Chapter 3: Research Methodology

3.1 Introduction

In chapter-3, it explores the research methodology employed and is divided into three main parts: Research design, problem formulation and datasets, all components considered highly important to a study's outcome. Research design describes the overall concept and its elements employed to realize the research goals and objectives. The detailed information about the employed methods and techniques, as well as, the justification of the choice of certain sentiment analysis algorithms and machine learning models is provided. This section also looks at the procedures followed in gathering data as well as in tabulating and analyzing information so that the research is conducted systematically and scientifically. Formulating the problem can be described as the process of identifying and defining the research problem. t contains the methodology description of research questions and hypothesis, objectives in order to provide the basis for further detailed examination of the connection between news sentiment and stock price fluctuations. Databases give a comprehensive description of the data sources that were used in the research work. This entails providing a comprehensive account of the main data source as are the credible online news portals in Malaysia that include New Straits times, Bursa Malaysia and the Edge Market among others. It also provides information about the data collection process and focuses on the criteria for selection of the articles, and on the techniques used for the classification of sentiments on the granular level. The datasets section increases the degree of transparency while using the given data and highlights the reliable sources used for analysis.

3.2 Research Design

The study will use both quantitative and qualitative methods in the analysis of the given research questions, which are essential in examining the influence of sentiment expressed and found in financials news headlines on stock prices in the Malaysian market. Research framework for this proposal as shown in Figure 3.1.

In phase one there is problem formulation and the identification of the background to the study through literature review. In Phase 2, the researchers will gather all relevant data such as news articles retrieved from appropriate online sources concerning the Malaysian market, headlines of the articles, and the historical stock price data. Phase 3 will be involved in the main analysis processes where the extracted news headlines will be subjected to sentiment analysis both using lexicon-based and machine learning based sentiment analysis to obtain sentiment scores. This sentiment data will then be combined with the stock price data of the specific firm in question. In Phase 4, the features from the integrated dataset will be utilized to build models that are regression based and deep learning based such as LSTM and GRU to forecast the movement of the stock prices. Finally, in Phase 5 the results coming from the two phases of the sentiment analysis and stock price modelling phases will be analysed and discussed emphasising on the comparison of the most effective approaches to the sentiment analysis, revised list of relevant sentiment related features, and concluding remarks as to the practical implications of the results for multiple stakeholders. The results of the conducted research will be described and exemplified utilizing multiple forms of visualization and reporting tools for better information sharing with regard to the outcomes of the study.

