

Jian Yang

Senior Software Engineer · Shanghai

☎ (+86) 18516191607 | ✉ ernestyj@outlook.com | 📱 Ernestyj

M.Eng. in Software Engineering, Shanghai Jiao Tong University

B.Eng. in Software Engineering, University of Electronic Science and Technology of China

Summary

Current data engineer at Paypal with 3 years work experience. Developed C++ based large distributed NoSQL database; experienced with C++ developing. Designed and developed financial transaction risk control system, including decision rules analysis, automatic risk control platform, real time transaction data monitoring and alerting; experienced with Python full stack development, data analysis, model building and visualization.

Work Experience

PayPal · Decision Analytics and Management

Aug. 2017 - Now

Rule360 Decision Rule Analytics - Senior Software Engineer

- Cooperated with global analysts and analyzed global transactions data (SQL, Pandas). Built Decision Rule Analytics System (Flask, React), including rule parsing (XML), near real time decision data monitoring, rule performance metrics evaluate, decision factors reference analysis (Neo4j). Saved about 80% decision analytics cost.
- Optimized read/write performance on risk big data structural storage (MySQL), supported full data update.

HyperNet 自动化风控平台 - 数据分析与全栈开发

- 构建统一数据预处理平台, 对不同源数据 (HDFS, Teradata) 进行 Spark ETL, 进行特征统计分析与特征独热编码, 整合生成训练与测试数据集
- 使用 XGBoost 构建决策树模型, 按风控经验指标过滤决策规则生成风控方案; 使用 React 构建前端损失报警与自动风控方案监控管理平台
- 构建 SQL 风控规则翻译引擎, 支持用户从任意 SQL 规则翻译成 IBM 决策引擎规则
- 有效缩短 2-3 周生成风控方案的周期, 节省约 70% 人力成本, 每月止损收益约 \$80k, 荣获 PayPal Spot Award(5%)

Spider 实时交易数据监控报警平台 - 后端开发

- 开发报警任务执行系统, 消费交易决策引擎产生的警报 (Apscheduler), 并调度执行事后聚合分析 (Pandas, Spark)、可视化 (Matplotlib)、邮件报警 (Jinja2) 等任务
- 基于 Celery, RabbitMQ 实现任务链和异步优先级任务调度, 提高报警服务性能及可靠性

百度 · 基础架构部

2017 年 3 月 - 2017 年 6 月

Mola 大型分布式 NoSQL 数据库 - 研发工程师

- 研究 Mola 分布式架构, Meta Server 选主机制 (Lease), Node 数据分片自动恢复
- 通过加锁解决数据存储节点 Node 中的分片在自动恢复过程中, 异步修改 Recover Task 引发的恢复任务错误异常, 降低 Mola 服务的故障率

百度 · 生态数据产品部

2016 年 8 月 - 2017 年 2 月

精益广告投放数据分析 - 数据工程师 (实习)

- 精益对凤巢、Holmes 等广告数据进行监测与评估, 建立漏斗模型 (人群转化、流动、归因分析) 分析广告营销活动效果
- 负责对上游广告源数据进行 ETL(Hive) 及数据拼接、编写 Spark UDF, 优化 ETL 速率, 并线上验证数据分析准确性

项目经历

Alexa 智能家居项目 - 负责人

2018 年 8 月 - 2018 年 10 月

- 参与 PayPal 创新实验室智能家居组, 负责 Alexa 与飞利浦智能灯具集成开发, 组织创新日活动并发表 Alexa 与智能家居相关演讲

量化投资策略分析 - 算法研究

2016 年 2 月 - 2016 年 10 月

- 研究并提出基于最大信息熵的股票特征选取方法, 对比主成分分析、奇异值分解等方法高出约 4.5% 的预测精确度; 研究并提出 SRA-Voting 混合二元分类模型, 对比支持向量机、随机森林、AdaBoost 高出约 3.125% 的预测精确度

技能

编程语言	Python, C++, Java, SQL, Shell
框架技术	Pandas, Flask, Celery, React, Spark
工具	Teradata, MySQL, MongoDB, Emacs, Git, Jenkins, Docker, Excel, Tableau
英文	CET6, GRE(314), 良好听说读写能力