# 杨旭程

+86-18391029910|xy224@ic.ac.uk

## 教育经历

**硕士-帝国理工大学**专业: 2024年 9 月 – 2025年 1 1 月

金融科技 | 一等学位

相关课程：金融计量学实战，资产组合管理，金融数学，金融大数据，区块链金融，金融随机微积分等

**本科-南开大学：理学学士 | 经济学学士** 2020 年 9 月 – 2024 年 7 月

主修：数据科学与大数据技术 GPA: 3.67/4

双修：金融学 GPA: 3.61/4

相关课程：随机过程，数理统计，数据挖掘与机器学习，运筹与优化，实变函数，概率论等

## 实习经历

**北京小米科技有限公司** 人工智能算法实习生 2024 年 1 月 – 2024年 7 月

- 调研并复现大型语言模型及人工智能的最新进展，撰写并阐释了五份研究报告作为部门同事的参考；
- 负责小米AI笔记项目，通过prompt engineering。在若干笔记场景中设计不同的提示词，分别评估了四种不同的大语言模型在不同参数和提示词下的表现，并相应调整，选择最合适的提示词和参数；
- 参与智能家居项目，利用随机森林和聚类等算法挖掘和研究用户的潜在习惯，学习用户的日常习惯，据此进行主动化的操作推荐

**北京快手科技有限公司** 数据分析实习生 2023 年 6 月 – 2023 年 1 2 月

- 使用SQL，从数据库中按需聚合各种维度的数据，根据视频池中的点赞率，评论率，投票情况等已有指标为新用户推荐高质量的视频，并对过去的视频推荐策略加以改进，以增加用户留存的可能性；
- 在缺乏先验知识的情况下，用主成分分析(PCA)的方法对人群的所有特征做降维，使用降维后的特征对人群做聚类分析，将目标人群分为了有显著差异的五大群体，为后续分析提供借鉴；
- 使用随机森林模型，根据人群特征对高价值人群建模，获得各特征在该人群上的重要性排序，并使用shap进一步得到不同特征对目标人群的正负向影响；
- 根据已有的认知，对人群的静态特征和互动行为做特征提取，并对覆盖率不均衡的特征做下采样处理。进而使用泊松回归模型获得上述特征与用户未来一周活跃情况的关系，量化了过去认知中的各个指标，使不同特征对活跃度的影响在量上有了一定的可比性。

**安永(中国)企业咨询有限公司** 实习生 2023 年 1 月 – 2023 年 2 月

- 参与建筑、冶金行业央企与其关联企业的年审，参与了企业的抽凭、固定资产盘点工作；
- 负责会审中全部 20 家主体的账表核对工作，并完成货币资金、固定资产、应付职工薪酬与权益类的底稿；
- 与数据分析团队对接，收集 65 家子公司的财务数据，用 EY random 与 EY Microstart 做抽样，分析企业经营状况。

**前海维港创业投资基金（深圳）有限公司** 投行部实习生 2022 年 7 月 – 2022 年 8 月

- 参与了目标公司的尽职调查工作，根据目标公司的财务报表，分析了公司的营业收入情况和负债情况，发现该公司可能存在的问题；
- 收集并阅读了大量新能源光伏行业的行业研究报告，研究该行业的发展情况、产业链上下游、市场规模和市场格局，并以隆基绿能科技股份有限公司为例，分析目前其经营状况以及融资来源及特点，提出合适的发展建议；
- 参与撰写了目标公司的商业报告书，完成网络安全行业的行业分析，利用估值模型对公司收益进行预测，并制定融资计划，最终参与了路演。

