STOCK PRICE FORECASTING USING NEWS SENTIMENT AND DEEP LEARNING: EVIDENCE FROM THE MALAYSIAN BANKING SECTOR

NUR AINA FARRAAIN BINTI ZAHANIZAM

UNIVERSITI TEKNOLOGI MALAYSIA

PSZ 19:16 (Pind. 1/23)

UNIVERSITI TEKNOLOGI MALAYSIA
DECLARATION OF Choose an item.

Author's full name	:	NUR AINA FARRAAIN			
Student's Matric No.	:	MCS 241053	Academic Session	:	
Date of Birth	:	28 APRIL 2000	UTM Email	:	nurainafarraain@graduate.utm.my
Choose an item. Title	:	STOCK PRICE FORECASTING USING NEWS SENTIMENT AND DEEP LEARNING: EVIDENCE FROM THE MALAYSIAN BANKING SECTOR			

I declare that this Choose an item. is classified as:

☒	**OPEN ACCESS**	I agree that my report to be published as a hard copy or made available through online open access.
☐	**RESTRICTED**	Contains restricted information as specified by the organization/institution where research was done. *(The library will block access for up to three (3) years)*
☐	**CONFIDENTIAL**	Contains confidential information as specified in the Official Secret Act 1972)

(If none of the options are selected, the first option will be chosen by default)

I acknowledged the intellectual property in the Choose an item. belongs to Universiti Teknologi Malaysia, and I agree to allow this to be placed in the library under the following terms :

1. This is the property of Universiti Teknologi Malaysia
2. The Library of Universiti Teknologi Malaysia has the right to make copies for the purpose of research only.
3. The Library of Universiti Teknologi Malaysia is allowed to make copies of this Choose an item. for academic exchange.

Signature of Student:

Signature :

Full Name NUR AINA FARRAAIN
Date : 30/6/2025

Approved by Supervisor(s)

Signature of Supervisor I:	Signature of Supervisor II
Full Name of Supervisor I	Full Name of Supervisor II

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

Note: This page is only applicable for theses classified as Restricted/Confidential. Please delete this page from the template if your thesis is not classified as Restricted/Confidential. This letter should be written by a supervisor and addressed to Perpustakaan UTM. A copy of this letter should be attached to the thesis.

Date:

Librarian
Jabatan Perpustakaan UTM,
Universiti Teknologi Malaysia,
Johor Bahru, Johor

Sir,

CLASSIFICATION OF THESIS AS RESTRICTED/CONFIDENTIAL

TITLE: Click or tap here to enter text.

AUTHOR'S FULL NAME: Click or tap here to enter text.

Please be informed that the above-mentioned thesis titled _________ should be classified as RESTRICTED/CONFIDENTIAL for a period of three (3) years from the date of this letter. The reasons for this classification are

(i)
(ii)
(iii)

Thank you.

Yours sincerely,

SIGNATURE:
NAME:
ADDRESS OF SUPERVISOR:

"Choose an item. hereby declare that Choose an item. have read this Choose an item. and in Choose an item.
opinion this Choose an item. is sufficient in term of scope and quality for the award of the degree of Choose an item."

Signature	:	______________________________
Name of Supervisor I	:	
Date	:	

Signature	:	______________________________
Name of Supervisor II	:	
Date	:	

Signature	:	______________________________
Name of Supervisor III	:	
Date	:	

Declaration of Cooperation

This is to confirm that this research has been conducted through a collaboration Click or tap here to enter text. and Click or tap here to enter text.

Certified by:

Signature :
Name :
Position :
Official Stamp
Date

* This section is to be filled up for theses with industrial collaboration

Pengesahan Peperiksaan

Tesis ini telah diperiksa dan diakui oleh:

Nama dan Alamat Pemeriksa Luar :

Nama dan Alamat Pemeriksa Dalam :

Nama Penyelia Lain (jika ada) :

Disahkan oleh Timbalan Pendaftar di Fakulti:

Tandatangan :

Nama :

Tarikh :

STOCK PRICE FORECASTING USING NEWS SENTIMENT AND DEEP LEARNING: EVIDENCE FROM THE MALAYSIAN BANKING SECTOR

NUR AINA FARRAAIN BINTI ZAHANIZAM

A Choose an item. submitted in Choose an item. of the
requirements for the award of the
Choose an item.

