

# 简历

## 联系方式

- 姓名：陈晨
- 电话：15921596064
- 邮件：me@doudou0o.com
- 博客：www.doudou0o.com

## 个人信息

- 陈晨/男/1991
- 研究生/华东师范大学计算机系
- 所在地：上海 浦东新区
- 期望职位：后端研发 (Python & Java)

## 自我评价

具备一定机器学习和NLP能力，兼具数据分析能力和ETL能力的，一名后端研发。

对于后端系统研发(Python & Java)有丰富的经历和扎实的基础。有清晰的系统分析和研发逻辑，具备独立承担应用研发和搭建的能力。对机器学习有着浓厚的兴趣，能够使用一些简单的算法来解决具体问题。具有较强的学习能力和执行能力，主动发现解决问题，提出改进优化以及可行方案。在招聘领域，数据标注领域和互联网餐饮领域有过一定经验积累。

## 教育背景

- 2013.09 -- 2016.01 华东师范大学 计算机科学技术 硕士
- 2009.09 -- 2013.07 华东理工大学 计算机科学技术 本科

## 工作经历

### Appen 澳鹏数据（上海）有限公司

Senior Manager Senior Software Engineer

2020年06月 -- 至今

1. 负责3D点云车道线辅助标注，提升了50%车道线标注效率
2. 负责新数据收集平台的整体设计和搭建，日均收集20W+数据，提升50%项目管理效率，提升20%数据质量
3. 负责语音标注平台Ampersand在AWS ECS的日常研发维护，包括定时交付，批量归档等功能
4. 负责PE事务以及搭建PE服务平台，减少了30%日常工单量

### 支付宝口碑（上海）信息技术有限公司

高级开发工程师

2017年08月 -- 2020年06月

1. 负责口碑商品库存的日常研发，包括商品去重，次卡研发，商品延时发布等功能
2. 负责智能菜品录入整体工程链路设计研发与菜品实体提取，包括语音录菜/拍菜谱录菜/极速录菜
3. 业务监控效果提升的规划和分析，利用3 $\sigma$ 原则和自适应方式设计监控准确率提升方案

- 负责服务口碑VOC(客服声音等)分析平台整体建设, 对反馈数据进行文本分析和分类, 映射产品问题促进业务优化

## 上海逸橙信息科技有限公司

### Java算法工程师

2015年12月 -- 2017年08月

- 负责公司简历解析系统的研发以及维护, 提升性能和优化解析算法
- 负责简历解析应用Python版转移至Java版本整体框架设计
- 负责简历解析中重要字段的识别提取模型, 使用频繁项集和实体分类器辅助识别
- 负责设计固定模板简历解析模块, 能够自动提取简历内容并自适应模板的改动

## 项目经历

### 新收集平台设计和搭建

2022.07 -- 至今 | 角色: 研发/架构

- 带领一个团队从0开始设计及搭建整个数据收集平台
- 优化数据分发逻辑, 提升数据分发性能和分发灵活性
- 实现自动化实时数据校验, 以及易扩展的算法接入方式
- 实现数据质量流, 包括人工质检以及打回重做逻辑
- 设计并实现实时数据报表, 提供分钟级实时数据信息及统计指标

业绩: 提升80%项目管理效率, 提升20%数据质量, 缩减30%以上的项目成本。日均能接收20W+数据。

### 3D点云车道线辅助标注

2021.05 -- 2021.07 | 角色: 算法/研发

- 设计并实现了一套实时后端算法辅助车道线标注功能, 并搭建一套Flask来提供服务
- 利用点云反射率进行数据二值化来提取车道点云候选集
- 利用DB-SCAN聚类以及异常值检测进行车道线拟合
- 实现了实线, 曲线, 双实线等辅助标注, 并作用于多个不同项目中

业绩: 大幅提升车道线标注效率50%以上, 缩减30%以上的标注成本。

### 口碑VOC系统搭建

2019.05 -- 2019.08 | 角色: 研发/架构

- 设计整体口碑VOC平台(客户之声/商户反馈)的系统框架和业务应用, 以及产品落地方案
- 将咨询反馈记录数据进行预处理并逻辑回归分类, 映射至产品问题, 促进产品优化
- 提炼二清交易数据, 精准匹配口碑门店, 下放BD任务, 进而减少二清交易投诉
- 提出线下模型训练, 线上管理模式, 对VOC分类模型进行统一收口管理
- 提供实时分析能力, 并对于问题爆发主动拉铃预警触达各相关方

业绩: 咨询服务记录分类准确率达91%, 减少10%的二清交易, 降低20%商品拒用投诉。

### 口碑掌柜智能录菜

2018.04 -- 2018.11 | 角色: 研发/架构

1. 第三方菜品批量导入功能，离线收集清洗商家菜品数据，并使之能够一键导入到口碑菜品库中
2. 语音录菜功能研发，包括录菜调用后台研发，从语音文字部分提取价格，识别菜品名
3. 极速录菜功能研发，包括店铺轻品牌词挖掘，菜品库清洗，菜品搜索推荐逻辑等

业绩: 通过智能录菜5个月内录入80W+菜品数据，服务超过5W个商户。从原先手动录菜时代的5s/菜的效能，提升至2s/菜。

## 简历解析

2016.09 -- 2017.08 | 角色: 研发/算法

1. 设计简历解析的系统架构和产品功能，搭建一套完整的 Flask 系统
2. 构建频繁项集优化简历字段识别规则，提升识别覆盖率
3. 调研训练LR分类器辅助识别字段，提升识别准确率
4. 使用 Gearman 进行服务RPC响应和处理
5. 实现完整简历解析全部功能，包括文本提取，词语解析，文本分类，模块聚合等

业绩: 简历解析使用率为每天超过5W个简历解析，可最高支持200tps调用量，每次解析时间为5s之内。支持90%以上的简历格式。简历解析重要字段识别成功率为95%以上，识别错误率在5%以内。整理简历解析解析率保持在90%以上。

## 模板生成学习算法

2017.01 -- 2018.08 | 角色: 算法

1. 设计模板提取算法，能够识别多种模板简历，并自动提取其中的字段
2. 能够不依赖人工标注的情况下主动训练得到可靠的模型
3. 利用算法的自学习能力，能够学习模板简历的小幅改版

业绩: 支持了两种固定模式的简历生成方式进行解析优化，在小幅改动的情况下，几乎不需要认为参与则进行模板修正。