UNIVERSITI TEKNOLOGI MALAYSIA

PREDICTING STOCK MARKET TRENDS USING MULTI-SOURCE SENTIMENT ANALYSIS AND ADVANCED DEEP LEARNINGALGORITHMS

PREPARED BY : RAIAN HAFIZ NILOY
PREPARED FOR: PROF. MADYA.TS.DR.MOHD SHAHIZAN BIN OTHMAN

Innovating Solutions

OUTLINES

- Introduction
- Literature Review
- Methodology
- Initial Findings
- Discussion and Future Works
- References

INTRODUCTION

- Accurate prediction of stock market movement is very crucial for decreasing investment risks and increase profit
- Stock prices are affected by economic, political, social and psychological factors
- Traditional models rely on historical data but ignore sentiment-driven fluctuations
- Sentiment analysis from news articles and social media posts can help to predict market movements

Problem Statement

- Traditional stock prediction models rely solely on numerical data, missing sentiment impact.
- Some Machine Learning models are computationally efficient but often cannot handle non-linear high-dimensional multi-featured data efficiently
- Some advanced Deep Learning models are expert in handling complex data but show less accuracy
- So, there is always a dilemma choosing between models

Research Objectives

- To enhance the comprehensiveness of prediction by collecting and preprocessing sentiment data
- To implement FinBERT for detailed sentiment analysis and develop LSTM networks for improving prediction accuracy
- To rigorously assess the model's performance with a number of evaluation metrics such as Accuracy, Precision, Recall, F1 score etc

Literature Review

Traditional Stock Market Prediction:

•Early models: ARIMA (Box & Jenkins, 1970)
•GARCH (Engle, 1982)
•Linear Regression (Fama & French, 1992)

•Limitations:
•Cannot capture sentiment,
•poor adaptability to financial market fluctuations.

Machine Learning Approaches:

(Huang et al., 2005; Chen et al., 2013).
•Decision Trees,
•Random Forests, and
•SVMs

•These models work well with structured data
•fail with textual sentiment.

Literature Review

Sentiment Analysis & Financial Markets:

•Sentiment plays a crucial role in stock price movement
(Loughran & McDonald, 2011)

•Studies show financial sentiment can predict stock market trends
(Bollen et al., 2011).

Deep Learning & NLP Models:

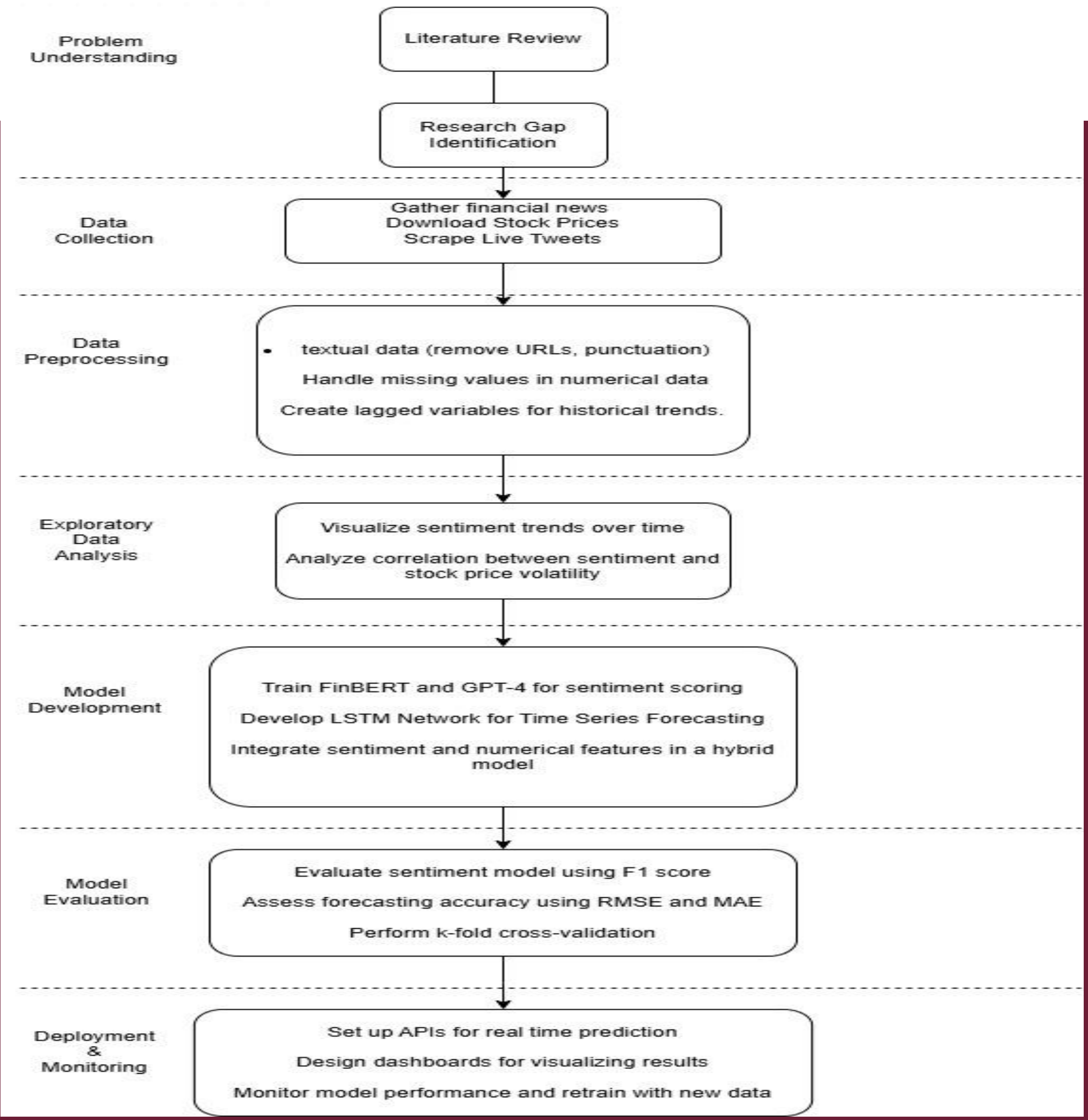
•**FinBERT** outperforms traditional sentiment models
(Araci, 2019).

•**GPT-4** can extract complex sentiment cues but has computational constraints
(Brown et al., 2020).

•**LSTM** effectively captures sequential dependencies in stock trends
(Fischer & Krauss, 2018).

Research Methodology

- **Step-by-step approach followed in this study:**
  🔷 **Step 1:** Data Collection (**Stock Prices + Financial News**)
  🔷 **Step 2:** Data Preprocessing (**Cleaning, Feature Engineering**)
  🔷 **Step 3:** Sentiment Analysis (**FinBERT**)
  🔷 **Step 4:** Machine Learning Model
  🔷 **Step 5:** Evaluation & Results



Research Methodology

Data Collection

- Stock Market Data: 15 NYSE-listed companies stock price for a period of 20 years
- Sentiment Data: Financial news headlines from CNBC

- Challenges Faced
  Original plan was to collect both data for 20 years, but sentiment data spanning 20 years was not found. So sentiment datset limited to 2018-2020

Research Methodology

Stock Data Preprocessing

- Handling missing values, outliers
- Creating technical indicators & lag features

Sentiment Data Preprocessing

- Text Cleaning: Removing special characters, stopwords
- Tokenization and Lemmatization

Initial Findings(EDA)

Feature Engineering

1. Stock Based Features

- Lagged Prices – Previous stock prices as inputs
- Technical Indicators – Moving averages, volatility, momentum indicators

2. Sentiment-Based Features

- Aggregated Sentiment Scores over rolling time windows
- Sentiment Trends - Capturing fluctuations in market mood