School of Computing
Faculty of Computing
Universiti Teknologi Malaysia

JUNE 2025

DECLARATION

I declare that this thesis entitled "STOCK PRICE FORECASTING USING NEWS SENTIMENT AND DEEP LEARNING: EVIDENCE FROM THE MALAYSIAN BANKING SECTOR " is the result of my own research except as cited in the references. The project has not been accepted for any thesis and is not concurrently submitted in candidature of any other thesis.

Signature	:	....
Name	:	
Date	:	30 JUNE 2025

ACKNOWLEDGEMENT

In preparing this thesis, I was in contact with many people, researchers, academicians, and practitioners. They have contributed towards my understanding and thoughts. In particular, I wish to express my sincere appreciation to my main thesis supervisor for encouragement, guidance, critics and friendship. I am also very thankful to my co-supervisor for their guidance, advices and motivation. Without their continued support and interest, this thesis would not have been the same as presented here.

My fellow postgraduate student should also be recognised for their support. My sincere appreciation also extends to all my colleagues and others who have provided assistance at various occasions. Their views and tips are useful indeed. Unfortunately, it is not possible to list all of them in this limited space. I am grateful to all my family member.

ABSTRACT

This study explores the integration of news sentiment analysis with deep learning models to forecast stock prices in the Malaysian banking sector. The main focus on CIMB and Maybank. As the public sentiment is believed to influence the financial market, this research is to use the sentiment extracted from financial news headlines to improve the prediction accuracy. News data were scraped from multiple reputable sources including The Star, Malay Mail, and The Edge Markets, while historical stock prices were obtained from Yahoo Finance for the period 2019–2025.

Sentiment analysis was conducted using the VADER to classify headlines as positive, neutral, or negative. These sentiment scores were aligned with stock price data. Then, it used as input features for three deep learning models: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and a hybrid Attention-based CNN-LSTM (ACNN-LSTM). The models were trained and tested on the prepared dataset. The model was then evaluate using RMSE, MAE, and MSE.

Findings revealed that the ACNN-LSTM model outperformed the other models. It shows the model has better accuracy in capturing both short-term sentiment signals and long-range price trends. While sentiment was found to have a moderate impact on stock prices—more notably for Maybank than CIMB—trading volume remained the most influential predictor. This study contributes to the emerging field of sentiment-enhanced financial forecasting in Southeast Asia by offering a replicable framework and highlighting the relevance of local sentiment data in investment decision-making.

ABSTRAK

Kajian ini meneroka integrasi analisis sentimen berita dengan model pembelajaran mendalam untuk meramalkan harga saham dalam sektor perbankan Malaysia. Fokus utama kepada CIMB dan Maybank. Memandangkan sentimen orang ramai dipercayai mempengaruhi pasaran kewangan, penyelidikan ini adalah untuk menggunakan sentimen yang diekstrak daripada tajuk berita kewangan untuk meningkatkan ketepatan ramalan. Data berita dikikis daripada pelbagai sumber ternama termasuk The Star, Malay Mail dan The Edge Markets, manakala harga saham sejarah diperoleh daripada Yahoo Finance untuk tempoh 2019–2025.

Analisis sentimen telah dijalankan menggunakan VADER untuk mengklasifikasikan tajuk sebagai positif, neutral atau negatif. Markah sentimen ini diselaraskan dengan data harga saham. Kemudian, ia digunakan sebagai ciri input untuk tiga model pembelajaran mendalam: Memori Jangka Pendek Panjang (LSTM), Unit Berulang Berpagar (GRU) dan CNN-LSTM berasaskan Perhatian hibrid (ACNN-LSTM). Model telah dilatih dan diuji pada set data yang disediakan. Model tersebut kemudiannya dinilai menggunakan RMSE, MAE, dan MSE.

Penemuan mendedahkan bahawa model ACNN-LSTM mengatasi model lain. Ia menunjukkan model mempunyai ketepatan yang lebih baik dalam menangkap kedua-dua isyarat sentimen jangka pendek dan arah aliran harga jarak jauh. Walaupun sentimen didapati mempunyai kesan sederhana ke atas harga saham—lebih ketara bagi Maybank berbanding CIMB—jumlah dagangan kekal sebagai peramal yang paling berpengaruh. Kajian ini menyumbang kepada bidang baru muncul ramalan kewangan yang dipertingkatkan sentimen di Asia Tenggara dengan menawarkan rangka kerja yang boleh ditiru dan menyerlahkan kaitan data sentimen tempatan dalam membuat keputusan pelaburan.