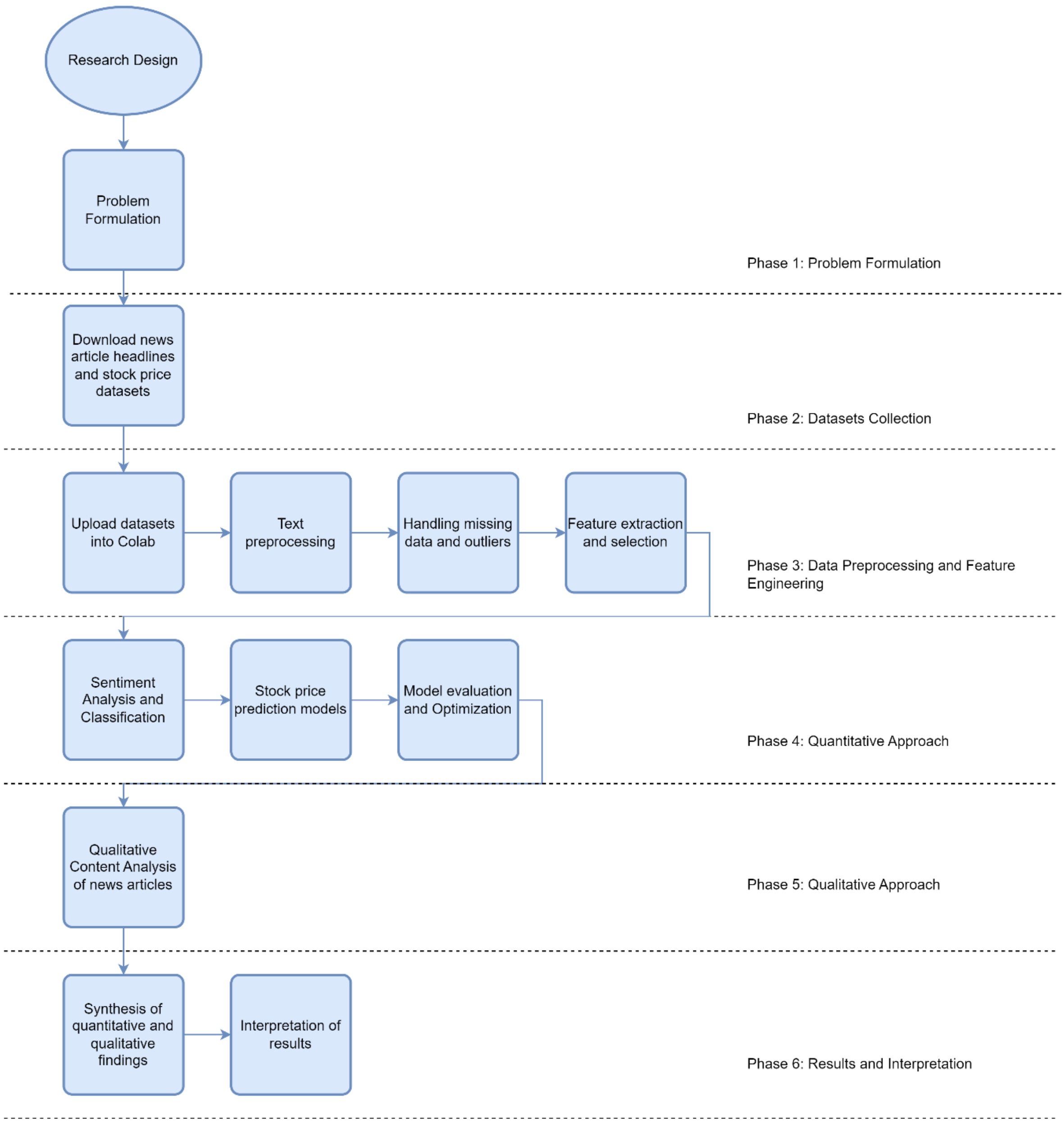


Figure 3.1: Research Framework of Proposal

3.3 Problem Formulation

Specific problems that the study aims to address are highlighted in Table 3.3.

No.	*Research Questions*	*Research Objectives*	*Proposed solutions*
1	To what extent can specific sentiment in the headlines of financial news affect the stock prices of Malaysia?	To examine how specific sentiments vividly articulated in headlines of financial news affect stock price changes in the Malaysian stock market.	• Conduct sentiment classification at the sentence level, categorizing each sentence in the news headlines as either 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 sentiment expressed in the news headlines at the granular level. • Investigate the relationship between the classified sentiment and the corresponding stock price movements to uncover the nuanced impact of specific sentiments.
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	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.	• 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

	LSTM networks to overcome those difficulties and enhance its effectiveness?		data extracted from news headlines. • Compare the performance of the traditional and advanced models using evaluation metrics, such as mean absolute error (MAE) and root mean squared error (RMSE). The aim is to identify the most effective techniques for the Malaysian market.
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?	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.	• Analyze the time-series relationship between the sentiment expressed and extracted from news headlines and the corresponding related 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.

Table 3.1: Problem formulation

3.4 Datasets

In sum, it is evident that the methodology of any stock price prediction model is highly dependent on the data used for testing and training the model. In the process of implementing this concept, the data accumulation phase entails the acquisition of information pertinent to the current project's analysis and modeling components. The description of practical datasets is shown in the following Table 3.5.

Dataset	Description	Data source
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 make their article content available for download. Financial Reports and Press Releases.
Sentiment Analysis Scores	Numerical score representing the sentiment of the article content, typically derived from a machine learning model, Sentiment score calculated using a lexicon-based approach	Sentiment scores could be generated using machine learning models trained on labeled datasets, Lexicon sentiment scores might come from predefined sentiment lexicons such as AFINN, VADER (Valence Aware Dictionary and Sentiment Reasoner), or the NRC Emotion Lexicon, which assign sentiment values to words and phrases. Commercial and open-source sentiment analysis tools and APIs, such as those provided by Google Cloud Natural

		Language API, IBM Watson, or Python libraries like TextBlob and NLTK, could also be used to derive these scores
Metadata	Names of the authors of the articles, dates and times when the articles were published.	Web Scraping or news aggregators

Table 3.2: Datasets

3.4.1 News Article Data

The primary data for this research shall be the headlines that relate to financial news in Malaysia, which may be obtained from the following Malaysians reputable online newspapers; Malaysiakini, New Straits Times, Bursa Malaysia, and The Edge Market. The following news sources were identified and chosen from the list of those most prominent and credible in the Malaysian financial sphere: The news headlines will be accumulated for half a decade: From January 2018 to June 2024, so as to have a standard and sufficient amount of material to make an analysis. Such period was selected to test how different economic and market events can affect the correlation between the news sentiment and stock price fluctuation.

