

### 95d631a337ad21861Xd43dm-GVFZwYS-U\_KaWOGrlvbTNhFm2A---PERSONAL RESUME 相信自己,实现自我价值

## 🎝 基本信息

❖ 求职意向: NLP 算法工程师、算法工程师、数据挖掘工程师、数据分析师

◆ 工作经验: 4年

◆ 年 龄: 28岁

◆ 籍 贯: 山西运城

◆ 电 话: 18515880196

◆ 薪 资: 面议◆ 民 族: 汉族

◆ 现 居: 杭州

◆ 邮箱: zengshuaiguo@163.com



### 🎓 教育背景

♦ 2016.09-2020.07 北京科技大学天津学院 计算机科学与技术(本科)

# 专业技能

- > 熟练使用 Python 语言, 熟悉 Java 语言, 具有良好的编程基础;
- > 熟练使用深度学习框架 Tensorflow、Pytorch、Keras, 熟悉 Caffe、Scala;
- ▶ 熟练使用机器学习、集成学习、深度学习、NLP等算法;
- > 熟练使用 sklearn、jieba、gensim、nltk、pyhanlp 等算法库;
- 熟练使用关系型数据库,熟悉 Mysql、Oracle、Neo4j等数据库原理;
- 熟练使用虚拟机操作系统,熟悉 Linux、Ubuntu、CentOS 等系统原理;
- > 熟悉大数据架构,如 Hadoop、Spark、Hive、Zookeeper、Mapreduce等;
- > 熟练使用 Pycharm、Jupyter Notebook、Anaconda、Eclipse 编程工具。
- 熟悉 Django Web 框架、Scrapy 爬虫框架等。

## 🚊 工作经历

2020.07-2022.07

彩虹无线 (北京) 新技术有限公司

数据挖掘工程师

## 主要职责:

参与执行公司多项数据挖掘等项目,涉及电商、金融、通信等领域;根据相关业务需求和场景,熟悉产品开发流程,包括产品规划、需求分析及报告;数据清洗,衍生特征构建,数据规范化,数值编码。使用统计分析机器学习方法进行特征选取,对处理好的数据进行建模,参数优化和模型评估。

2022.07-2024.11

杭州乐科锐网络科技有限公司

NLP 算法工程师

## 主要职责:

参与过多项自然语言处理算法等项目,涉及金融、通信、媒体等领域; 参与过项目周期规划、需求分析,包括算法调研、项目流程、项目汇总; 负责过项目方案实施,项目开发,模型部署,推动完善项目投入应用; 负责过数据处理、特征工程、构建模型、参数优化、模型评估; 负责过项目上的部门之间的工作同步执行。

# 🚊 项目经验

**—**、2024.06-2024.11

BERT 新浪网新闻分类

#### 项目描述:

基于每天产生的新闻,包括财经、房产、股票、教育、科技、社会、时政、体育、游戏、娱乐等栏目,为了让新闻爱好者获取和分享今日头条实事,新浪网作为重要媒体交互平台,把每一篇新闻划分到对应栏目下,用户能准确地搜索到想要关注的最新消息,拓宽了信息交流渠道,让天下事尽在新浪网。

#### 项目流程:

- 1、前提调研,商业论证,成立项目组
- 2、项目规划,确定项目成员和职责,并召开开工大会
- 3、项目开发,设计,加工

### 主要职责:

- 1、文本预处理
- 2、构建 BERT、ERNIR 模型
- 3、定义损失函数、梯度优化器
- 4、训练模型,评估模型
- 5、选择评估指标高的模型
- 6、模型部署应用

二、2024.01-2024.06

网易有道词典中英文翻译

### 项目描述:

随着网络信息的不断庞大,网易公司继主流产品 163 网易邮箱、微博后,逐一研发了网易音乐、有道词典、理财、直播、搜索引擎等系列产品,有道词典作为现在网络广为流传的翻译软件,深受各界人士的追捧和喜爱,不仅开发了一种语言的翻译,而且还有许多国家的语言翻译,可供使用。

## 项目流程:

- 1、项目启动,项目计划
- 2、项目执行,项目监控
- 3、项目收尾,项目维护

## 主要职责:

- 1、文本向量化
- 2、构建 Transformer 模型
- 3、定义损失函数、梯度优化器
- 4、训练模型,评估模型
- 5、选择评估指标高的模型
- 6、中英文翻译

