

# 简历

## 个人资料

姓名：胡智杰

性别：男

手机：18124189052

邮箱：hhu3301@163.com

微信/QQ：247542266

求职意向：数据分析师

## 教育背景

2007/09 – 2010/01	北京航空航天大学(985)	材料科学与工程	研究生(统招)
2003/09 – 2007/07	北京航空航天大学(985)	材料科学与工程	本科(统招)

## 技能描述

1. 熟练掌握 python 语言，能用 python 进行相关项目的开发。
2. 掌握常见 SQL 语句，熟悉 mysql 数据库的语法，了解非关系型数据库 mongodb、redis 的日常操作。
3. 熟悉 numpy、pandas 等数据分析常用库，可进行数据清洗、分析等操作，能使用 matplotlib 进行数据可视化。
4. 熟悉机器学习的基本模型，如线性回归、逻辑回归、决策树、随机森林等。
5. 能用 scikit-learn 进行特征工程、调参优化、模型建立。
6. 了解 HTML，能使用 requests、Selenium 等爬虫库从网页中获取响应，并使用正则表达式、css、XPath 提取信息。
7. 熟悉 scrapy 爬虫框架，能进行爬虫工作开发。
8. 熟悉 Linux 开发环境，掌握 Linux 常用命令，了解 shell 脚本，可以进行简单的开发。
9. 熟练掌握 EXCEL、PPT 等办公软件。

## 工作经历

1. 2015/10 – 至今                      深圳市中京信息技术有限公司                      数据分析师  
收集和处理客户提供的数据并整理和清洗。  
不定期进行数据分析和相关程序优化维护。  
根据客户要求，对项目数据进行建模及预测分析（机器学习），基于数据分析的结果推动相关问题的解决。

2. 2010/04 – 2015/09

富泰华(深圳)工业有限公司

产品工程师

主要负责对产品生产过程中各种数据、参数工艺、异常问题进行分析、处理、提出优化改善，维持生产。  
根据现有数据，对目前的生产营收、支出做统计、预算等工作。  
收集分析数据，对未来的产品、前言技术进行预言。

## 项目一·

### 项目描述

项目的其中一个功能是将公众提交的法律咨询问题自动进行分类，方便将咨询题目推送给对应人员进行解答。

1. 原始问题数据主要来自爬虫爬取，目标是将数据分为婚姻家庭、消费维权、劳动纠纷、债权债务、房产纠纷、交通事故、医疗纠纷等十五个类别。
2. 用 jieba 将问题进行分词，去除数字、标点符号等无用词，过滤掉停用词。
3. 对数据进行探索性分析，对样本按比例化分成训练集、测试集，然后用 TF-IDF 统计词频，进行量化，进行特征提取。
4. 使用平滑处理和朴素贝叶斯分类器对训练数据集的特征进行训练。
5. 对测试数据集进行预测，用准确率评估模型，准确率达到 93%。

## 项目二·

### 项目描述

由客户提供部分脱敏的生产参数数据，要求通过输入不同槽位的化学药品种类、浓度、浸泡时间、提拉速度、基材材料、塑料类型等参数，对最终成型产品的结合强度（拉拔力大小）进行预测。

1. 原始数据主要来自客户生产、调试产生的数据。
2. 进行数据探索，查看数据各特征的分布，以及特征之间是否存在相关关系等冗余，用直方图、散点图等对单变量的分布进行分析，获取各属性的统计特性。
3. 对各特征数据进行标准化预处理。
4. 采用交叉验证的方法，用线性线性回归模型进行学习。
5. 对生产参数数据进行测试、用 `r2_score` 评估模型。
6. 根据分析结果，预测不同参数下的结合强度，完成客户的要求。

## 项目三·

### 项目描述

客户提供生产参数和最终产品颜色，要求输入生产参数时软件能预测其产品颜色，输入颜色数据时软件能返回生产此颜色所需要的生产参数。

1. 根据客户提供的历史数据，用 numpy 和 pandas 对产品数据进行数据清洗。
2. 对参数数据进行标准化处理。

3. 使用欧氏距离公式，通过 KNN 算法对参数数据进行分类预测。
4. 针对输入的颜色数据，输出最接近的几条生产参数数据。

## 个人评价 ·

---

1. 工作认真负责，做事严谨，有良好的学习能力和团队协作意识；
2. 具有良好的逻辑分析能力；
3. 在闲暇时间，不断学习充实自己，提升自己的能力。