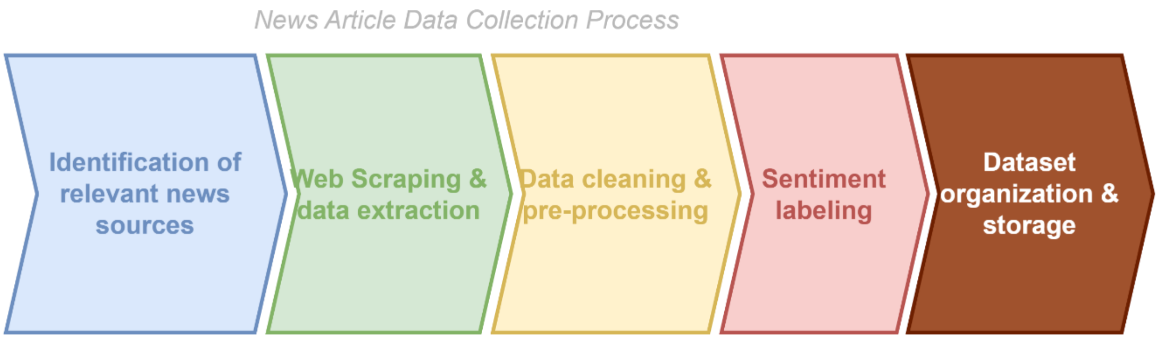


Figure 3.2: News article data collection process

Process steps	*Description*

Relevant news sources identification	The trustable Malaysian online news websites that regularly cover financial and stock market news will be identified. The sources will be chosen depending on their standing, audience, and emphasis on the Malaysian financial market.
Web scraping & data extraction	Automated web scraping techniques will be used to extract financial news headlines from chosen online news portals. This procedure will require creating scripts or using web scraping tools to gather headlines and related metadata (such as publication date and article URL) systematically over 5 years timeframe.
Data cleaning & preprocessing	The news headlines that have been gathered will be subjected to a thorough data cleaning and preprocessing step. This will involve actions like eliminating duplicate entries, addressing missing data, and standardizing the format and structure of the headlines to ensure consistency across the dataset.
Sentiment labeling	Each news headline will be manually reviewed and categorize each news headline based on its sentiment (positive, negative, or neutral). This manual labeling process will act as the foundation for the following sentiment analysis and model training.
Dataset organization & storage	The cleaned and labeled news headline dataset will be organized and stored in a structured format, such as a CSV file or a relational database, making it easier to manage and analyze the data effectively.

Table 3.3: Data collection process steps

Web scrapping step	*Description*
1. Define the URL Range	Articles were scraped from the news section of the Malaysiakini website, specifically targeting URLs within a specified range. The range covered articles from

	https://www.malaysiakini.com/news/405000 to https://www.malaysiakini.com/news/710000.
2. URL Construction and Looping	A loop was set up to iterate through each URL in the defined range. For each URL, the newspaper3k library was used to download and parse the article.
3. Extraction of Data	The following information was extracted from each article: - Title: The news articles' headline. - Author: Name of author of the article. - Published Date: The date on which the article was published. - Content: The main body of text of the article.
4. Filtering Criteria	Articles were filtered based on their publication dates to include only those published between 2018 and 2024. This ensures that the analysis focuses on recent and relevant articles for past 6 years.
5. Error Handling and Data Storage	Log will skip any URLs that failed to return a valid response and extracted data was stored in a panda DataFrame and subsequently saved to a CSV file with UTF-8 encoding to handle text data, including Chinese characters, accurately.

Table 3.4: Web scrapping steps

3.4.2 Stock Price Data

In the same regard, historical data concern to the index of stock price of the Malaysian stock market also will form part of the result headline. This data will be extracted from the Yfinance which includes mandatory columns namely Date, Open, High, Low, Close, Volume, and Adjusted Close prices. From reliable sources like the Bursa Malaysia, which is the Malaysian stock exchange or even from sources like yahoo finance or bloomberg. The following key stock market data will be collected:

1. Stock Prices, as shown in Table 3.8.

2. Trading Volume: The number of shares traded for a particular stock on a given day.

3. Stock Indices: Relative stock market indexes, for instance, the FTSE Bursa Malaysia KLCI often abbreviated as FBMKLCI, is the main stock market index in Malaysia.

Related stock price data of the particular stock will also be extracted for the same time period as the news article data so that both could be integrated for analysis. The data of the stock prices will be again cleaned and preprocessed so as to handle any missing data, any outliers, or else. Through the inclusion of the Malaysian stock price data, the research will be able to analyse the relationship that exists between news sentiment and actual stock price in the Malaysian market. This will be helpful to investors, traders and especially the financial analyst in a way that they will be able to comprehend the relationship between news sentiment and stock return.

Attribute	Meaning
Date	the trading date.
Open	the stock's opening price on the given date.
High	the highest price of the stock on the given date.
Low	the lowest price of the stock on the given date.
Close	the closing price of the stock on the given date.
Volume	the number of shares traded on the given date.
Adjusted Close prices	the stock's closing price adjusted for corporate actions.

