Sentiment Analysis of News Title on the Movement of Stock Prices in Malaysia Quantitatively and Qualitatively

Name: Yap Qi Yuan

Matric number: MCS231025

CHAPTERS

Introduction

Introduction, objectives, scope

Literature Review

Overview, relationships, sentiment analysis, advanced predictive models

Research Methodology

Description of data sources, data collection methods, and data pre-processing steps

EDA/ Initial Results

Visualizations, descriptive statistics, initial insights & results, ML, and future work

Chapter-1

OVERVIEW

- Sentiment analysis, or opinion mining. Process of identifying **individuals‘ opinions, emotions, attitudes, and feelings in financial news**.

- Sentiment analysis **provide valuable insights for their decision-making** processes in stock market prediction.

- Technologies like LSTM networks can **improve stock price prediction accuracy**.

- Research aims to **explore sentiment analysis of news headlines and its impact on stock prices in Malaysia**.

Chapter-1

RESEARCH BACKGROUND

- Sentiment analysis of news articles has emerged as a **powerful tool** to understand market sentiment and its influence on financial markets.
- Previous studies have explored the use of sentiment analysis in the Malaysian context, focusing on techniques like **Naive Bayes** and **lexicon-based approaches**.
- However, there is a need to investigate **more advanced sentiment analysis methods**, such as **Long Short-Term Memory (LSTM) neural networks**, to enhance the accuracy of sentiment classification.
- **Incorporating news sentiment into stock price prediction models** can potentially **improve forecasting performance** compared to using historical stock data alone.

Chapter-1

STATEMENT OF THE PROBLEM

- **Accurately predicting stock price movements** is a challenging task due to the complex and dynamic nature of financial markets.
- **News sentiment analysis can provide valuable insights** into the factors driving stock price changes, but its impact on the Malaysian stock market is not well-understood.
- Existing research in Malaysia has **primarily used traditional sentiment analysis** techniques, leaving room for exploring more advanced methods like LSTM.
- This study aims to **bridge the gap by investigating the influence of news sentiment**, extracted using **both lexicon-based and LSTM-based approaches**, on stock price predictions in the Malaysian context.

INTRODUCTION

Research Questions

1. To what extent can specific sentiment in the headlines of financial news affect the stock prices of Malaysia?

2. What are the difficulties for implementing the sentiment analysis to forecast the stock prices in the Malaysian market and how to improve the sophisticated models such as LSTM networks to overcome those difficulties and enhance its effectiveness?

3. Regarding various sentiment analysis techniques such, Hybrid Naive Bayes and Opinion Lexicon, Malaysia's stock price change prediction can be determined in what way and how approach can be assessed and advanced to achieve higher accuracy?

INTRODUCTION

Research Objectives

1. To examine how specific sentiments vividly articulated in headlines of financial news affect stock price changes in the Malaysian stock market.
2. To identify and evaluate challenges that hinder the efficient analysis of stock prices in the Malaysian market using sentiment analysis techniques and enhance superior models like LSTM networks to enhance the accuracy of the forecast.
3. To investigate on how some of the sentiment analysis methods such as Hybrid Naive Bayes and Opinion Lexicon-based will affect the forecast of stock price changes in Malaysia and how to improve upon the algorithms to increase accuracy.

INTRODUCTION

Scope of the Study

1. To investigate the impact of sentiment analysis from financial news headlines on stock price movements in the Malaysian stock market.
2. To examine the impact of feelings (positive, negative, and neutral) on stock prices in Malaysia at the sentence level, utilizing trusted Malaysian online news portals like the New Straits Times, Bursa Malaysia, and The Edge Market as main sources of data.
3. Apply traditional SA techniques and advanced machine learning such as Long Short-Term Memory networks to predict the movements of stock price using news sentiment.
4. Gather and analyze data over a period of 5 years and extend helpful insights to traders, investors and financial analyst in Malaysia for investment heading.

INTRODUCTION

Significance of Research

1. Provides insights into financial news sentiment for traders and investors.
2. Enhances predictive models by integrating sentiment analysis, improving stock price prediction accuracy.
3. Supports automated trading systems and offers practical implications for financial stakeholders in Malaysia.