三、2023.07-2024.01

招商银行 rasa 知识图谱问答系统

#### 项目描述:

招商银行业务种类增长迅速,市场竞争形势严峻,客户需求量日益增长,特 此推出一款问答机器人,供广大人民群众前来享用,对于每个使用者提出的 问题,回答或者保留宝贵的建议,不定时更新机器人系统对每个人的专属特 定服务, 保证客户的高级体验感。

#### 项目流程:

- 1、明确项目目标,项目需求分析
- 2、制定初步流程图,团队讨论与反馈
- 3、实施与执行,监控与调整

#### 主要职责:

- 1、用户界面语音获取
- 2、自动语音识别
- 3、语音牛成文本
- 4、自然语言理解,分层意图识别,意图的实体获取
- 5、对话管理
- 6、文本生成
- 7、文本转语音
- 8、用户界面语音对话

四、2023.01-2023.07

博乐达产品推荐系统

## 项目描述:

通过用户已经浏览、收藏、购买的记录, 更精准的理解用户需求, 对用户进行 聚类、打标签,推荐用户感兴趣的商品,帮助用户快速找到需要的商品,适时 放大需求,售卖更加多样化的商品。甚至在站外推广时,能够做个性化营销。

### 项目流程:

- 1、根据客户提出的要求,并对指定商品做出高质量的推荐,从而提高营业额
- 2、根据客户购物车中的商品和客户可能感兴趣的商品推荐类似的商品
- 3、根据客户的喜好和风格向客户推荐最有可能吸引客户的商品
- 4、根据客户的相似度推荐其他客户购买的商品

## 职责描述:

- 1、对已有数据进行分析,然后根据项目进行算法调研及周期制定;
- 2、提取数据,特征主要来源:用户注册信息、商品评论、用户评分、用 户浏览记录、用户购物车等;
- 3、对数据进行清理,过滤,转换,提取等操作,将非结构化的数据处理 成半结构化或者结构化的数据;
- 4、加入嵌入层将用户特征及商品信息向量化,得到其用户和商品的特征 矩阵;



- 5、使用协同过滤算法将用户特征与商品特征相关联,预测用户对商品的 偏爱程度;
- 6、使用余弦相似度、皮尔逊相关系数等方法计算物品相似度;
- 7、根据用户需求及预测的用户对商品的偏爱程度完成推荐。

**万**、2022.07-2023.01

博乐达数据决策系统

#### 项目描述:

根据各部门征求的博乐达产品数据需求,建立博乐达数据决策系统平台,首先 建立博乐达产品销售数据的数据库,接着将历史记录的所有博乐达产品销售数 据分别转储到各类 sql 表中,然后根据业务需求汇总数据制作报表。

#### 项目流程:

- 1、需求受理,制定解决方案
- 2、数据处理,数据入库
- 3、数据分析及预处理
- 4、汇总数据,制作报表

#### 职责描述:

- 1、根据制定的决策方案和各部门收集到的数据,解析数据结构;
- 2、按照部门主管要求新建不可缺少的多个 sql 表;
- 3、将所有数据分别存储到特定的 sql 表中;
- 4、确保建立的数据库能够长期使用;
- 5、成功建立完数据库,再根据其它部门需求可再扩展 sql 表和本部门要求制 作数据决策系统报表;
- 6、将各部门积累的数据,根据 sql 表处理完,实时导入数据库。

六、2022.01-2022.07 吉利商用车蓄电池分析平台

## 项目描述:

根据车主驾驶蓄电池商用车产生的数据,探索应用模块:电池寿命分析、

电池故障预警、画像、续航里程评估、电池更换提醒。

电池寿命分析: 电池实际电量评估、电池容量衰减趋势预测。

画像:用户画像、电池画像。

续航里程评估:车辆实际续航里程、车辆续航里程预测。

电池更换提醒: 电池更换提示、电池更换统计。

分析报告模块:报告生成、报告导出。

## 项目流程:

- 1、需求受理,规划解决方案
- 2、数据分析及预处理
- 3、数据特征选取及建立模型
- 4、模型优化及项目实际效果

## 职责描述:

- 1、预测每辆车蓄电池未来两个月的容量衰减百分比;
- 2、预测每辆车蓄电池未来两个月的实际耗能;
- 3、描述蓄电池、用户画像,对蓄电池、用户进行评价;

- → 项目经验
  - 4、预测每辆车蓄电池未来两个月的实际续航里程;
  - 5、判断蓄电池是否需要更换并提供原因;
  - 6、统计需要更换蓄电池车辆的占比;
  - 7、预估未来1个月需要更换蓄电池的占比情况;
  - 8、模型汇总报告。

