

基于情绪分析和时序模型的股价预测

Yuting Feng

摘要

股票市场因其波动性和非线性特性，使得股价预测成为一项具有挑战性的任务。本研究通过结合社交媒体情绪分析与时序模型，探索其在股票趋势与价格预测中的应用。研究中使用了两种基于 Transformer 的预训练模型（FinancialBERT 和 DistilRoBERTa），对金融推文进行情绪分析，并结合股票量价数据构建综合数据集。传统机器学习模型被用于趋势预测，而长短期记忆网络（LSTM）则被用于价格预测。实验结果表明，DistilRoBERTa 模型在情绪分析任务中的表现优于其他模型，而 LSTM 模型在结合情绪特征后，对开盘价的预测效果最佳。本研究证明了市场情绪特征对股票开盘价的显著影响，并为结合情绪和量价特征的预测任务提供了有力的技术支持。

关键词：情绪分析，Transformer 模型，LSTM 模型

1 研究背景及意义

1.1 研究背景

股票市场作为现代经济的重要组成部分，一直是投资者关注的焦点。然而，股票价格的波动性、非线性及其高度噪声特性，使得股价预测成为一个复杂且具有挑战性的问题。社交媒体情绪分析逐渐成为了解市场情绪并预测股价的重要工具，尤其是 Twitter，因其即时性和直接性而成为投资者表达观点的热门平台。许多重要人物和金融专家的言论常通过 Twitter 传播，并引发股票价格剧烈波动。近年来，深度学习的快速发展，尤其是在自然语言处理（NLP）领域的突破，为这一领域提供了新的研究方向。例如，利用 NLP 对社交媒体和新闻数据中的情绪信息进行分析，能够揭示投资者对市场的情绪反应，并有效预测股价变化。

1.2 研究意义

本项目的主要目标是开发一个基于 Transformer 的神经网络的 NLP 模型，通过分析 Twitter 上的情绪，结合股票量价信息使用时序模型如 LSTM 预测特定股票的走势，

并达到合理的准确性。如上所述，已有许多研究尝试利用新闻和 Twitter 情绪分析来辅助股票市场预测。尽管这些模型在一定程度上取得了准确的预测，但它们使用的是较为传统的机器学习技术，效率远不及当前学术界和工业界推崇的基于 Transformer 的方法。因此，本项目的另一目标是利用这些 Transformer 方法，构建速度更快、效率更高的模型，同时保留先前模型所达到的预测准确性。这一研究不仅能够帮助投资者更好地理解市场情绪对价格波动的影响，还能为算法交易和量化投资提供技术支持。

2 研究问题

在此背景下，本项目的具体研究问题主要包含以下几个方面：

1. 如何通过 Transformer 方法构建效率更高、准确性更高的情绪文本分析模型？

Transformer 模型在情绪文本分析中的应用具有显著优势，但选择合适的预训练模型和数据集至关重要。本研究将探讨以下问题：

- **模型选择：**应选择哪种预训练的 Transformer 模型（如 BERT、RoBERTa 或其金融领域变体）以适配金融情绪分析任务？
- **数据集选择：**应使用何种数据集（如 Twitter 金融情绪数据集）进行模型训练和微调？如何确保数据集的多样性和标注质量以提高模型泛化能力
- **数据预处理：**Twitter 情绪文本往往包含噪声，如拼写错误、表情符号和 URL 等。研究将设计清理与规范化流程，优化文本分词方法，以减少噪声对模型训练的干扰

2. 如何结合股票量价数据，通过传统机器学习模型预测股价趋势？

- **趋势特征生成：**如何从股票的历史价格、成交量等特征中生成反映趋势的信息以提高趋势分类的准确性？
- **模型选择与评价：**哪种传统机器学习模型（如支持向量机、随机森林、逻辑回归等）最适合结合这些特征进行趋势分类？如何通过交叉验证和性能指标（如准确率、F1 分数）评估模型的有效性？

3. 通过 LSTM 模型是否能较为准确地预测股票具体价格？

时间序列数据的特性使得 LSTM 成为股票价格预测的有力工具。本研究将回答以下问题：

- **LSTM 模型的原理：**长短期记忆网络（LSTM）如何通过记忆和遗忘机制有效捕捉股票价格序列中的长期依赖关系？

- **预测效果评估：**如何结合情绪特征与历史交易数据对 LSTM 模型的性能进行评估？采用哪些指标（如均方误差 RMSE 或平均绝对误差 MAE）衡量其预测准确性？

