**数据挖掘方向简历分析**

**专业方向：**软件工程、数理金融等

**主修课程：**算法分析与设计、神经网络程序设计、程序设计与计算机系统、高级软件工程、信息安全、线性代数、随机过程、概率论与数理统计、回归分析、组合数学等

**相关技能：**

* 掌握C / C++ / Linux 、Python、Java、MATLAB等开发语言
* 常用数据结构与算法
* Machine Learning / Deep Learning / CV
* Hadoop
* 自然语言处理
* 神经网络模型、熟悉GAN、LSTM等神经网络模型
* 熟悉 Tensorflow、Keras 等深度学习框架的使用
* 数据处理与建模
* 熟悉 ECharts 等数据可视化工具
* 图像处理

**项目经历：**

* **影视大数据分析：**在对影视相关数据的爬取基础上，通过自然语言处理和机器学习等技术给出评判影视作品的相关指标，最后通过数据可视化的方式将分析结果展示出来。（**承担工作：**对电影评论的分析:利用机器学习中的朴素贝叶斯，SVM 以及深度学习中的栈式 Bi-LSTM 来训练模型，使用训练的模型对每条评论只针对其内容进行情感分析并打出一个总体分数。还可以将每条评论根据其内容细化成不同维度，给每个维度都打出一个分数。综合某部电影的所有评论的分析结果得出这部电影观众的大致评价，给还没有观看的观众提供一些观影指导。）
* **信用卡评级模型：**利用银行提供的客户数据，进行数据整理，抽样及特征选取，建立信用卡申请模型评分模型及信用卡行为模型，评估客户将来的违约风险。 （**承担工作：**主要使用逻辑回归，支撑向量机，决策森林及Xgboost 等算法。）
* **基于生成式对抗网络的停车场数据修补：**以地理空间信息为依据，使用贝叶斯公式确定在停车场空间相似条件下数据相似的概率；之后以同类型现有停车数据为学习样本，使用GAN生成新数据完成同类型无数据停车场的数据修补，并设计贝叶斯过滤器实现生成结果的筛选。
* **基于ECharts和Leaflet的前端地图数据展示：**使用ECharts和Leaflet等开源库对深圳市和太原市停车场分布、分类情况做前端渲染，同时根据车辆出入停车场的规划路径做光线特效渲染和热力图展示。

**实习经历：**

* **百科实体类别标签挖掘：** 挖掘百度百科实体所属类别，丰富知识库实体概念，用于百度开放分类体系的构建。（**承担工作：**1、模板挖掘：利用基础的句式构建的模板和部分扩展模板直接进行挖掘。2、质量控制：训练 SVM模型提升(实体，标签)准确性。）
* **开放自由文本二元关系挖掘：**在开放自由文本中，比如百度百科，挖掘文本句子中存在的主语(s)、谓语(p)、宾语(o)三元组，并将部分数据用于精准问答和实体卡片属性的建设。（**承担工作：**1、特定领域关系分类:针对特定领域，比如人物关系，专门训练此领域的 svm 模型，配合实体识别的方式来挖掘 spo 三元组。这一部分数据可以用来补充实体卡片缺少的属性值。2、开放领域关系抽取:在纯开放领域，设计以 bi-lstm 为基础的深度学习模型，模型整体是 pipe-line 结构，由两个 bi-lstm 模型串行组成，挖掘出无领域限制的 spo 三元组。部分热门领域的三元组用于精准问答。）
* **策略开发：**参与到公司多因子平台的搭建。利用神经网络、深度学习、小波分析等技术。通过因子分析及筛选，利用 tensorflow 编写程序，开发 LSTM 多因子选股策略。利用 python 分析可视化客户数据、协助制作公司季报年报。开发国债期货趋势套利策略。
* 分析公司各地区、分店报告的客户数据、通过合理提取特征，利用逻辑回归等机器学习算法、预测客户受保概率及公司资金流向。 协助新型保险产品开发，分析数据、为开发决策提出意见

**加分项：**奖学金、英语证书、相关学术论文、熟练使用效率型工具、中基软件工程证书