**Predicting Stock Market Trends based on Historical Data and News Sentiment Analysis (Enhancing Stock Market Predictions Through Historical Data Including Both Prices and Patterns, Multilingual News Sentiment Analysis along with Emotion Detection for Misinformation and Financial Indicators Status Using Limited Labeled Data)**

**Content of the paper**

**Chapters**

**Introduction**

* 1.1 Context of the Financial Market and Importance of Prediction
* 1.2 Objectives of the Paper
* 1.3 Contribution of the Paper to Sentiment Analysis,Financial Data and Fake News Detection
* 1.4 Structure

**Literature Review**

* 2.1 Traditional Methods for Market Prediction
* 2.2 Sentiment Analysis in Context of Stock Market Prediction
  + 2.2.1 Sentiment from News and Its Impact on Markets (Wang & Ma, 2023; Samuels & McGonigal, 2023)
  + 2.2.2 Emotion Detection for Misinformation (Liu, 2023)
* 2.3 Challenges of Using Limited Labeled Data and Proposed Solutions
* 2.4 Recent Studies on Integrating Sentiment into Prediction Models (Talazadeh and Perakovic, 2023)

**Methodology**

* 3.1 Data Collwction
  + 3.1.1 Historical Price Data and Patterns
  + 3.1.2 Sentiment from Multilingual News and Emotional Analysis
  + 3.1.3 Relevant Financial Indicators
* 3.2 Data Processing and Analysis
  + 3.2.1 Data Preprocessing Techniques
  + 3.2.2 Sentiment and Emotipn Analysis Methods
* 3.3 Prediction Models Used
  + 3.3.1 Machine Learning Models and Data Integration
  + 3.3.2 Use of Optimization Techniques (Sha, 2024)

**Experiments and Results**

* 4.1 Experimental Setup and Evaluation Methodology
* 4.2 Model Evaluation Based on Historical Data and Sentiment
* 4.3 Comparison of Results with Previous Studies
* 4.4 Impact of Using Limited Labeled Data on Model

**Discussion**

* 5.1 Interpretation of Results for Investors
* 5.2 Limitations of the Study
* 5.3 Future Research Directions

**Conclusions**

* 6.1 Summary of Key Findings
* 6.2 Contributions to the Field of Financial Market Prediction
* 6.3 Practical and Theoretical Implications of Results

**Appendices**

* 8.1 Code and Algorithms Used
* 8.2 Additional Data and Detailed Results

**Refferences**

* 7.1 Refferenced Articles

**Plan for the Application Section of the Paper:**

Hypothesis: Integrating historical stock price data with multilingual sentiment analysis and emotion detection based on some news, also integrating the misinformation (fake news) detection enhances the accuracy of stock market trend predictions, potentially even with limited labeled data or at some point unlabeled data

List of funcitonalities:

**Stock market trend prediction**: Utilises hystorical price data and patterns to forecast future market trends

**News sentiment analysis**: Processes and analyzes news articles from multiple sources and languages to extract general sentiment (positive, negative, neutral).

**Emotion detection for misinformation**: Identifies emotions like fear or excitement in news to signal potential misinformation or exaggerated news.

**Financial indicators status**: Correlates predictions with financial indicators such as GDP, interest rates, and unemployment data

**Trading signals generation**: Issues "buy" or "sell" signals based on combined analyses of historical data and current sentiment and maybe a graph with the future trend for a specific stock price

**Multilingual support**: Analyzes news articles in multiple languages for global market coverage.

**Access to historical data and forecasts**: Offers access to historical datasets and short to mid-term forecasts for various financial markets.

Methodology

1. **Data Collection:**

-Gather historical stock price data from financial databases

-Collect diverse multilingual news articles related to financial markets from various sources;

-Extract relevant financial indicators (volatility, trading volume);

-Use historical patterns

1. **Data Preprocessing / Processing and Analysis:**

-Data Preprocessing Techniques: Clean and normalize historical price data; preprocess text data for easier sentiment analysis (tokenization)

-Sentiment and Emotion Analysis Methods: Implement sentiment analysis algorithms to assess news articles and detect emotions (for instance positive, negative, neutral etc.)

1. **Prediction Models Used:**

-Machine Learning Models: Explore various models, such as Random Forest, LSTM (Long Short-Term Memory), and XGBoost (implementation of gradient boosted decision trees – following patterns), to predict stock movements based on the processed data;

-Integration of Sentiment Analysis: Incorporate sentiment scores and emotional indicators as features in the models;

-Integration of misinformation detection (detecting exaggerated conclusions using developed means or standard devations)

**4. Description of the Original Approach**

-The approach combines advanced sentiment analysis techniques with traditional financial indicators, creating a general prediction model that combines both quantitative and qualitative data.

-Emphasizing multilingual sentiment analysis addresses the global nature of financial markets and provides a broader perspective on market sentiment

**5. Experimental Setup**

-Define evaluation metrics (accuracy, precision, recall) to assess model performance.

-Conduct experiments comparing traditional prediction models against those integrated with sentiment analysis and emotion detection.

-Perform a long-term experiment on a specific stock price and check the accuracy of the prediction to prove the efficiency of the general prediciotn model and the usage of the app to the investors

**Unique Aspects Compared to Existing Applications**

My research paper implies an Enhanced Multilingual Capability while most of the applications focus on a specific language sentiment analysis and especially on the prediction functionality and accuracy, but this research integrates multilingual news sentiment, offering a more complex understanding of global market influences.