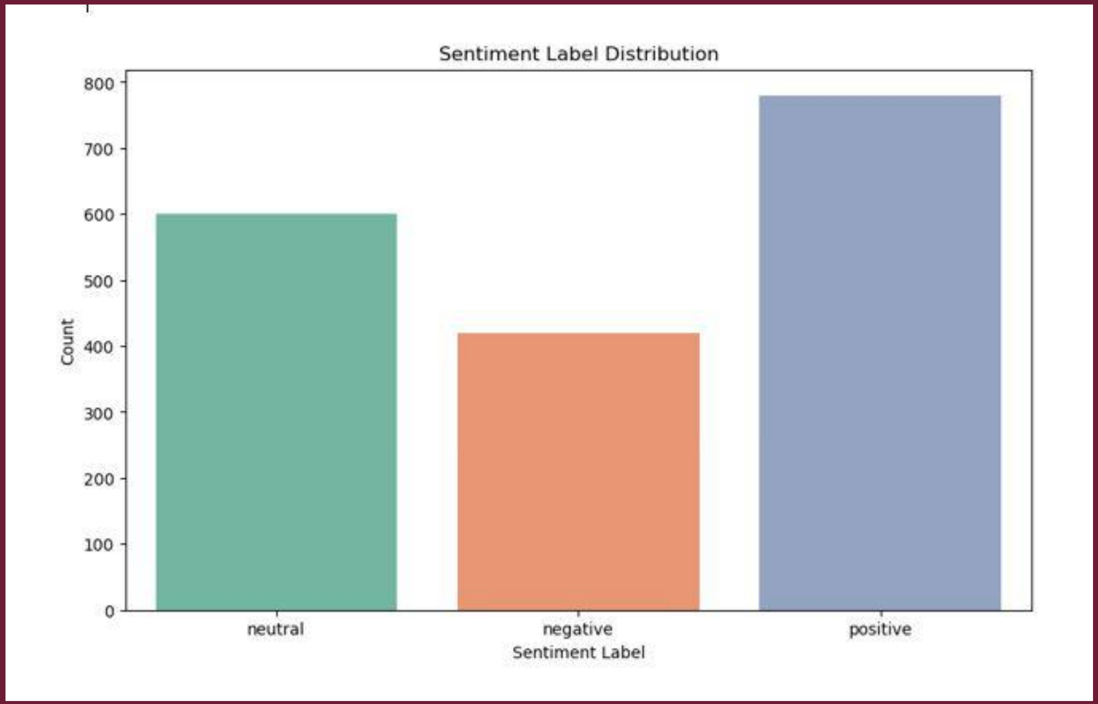
Initial Findings(EDA)

Sentiment Classification Using FinBERT

- FinBERT assigns **positive, neutral, or negative** labels to headlines
- Helps understand **market sentiment trends**

Why FinBERT?

- Pretrained on **financial texts**, outperforming general sentiment models.
- **Understands market-specific language** better than traditional NLP.



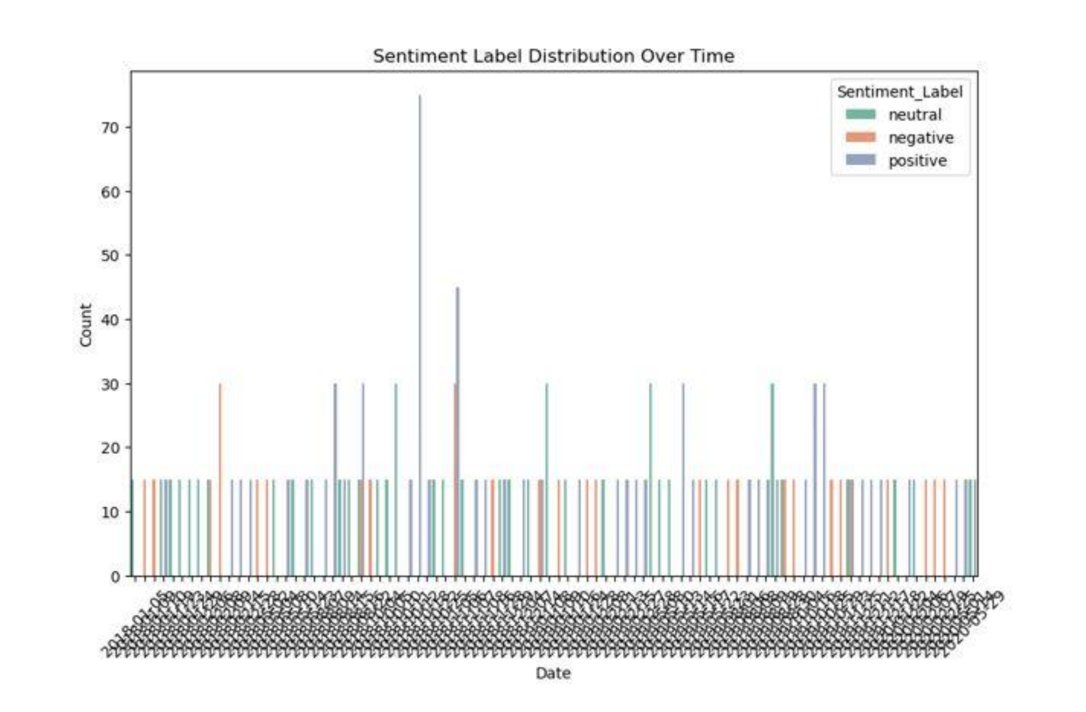
Initial Findings(EDA)

