



## 学历

- 2017-2020 计算机技术专业（数据分析方向），硕士，都柏林理工大学  
2013-2017 应用化学专业，学士，北京化工大学

## 项目经验

- 2019.09 毕业论文, A Comparison Study on State-of-the-art Minority Class Data Oversampling Techniques for Imbalanced Learning .  
-  
2020.01 项目内容和总结：  
- 阅读文献，研究多种应对不均衡数据分布时采用的过取样方法原理（以 SMOTE 及其改进算法为主）。  
- 在多个数据集上计算多种过取样算法的表现，比较它们表现的差别并进行统计学检验。  
- 根据算法原理，分析改进算法的表现和数据分布形式之间的关系。  
- [项目链接: Msc Dissertation Han Tang](#)
- 2019.08 实习经历, 中国邮政储蓄银行数据分析实习。  
- 通过过往客户贷款数据，分析预测客户下个季度信贷违约情况。
- 2019.09 使用 Python 分析特征相关关系，进行特征工程操作。  
- 通过 Xgboost 和 Random Forest（随机森林）等多种算法构建模型。  
- 预测模型在测试数据中召回率达到 0.82。
- 2019.03 课程作业, 多标签鸟类种类识别。  
- 数据集包含多种鸟类叫声的录音。样本量大小为 1000，包含 87 个种类。
- 2019.04 运用 Python 中的 librosa 库将音频文件转化为梅尔波频谱，并修正为统一大小。  
- 通过 fastai 库构建卷积神经网络训练模型，模型召回率为 0.87。  
- [项目链接: NIPS2013 鸟类种类识别](#)
- 2019.04 项目经历, 街景路牌数字识别。  
- 构建卷积神经网络模型识别图片中的号码牌数字，并判断数值。
- 2019.06 数据包含 10 万张街景图像，其中各包含一个或多个数字。  
- 使用 Python 中的库 TensorFlow 和 Keras 构建多个不同架构的卷积神经网络模型，其中最优模型在测试集中预测 F1 值为 0.9438。  
- [项目链接: Steet View House Number 数字图像识别](#)

## 专业能力

- 熟练 Python, R, SAS, 机器学习算法, 数据分析和可视化, 数据科学相关的常用库 (Scikit-learn, Pandas, Numpy, Scipy 等)  
熟悉 SQL, Web 开发, Java, 深度学习算法 (CNN, RNN), 深度学习常用库 (Pytorch, TensorFlow), 网页爬虫

## 语言

- 英文 熟练的听说读写能力

---

## 课外活动

足球 参加多支足球队。

- 北京化工大学理学院足球队。
- 北京化工大学足球队。
- 参加北京市足球业余联赛。
- 参加爱尔兰地区级足球业余联赛 UCD Super League。