

# 许文漪

个人网站 | [Github](#) | [领英](#) | [wenyi.xu@columbia.edu](mailto:wenyi.xu@columbia.edu) | (+1) 646-881-2612 | 微信: xukeke13270805582

## 教育经历

哥伦比亚大学 (CU) 美国, 纽约  
商业分析硕士, 3.83 / 4.00 (专业前 10%) 2019.08 – 2020.12

- 核心课程: 统计、仿真、最优化、数据分析、数据可视化、机器学习、人工智能、项目管理、商业分析

上海交通大学 (SJTU) 中国, 上海  
集成电路工程硕士, 3.50 / 4.00 (专业前 20%) 2016.09 – 2019.03

- 荣誉: 全国研究生数学建模竞赛国家级二等奖, 京贸创客杯投研分析大赛全国二等奖
- 核心课程: 数据挖掘、数据库技术、大规模数据处理、随机过程

南京大学 (NJU) 中国, 南京  
微电子工程学士, 4.40 / 5.00 (专业前 5%) 2012.09 – 2016.06

- 奖学金: 兴全基金社会责任奖学金, 人民奖学金, 南京大学优秀学生
- 核心课程: 微积分、概率论、线性代数、面向对象的程序设计

## 专业技能

- 数据处理: MySQL (Queries、Subqueries、窗口函数), Excel (函数、图表、透视表), Python (Numpy、Pandas、Scipy、scikit-learn), R (dplyr), ETL (SSIS)
- 数据可视化: Tableau、Power BI、Python (Matplotlib、Plotly、pydash)
- 统计 (概率分布、假设检验、最大似然估计), 机器学习 (回归、分类、聚类、特征选择、特征工程)

## 实习经历

松下电子北美总部 美国, 新泽西  
数据科学家实习生 (商业智能团队) 2020.08 – 2020.12

- 建立了松下电子元器件部门在未来三个月的月收入预测模型, 通过根据产品的分销商和类别对电子组件进行细分来提高收入估算的可预测性, 并在 Azure 云平台上部署了该模型以方便内部客户在 Power BI 端进行使用
- 使用 Python 中的 scikit-learn 包实现了 XGBoost 回归模型, 以分析模型的输入变量 (包括制造商所在地、运输方式) 对输出变量 (元器件产品的交付时间) 的影响, 和经验模型相比将预测误差从 30% 降低至 15%
- 根据客户的要求建立 Power BI 交互式可视化仪表盘, 用来描述预测逻辑并跟踪预测错误, 并通过在 Power BI 内部实现多个 DAX 函数实现高效的数据可视化

Balyasny 资产管理 美国, 纽约  
数据分析师实习生 (股票研究团队) 2020.01 – 2020.05

- 使用交易信号 (例如移动平均线、布林带等) 训练随机森林模型以预测罗素 3000 股票的未来三日价格走势
- 对模型性能进行压力分析, 通过对比标准普尔 500 指数和 VIX 指数的相对关系来确定历史上的重大金融危机和其后续的恢复期, 从而从这些时期的股价波动特征中了解 Covid-19 大流行的影响
- 向内部数据分析师介绍了交易策略的原理和分析的结果, 并且获得了分析大量金融数据的丰富经验

德勤 中国, 上海  
数据科学家实习生 (数据智能团队) 2019.12 – 2020.01

- 与数据智能团队内部的 4 人团队合作, 创建每日头条新闻电子专栏, 并使用 NLP 自动化其内容的生成流程
- 通过进行网页爬虫 (BeautifulSoup4) 和文本挖掘 (NLTK) 来识别每天热度最高的前十个的新闻主题和相应的新闻文章内容; 每周与内部咨询团队讨论并根据业务要求调整 Python 脚本

安信证券 中国, 上海  
研究员实习生 (股票分析团队) 2018.05 – 2018.07

- 使用从 Bloomberg 和 Wind 获得的公共数据集、产品供需数据和行业趋势指数, 对中国集成电路行业进行了详细的基本面分析, 并撰写了一篇关于 IC 载板行业的深度研究报告, 在公司的微信公众号上阅读量超过 1800
- 参加每月的客户路访会议, 以了解客户的需求变化和项目期限; 使用 Excel 函数 (如数据透视表) 和 Bloomberg 的 Excel 加载项等函数构建上市公司的估值模型 (例如 DCF 估值模型)

## 实践项目和兴趣爱好

电影票房预测模型 美国, 纽约  
哥伦比亚大学 2019.10 – 2019.12

- 使用网络爬虫获得 IMDB 电影特征数据, 将 30000 部电影的获利能力分为 3 个级别, 分类的准确度为 64.6%
- 将电影的数字和文字特征数据与经济指标时间序列数据结合起来, 用于训练决策树、KNN、随机森林分类模型

兴趣爱好: 绘画 (静物素描、卡通), 图片编辑 (Photoshop), 视频编辑 (用于活动宣传、项目报告)