**2024 年华南师范大学研究生数学建模赛题**

**A题 基于舆情因子的因子选股策略构建**

【赛题背景介绍】：

据统计，中国股票市场中散户投资者数量约占投资者总数的95%，中国股票市场也被称为“散户市场”。由于散户投资者的情绪极易受到市场形势与氛围的影响，因此中国股票市场具有很明显的情绪化现象。

伴随着科技的发展，互联网搜索引擎以及各类社交网站不断进入人们的生活，越来越多投资者倾向于在互联网平台交流投资建议与心得，股吧为投资者提供了良好的交流平台，每个人既是信息的制造者，也是信息的传递者。投资者往往在互联网平台上发表个人观点，这一行为会对其他交易者的交易行为产生影响，从而造成股价波动。从投资者情绪的角度出发，构造舆情因子并进行选股策略的构建能够启发投资新思路，为因子选股策略的构建提供更多的选择。

【赛题数据描述】：

舆情对股价的影响可以从两个方面刻画:股吧评论数量和股吧评论体现的情绪，也称为情感值（分为正向情感和负向情感）。股吧评论数量对股价的影响可能更偏短期；而情感值分析由于涉及了对公司业绩、经济环境的中长期展望等情感，所以对股价会产生持续的影响。本题主要通过分析情感值来构造舆情因子。

舆情因子的构造可以通过现有的库与函数算得情绪值，或者做一定地映射，也可以使用自己的语料库构建情感分析模型（赛题提供部分与金融相关的训练语料库）。参赛者需要自行通过爬虫技术爬取股吧的股票评论，利用情感分析模型为每一条评论进行情感赋值。

1. 构建传统的多因子选股模型，从中国A股市场选择不少于10只股票，给出股票代码，并解释选择股票的逻辑思路。
2. 根据第1问构建选定的股票数据，选定一个或多个股吧，获取股吧的文本数据（例如股票评论、财经新闻等），并将获取的数据储存至文件中，数据内容，格式不限。

要求：获取2021.01.01 – 2021.06.01的数据；

1. 基于第2问获取的文本数据建立数学模型，构造舆情因子，并检验构建证舆情因子在A股市场的有效性，然后根据有效性验证结果进一步优化构建的舆情因子；
2. 基于第3问构建的舆情因子，构建基于舆情因子的单因子选股策略，将第1问挑选出的股票作为股票池，利用Auto-trader平台，在2021.01.01 – 2024.06.01的时间区间进行回测，并对构建的策略进行分析；