Chapter-2

Literature Review

SA in Financial Market

Tool for understanding and predicting market movements by using textual data from financial news, articles, and social media.
Using prominent models such as FinBERT and eXplainable Lexicons (XLex).
FinBERT: Effective but requires extensive data and computational resources; XLex: Combines lexicon-based methods and transformer models for better efficiency and interpretability.

Sentiment Analysis in Financial Market

SA with Stock Price Movements

SA & Stock Price Movements

Investor sentiment significantly influences stock performance.
Models used:

- Long Short-Term Memory (LSTM)
- Gated Recurrent Units (GRU)
- Mixed frameworks

Predictive Models for Stock Price Forecasting

Predictive Models for Stock Price Forecasting

Sentiment analysis helps predict stock movements amidst market volatility.
Models tested such as LSTM & GRU.

Gaps & Opportunities

Gaps & Opportunitie s

Limited focus on sentiment in news article titles
Computational intensity of transformer models.
Research opportunities:

- Analyzing news headlines.
- Enhancing sentiment classification techniques.
- Exploring new methodologies.

Chapter-2

Literature Review

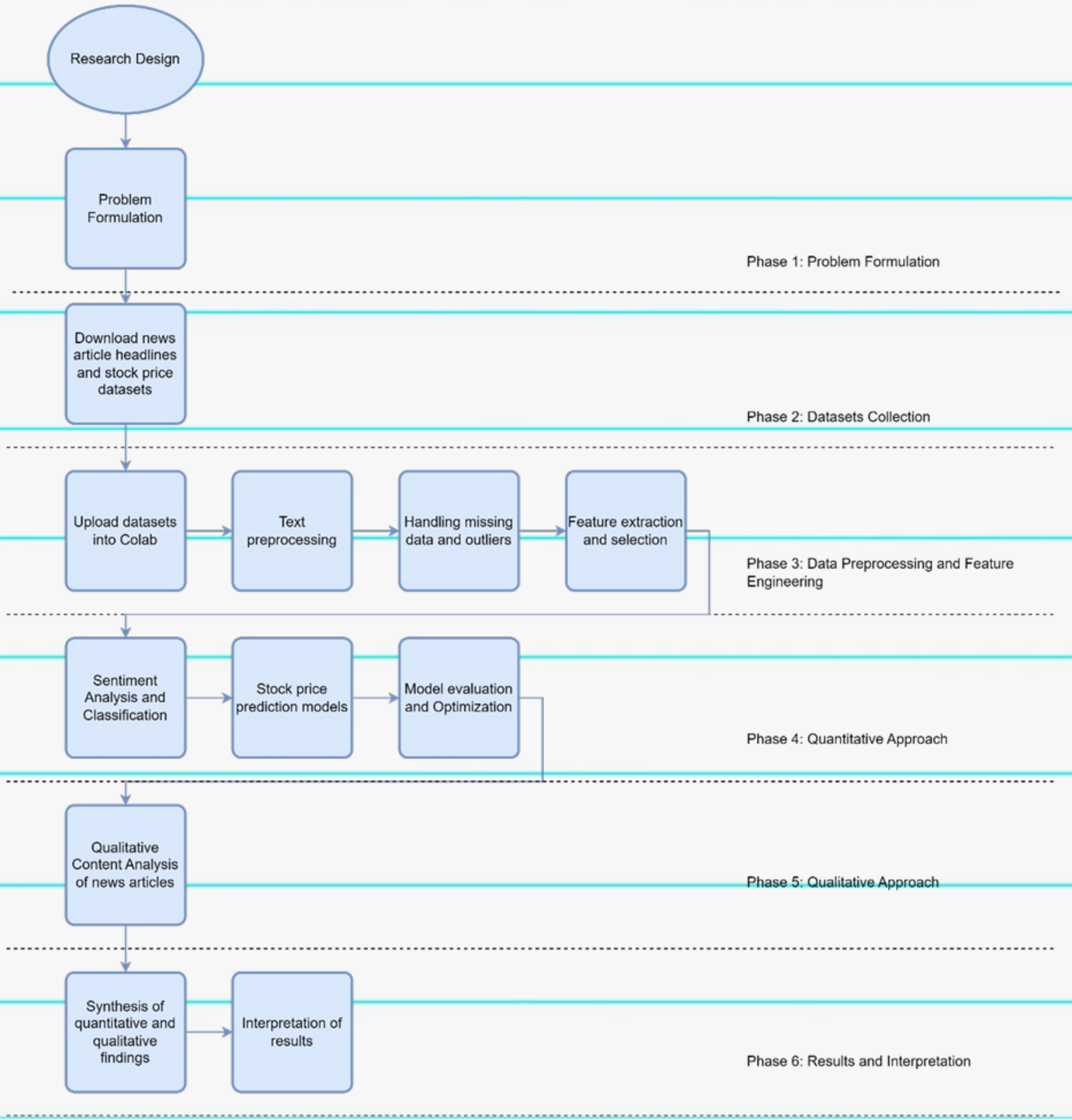
References	*Gap*	*Description*
Jiang & Zeng, 2023; Jiang et al., 2023	Limited focus on sentiment in news article titles	Previous work has discussed sentiment analysis in a number of textual sources, but the influence of news titles (the most important source of information for investors most of the time) has been explored in particular rather limitedly. Exploring the link between the sentiment of news titles and the fluctuations in the stock prices would be useful as it helped investors.
Rizinski et al., 2023, 2024	Computational and data-intensive nature of transformer models	Advanced transformers such as FinBERT used for sentiment analysis, which requires large data and computational power. This makes them unsuitable for reactive use or in systems that allow for only a small amount of processing power. Developing new and less complex approaches for sentiment analysis, for instance, eXplainable Lexicons (XLex), could improve financial decisions made as a result of the sentiment analysis and the application of the method as a whole.