3 技术思路

本研究构建的模型包括两个部分：一个是用于推文的情绪分类器，另一个是股票预测模型。

3.1 情绪分析

3.1.1 推文情绪的三类分类

- Bullish: 买入股票
- Bearish: 卖出股票
- Neutral: 不采取行动

模型将为这三种标签输出概率值，并通过 `argmax` 方法选择每条推文的“情绪”标签。与通用的情绪分类器不同，本研究选择了专为金融领域的情绪优化的模型。

3.1.2 数据集选取

本项目用于模型训练的数据集是“zeroshot/twitter-financial-news-sentiment” Face (2025b)。该数据集由 Hugging Face 提供，包含 9938 条标注推文作为训练集是，以及 2468 条推文作为验证集。是一个包含多种与金融相关推文的英语语料库，旨在通过对推文的情绪分类，帮助研究者更好地理解金融市场中情绪对股票价格波动的影响。

3.1.3 预训练模型和分词器

本研究参考了 Palomo 的项目 Tweet Sentiment Analysis to Predict Stock MarketPalomo (2025)，该项目选用了 FinancialBERT-Sentiment-AnalysisHazourli (2025) 模型，该模型是专门为金融领域设计和优化的情绪分析模型。基于 BERT(Bidirectional Encoder Representations from Transformers) 架构，针对金融文本进行了微调，尤其擅长处理与市场动态相关的推文、新闻等内容。但考虑到 Transformer 模型的迭代和优化，本研究同时选用了 distilroberta-finetuned-financial-news-sentiment-analysisFace (2025a) 模型，该模型是一个基于 DistilRoBERTa 轻量化的情绪分析模型，同样专注于金融领域。作为 DistilRoBERTa 的轻量化版本，具有更少的参数量和更快的训练速度，这显著

降低了模型在训练和推理过程中的计算成本。尽管 distilroberta 模型较小，但在金融领域的情绪分析任务中，其表现与 FinancialBERT 相近，能够在准确性和效率之间取得良好的平衡。且两个模型用于训练的是同一个数据集 “takala/financial_phrasebank”，故本研究采用两个 Transformer 模型用于情绪文本分析的训练，最后评估出较优模型。

3.1.4 微调 (Finetuning)

为了提升模型在金融情绪分析任务中的性能，本研究对预训练模型进行了微调 (Finetuning)。通过微调，模型能够针对特定任务和领域 (如本研究即为股票市场) 的数据进行参数优化，使其更好地捕捉目标任务中的关键特性。

对于 FinancialBERT-Sentiment-Analysis 模型：

参数设置如表 1 所示：

超参数	设置	描述
学习率 (Learning Rate)	2e-5	用于控制模型的更新速度
训练批次大小 (Batch Size)	训练: 16, 验证: 16	训练和验证的批次大小
训练轮数 (Epochs)	2	设置为 2 轮，以避免过拟合
权重衰减 (Weight Decay)	0.01	设置为 0.01，用于正则化模型
优化器 (Optimizer)	Adam (beta1=0.9, beta2=0.999, eps=1e-08)	选用 Adam 优化器，包含参数设置

表 1: FinancialBERT-Sentiment-Analysis 模型的参数设置

对于 distilroberta-finetuned-financial-news-sentiment-analysis 模型：

参数设置如表 2 所示：

超参数	设置	描述
学习率 (Learning Rate)	2e-5	用于控制模型的更新速度
学习率调度器 (Scheduler)	linear	采用 linear 学习率调度器，使学习率逐步下降
训练批次大小 (Batch Size)	每设备训练: 8, 验证: 8	训练和验证的批次大小
训练轮数 (Epochs)	1	设置为 1 轮
随机种子 (Seed)	42	设置为 42
优化器 (Optimizer)	Adam (beta1=0.9, beta2=0.999, eps=1e-08)	选用 Adam 优化器，包含参数设置

表 2: distilroberta-finetuned-financial-news-sentiment-analysis 模型的参数设置

3.1.5 数据预处理

在对被分类对象推文进行分词和评估之前，为了降低数据噪声、提升模型性能，本研究进行了如标准化每日推文数量，删除缺失值和无关列，去除推文中的表情符号等预

处理步骤。

3.2 股票预测

通过使用 yfinance 库, 从 Yahoo Finance 获取与情绪特征相同时间区间的量价数据, 包括开盘价 (open)、收盘价 (close)、最高价 (High)、最低价 (Low) 和交易量 (Volume)。将这些量价数据与前期提取的情绪特征进行合并, 形成一个综合数据集。该综合数据集将作为后续用于趋势预测的传统机器学习模型和用于股票具体价格预测的 LSTM 模型的基础。