Table 3.5: Attribute and meaning

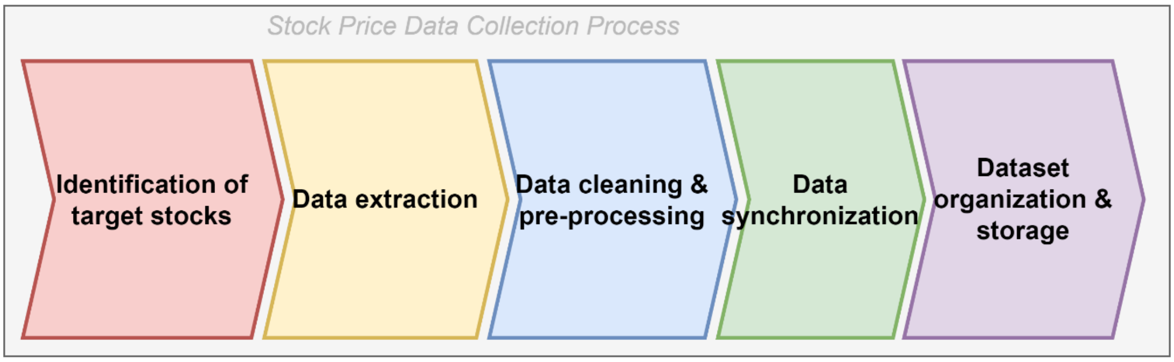


Figure 3.3: Stock price data collection process

3.5 Data Preprocessing and Feature Engineering

In the news article collection from the different online source, the data will go through a rigorous process of preprocessing to increase its quality for the sentiment analysis and the stock price prediction modeling to be done on it. In feature engineering, the preprocessed news article data is formatted to a set of useful features that can be used for the analysis of sentiments and stock prices. The feature engineering step will make the dataset ready for the steps that follow sentiment analysis and stock price prediction modeling using both text data and data from the financial market to map the sentiment of news with the volatility of stock prices. Data Preprocessing steps and Feature Engineering as shown in Table 3.6 and Table 3.7, respectively.

Data Preprocessing Steps	*Description*
Text Cleaning	As for the data preprocessing of the news article content, they will be preprocessed by stripping off HTML tags, eliminating other characters that include, but are not limited to ‘&’, ‘@’, ‘£’, ‘$’, ‘,’, ‘.’, ‘?’, ‘/’, ‘(‘, ‘)’, ‘[‘, ‘]’, ‘{‘, ‘}’, ‘:’, ‘;’, ‘+’, ‘=’, This will assist in normalizing the text and pay more attention on the useful information.
Standardization	The text will be then transformed into a simpler lower case letters type because such variations as upper and lower cases are distinguished by the computer system.
Tokenization	The text of the news article will be preprocessed by dividing the text into individual words or tokens, which is going to be an important stage in the natural language processing procedures.
Stopword Removal	Special specific inflexional words that have low semantic importance for texts, including “the,” “a,” and “and,” will be eliminated from the text to highlight important terms.
Lemmatization/Stemming	Lemmatization or stemming of the words will take place where all the words will be converted to their base or root form. This contributes to building the consistency and decreasing the number of features to reduce the dimensionality of text data.

Table 3.6: Data Preprocessing steps

Feature Engineering key steps	Description
Sentiment Score Calculation	The sentiment of each news article headline will be determined as positive, negative, or neutral depending on the scores derived from both the lexicon-based methods and machine learning sentiment analysis methods. The sentiment scores to be obtained will incorporated in the modeling process to come up with the features.
Textual Feature Extraction	Additional textual features will be extracted from the news article headlines, such as: • Term Frequency-Inverse Document Frequency (TF-IDF): Estimating the position of words in an article in regard to the significance degree of the words versus the total number of articles. • N-grams: Splitting the text into n-grams, which are two (bigrams), three (trigrams), four or more consecutive words to combine contextual data, and frequently appearing terms. • Named Entity Recognition: Selecting and extracting proper nouns like the names of the companies from the headlines.
Temporal Feature Engineering	To capture the temporal dynamics of news sentiment and its impact on stock prices, the following features will be created: • Publication Date and Time: Adding the information on when the article was published and then use temporal data analysis. • Lagged Sentiment Scores: Including leads for the sentiment scores such as sentiment score of the previous day or the previous week due to time lags of sentiments on stock prices.
Categorical Feature Engineering	Non numerical variables like the source of articles for example from New Straits Times, Bursa Malaysia, The Edge Market as well as the author of the articles will be coded to help in debiasing the data collected.
Stock Market Feature Integration	Other features of the stock market which will be incorporated into the set of predictors for the stock price prediction models include historical stock prices, trading volumes, and market indices.

Table 3.7: Feature Engineering key steps

3.6 Quantitative Approach

The quantitative component of the research design will involve the following key elements illustrated in Table 3.8.

Quantitative component	*Description*
Data Collection	• The main sources of data will be collected from trusted online news portals in Malaysia, such as the New Straits Times, Bursa Malaysia, and The Edge Market. • Financial news or article headlines will be collected for a period of 5 years to create a complete dataset for analysis. • Historical data on stock prices for the same period will be sourced from reputable sources such as Bursa Malaysia.
Sentiment Analysis	• Classification of sentiment will be conducted on a sentence-by-sentence basis, identifying whether each sentence in the news headlines is positive, negative, or neutral. • Traditional sentiment analysis methods like Naive Bayes and Lexicon-based approaches will be used alongside more sophisticated machine learning models like Long Short-Term Memory (LSTM) networks for sentiment analysis. • The effectiveness of these sentiment analysis strategies will be assessed and compared to identify the most effective methods for predicting stock price movements in the Malaysian market.
Stock Price Prediction Models	• The relationship between the sentiment expressed (positive, neutral, negative) in news headlines and historical stock prices will be analyzed to uncover patterns and temporal dependencies. • Traditional forecasting models, such as ARIMA, will be employed as a baseline for comparison. • Advanced predictive models, including LSTM and Gated Recurrent Unit (GRU) networks, will be developed and

	optimized to improve the precision when forecast stock price based on news sentiment. • The performance of these models will be evaluated using metrics such as root mean squared error (RMSE) and mean absolute error (MAE).