References	*Gap*	*Description*
Liu et al., 2022, 2023	Challenges in accurately classifying neutral comments	Indeed, as documented in prior literature, it is challenging to classify neutral posts in the context of SA specifically for social media platforms such as Stocktwits. Neutral sentiment is essential to stock prices; it is possible to advance the models and techniques that have been designed for categorizing words in this context and improve the overall applicability of the model to stock price prediction.
Liu et al., 2022, 2023	Research Opportunities	These gaps highlighted above are the areas of research that offer grounds for further studies concerning the contribution of sentiment analysis to the Malaysian stock market. This could be relevant to the investor, policy-maker and the analyst in the financial market.

Table 2.1: Research gap analysis

Chapter-3

Research Framework



Chapter-3

Research Methodology

Quantitative Approach	Qualitative Approach
Data Collection: Uses news headlines from reputable Malaysian sources and historical stock prices over five years. **Sentiment Analysis:** Employs both traditional algorithms and advanced machine learning models to classify sentiment in news headlines. **Stock Price Prediction Models:** Analyzes the relationship between news sentiment and stock prices using models like ARIMA, LSTM, and GRU.	**Interviews with Financial Experts:** Gathers insights from professionals on sentiment analysis in stock price forecasting. **Content Analysis of News Articles:** Provides deeper understanding by analyzing full text of selected news articles.

Datasets

01 Textual Data - Full text of the news articles.
News websites such as Malaysiakini, The New York Times, The Washington Post, BBC, CNN provide APIs or their article content available for download.

02 Sentiment Analysis Scores - Numerical score from ML model or Sentiment score from Lexicon-based approach (AFINN, VADER or NRC Emotion Lexicon)

03 Metadata - Names of the authors of the articles, dates and times by Web Scraping or news aggregators.

Chapter-3

PROBLEM FORMULATION

	Research Questions	Research Objectives	Proposed Solutions
1.	How does the specific sentiment expressed in financial news headlines impact the movement of stock prices in Malaysia?	To analyze the nuanced impact of specific sentiments expressed in financial news headlines on stock price movements within the Malaysian stock market context.	• Conduct sentiment classification at the sentence level, categorizing each sentence in the news headlines as positive, negative, or neutral. • Employ both traditional sentiment analysis algorithms (e.g., Naive Bayes, Lexicon-based) and advanced machine learning models (e.g., LSTM networks) to capture the granular-level sentiment expressed in the news headlines. • Investigate the relationship between the classified sentiment and the corresponding stock price movements to uncover the nuanced impact of specific sentiments.

