app 数据可视化 墨尔本 | 2018 年 10 月-2018 年 11 月

- 项目描述: 对澳洲各个行业/职位的从业人数历史趋势进行可视化, 分析男女从业人员的趋势以及比例;
- 独自完成数据的选择, 清洗, 分组, 建立交互式图表;
- 通过交互式图表发现, 女性在大多数行业中仍占据少部分比例, 尽管一些行业男女员工总数量相近, 但男性更可能持有全职工作; 通过 Shiny app 建立可交互式轻应用, 使可视化可以通过 url 进行分享;
- 项目网址: <https://leogao.shinyapps.io/assignment3/>; 成绩在全班前 5%;

MapReduce 算法及 cluster 数量对表现影响的比较, 2 人 墨尔本 | 2018 年 09 月-2018 年 10 月

- 分析 Pair 和 Stripe 的文字统计算法在是否利用 combiner 及 partitioner 的情况下处理结果的速度差异; 同时分析两种算法 (pair 和 Stripe) 在不同数据大小以及不同 Cluster 数量不同的影响;
- 使用 Java 编写 MapReduce, 比较 combiner、partitioner 数据大小及 Cluster 数量对两种不同算法的影响;
- 选择使用数据, 清洗非英语字符, 实现两种不同算法, 并制作可视化图表;
- 项目结果: ①Pair 算法在小型数据与 Stripe 算法表现类似, 随着数据量增大, Stripe 算法的优势越明显; ②Combiner 对 pair 的处理速度优化明显; ③Cluster 数量升高, 对 Stripe 算法的速度优化强于 Pair 算法;
- 分析实验结果, 撰写汇总报告并答辩, 最终获得满分;

时间序列分析——比特币价格预测, 4 人 墨尔本 | 2018 年 03 月-2018 年 05 月

- 负责提取比特币历史价格, 用 R 语言进行时间序列分析, 建立 ARIMA+Garch 模型, 预测未来 10 日价格;
- 完成数据的协相关分析, 提取 arima 以及 garch 模型所需参数, 进行模型优选, 预测;
- 难点: 比特币价格变化幅度大, 包含有趋势以及方差变化大等;
- 项目结果: 在作业提交后 10 天, 价格预测的精度很高 (± 30), 得到满分;

竞赛

2019 年安永数据科学挑战赛

- 来自 15 个国家, 470 所大学的 4500+ 参赛者, 最终获得中国区第 5 名, 世界区第 90;

技能

语言: 中文 (母语), 英文 (熟练);

计算机语言: Python (2 yrs)、R (2 yrs)、Java(2 yrs)、SQL(2 yrs); 统计软件: SaS;