Table 3.8: Quantitative component

3.6.1 Lexicon-based Approach

To aid the sentiment analysis aspect of the research, sentiment lexicons will be used. These lexicons are collections of words and their corresponding sentiment ratings. These dictionaries will act as a basis for the conventional algorithms used in sentiment analysis, such as the Lexicon-based method. The study will investigate the effectiveness of different sentiment lexicons, such as general-purpose and finance-specific lexicons, to identify the most suitable resources for the Malaysian financial context.

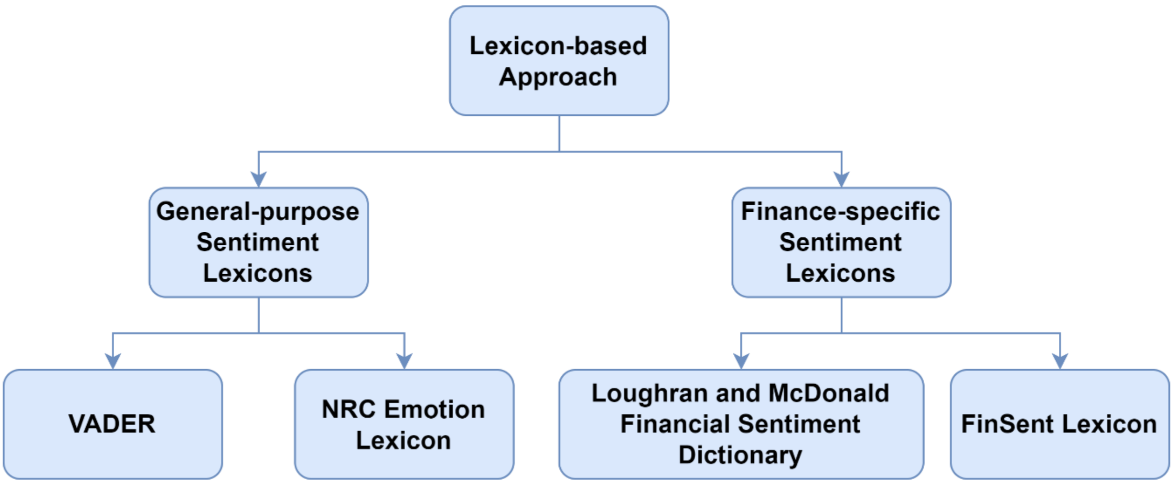


Figure 3.4: Lexicon-based approach

Lexicon-based approach		Description
	VADER (Valence Aware Dictionary	VADER is a sentiment analysis tool that focuses on emotions conveyed in social media, using

General-purpose Sentiment Lexicons	and Sentiment Reasoner)	lexicons and rules. It offers an extensive compilation of words along with their corresponding sentiment scores.
	NRC Emotion Lexicon	The NRC Emotion Lexicon is a popular sentiment lexicon that links words with eight fundamental emotions (anger, surprise, anticipation, fear, trust, sadness, disgust, and joy) and two sentiment polarities (positive and negative).
Finance-specific Sentiment Lexicons	Loughran and McDonald Financial Sentiment Dictionary	This specialized dictionary was created for the financial domain and includes a comprehensive list of words with their corresponding sentiment ratings in the context of financial reporting and news.
	FinSent Lexicon	The FinSent Lexicon is a sentiment lexicon specifically designed for analyzing sentiment in the financial domain, using financial text data for its creation and validation.