3.2.1 股票趋势预测-传统机器学习模型

为了量化股票的价格变动趋势, 本研究通过每日收盘价的变化计算趋势特征。具体而言, 若当天的收盘价较前一交易日上升, 则趋势特征值标记为 1 (上涨); 否则标记为 0 (下跌或无变化)。这一特征通过以下公式定义:

$$Trend_t = \begin{cases} 1, & \text{if } (Close_t - Close_{t-1}) > 0 \\ 0, & \text{if } (Close_t - Close_{t-1}) \leq 0 \end{cases} \quad (1)$$

为了预测股票的趋势 (上涨或下跌), 本研究采用了多种传统机器学习模型进行分类任务。这些模型包括逻辑回归 (Logistic Regression)、支持向量机 (Support Vector Machine, SVM)、决策树 (Decision Tree)、XGBoost (Boosted Tree) 以及随机森林 (Random Forest)。最后, 通过分类报告和混淆矩阵进行模型表现评估。

3.3 股票价格预测-LSTM

LSTM 是一种特殊的循环神经网络 (RNN), 通过引入记忆单元 (Cell State) 和门机制 (Gate Mechanisms), 有效解决了传统 RNN 在处理长序列时的梯度消失和梯度爆炸问题。LSTM 的核心在于其能够选择性地记住或忘记输入序列中的信息, 使其非常适合处理时间序列数据。这使得 LSTM 能够有效应对股票价格预测中的非线性、高波动性和多维特征挑战, 为复杂的金融预测任务提供了可靠的解决方案。

LSTM 的结构如图 1 所示:

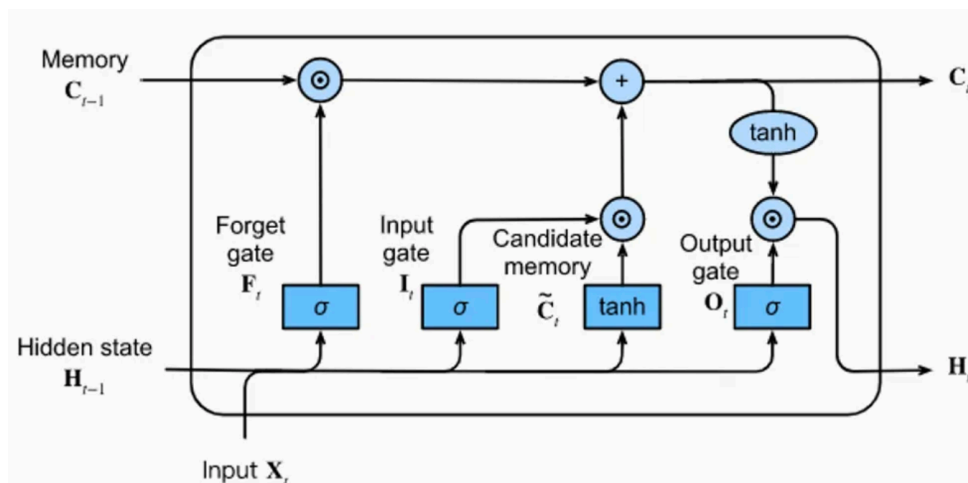


图 1: LSTM 结构 (https://d2l.ai/chapter_recurrent-modern/lstm.html)

LSTM 结构中的关键部分：

1. 遗忘门 (Forget Gate)

遗忘门控制从上一时刻的记忆中应该保留哪些信息，忘记哪些信息。其表达式为：

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \quad (2)$$

其中：

- f_t : 遗忘门的输出，值域为 $[0,1]$ ，表示遗忘程度
- $\sigma(\cdot)$: Sigmoid 激活函数，将结果压缩到 $[0, 1]$
- $W_f b_f$: 遗忘门的权重和偏置
- h_{t-1} : 上一时刻的隐藏状态
- x_t : 当前时刻的输入

2. 输入门 (Input Gate)

输入门决定当前时刻的输入信息如何影响记忆单元状态，其表达式包括两部分：

1. 输入门的激活值 (决定更新的比例)：

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \quad (3)$$

2. 候选记忆信息 (决定引入的新信息)：

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_C) \quad (4)$$

其中：

- i_t : 输入门的输出, 值域为 $[0, 1]$, 表示输入的重要程度
- $\tilde{\mathbf{C}}_t$: 候选记忆信息, 值域为 $[-1, 1]$, 表示当前时刻的新信息
- $\sigma(\cdot)$: Sigmoid 激活函数
- $\tanh(\cdot)$: 双曲正切激活函数
- $\mathbf{W}_i \mathbf{b}_i \mathbf{W}_C \mathbf{b}_C$: 输入门和候选记忆的权重与偏置

