

本项目构建了一套自动化 **金融新闻分析与价格预测系统**，涵盖 **新闻抓取、实体提取、情感分析、主题建模、价格预测** 等多个 NLP 任务，结合 LLM、LangChain 和本地知识库，探索新闻对债券价格波动的影响。

- **文本采集清洗**: 开发自动化爬取脚本，从订阅邮件账户提取新闻文章，进行结构化解析与清洗。
- **NER**: 基于 Transformer 模型（如 BERT、RoBERTa）训练 NER 任务。
- **实体匹配**: 利用 BERT 微调 + 向量检索（FAISS），模糊匹配，将提取实体与数据库对齐。
- **RAG**: 基于 LangChain 搭建 RAG 框架，结合本地知识库。
- **主题建模与情感分析**: 采用 BERTopic + FinBERT + LLM，提取主题并评估市场情绪。
- **Prompt、LLM设计优化**: 评估、优化 LLM 处理金融文本的能力，提高生成质量与规范性
- **统计分析**: 价格走势分析，假设检验、生存分析、回归分析、A/B 测试

基于 LSTM 的债券价差预测与自动化部署系统

数据科学家，机器学习工程师

2021.12 - 2022.10

探索、构建债券价差预测模型，使用 LSTM、XGBoost、LightGBM 进行实验，并通过 MLflow 进行版本管理，结合 CI/CD、Airflow、Jenkins 和 OpenShift 实现自动化训练、微调和部署，同时设计实时监控系統确保长期稳定性。

- **数据处理**: 收集并清理文本、时序数据，EDA 和特征工程。
- **模型开发**: 实验 LSTM、XGBoost、LightGBM 预测债券价差，利用 MLflow 进行版本管理、超参优化与模型评估。
- **自动微调**: 使用 Airflow 进行每日自动微调，并结合 Evidently AI 进行数据漂移检测。
- **自动化部署**: 通过 Jenkins + OpenShift 实现模型自动更新、测试与生产部署。
- **实时监控**: 采用 Prometheus 监测推理结果，异常时触发数据漂移检测并生成报告。

实时 RFQ 分析与债券偏向评估平台

数据工程师，数据分析师

2024.03 - 2024.12

构建了一个实时 RFQ(询价单)分析平台，用于评估 债券 Skew 和对手方购买意愿，帮助交易员在谈判中优化定价策略。系统基于 WebSocket 进行实时数据流处理，并通过 Flask RESTful API 计算 债券库存、市场需求、历史成交数据 等关键指标。

- **实时数据处理**: 基于 WebSocket 低延迟处理 RFQ 订单流，实现实时数据展示和计算。
- **债券 Skew 计算引擎**: 基于 MariaDB 设计多表查询，实现 Skew 计算引擎，根据 库存水平、成交量、市场波动 计算债券购买意愿评分，为交易员提供定价策略参考
- **高并发 API 开发**: 采用 Flask + RESTful API 支持交易数据查询、库存分析、Skew 计算等关键功能。
- **数据自动分析**: 开发 Auto-Profiling 工具，实现 EDA、特征工程和数据可视化 自动化。

5G流量预测与异常检测系统

算法工程师

2021.06 - 2021.09

本项目旨在构建一个分布式时序预测和异常检测系统，利用 Spark 进行大规模数据处理和分布式机器学习建模，并结合 Elasticsearch 实现实时异常检测。

- **分布式时序预测**: 基于 Spark 并行计算，训练大规模 5G 网络流量预测模型，提升处理效率。
- **机器学习与深度学习建模**: 采用 深度学习（LSTM）或回归模型（XGBoost、LightGBM）预测未来 48 小时流量，优化资源调度。
- **增量学习与模型自适应**: 构建 历史学习 + 增量学习 流程，动态更新模型，提升预测精度。
- **误差学习优化**: 基于 回归分析 + 异常检测，分析预测误差，优化模型稳定性。