



# 杨睿

✉ [yangruipis@163.com](mailto:yangruipis@163.com)    📖 [cnblogs.com/yangruiGB2312](http://cnblogs.com/yangruiGB2312)  
☎ (+86) 188-1823-1378    🌐 [github.com/Yangruipis](https://github.com/Yangruipis)  
🏠 上海市-松江区-文汇路 600 弄

## 🎓 教育经历

上海对外经贸大学	2012.09 – 2019.03
研究生 (保研)	数量经济学
本科	统计学 (商务统计方向)
	专业排名:4/17 GPA:3.31, 专业排名:2/54

## 🏆 荣誉和奖项

“东证期货杯”统计建模大赛	一等奖 ( <i>top 10 in 1000+</i> )	2018.03
“华为杯”全国研究生数学建模大赛	二等奖 ( <i>top 15%</i> )	2017.09
中国大数据创新行动大赛	一等奖 ( <i>top 1</i> , 奖金 5 万元)	2016.12
美国大学生数学建模竞赛 ( <i>ICM</i> )	二等奖	2016.02
上海市优秀毕业生、校十佳学术标兵、爱建企业奖学金一等奖 (全校 <i>top10</i> )		

## 💼 工作经历

平安好房, 数据挖掘实习生      2018.06 – 2018.8

- 挂牌与成交房源的特征工程, 以及房价走势预测的指标体系构建
- 小区均价模型构建与上线 (主要是问题回溯), 包括了小区相似度模型 (用以补全缺失小区数据) 和小区均价预测模型 (*GBDT*, *LSTM* 等)
- 基于小区均价的一房一价模型构建 (虚拟变量回归、分位数回归等)

上海通金投资有限公司 (通联数据旗下私募), 量化研究员      2017.04 – 2017.10

- 基于机器学习的期权策略, 包括了 50etf 期权相关特征提取, 模型选择等 (*LR, SVM, LSTM*), 策略回测收益 12%
- *CTP* 接口 *Python* 版本再封装与测试, 大大简化了策略模拟盘和实盘操作 🔄 *ctp-python*
- 股指期货跨期套利策略模拟盘与实盘上线 (*python*)

Kantar Media CIC (中国领先的网络口碑咨询公司), 数据挖掘      2016.03 – 2016.10

- 文本挖掘软件开发 (两个月, *C#*), 包括了窗体设计和优化、情感分析 (*Naive Bayes*, *word2vec*), 文本匹配 (根据公司业务场景提出规则表达式方法)
- 爬虫 (*BS4*, *WireShark* 等); 数据可视化 (*ggplot2*, *ggmap*); 微博账号影响力评价体系构建

## 📁 项目经历

Kaggle Home Credit Default Risk      2018.06 – 2018.08


- 负责特征工程, 包括了数据预处理, 多表聚合, 特征选择 (*Lasso*, *LightGBM*, *FeatureTools*), 以及模型的选择和优化 (*GridSearch*, *BayesOpt*, *LightGBM*, *Stacking*)
- 截至 2018 年 7 月 25 日, 获得 AUC 值为 0.800, 在 5000+ 队伍中排名 7%(343)

“东证期货杯”全国大学生统计建模大赛, 信贷违约模型      2018.01 – 2018.02

- 负责利用 *python* 实现完整机器学习流程, 包括数据入库、特征工程、模型选择 (*LR*, *NB*, *CART*, *RF*, *GBDT*)、模型调参 (*bayes* 优化) 与融合 (*Stacking*), 获全国一等奖 (*top 10 in 1000+*)
- 难点 1. 数据不平衡 (*SMOTE+Tomek*); 难点 2. 需要降低授信成本 (引入 *Cost-Sensitive Matrix*)


## 个人机器学习库开发

2017.12 – 2018.01

- 个人项目  simple\_ml , 开发语言: *python*, 开发平台: *Manjaro*, 开发工具: *Emacs+PyCharm*
- 仅依赖 *numpy* 实现了常用的特征工程、模型评价、分类 (*LR, KNN, NB, DT, SVM, SVR, BP, AdaBoost, GBDT, RF, stacking*) 与聚类 (*K-means, 层次*) 算法、自动化数据预处理和特征选择方法, 测试覆盖率 43%

## 大学生挑战杯, 商科统计软件 VISAP 开发

2015.01 – 2015.06

- 团队项目  VISAP , 开发语言: *C#*, 开发平台: *Visual Studio*
- 负责统计功能 (空间作图、方差分析、回归分析、时间序列、灰色预测等), 抽样算法开发 (*ARS*)
- 爬虫对接, 包括: 国家统计局宏观数据、财经股票数据等一键获取; 界面开发和优化

## 技能 & 自我描述

---

- *python* 三年开发与建模经验, 熟练使用 *sklearn, pandas* 等库, 熟悉 *XGBoost, LightGBM, TensorFlow, Keras*, 面向对象编程
- 熟练掌握 *MySQL, Hive* 的查询和优化、*R* 语言数据可视化, *Java* 实现基本数据结构与算法, 了解 *Git, Spark, Docker*, 两年 *Linux* 使用经验, 长期使用 *Emacs* 和番茄工作法进行 *GTD*。
- 数理统计基础扎实, 熟悉计量经济, 多元统计, 时间序列, 贝叶斯统计等方面的知识。
- 熟练掌握机器学习相关算法及原理 (*SVM, LR, DT, RF, NB, GBDT* 等), 了解深度学和推荐系统基本算法 (*RNN, LSTM, SVD, ICF* 等)

热爱数据与编程, 有大量统计建模经验, 喜欢从底层实现算法。乐于挑战, 追求新技术。英语读写流畅 (CET6:541), 爱好有摄影 (校摄影大赛二等奖)、骑行 (270km/3 天) 等。