Chapter-3

PROBLEM FORMULATION

2.

Research Questions	Research Objectives	Proposed Solutions
What are the key challenges in accurately predicting stock prices in the Malaysian market using sentiment analysis techniques, and how can advanced models like LSTM networks be optimized to address these challenges?	To evaluate and contrast the predictive capabilities of traditional sentiment analysis techniques (e.g., Naive Bayes, Lexicon-based) against advanced machine learning models (e.g., Long Short-Term Memory networks) in forecasting stock price movements based on news sentiment.	• Develop and train traditional sentiment analysis algorithms, such as Naive Bayes and Lexicon-based approaches, to predict stock price movements based on news sentiment. • Construct advanced machine learning models, particularly LSTM networks, to forecast stock prices using the sentiment data extracted from news headlines. • Compare the performance of the traditional and advanced models using evaluation metrics, such as root mean squared error (RMSE) and mean absolute error (MAE), to identify the most effective techniques for the Malaysian market.

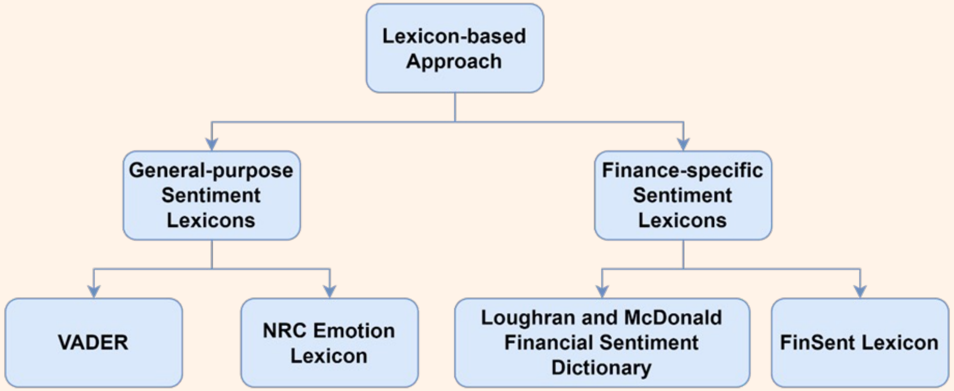
Chapter-3

PROBLEM FORMULATION

	Research Questions	Research Objectives	Proposed Solutions
3.	What impact do different sentiment analysis algorithms, such as Hybrid Naive Bayes and Opinion Lexicon-based approaches, have on predicting stock price movements in Malaysia, and how can these algorithms be compared and improved for more precise predictions?	To explore the temporal relationship between historical stock prices and sentiment data extracted from financial news headlines in Malaysia, uncovering patterns that influence stock price movements over time.	• Analyze the time-series relationship between the sentiment expressed in news headlines and the corresponding historical stock prices. • Identify patterns and temporal dependencies that influence stock price movements over time, leveraging techniques like time-series analysis and cross-correlation. • Incorporate the temporal insights into the development and optimization of the stock price prediction models, including LSTM and GRU networks, to enhance the accuracy of forecasts.