Table 3.9: Lexicon-based approach and its description

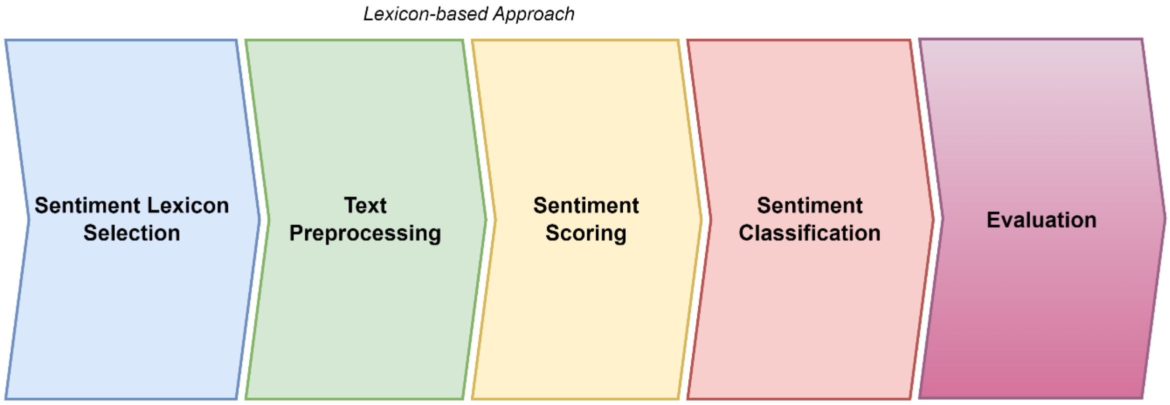


Figure 3.5: Lexicon-Based Approach steps

Lexicon-based sentiment analysis steps	Description

1.Sentiment Lexicon Selection	Several sentiment lexicons will be evaluated, such as the VADER (Valence Aware Dictionary and Sentiment Reasoner) lexicon. It is specifically designed for social media text, and the NRC Emotion Lexicon, which provides associations between words and eight basic emotions.
2.Data Pre-processing	The news article content will undergo standard text preprocessing steps, including: 1. Cleaning text– Remove unnecessary content like HTML tags, special characters, punctuations, and digits from text. 2. Standardization in lower case – Standardize text in the same lower case as the computer differentiates between lower case and upper case. 3. Tokenization – Convert sentences into words. 4. Stopword removal – Words that provide no meaningful information such as ‘this’, ‘a’, ‘there’, and ‘an’. 5. Lemmatization or stemming - to simplify words by stripping off affixes and returning them to their base form.
3.Sentiment Scoring	For each news article, the sentiment score will be calculated by the sum of sentiment scores of the individual words in the text, based on their association with positive or negative sentiment in the selected lexicon(s). For example, sum of sentiment of 1 is positive, while 0 is negative.
4.Sentiment Classification	The news articles will be classified into positive, neutral, or negative sentiment categories based on the calculated sentiment scores. This can be done by setting appropriate thresholds or using a rule-based approach.
5.Evaluation	The performance of the lexicon-based sentiment analysis will be evaluated using appropriate metrics, such as precision, recall, accuracy, and F1-score. This is able to provide insights into the effectiveness of the chosen lexicons and the overall reliability of the sentiment classification.

Table 3.10: Lexicon-based sentiment analysis steps (Srivastava et al., 2022)

3.6.2 Machine Learning Models

Training of a model in sentiment analysis in machine learning involves the use of a training dataset which has labels, and categorizing of a text or a document in terms of positive, negative or neutral sentiment. It is often easier and more accurate than the aforementioned lexicon-based approach because it can learn such patterns and relationships from the data. But, it needs a bigger and more marked up data corpus for training and may be slower.

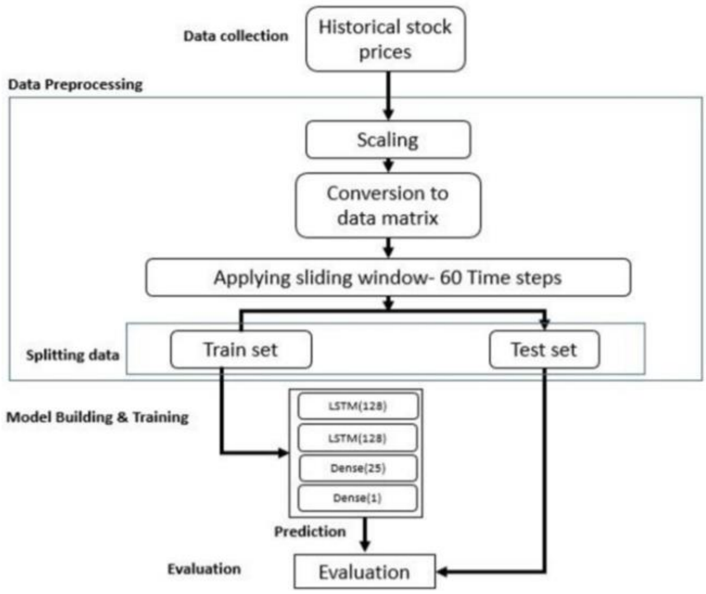
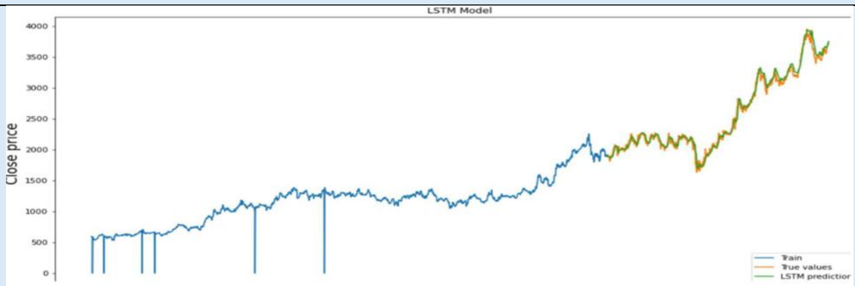


Figure 3.6: Machine learning-based sentiment analysis framework (Gangthade, 2024)

Process	Description
1.Development Phase 1	Collection of data in .csv text file.
2.Root Mean Square Error	Using formula “root mean square error (rmse) = np.sqrt(np.mean(predictions – y_test)**2))” to get value of root mean square error.
3.Plot predicted data	Plot the predicted data to examine how close is it to the actual values.
4.Example of plotted graph	 Fig: Prediction Graph

<table>
<tr>
<td>5.Example of close price and prediction</td>
<td>
<table>
<tr><th>Date</th><th>Close</th><th>predictions</th></tr>
<tr><td>2019-01-02</td><td>1923.300049</td><td>1903.230591</td></tr>
<tr><td>2019-01-03</td><td>1899.949951</td><td>1906.295898</td></tr>
<tr><td>2019-01-04</td><td>1876.849976</td><td>1904.251465</td></tr>
<tr><td>2019-01-07</td><td>1897.900024</td><td>1897.174561</td></tr>
<tr><td>2019-01-08</td><td>1893.550049</td><td>1896.034912</td></tr>
<tr><td>...</td><td>...</td><td>...</td></tr>
<tr><td>2021-12-27</td><td>3696.100098</td><td>3691.024414</td></tr>
<tr><td>2021-12-28</td><td>3706.550049</td><td>3709.432617</td></tr>
<tr><td>2021-12-29</td><td>3694.699951</td><td>3726.835449</td></tr>
<tr><td>2021-12-30</td><td>3733.750000</td><td>3737.803467</td></tr>
<tr><td>2021-12-31</td><td>3738.350098</td><td>3754.460938</td></tr>
</table>
741 rows × 2 columns
</td>
</tr>
</table>