3. 记忆单元更新 (Cell state Update)

记忆单元通过结合遗忘门和输入门的结果更新状态:

$$\mathbf{C}_t = f_t \odot \mathbf{C}_{t-1} + i_t \odot \tilde{\mathbf{C}}_t \quad (5)$$

- \mathbf{C}_t : 当前时刻的记忆单元状态
- \mathbf{C}_{t-1} : 上一时刻的记忆单元状态
- f_t : 遗忘门的输出, 决定需要遗忘的记忆比例
- i_t : 输入门的输出, 决定需要引入的新记忆比例
- $\tilde{\mathbf{C}}_t$: 候选记忆信息
- \odot : 逐元素乘法 (Hadamard Product)

本研究采用了 Keras 框架的 Sequential 模型, 结合 LSTM 单元、Dropout 和全连接层 (Dense Layer) 构建神经网络。在编译阶段, 本研究使用 Adam 优化器动态调整学习率以加速收敛, 采用 mean_absolute_error(平均绝对误差) 作为损失函数。

4 实验细节

4.1 数据集处理

- **Nvidia_Tweets**

自从 Twitter 更名为 X 以来, 该平台对数据爬虫的限制愈发严格。以往常用的爬取推文的库 (如 Tweepy 和 snsrape) 在实际使用中经常会遇到接口失效或频繁被限制的问题。此外, 使用官方的 Twitter API 需要付费订阅, 才能访问相关数据服务, 这对研究人员和开发者构成了一定的经济门槛。以上限制直接影响了本研究在推特数据获取方面的灵活性和全面性。因此, 本研究选择了公开数据平台 Kaggle 上的 100K Nvidia Tweets 数据集 Tehranipour (2023)。该数据集涵盖了从 2022 年 11 月 21 日到 2023 年 2 月 6 日, 共计 78 天的时间跨度, 包含了 10 万条与英伟达股票相关的推文。尽管数据量较大, 但由于覆盖的时间区间较短, 平均每天的推文数量较多, 这可能导致情绪分析中信息密度过高, 从而引入噪声。因

此，为了避免过拟合问题，本研究以日期为单位，对每天的推文随机选取 300 至 400 条推文作为进行情绪分析的数据池，限制了每日推文的数量。

- **Nivida_stocks**

通过 yfinance 得到相应时间段 (2022 年 11 月 21 日-2023 年 2 月 6 日) 英伟达股票的量价数据，再将 Nvidia_tweets 和 Nvidia_stocks 按照交易日合并，最终得到既含有情绪特征又有量价特征的综合数据集。其中，在对 Nvidia_tweets 的情绪分类阶段，我们并不能直接得到每日情绪分数 (Daily Sentiment)。我们通过对每条推文的情绪得分 (Sentiment) 和置信度 (Confidence) 相乘，计算加权情绪贡献。再将所有推文的加权情绪贡献求和，除以置信度的总和，以得到当天的加权平均情绪得分。**计算每日情绪得分的数学表达式为：**

$$\text{Daily Sentiment} = \frac{\sum (\text{Sentiment}_i \times \text{Confidence}_i)}{\sum \text{Confidence}_i} \quad (6)$$

最后，得到时间长度为 50 天的综合数据集，如图 2 所示：

	Datetime	Open	High	Low	Close	Volume	Daily_Sentiment
0	2022-11-21	15.132120243655514	15.461795982930148	15.0651856742172	15.301953315734863	404739000	0.547059
1	2022-11-22	15.312941004415727	16.042224307977705	15.107143062322645	16.02224349975586	472866000	0.554893
2	2022-11-23	16.082185426693993	16.51076484483677	16.032235307241766	16.502771377563477	427241000	0.580595
3	2022-11-25	16.301971969557428	16.47080468398519	16.156115473704112	16.254018783569336	167934000	0.631039
4	2022-11-28	16.01025486285479	16.34093096346319	15.709551881196315	15.811450958251953	303741000	0.617977

图 2: 综合数据集

4.2 模型调整

- **用于预测趋势的传统机器学习模型**

五个模型都选用默认参数设置

- **LSTM-股价预测模型**

1. 为了统一数据尺度，使用 MinMaxScaler 对输入特征和目标值进行归一化处理，将数据缩放至 [0,1] 范围，以提升模型的训练稳定性。
2. 本研究使用滑动窗口法提取时间序列特征，将每 sequence_length 天的特征组合作为输入，目标值为第 sequence_length+1 天的开盘价。其中，因为数据集时间长度过短，即设 sequence_length 为 2
3. 基于 Keras 框架构建了一个三层神经网络：

