1. 引言

第一节 研究背景与意义

1. 研究背景

（一）投资者情绪对资本市场的重要影响

投资者情绪作为一种难以量化但影响深远的因素，长期以来在资本市场中扮演着重要角色。早在1636年荷兰郁金香狂潮事件中，人们就已发现投资者情绪对资产价格的显著影响，投资者心理中存在的从众和恐慌心理往往导致市场非理性波动。在金融危机来临时，金融市场的急剧变化与投资者情绪密切相关，体现了情绪因素在市场运行机制中的重要地位。

近年来，学术界对投资者情绪的关注度不断提高。研究表明，互联网投资者情绪对于股票市场的波动具有指导意义。吴杰胜等（2019）基于多部情感词典结合规则集，对微博情感进行分析，实证结果显示该方法在微博情感分析领域效果较好。戴德宝等（2019）使用机器学习方法对文本进行分析，同样发现投资者情绪指标对股价走势具有一定的预测作用。

陈其安等（2017）的研究进一步证实，中国股票市场价格波动性与投资者情绪呈正相关关系，与市场利率呈负相关关系，且投资者情绪会弱化货币政策的调控作用。这些研究共同表明，投资者情绪已成为影响资本市场的重要因素，不可忽视其在价格形成和市场波动中的关键作用。

（二）自然语言处理技术在金融领域的应用前景

随着大数据和人工智能技术的迅猛发展，自然语言处理（NLP）技术在金融领域的应用前景日益广阔。近十年来，人工智能科技已成为全球最热门的科学技术领域，特别是模拟人工神经元网络的机器学习流派深度学习的兴起，极大地推动了通用人工智能方法的普及。

在金融文本分析方面，情感分析是从股票文本中获取投资者情绪的主要手段，其准确性直接影响后续股票价格波动分析的基础。不同的情感分析方法准确率各异，因此选择较高准确率的情感分析方法对研究结果至关重要。然而，需要注意的是，直接套用现成的情感分析软件和外部词典资源在金融领域难以取得很好的效果，需要针对金融特定领域进行模型调整和优化。

相比传统基于词典和机器学习的统计方法，基于深度学习的自然语言处理技术，如BERT模型，在情绪信息识别方面表现出色。通过对BERT模型进行金融领域特定文本的进一步预训练并调整其内部语义分布，可以显著提升舆情抽取和识别任务的准确性和可信度。这种技术进步为挖掘海量金融文本数据中的情绪信息提供了有力工具，为投资者情绪研究打开了新的可能性。

1. 研究意义

（一）理论意义

本研究从行为金融学的角度，探讨了互联网平台上投资者言论所蕴含的情绪信息与股票市场表现之间的关系，丰富了行为金融学的理论内涵。通过构建情绪指标体系，本研究有助于深化对投资者情绪如何传导并影响市场的理解，为非理性投资行为研究提供了新的分析框架。本研究还将NLP技术与金融数据分析相结合，探索了深度学习模型在金融文本分析中的应用效果，拓展了自然语言处理技术在专业领域中的应用边界。通过比较不同情感词典和数据集在股票预测领域的表现，本研究为金融领域的文本分析方法提供了重要的方法论参考。

（二）现实意义

首先，互联网评论作为投资者情绪的直接表达，对市场预期具有重要的指示作用。通过分析和理解网络评论中所包含的情绪信息，可以为投资决策提供参考，帮助投资者更好地把握市场动向。其次，随着"互联网+"政策的实施，网络媒体在信息传播过程中发挥的作用越来越大。正确识别新闻报道和网络评论中的舆情信息，并理解其与股票收益之间的关系，不仅有利于投资决策，还能为政府监管部门制定相关政策提供参考，维护市场秩序和稳定。此外，通过研究不同行业的投资者情绪特征及其对市场的影响，可以揭示行业间的情绪传导机制，为投资组合的风险管理和资产配置提供新的视角。