Figure 3.7: Process steps, graph, and outputs of machine learning-based sentiment analysis (Gangthade, 2024)

3.6.3 Sentiment Classification Refinement (Hybrid Approach)

The mixed strategy based on the usage of both the lexicon-selected benchmark and machine learning tools to take advantage of both approaches. In this case, the lexicon-based method is first applied to give a raw sentiment score that is then further adjusted by a machine learning model trained with sentiment labelled data. This can help in giving more accurate and refine sentiment analysis especially when dealing with more elaborate or vague text.

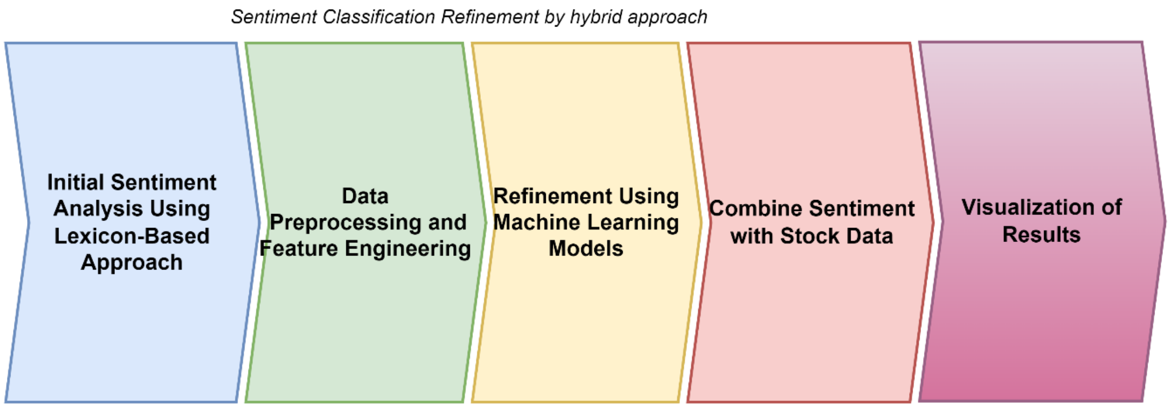


Figure 3.8: Sentiment classification refinement by hybrid approach

3.6.4 Deep Learning Techniques

As was mentioned before, the modern advancements in deep learning have contributed to the improvement of the sentiment analysis approaches. Some of these are the application of feed forward neural networks for instance convolutional neural networks (CNNs) and recurrent neural networks (RNNs) used in prevailing semantic and contextual nuances of the text data. Analyzing the results of the current deep learning models, a higher level of accuracy compared to traditional machine learning can be achieved, especially in tasks that require an understanding of the depth of the language and sentiment.

Among them, the necessity of the labelled data, the complexity of sentiment patterns identified in the text and the availability of computational power are to be taken into consideration.

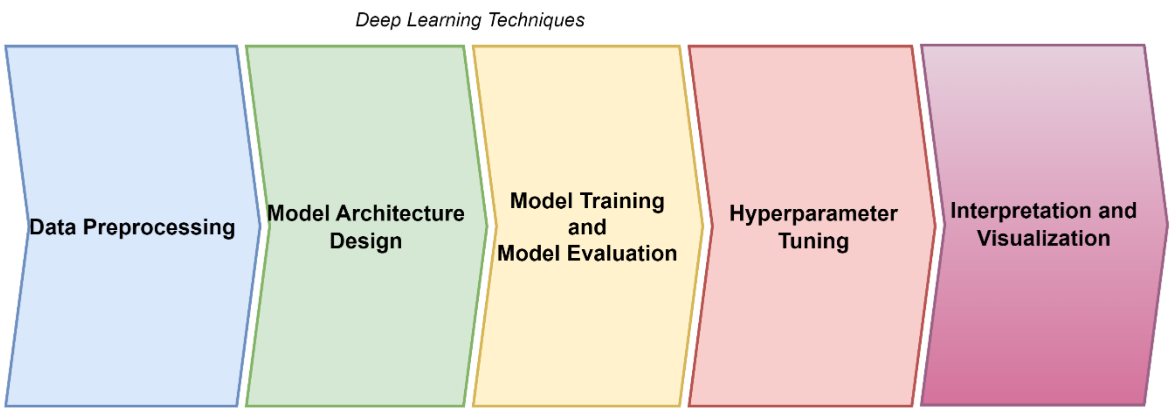


Figure 3.9: Deep Learning Techniques

3.7 Qualitative Approach

The qualitative components of the research design are described in Table 3.11.