- **LSTM 层:**
包含 20 个单元 (units=20),
return_sequences=False (表示仅输出最后一个时间步的隐藏状态)
- **Dropout 层:**
丢弃率为 0.5, 用于防止过拟合
- **Dense 层:**
输出一个单元, 用于预测目标值

模型编译使用 Adam 优化器, 并以平均绝对误差 (mean_absolute_error) 作为损失函数。训练时设置训练轮数 (epochs) 为 40, 批次大小 (batch_size) 为 8。

5 实验结果呈现与分析

5.1 情绪文本分析模型

FinancialBERT-Sentiment-Analysis 和 distilroberta-finetuned-financial-news-sentiment-analysis 的模型评估结果如表 3 所示:

Model	Accuracy	Precision	Recall	F1-Score
FinancialBERT-Sentiment-Analysis	0.8417	0.8387	0.8417	0.8396
distilroberta-finetuned-financial-news-sentiment-analysis	0.8664	0.8687	0.8664	0.8672

表 3: 情绪分析模型表现评估

由图可以看出, 用于评估的指标有准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1-Score。两个模型都优于传统的机器学习模型 (如逻辑回归、支持向量机等) 在情绪分析任务中的表现 (普遍准确率在 60 ~ 70%), 但 distilroberta-finetuned-financial-news-sentiment-analysis 在所有评估指标上均优于 FinancialBERT, 特别是在准确率和 F1-Score 上的提升尤为明显。这表明 distilroberta 模型在处理金融推文情绪分析任务时具有更强的鲁棒性和适应性。

综上, 出于性能优化的考虑, 本研究最终选择了 distilroberta-finetuned-financial-news-sentiment-analysis 作为情绪分析任务的核心模型, 为后续股票价格预测提供更加精准的情绪特征。

5.2 股票预测模型

5.2.1 趋势预测表现评估

各模型的评估结果如表 4 所示 (其中, 准确率 (Accuracy)、精确率 (Precision)、召回率 (Recall) 和 F1-Score 都是指的类别的加权平均数):

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.45	0.86	0.45	0.48
SVM	0.45	0.86	0.45	0.48
Decision Tree	0.64	0.88	0.64	0.68
XGBoost	0.64	0.88	0.64	0.68
Random Forest	0.64	0.88	0.64	0.68

表 4: 趋势预测模型表现评估

从图中可以看出, 在 5 个传统机器学习模型中, XGBoost、Decision Tree 和 Random Forest 的表现最好, 准确率有 0.64。但尽管这些模型的表现相对优于其它模型, 其整体性能仍然有限, 特别是较低的召回率和 F1-Score, 这体现出模型并没有很好的泛化能力。

模型在趋势预测任务中表现不佳的主要原因可能是: 首先, 由于推特数据获取受到限制, 本研究所能使用的最适合的数据集仅覆盖了 50 天的时间跨度。这种较短的时间覆盖期导致数据样本数量严重不足, 极大地限制了模型的学习能力, 使得模型可能无法充分捕捉类别之间的差异。其次, 数据集中不同类别的样本分布严重不平衡, 这进一步影响了模型的预测表现。分析结果显示, 各模型对类别 1 (看涨情绪) 的预测精确率都达到了 1.00, 而对类别 0 (非看涨情绪) 的预测精确率仅在 0.25 到 0.3 之间。这表明模型倾向于预测频率较高的类别 (如类别 1), 忽视了频率较低的类别 (如类别 0)。这种现象可能是由于样本分布不均导致的类别偏倚 (Class Bias), 使得模型在学习过程中无法平衡不同类别的重要性, 从而在类别 0 的预测上表现较差。

5.2.2 价格预测表现评估

由于数据集的局限性, 本研究侧重于分析加入情绪指标后的 LSTM 模型对哪个时间段的价格指标预测效果最好, 从而找到情绪特征对股票价格影响的规律。本研究采用了 4 轮 LSTM 对股票价格的预测, 分别是: 对收盘价、开盘价、最高价和最低价的预测, 预测结果如图 3-6(趋势可视化) 和表 5(量化指标) 所示:

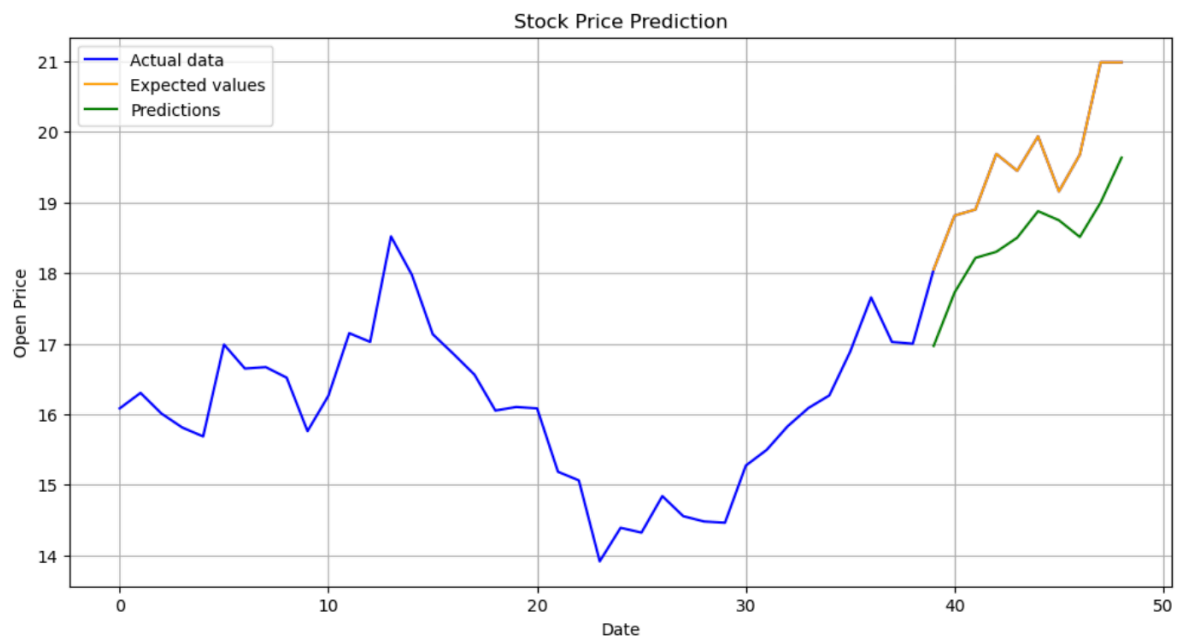


图 3: 对开盘价的预测

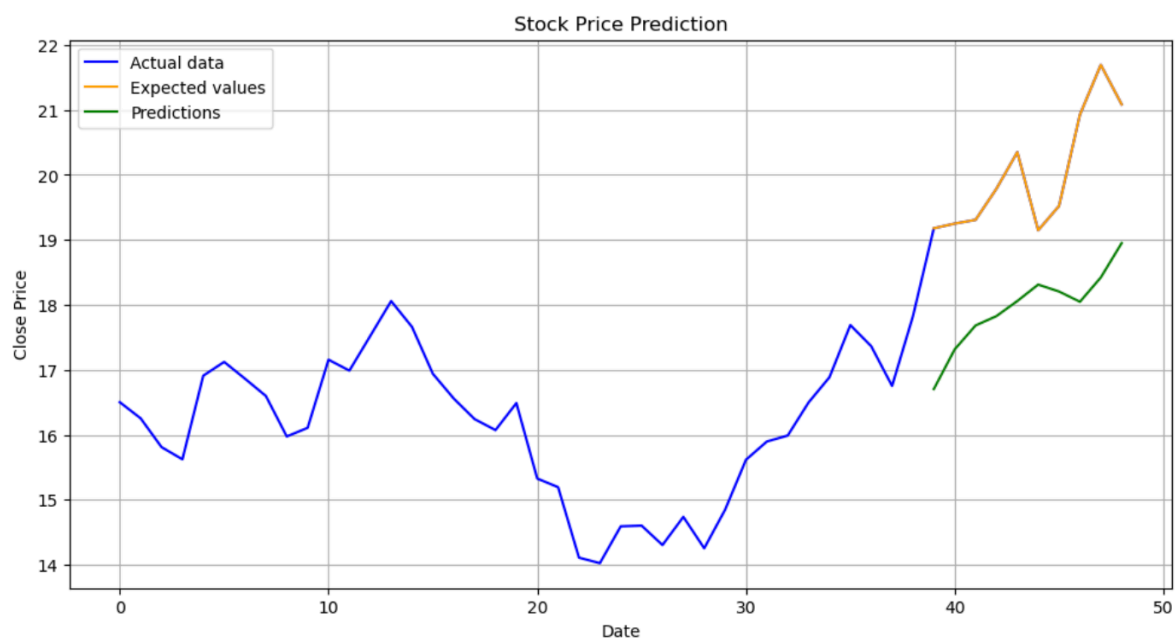


图 4: 对收盘价的预测

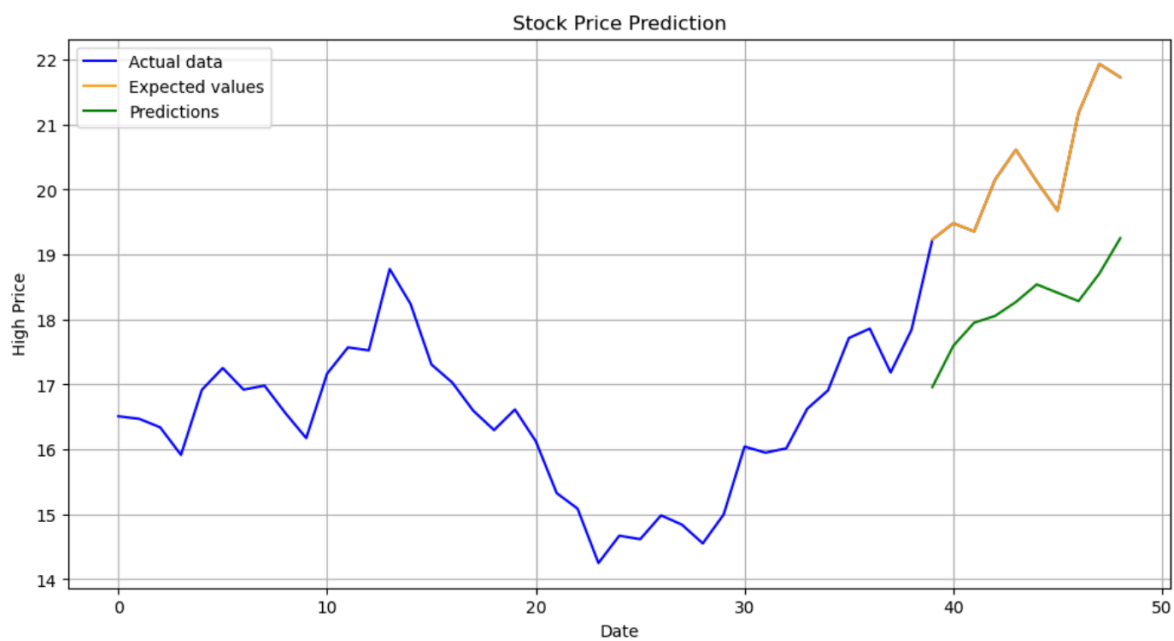


图 5: 对最高价的预测

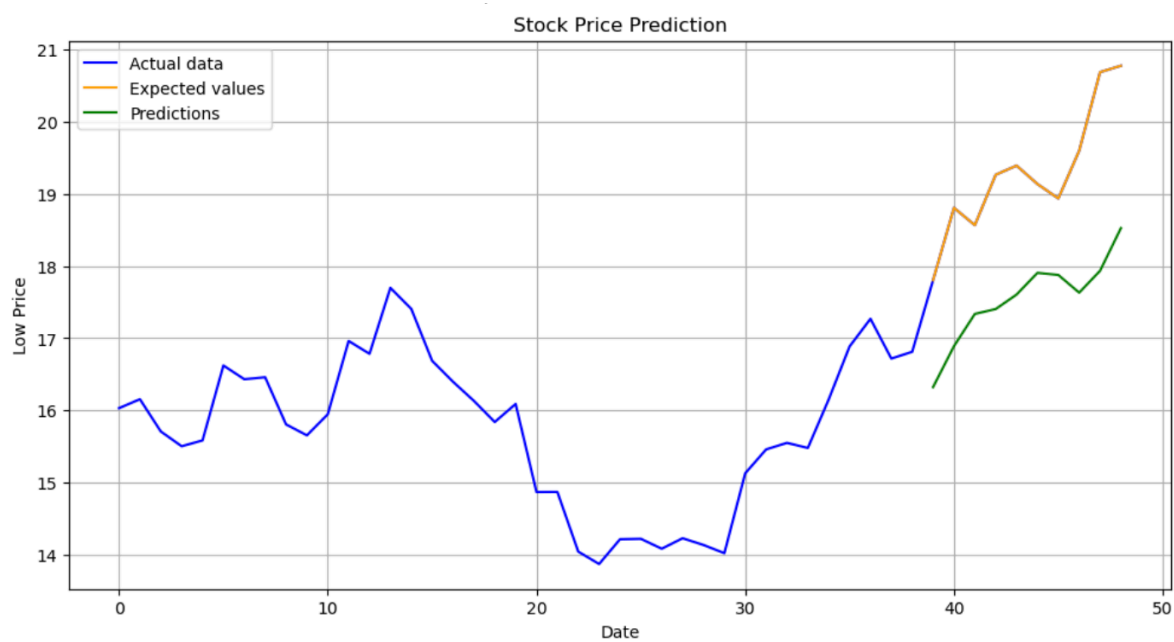


图 6: 对最低价的预测

预测对象	MAE	MSE	RMSE	R-squared
Open Price	1.1161	1.4063	1.1858	-0.8317
Close Price	2.0722	4.7576	2.1811	-5.1534
High Price	2.1446	4.9601	2.2271	-4.6981
Low Price	1.7525	3.3142	1.8205	-3.5007