(Note: Students are allowed to use either single or one-and-a-half spacing for the abstract, as long as it fits within one page. The chosen spacing style must be consistent across both the English and Malay sections.)

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACNN-LSTM	-	Attention Convolutional Neural Network
DL	-	Deep Learning
GRU	-	Gated Recurrent Units
LSTM	-	Long-Short Term Memory
MAE	-	Mean Absolute Error
ML	-	Machine Learning
MSE	-	Mean Square Error
RMSE	-	Root Mean Square Error

CHAPTER 1

INTRODUCTION

1.1 Introduction

The ups and downs of the stock market often tell us how healthy the wider economy really is, and that's why so many people want to figure out where share prices are headed next. In the real international, the ones charges deliver pointers about future cash flows, so by means of reading them it can get a quite top sense of ways production, profits, and even new investments would possibly trade throughout unique industries. Due to this, Wall Road and main avenue alike watch stock forecasts intently. Today, machine-learning equipment had been stepping into that area. From predicting market trends to checking credit risk, these systems are helping banks and individual investors make sharper calls than many older methods allowed. What makes them stand out is their ability to sift through mountains of data and spot tricky patterns that simpler formulas tend to overlook. Another up-and-coming technique adding advantages to those tools is sentiment analysis. By measuring the emotional tone of tweets, blog posts, and news headlines, sentiment programs give investors fresh clues about where prices might swing next.

Merging sentiment analysis with machine learning is proving useful for forecasting stock price movements. By means of turning tweets, news articles and forum posts into clear sentiment rankings, buyers get a range of they could honestly paintings with (Joshi & Younes, 2023). This approach topics a lot for Malaysian banks because giants like CIMB and Maybank hold the spotlight. When these two firms do well or struggle, the entire country's economic mood can shift on a dime.

The Malaysian stock market, just like any other exchange around the world, can swing wildly after an unexpected economic event or simply because traders are feeling optimistic—or not. That shaky temper makes the marketplace a splendid spot to test out new tech, in particular ormation technological know-how. with the aid of strolling sentiment evaluation on news headlines, researchers and traders can dig beneath the surface to unearth the feelings behind the words. Those hidden emotions often hold clues about where a stock is likely to move next.

This research focus on two of the biggest names on Bursa Malaysia: CIMB and Maybank. It seems at how device-mastering algorithms that examine news sentiment can help wager the future charge of these banking giants' stocks. By lining up headlines—positive, negative, or neutral—with actual trading numbers, the project adds to the growing toolkit available to investors and analysts working in Malaysia. More importantly, it tries to make those predictions sharper and decisions easier in a sector where even small errors can cost millions.
In short, better news-reading tools could give traders an extra edge, helping the Malaysian banking industry continue its run of steady progress and keep the economy.

1.2 Problem Background

The Malaysian banking industry, and especially the two big players, CIMB and Maybank, runs on a mix of economic signals, government rules, and worldwide money trends. The ones signals do not stop at the boardroom door; average market mood can leap up or down in a heartbeat and pull their stock fees in conjunction with it. Investor temper is greater than only a feeling. Research shows that it can drive decisions on corporate social responsibility projects and even affect a firm's true value.

The "market sentiment," roar of either optimism or gloom coming off trading floors, social feeds, and business chat rooms. That roar shapes marketplace situations and pushes stock values one way or the opposite, so know-how it quick will become task critical. for this reason, analysts have started walking sentiment snapshots from not anything extra than a headline to guess wherein CIMB and Maybank stocks may settle the next day. They are not alone. Big-data tools now comb through millions of tweets and posts every minute, pulling real-time mood readings that traditional reports often miss.

The tech behind this is straightforward yet powerful: machines read sentences, label them happy or sad, and then feed that learning into models that look forward instead of backward. The models then line up those readings against price charts, creating a fuller story in which numbers and words fit together. By doing so, it helps traders, analysts, and even the banks themselves sense early if the crowd is about to cheer or jeer.

