MULTI-MODAL STOCK INFORMATION RETRIVAL AND PREDICTION

Group 28

CONTENTS

Introduction to Stock Information Analysis

Significance of Automated Sentiment Analysis

- Analyzing financial texts, such as news articles and analyst reports, is crucial for making informed investment decisions.
- Automated sentiment analysis using NLP methods has gained popularity due to the vast volume of financial content produced daily.
- At the same time mathematical and machine learning models based on price data have been used for a long time but use of newer sequence modelling methods is now emerging.

Challenges in Financial Analysis

- Financial sentiment analysis specifically poses significant challenges due to the specialized terminology and the scarcity of labeled data specific to this sector.
- Standard models fall short due to the niche language employed in financial contexts.
- Traditional models of price analysis often utilize non-sequential approaches, treating each data point as independent of others.
- However, this method overlooks the inherent sequential nature of finan-
- cial data, where each price point can be significantly influenced by preceding events.
- This sequential dependency is pivotal for accurately predicting future market behaviors, as patterns tend to develop over time
- Reflecting the cumulative effects of trading behaviors and external factors on price movements

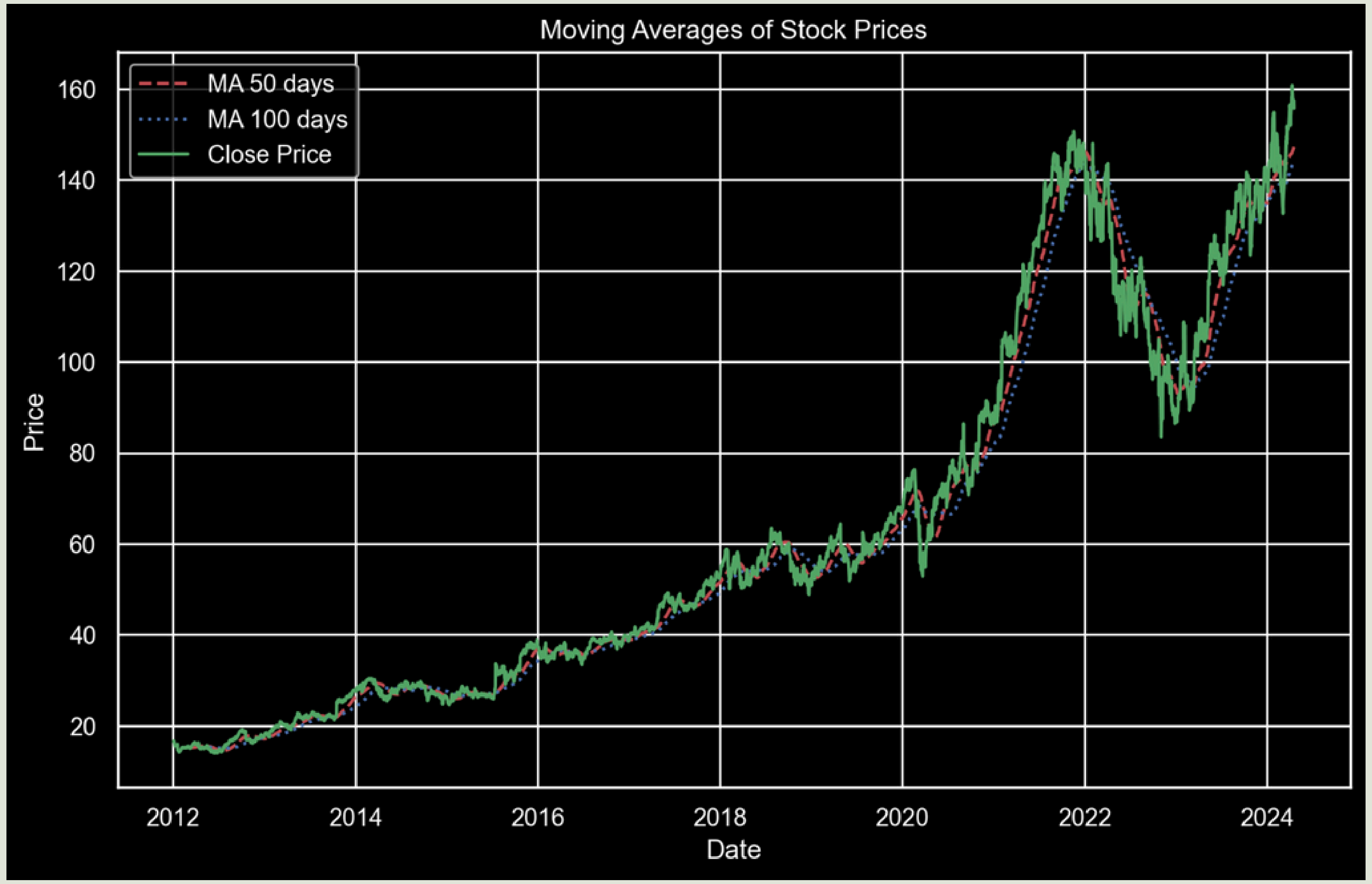
Novelty

- NLP transfer learning techniques provide a robust solution to the limitations of labeled datasets in financial sentiment analysis.
- ProMod, a language model derived from BERT, demonstrates superior performance in financial sentiment analysis.
- LSTMs are designed to process data sequences by maintaining a memory of past data points, which enables them to learn and recognize the temporal patterns that are critical in financial time series.
- This capability makes them highly effective for financial applications where past price actions are predictors of future activities, such as in trend following or mean reversion strategies.

Methodology and Model Development

Training of Predictive Models

- Two models were trained for predicting market sentiment and price trends to enhance market understanding.
- The use of moving averages and LSTM models was discussed for price trend prediction due to their capability to capture long-term dependencies.



ProMod Architecture and Training

- ProMod, a financial-specific BERT model, underwent pre-training using diverse financial datasets.
- The training process involves an initial unsupervised phase, where ProMod, our financial-specific BERT model, is trained on a large corpus of financial texts without requiring labeled data.