对投资者而言，本研究构建的情绪指标可以作为投资决策的重要参考。通过分析互联网评论中蕴含的市场情绪，投资者可以更全面地把握市场预期变化，避免盲目跟风，优化投资策略。对监管机构而言，本研究有助于构建金融市场情绪监测系统。互联网评论中的情绪波动往往能提前反映市场异常，通过实时监测投资者情绪，监管部门可以更早识别市场风险，及时采取干预措施，维护市场稳定。对研究机构而言，本研究提供了一种新型的市场分析工具。通过整合传统量化指标与情绪分析结果，研究人员可以构建更全面的市场预测模型，提高预测准确性。研究表明，情感分析能有效反映股票市场状况，并对股价走势产生强烈影响。

第二节 研究内容与方法

1. 研究内容

本研究旨在通过构建情绪指标体系，深入探究投资者情绪对股票市场表现的影响。本文的研究内容主要包括四个方面：首先，利用Python编程语言开发网络爬虫，从东方财富股吧获取8个行业111只股票在2025年2月的评论数据。通过数据清洗与预处理，构建高质量的文本分析样本。其次，基于先进的自然语言处理技术对股评文本进行情感分析，构建包括情感得分、情感极性、情感强度等在内的多维情绪指标体系。这些指标能够从不同角度刻画投资者情绪的特征。第三，通过akshare接口获取样本股票的日度交易数据，将情绪指标与股票行情数据进行匹配，构建包含情绪、交易等信息的面板数据集。第四，运用计量经济学方法，重点探究不同维度情绪指标对未来股票收益的预测能力。通过构建面板回归模型，分析个股和行业层面的投资者情绪对股市的影响机制。

二、研究方法

（一）文献研究法。通过系统梳理国内外关于投资者情绪与股市表现关系的研究文献，总结现有研究成果，为本文的研究提供理论基础和方法借鉴。

（二）数理分析法。运用数理统计和计量经济学方法，对文本情感分析结果进行统计描述，构建情绪指标体系，并通过相关性分析、回归分析等方法研究情绪指标与股票收益的关系。

（三）实证分析法。基于构建的面板数据，采用固定效应模型等计量方法，控制个体效应和时间效应，分析情绪指标对股票收益的影响。同时，为保证研究结果的稳健性，将进行一系列稳健性检验，包括采用不同的情绪指标进行对比分析，以及通过工具变量法解决可能存在的内生性问题，确保研究结论的可靠性。

第三节 研究创新点与贡献

在技术方法层面，本研究采用先进的自然语言处理模型进行情感分析。具体而言，通过应用预训练语言模型对金融文本进行深度语义理解，提高了对投资者评论情感的识别准确性。该模型能够有效捕捉文本中的细微情感差异，为后续构建情绪指标提供了可靠的技术支持。

在研究视角层面，本文创新性地构建了多维情绪指标体系。通过整合情感得分、情感极性、情感强度等多个维度的指标，全面刻画了投资者情绪的不同特征。这种多维度的情绪指标体系不仅能够更准确地反映投资者情绪状态，还可以深入分析不同维度情绪指标对股票市场的差异化影响。

第四节 结构安排

本文共分五章，论文的具体组织方式如下：

第一章为引言。该部分主要介绍研究背景与意义，阐述研究内容与方法，并说明本文的创新点与贡献。

第二章为文献综述。该部分回顾投资者情绪研究现状，包括情绪的定义与测度方法，以及投资者情绪对股市的影响机制。同时对自然语言处理技术的研究进展进行梳理。

第三章为研究设计。该部分详细说明数据来源与处理方法，构建情绪分析模型框架，设计情绪指标体系，并对模型设定和变量进行说明。

第四章为实证分析。该部分从个股和行业两个层面展开研究。通过描述性统计、单位根检验、协整检验等方法验证数据特征，运用面板数据回归分析投资者情绪对股市的影响，并进行异质性分析和稳健性检验。

第五章为研究结论。该部分总结主要研究发现，提出政策建议与启示，同时讨论研究局限，并对未来研究方向进行展望。