金融大数据课程个人报告

郭昭芃 2024100893

一、个人贡献

在本次《金融大数据与量化投资》课程大作业中，我主要负责因子构建和Transformer模型的设计与实现。以下是我在小组中的具体工作内容：

1. 因子构建

因子构建是量化策略的基础环节，我负责构建了多个类别的因子，包括基本面因子、另类因子、Barra因子和量价因子，力求在因子的多样性和创新性上取得突破。

**（1）基本面、量价与Barra因子：传统因子的多样性**

在基本面因子构建中，我根据质量、盈利、成长和估值的二级分类，构建了多样化的因子，如资产负债率、净资产收益率、净利润增长率、市盈率等，全面反映公司的财务健康状况和估值水平。量价因子方面，我设计了动量、反转、波动和流动性等类型的因子，如10日指数移动均线、相对强弱因子（RSI）、20日收益标准差等，捕捉市场微观结构的变化。Barra因子则涵盖了市场风险、动量、规模、盈利收益率等类型，如beta、momentum、size等，为策略提供了多样化的风险控制工具。这些传统因子从不同角度捕捉市场信息，为策略提供了坚实的基础。

**（2）另类因子：创新性与非传统数据**

除了以上传统因子的构建，我着重构建了非传统的另类因子。我注重从非传统数据中挖掘有效信息，并通过创新的方法构建因子。例如，我基于**新财富获奖分析师**和**分析师预测数据**，手动构建分析师的一致预期因子。此外，我还构建了**分析师共同覆盖关联网络**和**基金共同持仓关联网络**，捕捉关联股票之间的“领先滞后”效应。针对定增、回购等**事件**，我设计了不同的事件驱动因子，量化事件对股价的影响，并根据股票之间的Barra距离做横截面外溢，扩展因子覆盖度。这些另类因子通过引入非传统数据和创新构建方法，为策略提供了独特的超额收益来源。

**2. Transformer模型设计**

为了尝试提升Neural Network的预测能力，我设计并实现了Transformer深度学习模型。Transformer模型在自然语言处理领域表现出色，我将其创新性地应用于金融时间序列数据的预测。通过多头注意力机制和位置编码，模型能够捕捉金融时间序列数据中的长期依赖关系。

二、对课程的理解与思考

通过本次《金融大数据》课程的学习与实践，我对量化投资的核心理念和方法有了更深刻的理解。首先，金融大数据为量化投资提供了丰富的信息源，通过对海量数据的分析和挖掘，可以发现市场中的规律和机会。无论是传统的基本面数据、量价数据，还是非传统的分析师预测、事件驱动数据，都为我们构建有效的量化策略提供了重要支持。其次，量化投资的系统化方法减少了人为情绪和主观判断的影响，提高了投资决策的科学性和稳定性。通过多因子模型和机器学习技术的结合，我们能够更全面地捕捉市场信息，优化投资策略。

此外，课程让我认识到因子构建在量化策略中的核心地位。无论是传统因子还是另类因子，都需要结合市场特点进行创新设计，才能更好地捕捉超额收益。同时，模型的构建和优化也是量化投资的重要环节，如何将前沿技术应用于金融数据，是未来需要进一步探索的方向。

三、对课程的建议

1. 增加实践环节：建议课程引入更多实战性内容，如模拟交易平台或实盘操作机会，让学生在实践中巩固所学知识，提升对市场的敏感度和策略的实操能力。

2. 引入更多前沿技术：当前量化投资领域发展迅速，建议课程增加对深度学习、强化学习、自然语言处理等前沿技术的讲解，帮助学生了解最新的技术趋势和应用场景。

3. 加强案例教学：通过分析经典量化策略案例或失败案例，帮助学生更好地理解量化策略的设计逻辑和风险控制方法。