## 科研经历

**动量策略实证与改进研究** 2024 年 8 月 – 2024 年 1 0 月

- 基于轮动策略，使用过去十年间中国沪深300ETF，中证500ETF，黄金ETF以及外盘的纳指ETF构建轮动池。根据动量规则，每天买入前22个交易日内涨幅最大的ETF，获得基准夏普率0.98和基准累计收益率402.06%；
- 完善动量规则，使用前22个交易日内的最低价对最高价做线性回归，得到RSRS斜率。将RSRS斜率与拟合优度R方的乘积作为新基准选择取值最大的ETF，得到夏普率1.08和基准累计收益率616.28%；
- 以上述过程中选择占比最高的ETF，黄金ETF为例。使用22日均线，63日均线，126日均线，每日涨幅，每日换手率等因子分别构建线性回归模型，xgboost模型，神经网络模型作为基分类器，以RMSE比值作为权重构建线性可加模型以预测ETF的次日价格。获得夏普率1.94和基准累计收益率484.23%

**剑桥大学 AI+暑期研究项目** 2022 年 8 月 – 2022 年 9 月

- 通过计算得到账面市值比、市值、盈利能力和投资能力（收益率之差），建立 Fama-French 五因子模型；
- 使用 2×3 排序法，将沪深 300 成分股分为 6 个投资组合，用收益率构建反转因子；用换手率替代投资者情绪，对 2015-2021 年的沪深 300 成分股做了线性回归，使用岭回归克服多重共线性进一步建模，最终建立包含投资者情绪、反转因子，FF 五因子的七因子模型；
- 随机选取了 2018 年沪深 300 成分股中 10%的股票，按照收益率分为三组、构建随机森林模型，验证了七因子模型的有效性；
- 获得反转因子对股市的超额收益率有显著影响，但以换手率代替的投资者情绪对超额收益率的影响并不显著，需做进一步分析的结果。

## 比赛经历

**美国大学生数学建模比赛** 队长 2023 年 3 月

- 通过 Python 爬虫抓取单体船和两体船的特征，丰富原始数据集，通过逐步回归，选择 7 个特征建立了多元回归模型，并对模型做了误差诊断；使用方差分析方法，得出区域效应对船只价格有显著影响的结果，
- 使用 BP 神经网络算法，对多元回归模型做了非线性优化。并使用 ABC-人工蜂群算法，对 BP 神经网络做了进一步优化，防止局部最优的产生；
- 通过 Python 爬取香港二手船交易市场的相关信息，使用上述模型做了价格预测，得到了约 87%的准确率；
- 参照相关文献，开创性的使用了奢侈生活价格指数(CLEWI）对原始的截面型数据做了序列化优化，使结果更具实际意义，且准确率提高为 92%。

## 计算机和语言技能

**软件：** Python、R、SQL、Excel、Spark

**语言：** 英语 ( CET6 649）、雅思 7.5 、GRE 324

# XUCHENG YANG

+44-07776993230 | xy224@ic.ac.uk | in/xuchengyang011122

## EDUCATION

**Imperial College London** London, UK | Aug. 2024 – Dec. 2025

- **Major:** Financial Technology | Distinction degree
- **Relavent Module**: Financial Econometrics in R/Python, Investments and Portfolio Management, Mathematics for Finance, Big Data in Finance, Applied Quantitative Macro Strategies, Quantitative Equity Investing, Stochastic Calculus in Finance.

**Nankai University** Tianjin, China | Sep. 2020 – Jul. 2024

- **Degree:** Bachelor of Science|Bachelor of economics
- **Major:** Data Science and Big Data Technologies (GPA: 3.67/4)|**Minor:** Finance (GPA: 3.61/4)

## PROFESSIONAL EXPERIENCE

**Xiaomi Corporation|World TOP 500** Jan. 2024 –Jul. 2024
**AI Algorithm Engineer Intern**

- Researched and replicated 14 latest advancements in large language models and artificial intelligence, and wrote 6 research reports as references for department;
- Administered prompt engineering. Designed multiple prompts in 4 note-taking scenarios. Evaluated performance with 5 different parameters models and prompts, and chose most suitable prompt and parameters;
- Constructed proactive intelligence in smart homes, utilizing algorithms such as XGBoost and clustering to mine users' potential habits, enabling furniture to perform actions accordingly.