七、2021.07-2022.01

亿联银行贷前行为评分

#### 项目描述:

基于客户的基本信息,收入信息以及各类交易行为特征进行建模,通过对这些 特征的数值计算进行预测和评分,从而识别出信用卡逾期客户和非逾期客户, 并且得出客户的逾期概率, 使银行能比较准确地衡量风险、收益等各方面的交 换关系,找出适合自己的风险和收益的最佳平衡点。授信者可以根据客户的逾 期概率以及信用分数分析客户的还款能力,据此决定给予授信额度以及利率。虽 然授信者通过人工分析客户的历史信用资料,同样可以得到这样的分析结果, 但利用信用评分更加快速、更加客观、更具有一致性。既节约时间,减少投入 成本又提高了审批效率。

#### 项目流程:

- 1、需求受理,规划模型白皮书
- 2、数据获取及数据加工
- 3、数据特征选取和数据建模

#### 职责描述:

- 1、读取数据集,给数据加上标签列,以分辨违约用户和未违约用户;
- 2、找出有缺失值的变量给这些变量的缺失值进行填补;
- 3、将离散型特征独热编码,对数值型变量进行卡方分箱;
- 4、根据卡方分箱得出 woe 结果表,对 woe 进行分析,并进行 woe 转换;
- 5、分割数据集,建立逻辑回归,xgboost模型,建模过程中采用交叉验证, 选择最好的一个模型;
- 6、使用 GridSearchCV 算法进行参数搜索及调优;
- 7、使用 AUC, precision, recall, KS, ROC 曲线对模型进行评估;
- 8、按照评分卡刻度计算评分,根据计算出的评分计算 PSI;
- 9、计算变量的分数,根据变量的分数得到得分表,计算变量得分的 PSI;

八、2021.01-2021.07

光大银行理财用户画像

## 项目描述:

本项目通过理财公司和第三方提供理财数据,同时对客户进行全方面的分析和 考察,为用户设计产品或开展营销活动,提高用户体验,进而提高产品的流量, 提升产品的盈利额。

## 项目流程:

- 1、需求分析
- 2、数据获取和预处理
- 3、特征工程和建模
- 4、建模结果及项目实际效果

# 🔐 项目经验

#### 职责描述:

- 1、前期对数据的简单梳理和分析,然后找到相关资料;
- 2、使用 jieba 分词将得到的文本数据进行分词处理;
- 3、使用 CountVectorizer 计算词频,计算 TF-IDF 作为权重值,使用 word2vec 将文本转化为词向量格式;
- 4、分别使用 TF-IDF+LDA 主题模型和 TF-IDF+卡方检验进行特征选择;
- 5、使用决策树、xgboost 建模进行预测;
- 6、打印模型评估指标使用网格搜索进行参数调优;
- 7、使用调优后的模型完善用户的属性标签;
- 8、使用 K-means 聚类将用户划分为不同级别。

九、2020.07-2021.01

山西电信客户流失预警

#### 项目描述:

某电信通讯运营商近期用户流失严重,市场部希望能够通过数据挖掘建立用户流失预警名单,通过促销、用户关怀、服务质量提升等进行维系和挽留,减少由于客户离网带来的经济损失。

#### 项目流程:

- 1、数据准备、探索、清洗
- 2、建立模型、模型选择
- 3、模型应用

### 职责描述:

- 1、读取账单和通话详单数据,合并数据集;
- 2、进行数据选取、数据清洗和数据转换、生成新字段、去除重复数据、填补 缺失值等;
- 3、进行特征选取、结合图形数据探索,利用皮尔逊相关系数,卡方检验筛选变量;
- 4、按照 6:4 划分数据集,建立 GBDT 模型;
- 5、使用测试集对模型进行最后的评估(混淆矩阵,PRECISION,RECALL,F1-SCORE,ROC\_AUC);
- 6、模型上线,预测并找到有流失倾向的用户,由营销部门进行用户挽留。

## 国 自我评价

- ✓ 有良好的职业道德和素养,能承受较强的工作压力;
- ✓ 有很强的解决和分析问题的能力,乐于承担富有挑战力的工作;
- ✓ 有较强的适应和自学能力,对待工作谨慎认真富有责任感;
- ✓ 有较强的团队管理能力和合作精神,能够积极互动努力达成团队的目标;
- ✓ 对待工作,充满热忱,积极主动,义无反顾,恪尽职守。