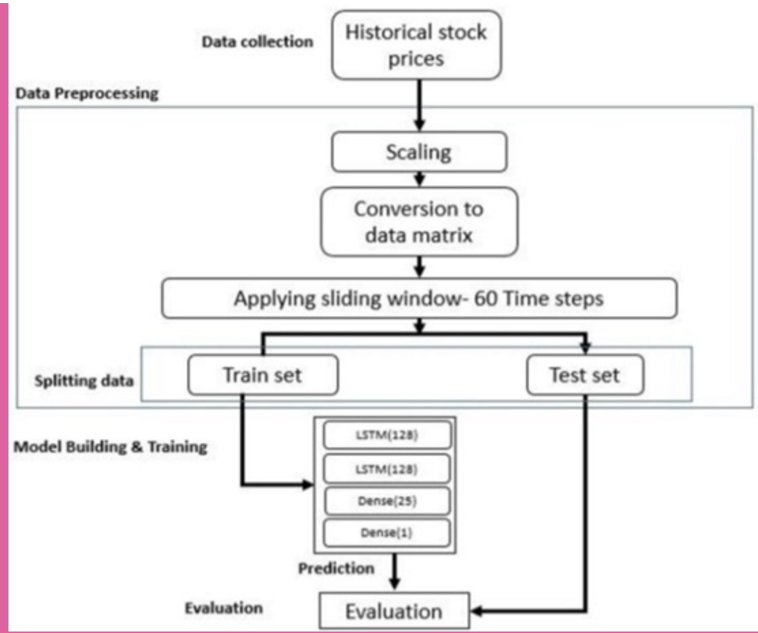
Chapter-3

Models



General-purpose Sentiment Lexicons:
- *VADER (Valence Aware Dictionary and sEntiment Reasoner)*
- *NRC Emotion Lexicon*

Finance-specific Sentiment Lexicons:
- *Loughran and McDonald Financial Sentiment Dictionary*
- *FinSent Lexicon*



Workflow for processing historical stock price data and building predictive models:

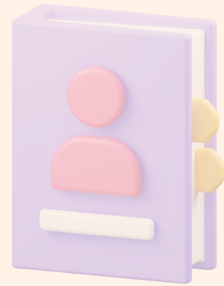
1. *Data collection of historical stock prices*
2. *Data preprocessing steps including scaling and conversion to a data matrix*
3. *Applying a sliding window of 60 time steps*
4. *Splitting data into training and test sets*
5. *Model building and training using LSTM (Long Short-Term Memory) neural networks*
6. *Prediction and evaluation of the model's performance*

Chapter 4

Feature Engineering and Expected results

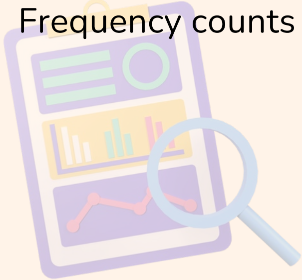
Visualization Tools

- Histograms to examine the distribution of sentiment scores.
- Box plots for median, quartiles, and outliers of sentiment scores.
- Time series analysis, observe trends and patterns in sentiment over time.
- Scatter plots and heatmaps for correlation analysis between sentiment and stock prices.



Descriptive Statisctics

- Mean, median, standard deviation, and variance to understand the central tendency and dispersion of sentiment scores.
- Frequency counts of sentiment categories (positive, negative, neutral) to quantify the overall sentiment landscape.



Expected Results

- Detailed sentiment categories: Identification and classification of specific sentiments (optimism, pessimism, fear, confidence) expressed in financial news headlines.
- Impact on stock price movements: Quantitative analysis showing how these specific sentiments impact stock price movements.
- Performance metrics: Comparison of traditional sentiment analysis techniques and advanced machine learning models in predicting stock prices.

