Tweets to Trades

### **Tweet Sentiment Analysis to Predict the Stock Market**

**Central Problem or Gap**

The project aims to address the gap left by previous news-based stock market predictors, which primarily focused on **statistical methods**, neglecting the potential of **Machine Learning (ML)** and **Natural Language Processing (NLP)** techniques. The central problem is the development of an NLP model capable of predicting the volatile stock market of specific stocks by analyzing **Twitter sentiment** with sufficient accuracy. The application of sentiment classification via NLP in the finance domain is a relatively new pursuit.

**Core Methodology**

The core methodology involves developing and fine-tuning the **Sentiment Analysis component** for stock prediction. A **transformer-based neural network** is employed to analyze Twitter sentiment and predict stock movements. For predicting the direction of a stock price, the key methodological feature is a **weighted-sentiment approach**. This method utilizes the number of **retweets** to adjust and emphasize the sentiment score, contrasting with a simpler approach of equally weighted sentiments. Historical financial data, specifically the adjusted closing price, is integrated into the prediction from sources like Yahoo Finance. The project compares the performance of the developed model against non-ML baselines.

**Challenges**

The inherent challenge for any stock prediction model is the market's nature: **volatile, unpredictable, and fast-moving**. The use of social media data, particularly Twitter, introduces the challenge of noise and varied influence. The proposed solution attempts to mitigate this by implementing a **weighted-sentiment approach**, recognizing that highly retweeted (and thus, more widely perceived) sentiment is likely more influential than unamplified opinions.

**Primary Contribution or Novelty**

The primary contribution is the successful development and demonstration of a **transformer-based NLP model** that uses **Twitter sentiment** to achieve reasonable stock market prediction accuracy. The principal novelty is the finding that a **weighted-sentiment approach**, specifically one where sentiment is weighted based on the number of **retweets**, yields better performance in predicting stock price direction than models using equally weighted sentiments. This suggests that incorporating a measure of public engagement or influence (like retweets) is a critical factor when leveraging social media for financial market forecasting.

CausalStock

### **CausalStock: Deep End-to-end Causal Discovery for News-driven Stock Movement Prediction**

**Central Problem or Gap**

The paper addresses two critical, unsolved issues in news-driven multi-stock movement prediction. First, existing models for stock relation discovery rely on capturing **correlation relations** (via attention or graphs), despite company relationships often being **unidirectional** (e.g., supply chain). This necessitates a shift to modeling **causal relations** to accurately depict the directional impact between stocks. Second, news data contains **substantial noise** which traditional text encoders (like GRU or LSTM) struggle to manage, making it difficult to extract genuinely effective predictive information.

**Core Methodology**

The paper proposes a novel framework named **CausalStock**, which performs deep end-to-end causal discovery for news-driven multi-stock movement prediction. The methodology integrates two core components:

1. **Temporal Causal Discovery:** A **lag-dependent temporal causal discovery mechanism** is used to model the distribution of the temporal causal graph between stocks. This mechanism defines the stock relationships as a **directed acyclic graph** and employs a **Functional Causal Model (FCM)** to encapsulate and optimize the discovered causal relations.
2. **Denoised News Encoder:** This component leverages the advanced knowledge, reasoning abilities, and text evaluation capabilities of **Large Language Models (LLMs)** to handle noise in the news data. The LLM scores each news text from multiple perspectives to extract useful and meaningful signals for prediction.

**Challenges**

The model explicitly tackles the inherent difficulties of correlation-based models failing to capture the **unidirectional nature of stock impact** and the pervasive problem of **noise** in financial news sources. A potential area for future refinement is the need for more complex distributions within the causal discovery mechanism to fully capture the intricate nature of causal relationships.

**Primary Contribution or Novelty**

The primary contribution is the **CausalStock framework** itself, which pioneers the integration of **temporal causal discovery** into multi-stock prediction, moving beyond traditional correlation-based methodologies. The significant novelty is the **LLM-based Denoised News Encoder**, which leverages the superior text evaluation capabilities of LLMs to effectively denoise news data and assimilate external knowledge, resolving a key limitation of prior text mining approaches. Furthermore, the resulting model offers a clear prediction mechanism with inherent **explainability**, a benefit derived directly from the discovered causal relations. The approach demonstrated superior performance against strong baselines across six real-world datasets spanning multiple global markets.

GNN + SA for stocks

### **Integrating sentiment analysis with graph neural networks for enhanced stock prediction: A comprehensive survey**

**Central Problem or Gap**

The paper identifies the need for a **systematic review** and **comprehensive roadmap** on the integration of **sentiment analysis** with **Graph Neural Networks (GNNs)** for stock prediction. Traditional stock prediction approaches, such as fundamental analysis and technical indicators, often fail to account for the crucial impact of **market sentiment** and collective investor behavior, which significantly influence price movements. The existing literature on combining the relational modeling power of GNNs with sentiment data required synthesis to advance the field.

**Core Methodology**

The methodology is a **comprehensive literature review** of the application of GNNs in conjunction with sentiment analysis for stock prediction. It first outlines the fundamental concepts and limitations of conventional methods. The study then systematically surveys diverse approaches, techniques, and methodologies employed in this interdisciplinary domain. Key elements examined include: various **graph structures** utilized (e.g., stock networks, investor networks), methods for incorporating sentiment information as node features, different sources of sentiment data (news articles, social media feeds, financial reports), and the evaluation metrics used to assess model precision.