**Beijing Kwai Technology Co., Ltd|Chinese TOP 200** Jun. 2023 –Dec. 2023
**Data Analyst Assistant**

- Used SQL to aggregate data of various dimensions from the database, recommended high-quality videos to new users and improved the previous recommendation strategy to increase the retention rate;
- Applied PCA method for dimensionality reduction and used the processed features to conduct cluster analysis on targeted groups, classifying them into 5 groups with significant differences;
- Constructed random forest model to obtain the rank of each feature for the high-value population as well as the positive and negative impacts of different factors on the target population;
- Conducted factor mining and feature extraction to illuminate the intricate relationship between user behaviors and their levels of activity and implemented the under-sampling method to address imbalances in the features, successfully developing a Poisson regression model, which ultimately yielded a 62.3% adj.$R^2$ and a 0.53 RMSE.

**Ernst & Young Global Limited** Jan. 2023 –Feb. 2023
**Audit Intern**

- Participated in the annual audit of state-owned enterprises and their affiliated enterprises in construction and metallurgy industries and conducted the work of withdrawal and fixed assets inventory;
- Responsible for checking the account statements of all 20 entities in the review, and completing the drafts;
- Interfaced with the data analysis team, collected the financial data of 65 subsidiaries, and used EY random and EY Microstart for sampling to analyze the business conditions of enterprises.

## RESEARCH EXPERIENCE

**Research of the Improvement and Validation for Momentum Strategy** Aug. 2024 –Oct. 2024

- Constructed a ETF pool and buy the highest increased one over the past 22 trading days from 2015-2024, using CSI 300, gold, and Nasdaq, etc. Achieved a benchmark 0.98 Sharpe ratio and 402.06% return.
- Performed a linear regression over the past 22 trading days to obtain the RSRS slope. Use the product of the that and R -square to select ETF, resulting in 1.08 Sharpe ratio and 616.28% return.
- Took the ETF with the highest proportion in the above process. Used factors such as 63-day MA, and turnover rates to construct linear regression, XGBoost, and neural network as base models. Used the RMSE as the weight to build a linear additive model and predict the price on the second day, achieving a 1.94 Sharpe ratio and 484.23% return.

**Cambridge Academic Programme (online) - Investment &Quantitative Finance Track** Aug. 2022 –Sep. 2022

- Collected stock data of CSI300, acquired the book-to-market equity, market value, profitability, and established a five-factor asset pricing model;
- Ranked CSI300 Constituent Stocks' earnings rate, used 2×3 ranking method to generate the reverse factor, used turnover rate as a sign of investor sentiment, and eventually established a seven-factor model;
- Performed linear regression and random forest model on the CSI300 component stocks from 2015 to 2021, and utilized ridge regression to copy with multicollinearity;
- Concluded the results that the reversal factor has a significant impact on the excess earnings of the stock market, while the investor sentiment replaced by the turnoverrate was insignificant.

## TECHNICAL SKILLS

- **Languages:** English (Fluent) ; Mandarin (Native)
- **Python:** Know basic functions and data structures. Can perform dataframe analysis and utilize various machine/deep learning models.
- **R:** Know how to use ggplot for visualization. Familiar with dataframe analysis and applying machine learning methods.
- **SQL:** Can select data under given conditions. Know how to aggregate data using window function or join dataset.
- **Statistics:** Master probability theory, various random distributions and their relavent properties. Familiar with stochastic process, machine learning models and hypothesis test methods. ODE & SDE (learning).
- **Time-Series:** Master relevant theories and test methods of time series analysis. Familiar with ARIMA and GARCH models.

## EXTRA-CURRICULAR & INTERESTS

- Chess enthusiast with one year studying experience, and also a 1200-rated player on chess.com
- Classical music lover, piano and violin pieces fan