表 5: 量化评估

从图 3 到图 6 可以明显地看出 LSTM 模型对于开盘价 (Open Price) 的预测效果最好。其中, 在量化评估表中, 本研究使用了四个评估指标:

- Mean Absolute Error(MAE):

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (7)$$

- Mean Squared Error(MSE):

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (8)$$

- Root Mean Squared Error(RMSE):

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (9)$$

- R-squared(R^2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

从量化指标评价表中也可以看出模型对开盘价的预测效果最好, 其原因可以从以下几个方面进行解释:

- 重大新闻的发布:

许多重大市场新闻 (例如公司公告、经济政策、国际事件等) 通常在股市收盘后发布, 此时市场的交易活动暂停, 但投资者的情绪波动已经开始显现。这些新闻直接影响投资者对股票价格的预期, 并通过社交媒体 (如 Twitter) 迅速传播。而第二天的开盘价在集合竞价阶段则是最直接反映这些情绪波动和预期时间段。

- **社交媒体的实时情绪反映**

收盘后社交媒体上的情绪信息往往更能反映投资者对新闻事件的解读，而这在很大程度上影响了第二天开盘价的涨跌方向。

- **情绪数据的分歧性：**

本研究的情绪分析模型是基于的 T-1(当天前) 的推文数据进行的预测，而随着时间的推移，各种推文和新闻的时效性会下降，所以无法捕捉情绪在当天所有交易时段的发酵和分歧。这种基于历史情绪的分析，更多地反映了市场在前一天收盘后受到新闻和情绪驱动的即时反应，而这些情绪的直接影响通常体现在第二天的开盘价上。由于模型无法实时获取当天交易时段内情绪的动态变化，模型的预测能力对开盘价的效果较好，但对全天价格波动的预测能力可能受到一定限制。

6 总结

本研究利用社交媒体情绪分析与时序模型，探讨了其在股票价格预测中的潜力。通过情绪分析模型（FinancialBERT 和 DistilRoBERTa）的微调和应用，研究发现情绪特征对开盘价的预测效果尤为显著。此外，结合股票量价特征的综合数据集进一步提升了预测性能。

1. **情绪分析模型：**实验表明，DistilRoBERTa 模型在准确率和 F1-Score 上均优于 FinancialBERT，展示了较高的情绪分类能力。
2. **趋势预测：**传统机器学习模型（如 XGBoost 和 Random Forest）尽管能够捕捉部分趋势特征，但因数据样本不足和类别分布不平衡，其泛化能力受限，预测性能较低。
3. **价格预测：**LSTM 模型在结合情绪和量价特征后，对开盘价的预测效果最佳。这主要归因于情绪信息在市场收盘后的集中发酵对开盘价的直接影响。

本研究的局限性在于数据集时间跨度较短、样本数量有限，这可能影响模型性能的全面评估。且本研究的情绪分析模型无法捕捉当日交易时段内情绪的动态变化和发酵。这种滞后性可能限制了模型在实时预测场景中的应用。未来的研究可以通过扩展数据集规模、改善数据质量，探究实时分析市场情绪的方法以及更复杂的深度学习模型（如多模态 Transformer），进一步提升情绪分析和股价预测的准确性。

参考文献

- Face, H. (2025a). distilroberta-finetuned-financial-news-sentiment-analysis. <https://huggingface.co/mrm8488/>

distilroberta-finetuned-financial-news-sentiment-analysis. Accessed: 2025-01-05.

Face, H. (2025b). Twitter financial news sentiment dataset. <https://huggingface.co/datasets/zeroshot/twitter-financial-news-sentiment>. Accessed: 2025-01-05.

Hazourli, A. R. (2025). Financialbert-sentiment-analysis. <https://huggingface.co/ahmedrachid/FinancialBERT-Sentiment-Analysis>. Accessed: 2025-01-05.

Palomo, C. (2025). Tweet sentiment analysis to predict stock market. <https://stanford.edu>. Stanford CS224N Custom Project.

Tehranipour, S. (2023). 100k nvidia tweets. <https://www.kaggle.com/datasets/soheiltehranipour/100k-nvidia-tweets>. Accessed: 2025-01-05.