It also supposes Emotion Detection Integration, so the inclusion of emotion detection adds depth to sentiment analysis, potentially improving the model's ability to respond to misinformation and emotional market reactions.

The uniquenss of my research paper comes from combining all the methods and funcitonalities together to get something as complex and useful as possible for more people, such that predicting stock prices is done more accurat than using only one method at a time. A section of the app includes Emotion Detection for misinformation too, which is critical in predicting stock prices, since misinformation spreads faster than ever before daily.

**Contribution to Research in the Field**

* This research contributes to the understanding of how sentiment analysis can be effectively integrated into stock market predictions, particularly in a multilingual context.
* It addresses critical research questions about the interaction between news sentiment, emotional responses, and stock market behavior, filling gaps in current literature.
* By demonstrating the efficiency of combining historical data with advanced sentiment analysis techniques for understanding the sentiment of news, the findings may encourage further exploration of interdisciplinary approaches in financial market prediction.

**References:**

1. <https://arxiv.org/abs/2409.05698>

Wang, M., & Ma, T. (2023). MANA-Net: Mitigating aggregated sentiment homogenization with news weighting for enhanced market prediction. *Proceedings of the 2023 International Conference on Artificial Intelligence and Data (ICAIBD)*

1. [[2007.02238] News Sentiment Analysis (arxiv.org)](https://arxiv.org/abs/2007.02238)

Samuels, A., & McGonigal, J. (2023). News sentiment analysis. *Proceedings of the 2023 International Conference on Data Science and Advanced Analytics (DSAA)*

1. [[2410.07143] SARF: Enhancing Stock Market Prediction with Sentiment-Augmented Random Forest (arxiv.org)](https://arxiv.org/abs/2410.07143)

Talazadeh, S., & Peraković, D. (2023). SARF: Enhancing stock market prediction with sentiment-augmented random forest. *Proceedings of the 2023 IEEE International Conference on Data Mining Workshops (ICDMW)*

1. <https://arxiv.org/abs/2206.09591>

Li, T., Chen, X., Dong, Z., Yu, W., Yan, Y., Keutzer, K., & Zhang, S. (2023). Domain-adaptive text classification with structured knowledge from unlabeled data. *Proceedings of the 2023 IEEE International Conference on Data Mining (ICDM)*

1. <https://arxiv.org/abs/2410.00024>

Arshad, S., Azhar, N., Sajid, S., Latif, S., & Latif, R. (2023). Cross-lingual news event correlation for stock market trend prediction. *Proceedings of the 2023 International Conference on Computer and Communication Systems (ICCCS)*

1. [*https://arxiv.org/abs/2312.04715*](https://arxiv.org/abs/2312.04715)

*Kocoń, J. (2023). Deep emotions across languages: A novel approach for sentiment propagation in multilingual WordNets. Proceedings of the ICDM Workshop: SENTIRE 2023*

1. [*https://arxiv.org/abs/2405.14535*](https://arxiv.org/abs/2405.14535)

*Mousi, B., Durrani, N., Dalvi, F., Hawasly, M., & Abdelali, A. (2024). Exploring alignment in shared cross-lingual spaces. Proceedings of the ACL 2024*

1. [*https://arxiv.org/abs/2311.14727*](https://arxiv.org/abs/2311.14727)

*Masson, M., Agerri, R., Sallaberry, C., Bessagnet, M.-N., Lacayrelle, A. L. P., & Roose, P. (2023). Optimal strategies to perform multilingual analysis of social content for a novel dataset in the tourism domain. arXiv:2311.14727 [cs.CL]*

1. <https://arxiv.org/abs/2311.00671>

*Liu, Z., Zhang, T., Yang, K., Thompson, P., Yu, Z., & Ananiadou, S. (2023). Emotion detection for misinformation: A review*

1. <https://arxiv.org/abs/2410.06935>

*Hafid, A., Rahouti, M., Kong, L., Ebrahim, M., & Serhani, M. A. (2024). Predicting Bitcoin market trends with enhanced technical indicator integration and classification models*

1. [*https://arxiv.org/abs/2405.03151*](https://arxiv.org/abs/2405.03151)

*Sha, X. (2024). Time series stock price forecasting based on genetic algorithm (GA)-long short-term memory network (LSTM) optimization*

1. [*https://arxiv.org/abs/2403.14063*](https://arxiv.org/abs/2403.14063)

*Daiya, D., Yadav, M., & Rao, H. S. (2024). DiffSTOCK: Probabilistic relational stock market predictions using diffusion models*

1. [*https://arxiv.org/abs/2307.05719*](https://arxiv.org/abs/2307.05719)

*Sakowski, P., Sieradzki, R., & Ślepaczuk, R. (2023). Systemic risk indicator based on implied and realized volatility*

1. [*https://arxiv.org/abs/2306.03763*](https://arxiv.org/abs/2306.03763)

*Chen, Z., Zheng, L. N., Lu, C., Yuan, J., & Zhu, D. (2023). ChatGPT informed graph neural network for stock movement prediction*

1. [*https://arxiv.org/abs/2410.12807*](https://arxiv.org/abs/2410.12807)

*Chakraborty, A., & Basu, A. (2024). A hierarchical conv-LSTM and LLM integrated model for holistic stock forecasting.*