Sentiment analysis has become a popular tool for trying to forecast where stock prices might head next. Several studies show it can give investors an edge, no matter the country they're trading in (Li et al., 2014). In Malaysia, understanding market mood is especially crucial because local news or chatter can quickly sway buying and selling decisions (Binti Alyasa Gan et al., 2022). Social media posts—and the emotions they carry—can also serve as a rough report card for a company, letting analysts compare that feeling with more traditional time-series forecasting models (, 2021). This comparison matters because most feedback online is messy, written in casual language, forcing tech tools to sift through raw text and pull out clear, usable insights (Rahim et al., 2021). That's why investors are leaning more on automated, machine-learning-based methods. By scanning the mood of today's headlines, traders hope to read the crowd's pulse and spot the next likely move in stock prices.

Sentiment analysis is a quick way to sort through huge piles of text and identify the useful insights. Because stock-price guessing is a long-standing topic in finance, researchers have turned to newer tools like machine learning as social media made public chatter more valuable than ever. Predictions are not limited to Wall Street; the sentiment models can be used for forecasting elections or even tracking patient outcomes in healthcare. through scanning information headlines for hidden feelings, investors get a sharper image of market mood and, with it, a better risk to shop for or sell accurately. Of path, the flood of tweets, weblog posts, and online articles can sense overwhelming, but that identical flood is what offers the analysis its energy. Machine-learning techniques thrive on messy, fast-moving data, so they turn that challenge into an opportunity. They pick apart breaking stories, match them to specific stocks, and tease out key drivers like investor faith or fear that push prices up or down.

Investors don't decide to buy or sell a stock in a vacuum. They read news headlines, listen to podcasts, and scroll through social media before committing their cash. Due to this, information the overall temper or "sentiment" round a enterprise is precious. Sentiment evaluation shall we buyers and analysts measure that temper in numbers, turning tweets and articles into information they could use to spot developments and form their next circulate (Mokhtari et al., 2023). When enough people act on that sentiment, it can tilt an entire market, so strategies that layer sentiment on top of traditional analysis quickly gain popularity.

1.3 Problem Statement

Predicting where a stock price will land tomorrow—or next hour—remains the holy grail of investing. A correct guess can mean profit that buys a vacation; a wrong one can wipe out hard-earned savings. To stay ahead, traders need clear signals, yet financial charts are anything but smooth. Prices jump after earnings surprises, wiggle when a CEO tweets, and drift slowly during holiday weeks. Combine that with the noisy, ever-changing nature of economic data and outside events, and it is easy to see why many say price forecasting is an art as much as a science.

Stock markets can feel like they are running on rumour and mood swings. A tweet or a trending headline can send prices up one minute and crashing the next. due to this regular chatter, it is hard for buyers to get a clear examine on in which a share is surely headed. On pinnacle of that, most of the statistics we see—tweets, blogs, or maybe flashy news banners—is messy and unstructured. Pulling useful numbers from that jumble is a real headache for analysts.
Machine learning, and especially sentiment analysis, is one of those tools. By training algorithms to read headlines the same way humans do—spotting excitement, worry, or anger—they can flag patterns traders might miss. These models dig deeper than numbers on a spreadsheet and catch hidden links between a tweet's tone and how shares move.

Predicting where a stock price will land next is never easy. Markets are messy, shaped by countless factors that don't always follow a clear pattern. still, equipment like system studying and deep getting to know have started to make forecasts a bit sharper (Alamu & Siam, 2024). one of the extra interesting directions blends these algorithms with sentiment analysis, which reads the mood of traders in real time (Gude & Hsiao, 2025). By doing so, researchers hope to move beyond the old trick of looking only at past prices. Instead, they add the tone of fresh news headlines to the mix, allowing the "feeling of the crowd" to steer the numbers (Wu et al., 2021).

1.4 Research Question

1. How do positive, negative or neutral news headlines influence the next closing price of stock market for Malaysian bank.
2. Can the next closing price of CIMB and Maybank be predict based on sentiment analysis of news headlines?
3. How does the predictive accuracy of news sentiment based on the traditional stock market features such as historical prices and trading volume?

1.5 Research Objectives

1. To perform sentiment analysis on news headlines related to CIMB and Maybank
2. To examine the relationships between sentiment in the news headlines and the next closing prices of CIMB and Maybank
3. To develop predictive model based on sentiment features to predict stock price trend.