**Challenges**

The survey highlights practical and theoretical challenges prevalent in this research area. Practical challenges center on the difficulties associated with **data collection, preprocessing, and annotation** of sentiment data from diverse sources. Theoretically, the article discusses the difficulty in **modeling evolving relationships** and potential limitations in the **generalization** capabilities of integrated sentiment-GNN models. A significant, outstanding challenge identified is the need to enhance the **interpretability** of GNN-based models, as their "black-box" nature makes them unsuitable for sensitive domains like finance and healthcare.

**Primary Contribution or Novelty**

The primary contribution is the provision of a **comprehensive roadmap** for utilizing the combined strengths of GNNs and sentiment analysis to improve stock prediction accuracy. The novelty resides in the structured synthesis of knowledge, which articulates the inherent advantages of uniting the relational reasoning capabilities of GNNs with the sentiment capturing power of text analysis, leading to the development of more robust models. The resulting work serves as a valuable resource for guiding future investigations by clearly identifying the field's current limitations and unanswered research questions.

### 

LSTM Model on DSE

### **Stock market prediction of Bangladesh using multivariate long short-term memory with sentiment identification**

**Central Problem or Gap**

Stock market prediction is a volatile and challenging task, especially in **developing nations like Bangladesh** (specifically the **Dhaka Stock Exchange or DSE**). The existing gap is the difficulty of achieving accurate predictions because the DSE is heavily influenced by **multiple external factors** (macroeconomic and sentiment-driven) that are typically neglected by models focused solely on technical indicators. The absence of a holistic dataset incorporating all these factors makes prediction suboptimal.

**Core Methodology**

The study proposes a prediction model for the DSE trend based on a **multivariate Long Short-Term Memory (LSTM) neural network**. The methodology's central feature is the creation of a **novel dataset** by integrating three distinct data streams for the period 2014-2021:

1. **Technical stock market data** from the DSE.
2. **External economic factors** include inflation, Gross Domestic Product (GDP), exchange rate, interest rate, and current balance.
3. **News sentiment** derived from financial news articles. To identify sentiment, I modified the **BTSC algorithm** to process **financial Bangla news**. The study then compares the prediction accuracy of the multivariate LSTM model using both technical factors alone *versus* the combined set of technical and external factors. The objective is to forecast the next day's opening price.

**Challenges**

The primary challenge is the increased difficulty of predicting market trends in a developing economy due to the numerous and influential external factors that are difficult to identify. A linguistic challenge was overcome by adapting a specific sentiment identification algorithm (**BTSC**) to the **Bangla language**, necessary for processing local financial news.

**Primary Contribution or Novelty**

The primary contribution is the development and validation of a **multivariate LSTM model** for the DSE that proves the necessity of a holistic data approach. The core novelty is the **novel dataset** created, which is specific to the Bangladeshi market and comprehensively merges technical, macroeconomic, and local news sentiment data. The quantitative result provides the key contribution: the inclusion of external factors **improved the accuracy of the LSTM-based predictions by approximately 24%**. This decisively supports the argument that external and sentiment-based information is critical for accurate stock market forecasting in emerging economies.

Temporal CNN

### **A Dual-Output Temporal Convolutional Network With Attention Architecture for Stock Price Prediction and Risk Assessment**

**Central Problem or Gap**

The paper addresses the challenge of simultaneously performing two essential financial tasks: accurately predicting stock prices and effectively assessing risk. Existing models, such as ARIMA and LSTM, struggle to capture the inherent **nonlinearity, noise, and especially long-term dependencies** present in financial time series data, resulting in suboptimal performance for both forecasting and evaluating critical risk metrics.

**Core Methodology**

The methodology introduces a novel deep learning architecture: a **Temporal Convolutional Network (TCN) with an Attention mechanism** structured as a **Dual Output** model.

1. **TCN for Dependencies:** The TCN component utilizes dilated convolutions to efficiently capture both short-term and long-term dependencies in the stock price sequence, thereby addressing the inefficiencies and dependency issues inherent in recurrent architectures like LSTM.
2. **Attention Mechanism:** The attention layer is integrated to enhance predictive accuracy by **selectively focusing on critical time steps** in the sequence, thereby helping to filter out the noise and nonlinearity inherent in financial data.
3. **Dual Output Design:** The model is explicitly configured with a dual output to concurrently forecast future stock prices (Open, Close, High, Low) *and* predict risk metrics, specifically **volatility and the Sharpe Ratio**. The model was validated on over 15 years of data (2008–2024) for MasterCard (MA) and Visa (V), demonstrating outperformance over LSTM and ARIMA baselines.

**Challenges**

The core challenge is rooted in the complex nature of financial time series data—its noise, non-stationarity, and critical long-range dependencies. The TCN-Attention architecture is specifically proposed to overcome the limitations of recurrent models. A current limitation noted by the authors is the model's reliance on the assumption that **past stock prices and risk metrics contain all predictive information**, suggesting that future work should explore the integration of external data sources.

**Primary Contribution or Novelty**

The primary contribution is the **Dual Output Temporal Convolutional Network with Attention Architecture**. The principal novelty is the **unified, dual-output structure**, which efficiently provides investors and risk managers with both **future price forecasts** and **simultaneous, interpretable risk assessments** (volatility and Sharpe Ratio) from a single model. By utilizing the TCN, It overcomes the historical limitations of modeling long-term dependencies in financial data, offering a robust and high-performing alternative to recurrent and statistical methods