- This step helps the model grasp the nuanced language of finance.
- Subsequently, supervised training refines ProMod's abilities, focusing on specific sentiment analysis tasks within the financial domain.
- Finally, to adapt to the evolving financial language, we continuously expand ProMod's vocabulary. New tokens that emerge in financial communications are added.

Data Collection and Code

- Data was collected from Yahoo Finance for training and validation of predictive models.
- Financial Phrase Bank: Comprises 4,840 sentences selected from financial news and meticulously labeled by 16 experts in financial markets.
- AnalystTone dataset: Contains 10,000 randomly selected phrases from analyst reports in the Investext database, with anno- tations for positive, negative, and neutral sentiments.
- Earning Calls, SEC Forms: Also included to enrich the model's learning and adaptability to real-world financial documents.
- A webpage application was constructed for real-time stock data retrieval and sentiment analysis report generation.
- The application provides actionable insights for investors based on predicted price trends and market sentiment.

Performance Evaluation and Future Implications

3.1: Superior Performance of ProMoD

- ProMod demonstrated superior accuracy and weighted F1 score in sentiment classification compared to existing models.
- The integration of ProMod with market price data shows potential for users to make better investment decisions.

Dataset	BERT cased	ProMod cased	ProMod+vocab
PhraseBank	0.64	0.74	0.83
AnalystTone	0.77	0.74	0.81

Table 1: Performance comparison of BERT and ProMod on various datasets

```
classification report
              precision    recall  f1-score   support

           0       0.82      0.75      0.79       156
           1       0.81      0.90      0.85       222
           2       0.88      0.83      0.85       172

    accuracy                           0.83       550
   macro avg       0.84      0.83      0.83       550
weighted avg       0.84      0.83      0.83       550
```

3.2: Implications and Future Directions

- The study's findings have implications for enhancing stock market analysis through automated sentiment analysis.
- The continuous adaptation of ProMod's vocabulary ensures its relevance in processing and understanding new financial terms.
- Future research directions include further refinement of NLP transfer learning techniques for stock information analysis. - The portrayal of ProMod's performance highlights the potential for advancements in stock market prediction and sentiment analysis.