1.6 Significance of the Study

Malaysian banking sector still lacks attention in the AI-driven finance literature. Most algorithms are trained on Western data, leaving Southeast Asian markets in the dark. Second, CIMB and Maybank are integral to the region's economy, so small shifts in their stock prices can ripple through suppliers, borrowers, and even everyday customers. Finally, by studying local figures with local news, the project hopes to provide traders, regulators, and finance students here with tools that speak their language and fit their market.

This study is not just academic; it has real-world value for anyone working in Malaysia's financial markets—investors, analysts, and day traders alike. When they add sentiment analysis to their forecasting tools, these market players can base their decisions on clearer insights. Sentiment analysis listens to the emotional language of news reports, which makes it easier to guess how stocks will react during major announcements. Because of these results, investors stand a better chance of spotting price swings before they happen, helping them fine-tune their trading plans.

This research takes a big step forward in how we make financial predictions by bringing machine learning into the mix. It tests out three cutting-edge models—Long Short-Term Memory (LSTM) networks, Gated Recurrent Units (GRU), and a special Attention Convolutional Neural Network-LSTM (ACNN-LSTM)—to forecast stock prices based on the mood found in financial news articles. By blending sentiment analysis with these high-powered models, the study creates a strong forecasting tool that plays to the unique strengths of each method. LSTM and GRU are especially good at tracking the time-based ups and downs of stock prices, while ACNN-LSTM adds an extra layer of capability by using a convolutional neural network to pick up both quick text clues and longer trend patterns. Together, these models push predictive finance into new territory by showing how closely stock movements can be tied to the sentiment behind the headlines.

By combining machine learning models with sentiment analysis, this research does more than improve stock price forecasts. It also lays out a tested way of working that other financial markets can easily copy. Because of that, the study opens new doors for predictive analytics, especially in parts of the world that have yet to explore how news mood affects share prices.

1.7 Scope and Limitation

The focus here is on two of Malaysia’s biggest banks—CIMB and Maybank. Researchers will track their stock prices from roughly 2019 until early 2025. The main goal is to see how the tone of daily news headlines affects price changes, especially the next-day closing value. To do this, the team will gather stories from trusted local sites like The Star, Malay Mail, The Edge Market, and New Straits Times, while pulling price data from Yahoo Finance. Linking these datasets will give a clearer picture of how what people read shapes what they pay for shares.

This research focus on CIMB and Maybank, so its results might not transfer neatly to banks in other countries or to other industries. The findings mirror the precise conditions of the Malaysian economy and the 2 banks studied, that could behave otherwise than larger or extra evolved financial markets. The research also covers a narrow window—2019 through 2025—so it will be harder to apply the conclusions if major economic or political surprises hit after 2025.

Another hurdle is the quality of the news data the study relies on. Sentiment readings rest heavily on how clean and representative the underlying stories are; if there are gaps or biases in the articles, the sentiment score will likely drift off course. Even though the deep-learning models, like LSTM and GRU, are cutting-edge, they don’t always nail the subtleties of the market, especially during wild swings or unexpected crises.

CHAPTER 2

LITERATURE REVIEW

2.1 Introduction

Over the years, traders, economists, and policy-makers have all dreamed of nailing down tomorrow's stock price. Dozens of classic models once dominated that chase, from shifting averages to dividend discounts. extra currently, breakthroughs in machine mastering and natural language processing have opened the door to messier, tougher-to-quantify information. Tweets, earnings call transcripts, and blog posts can now feed directly into algorithms. This chapter will discuss about research landscape, sentiment, and deep-learning techniques. It lays out the tools others have used, points out what still feels missing, and explains why this study is conducted.

2.2 Stock Market Prediction: An Overview

Predicting stock market behaviour requires analysts to look forward into tomorrow's prices by weighing whatever historical price trends, financial benchmarks, and external pressures they can muster. for many years, the ones forecasts were built commonly on fairly truthful statistical workouts like Autoregressive integrated moving average (ARIMA) and Generalized Autoregressive Conditional Heteroskedasticity (GARCH). both fashions hinge on the assumption that the underlying records keeps a few degree of linearity and stationarity. Although they retain clear advantages in interpretability and computational speed, they struggle when faced with the nonlinear spikes, regime shifts, and multidimensional clutter that financial markets routinely generate (Alamu & Siam, 2024).