Key Observations

- Negative sentiment is more frequent during financial crises or downturns.
- Periods of high positive sentiment often precede stock price surges.
- Neutral sentiment dominates in stable market conditions.

Why This Matters?

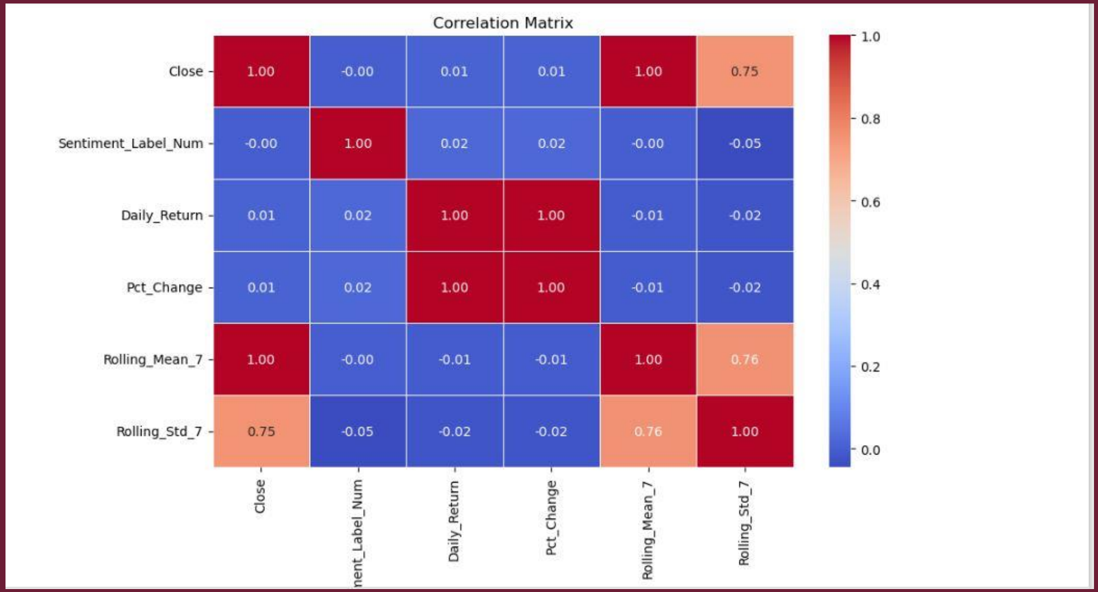
- Understanding sentiment trends helps predict market movements.
- Sentiment shifts can serve as an early indicator of volatility



Initial Findings(EDA)

Correlation Between Sentiment & Stock Returns

- Merged daily sentiment scores with corresponding stock returns.
- Analyzed positive, neutral, and negative sentiment impact.
- Positive sentiment correlates with short-term stock price increases.
- Negative sentiment correlates with stock price drops and volatility spikes.
- Neutral sentiment has minimal impact on daily stock returns.