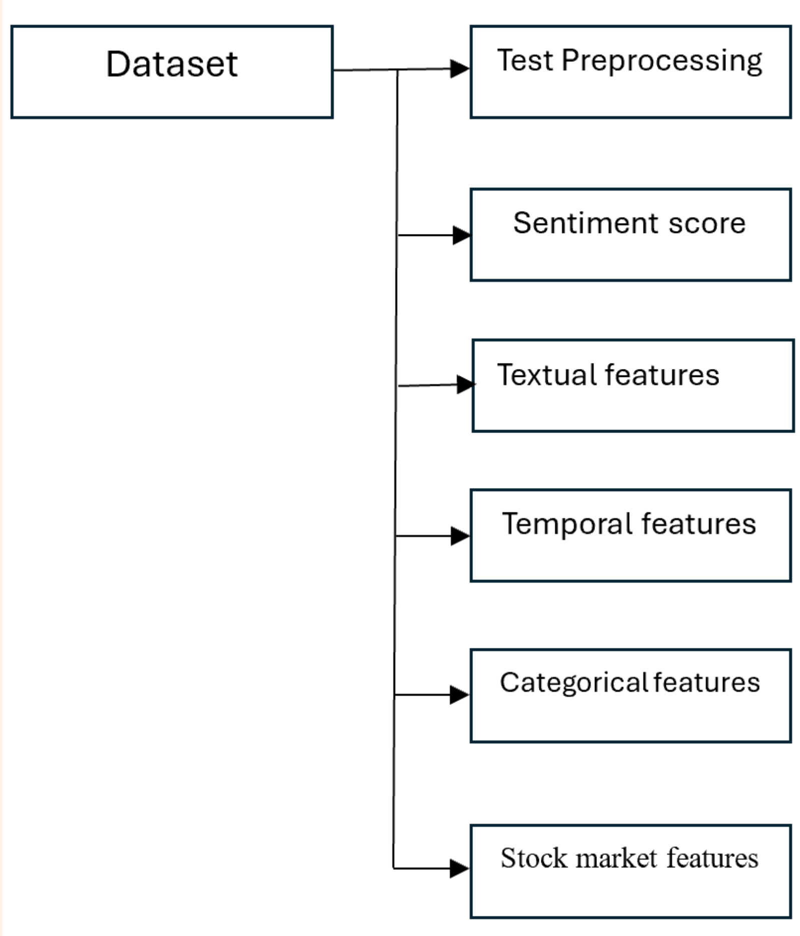
Chapter-4

Feature Engineering and Expected results

Feature Engineering

Incorporation of relevant stock market features:

- Stock prices (closing, opening, high, low).
- Trading volume.
- Stock price trends (moving averages, momentum indicators).
- Transformation of raw data into a format suitable for predictive modeling.



Initial Results

```
Processing file: cleaned_1.csv
Original shape: (1671, 4)
Sentiment analysis completed. Results saved as: sentiment_analyzed_1.csv
Final shape: (1671, 6)
                                           Title  Sentiment Score
0                 Happy New Year from Malaysiakini          0.9451
1                    500人赴双下集会，跨年喊“油价纳吉都要下”          0.0000
2              Seorang diplomat di Kota Darul Naim          0.8555
3  Yoursay: Non-Muslims pay taxes, but can’t be i...          0.9960
4  Najib praises Hadi's 'better way'; 500 rally o...          0.5267

Processing file: cleaned_2.csv
Original shape: (3039, 4)
Sentiment analysis completed. Results saved as: sentiment_analyzed_2.csv
Final shape: (3039, 6)
                                           Title  Sentiment Score
0  Say goodbye to elections should Najib win, Mah...         -0.1082
1         Maria’s candidacy breathes hope for GE14          0.9985
2                   “雪槟即大马未来”，阿兹敏冠英化身CEO卖政绩          0.0000
3   Trump ready to meet N Korea’s Kim Jong-un by May          0.9678
4                    马哈迪促澳洲总理，趁东盟峰会向纳吉提一马案          0.0000

Processing file: cleaned_3.csv
Original shape: (2575, 4)
Sentiment analysis completed. Results saved as: sentiment_analyzed_3.csv
Final shape: (2575, 6)
                                           Title  Sentiment Score
0  Nurul Izzah: Nation may witness 'dirtiest poll...          0.0258
1                   TV3记者“倒戈”坦承，报道黑函前没向安华求证          0.0000
2  PM's 'chaos if gov't changed' remark undemocra...          0.9806
3  SAMM gives MACC documents on S'gor sand mining...          0.2023
4                              到底为什么要投废票?          0.0000
```

Datasets source: Yfinance, Timeframe: 5 years
Method: URL web-based Scrapping
Process: Data cleaning, Pre-processing

- Eliminating duplicate entries, addressing missing data, and standardizing the format and structure of the headlines.

```
Accuracy: 0.7046435516767063
Classification Report:
              precision    recall  f1-score   support

           0       0.70      0.48      0.57      8622
           1       0.70      0.86      0.77     12461

    accuracy                           0.70     21083
   macro avg       0.70      0.67      0.67     21083
weighted avg       0.70      0.70      0.69     21083
```

Sentiment labeling: from -1 to +1 (negative, neutral, to positive)
Get classification report by Machine Learning Model (Logistic Regression)
Overall Accuracy: 70.46%, class-0 (likely negative sentiment), class-1 (likely positive sentiment)

macro avg(average of metric), weighted avg(average of metric weighted by proportion)

THANK YOU!

Any Questions?