The scene has shifted dramatically during the last ten years as machine learning and, more recently, deep learning have claimed centre stage in financial forecasting. traditional algorithms—help Vector Machines, choice timber, and ensemble techniques consisting of Gradient Boosting—have time and again outperformed older techniques by way of uncovering diffused systems buried in tangled facts. aid Vector Machines, as an instance, are in particular popular because they restriction overfitting even as hopefully navigating curved decision limitations (Geetha et al., 2024). Crucially, those techniques can accommodate an eclectic blend of indicators: the whole thing from rate momentum and moving averages to macroeconomic releases and social media sentiment can be folded into a single version within the desire of lifting prediction precision.

Recurrent Neural Networks (RNNs) and their more advanced offspring, especially Long-Short Term Memory (LSTM) networks, have swiftly installed themselves as powerful gear for modeling sequential economic data. The LSTM architecture excels in time-collection forecasting mainly due to the fact it is able to successfully remember information over prolonged intervals whilst largely fending off the vanishing gradient problem that hampers traditional RNNs (Ansah et al., 2022). This advantage has been empirically established; in one current head-to-head assessment, LSTM outstripped both ARIMA and help Vector gadget (SVM) fashions at the imply Squared errors (MSE) metric throughout several markets, ranging from equities to cryptocurrencies such as Bitcoin (Al-Alawi & Alshakhoori, 2024).

Alongside these stand-alone networks, hybrid frameworks that meld different machine-learning techniques are increasingly popular in the field. By fusing LSTM with techniques like LightGBM or by employing CNN-BiLSTM configurations, researchers can effectively extract both spatial and temporal patterns from the data. Such integrative methods frequently deliver superior forecasting accuracy when pitted against any single architecture (Yu, 2024). Moreover, Transformer-primarily based designs, such as the Temporal Fusion Transformer (TFT) and the LSTM-mTrans-MLP, have these days demonstrated marked gains in prediction accuracy, particularly when tasked with high-frequency datasets like the S&P 500 index (Kabir et al., 2025).

Machine-learning (ML) and deep-learning (DL) systems have come a long way, yet they still grapple with several practical drawbacks. For one, the hazard of overfitting looms huge, particularly in rapid-moving monetary environments wherein past styles quick lose relevance. further, these techniques often demand sizable education datasets and hefty computing electricity. at the same time as deep architectures can outperform simpler procedures, their opaque selection-making—frequently known as the “black box” hassle—makes it hard for buyers and analysts to comprehend why a given name changed into made (Supendi et al., 2024).

In contrast, more traditional forecasting methods, such as autoregressive integrated moving average (ARIMA) and exponential smoothing, continue to be useful when the need for clarity and computational speed trumps raw power. These time-tested models shine in stable, short-horizon projections, allowing users to follow their reasoning step by step. Yet as turbulence in the markets rises, they often fall behind, ceding ground to ML techniques designed to exploit noisy, non-stationary data.

The latest studies stress the value of melding disparate data streams—real-time news feeds, social-media sentiment, option-implied volatility—into a single predictive framework. Evidence gathered across several investigations shows that when such sentiment-rich information feeds through long short-term memory (LSTM) networks or convolutional neural networks (CNNs), forecasting accuracy improves noticeably by accounting for investor mood and attitude as drivers of price movement (Kapgate & Chaturvedi, 2025).

To summarise, traditional statistical techniques remain the bedrock of financial forecasting, yet machine-learning and deep-learning methods now provide flexible alternatives that can better accommodate the intricate dynamics of contemporary markets. The rise of hybrid architectures and transformer-based systems points to a promising trajectory for improving accuracy and informing decision-making in equity analysis.

2.3 Role of Sentiment in Financial Market

Investor sentiment, appearing as a powerful mental undercurrent, frequently steers costs in guidelines that fundamental variables can't absolutely justify. within the framework of behavioural finance, sentiment is perceived as a collective emotional kingdom—moulded with the aid of headlines, professional remark, and social media chatter—that frequently triggers overreactions, distortions, and spikes in volatility (Nyakurukwa & Seetharam, 2024). In light of this insight, students have begun to embed sentiment rankings extracted from text into forecasting models, hoping to refine predictive electricity and render marketplace behaviour more understandable.