Model Development

- **Machine Learning Model Selection**

Reasons for choosing Random Forest:

- Handles structured stock data well.
- Can incorporate both numerical (stock indicators) & categorical (sentiment labels) features.
- Robust against overfitting.

Model Development

- **Model Training & Testing**

- Train-test split: 80% training, 20% testing.
- Feature Scaling: Standardized stock price data.
- Sentiment Labels: Converted into numerical values for ML models.
- Random Forest trained using stock indicators + sentiment features.

Model Development

Model Performance Evaluation

Key Metrics Used for Evaluation:

- Accuracy – Measures how often the model correctly predicts stock movement.
- Precision & Recall – Evaluates how well the model identifies positive and negative trends.
- F1 Score – Balances precision and recall for overall performance assessment.

Model Development

- Stock Price Forecasting with and without Sentiment
  Random Forest (without sentiment): Lower accuracy due to missing external influences.
  Random Forest (with sentiment): Improved accuracy as sentiment added predictive value.
- Accuracy Scores of Different Models
  Random Forest (Stock Data Only): 51.35%
  Random Forest (Stock + Sentiment Data): 71.83% (Improvement)

- Sentiment integration improved stock movement prediction accuracy.

Discussion & Future Works

Achievements

- Combines sentiment analysis with machine learning for better predictions.
- Uses FinBERT, a finance-specific sentiment model for accuracy.
- Prediction accuracy is improved.

Limitations

- Sentiment dataset limited to 2018-2020 (original plan was 20 years).
- Real-time market predictions not implemented yet.
- Stock market unpredictability remains a challenge despite improved accuracy.

Discussion & Future Work

Future Work

- Expand data sources: Include social media (Twitter, Reddit) for better sentiment tracking.
- Test deep learning models: Transformer-based models (GPT-4, BERT variants).
- Develop real-time prediction systems: Implement sentiment-driven trading strategies.
- Use alternative financial sentiment datasets: To improve training data diversity.

Potential Applications

- Investment firms and traders can use sentiment-enhanced models for better decision-making.
- Automated trading systems could incorporate real-time sentiment analysis.

References

•Araci, D. (2019). FinBERT: A pre-trained language model for financial communications. *arXiv preprint arXiv:1908.10063*.

•Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, 5(2), 157–166.

•Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, 2(1), 1-8.

•Box, G. E. P., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. Holden-Day.

•Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New avenues in opinion mining and sentiment analysis. *IEEE Intelligent Systems*, 28(2), 15-21.

•Daas, P. J., Puts, M. J., Buelens, B., & van den Hurk, P. A. (2015). Big data as a source for official statistics. *Journal of Official Statistics*, 31(2), 249-262.

•Fischer, T., & Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654-669..

References

•Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
•Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780
• Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and Practice*. OTexts.
•Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, 1(9), 389-399.
•Kim, H., & Won, J. (2020). Hybrid models combining sentiment analysis and machine learning for stock price prediction. *Expert Systems with Applications*, 143, 113085.
•Loughran, T., & McDonald, B. (2011). When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *The Journal of Finance*, 66(1), 35-65.
•Mittal, A., & Goel, A. (2012). Stock prediction using Twitter sentiment analysis. *Stanford University Research Paper*, 15(2), 1-5.

References

•Nassirtoussi, A. K., Aghabozorgi, S., Wah, T. Y., & Ngo, D. C. L. (2014). Text mining for market prediction: A systematic review. *Expert Systems with Applications*, 41(16), 7653-7670.

•Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance*, 19(3), 425-442.

•Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. *The Journal of Finance*, 62(3), 1139-1168.

•Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment
analysis. *Computational Linguistics*, 37(2), 267-307.

•Tsay, R. S. (2005). *Analysis of Financial Time Series*. John Wiley & Sons.

•Brown, T. B., Mann, B., Ryder, N., et al. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*.

UNIVERSITI TEKNOLOGI MALAYSIA

www.utm.my

THANK YOU

univteknologimalaysia utm.my utmofficial