Qualitative component	*Description*
Interviews with Financial Experts	• Conducted semi-structured interviews with financial analysts, traders, and investment professionals to gain insights based on their perspectives on the role of sentiment analysis in stock price forecasting and decision-making. • The interviews will explore the practical challenges, limitations, and potential applications of sentiment analysis in the Malaysian stock market.
Content Analysis of News Articles	• In addition to the quantitative sentiment analysis of news headlines, a qualitative content analysis of the full text of selected news articles will be performed. • This analysis will provide a deeper understanding of the contextual factors and nuances that may impact the relationship between the movements of stock price with news sentiment.

Table 3.11: Qualitative component

3.8 Results and Interpretation

In this phase, the major findings from sentiment analysis and the chosen stock price prediction models elaborated in the previous phase. The results and their interpretation will be structured as follows:

Type of Results	*Description*
Sentiment Analysis Findings	Starting with the Lexicon-based methods, the performance of the various sentiment analysis techniques will also include machine learning-based methods for comparison. This will include: • Presenting the accuracy, precision, recall, and F1-score that was obtained by each of the sentiment analysis techniques in categorizing the news headlines to positive, negative, and neutrality. • Considering the advantages and disadvantages of the various methods for the sentiment analysis comparing the Malaysian stock exchange to international markets. • Determining the type of sentiment analysis technique, or a combination thereof, that would be most suitable for the problem domain in question and the particular data set. • Discussing if any features, such as trends in the sentimental scores of the news articles over time, or entertained news sources could be observed.
Stock Price Prediction Modeling Results	Stock price prediction models, that include the news sentiment features, shall be analyzed and their performance assessed and discussed. This will include: • Presenting the overall metrics measurement indicators, including RMSE, MAE, and R-squared with respect to memory-based, traditional regression analysis, and deep learning methods, including LSTM and GRU. • Evaluating the performance of the models that incorporate sentiment features against the models that only depend on the target stock's historical prices. • Assessing the effectiveness of the proposed classification approach using the feature importance to identify that of the sentiment-related features and define which of them significantly affect the stock price (i. e. , certain words, sentiment score, etc.).

	• Recalling any prominent findings or trends that can be identified from the model results together with the relative relevance of positive, negative, and neutral sentiments impacting on stock prices.
Practical Implications	A comprehensive interpretation of the findings will be provided and discuss their practical implications for different stakeholders, including: • Investors and traders: Knowledge about various aspects that has to do with integrating news sentiment analysis in equity investment as a means of improving stock trading. • Financial analysts: Suggestions on how they could enhance their utilization of sentiment analysis approaches to enhance their predictions of stock rates and other forms of market research. • Policymakers and regulators: Information on how news sentiments work in the context of the stock market which would help decision makers crack the whip for formulating policies and implementing regulations in the stock market.
Limitations and Future Research Directions	The conclusion of the study and suggest possible improvement and further research recommendation relevant to the conduct of the research study. This may include: • Analyzing the generalization of the discoveries, especially in the context of transferring the obtained results to other sectors or industries in the Malaysian stock exchange. • Finding potential places for suggestion like possible new approaches to the sentiment analysis, new sources of data, or analyzing how specific news affect the stock prices. • In terms of future research, based on the current study, specific research topics one might consider proposing for investigation by other scholars include the improvement of the current and still rather simple methods of the predictive model, the investigation of investor sentiment, or the cross-market analysis of the news sentiment.
Visualization and Reporting	With regards to the presentation of the research results and the insights acquired by the study, the following visualization and reporting aspects

will be incorporated into the research study: These will assist in improving the presentation and analysis of the outcomes of the study.

1. For Sentiment Analysis Finding, it can be visualized as follows:
   - Sentiment Score Distribution: These headlines will be grouped under positive, negative and the neutral category and a pie chart or bar graph utilized in depicting the findings.
   - Sentiment Trends over Time: To show the temporal changes on the sentiment scores, line charts commonly known as area plots will be used and where there is significant fluctuation, the specific periods will be noted.
   - Sentiment by News Source: To compare the sentiment scores different news sources a bar graphs or heat maps would be used to determine if there are any bias or variation in reporting.
2. For Stock Price Prediction Modeling, it can be visualized as follows:
   - Model Performance Metrics: The main evaluation measures, including RMSE, MAE, and R-squared, of the method and all the compared models (e.g., LSTM, GRU, and conventional regression models) will be illustrated in tabular and line chart forms.
   - Actual vs. Predicted Stock Prices: Thus, line charts or scatter plots will be used to display the actual stock prices and the model's forecasted values to evaluate the models' calibration and fitness.
   - Feature Importance: Feature importance plots or bar charts will be used to display the contribution of the sentiment-related features and other stock market factors to the model's decision-making process.
3. For overall reporting, it can be visualized as follows:
   - Executive Summary: Summary of the most important fact, the relevance of the findings and, of course, the contribution to the academic literature.
   - Detailed Results and Interpretations: The elaborate explanation and discussion on the assessment of sentiment analysis and stock price modeling as described in the earlier sections of this paper.

	• Practical Recommendations: Guidelines and suggestions for different users including shareholders, analysts covering the firm and policy makers. • Limitations and Future Research: An evaluation of the study's restrictions and future recommendations about how the current work can be expanded upon by the next researcher.

Table 3.12: Type of Results and their Interpretation