Sentiment analysis has become an integral part of contemporary financial research, using natural language processing to extract and measure the views embedded in market-related texts. while early efforts regularly trusted primary sentiment lexicons, the field has moved in the direction of greater effective architectures consisting of Bidirectional Encoder Representations from Transformers (BERT), XLNet, and diverse hybrid deep-studying frameworks. A latest contribution by Fu and Zhang (2024) illustrates this fashion; the authors developed a BERT-derived model that converts streams of financial documents into sentiment time series, which are then paired with traditional technical indicators and fed into a two-layer long short-term memory (LSTM) network for stock forecasting. Their results revealed a clear edge over pure quantitative models, underscoring how discernible emotional signals can enhance predictive accuracy.

The blessings of sentiment evaluation enlarge beyond mere type of textual content polarity. with the aid of incorporating dictionaries crafted specifically for the finance region, researchers have all started to provide finer-grained insights. Wang (2023) leveraged SnowNLP alongside a bespoke sentiment lexicon designed for chinese inventory forums, in the end creating an index that, when mixed with past charge actions, appreciably sharpened trend forecasts. In a related study, Agarwal and Gupta (2024) showed that large language models like Llama 2, once fine-tuned on financial datasets, reached over 89 percent accuracy in unraveling sentiment subtleties, thereby leaving traditional lexicon-driven approaches at a disadvantage.

Sentiment data is increasingly being harnessed outside the traditional confines of financial reporting. Social platforms, and in particular Twitter and Reddit, now serve as immediately barometers for investor sentiment. Osman (2023) proven that when Twitter mood is fused with historical price facts thru an ensemble approach, subsequent-hour inventory forecasts attain an accuracy of seventy four.3 in step with cent. Their research observed that social sentiment, rather than structured news, exerts a greater influence on U.S. intraday price swings. Conversely, Nyakurukwa and Seetharam (2024) found that in emerging and frontier markets, signals from official news channels carry more weight than social chatter, highlighting a geographic discrepancy in sentiment's power.

As analytical models evolve, so too does their capacity for prediction. Jiang et al. (2024) unveiled AEformer, a transformer-based architecture that leverages asymmetric embeddings and tailored attention layers to track rapid sentiment shifts typical of high-frequency trading. In parallel, Li and Hu (2024) crafted a hybrid framework which extracts sentiment with XLNet, layers in popularity data and technical indicators, and channels the output into a dual BiLSTM-highway pipeline. This setup reported tangible gains not only in short-term accuracy but also in risk-adjusted performance metrics.

Even with recent progress in the field, sentiment analysis still faces several obstacles. Models performing these analyses often struggle when confronted with poor-quality data, sarcastic remarks, ambiguous phrasing, or language that is specific to a particular industry. The rapid turnaround of events in financial markets makes it even harder to interpret sentiment information in real time. On top of that, measuring how sentiment actually moves prices is tricky in markets that are already very efficient, since arbitrage traders quickly incorporate any public information, they see (Kaveri et al., 2025).

2.4 Sentiment Analysis in Stock Prediction

2.4.1 Sentiment Analysis Techniques

At its core, sentiment analysis is an essential challenge inside natural language processing (NLP) that targets to spot, pull aside, and label the opinionated pieces of text. while implemented to stock-marketplace forecasting, it seeks to turn investor moods and mental cues located in news articles, profits releases, or social-media chatter into numbers that models can work with. Those numbers are then fed into predictive systems as extra features, hopefully leading to better forecasts. To achieve this, practitioners typically lean on three broad method families: lexicon-based rules, traditional machine-learning classifiers, and modern deep-learning networks.

Lexicon-driven sentiment analysis depends on curated lists of words each assigned a positive or negative score. Because these tools are straightforward to build and light on processing power, they remain popular. Yet they frequently misread industry jargon present in financial writing. Broad lexicons such as VADER and AFINN routinely misclassify terms like “bullish” or “overweight,” failing to acknowledge that their meanings shift sharply in a trading context (Nagendra et al., 2024). To address this weakness, researchers create dedicated financial dictionaries. The catch is that assembling and updating such lists demands considerable manual effort, a chore that hampers growth. In response, recent studies have unveiled eXplainable Lexicons, or XLex, which marry transformer-driven contextual embeddings to traditional rule-based glossaries. The goal is to fuse high accuracy with clear reasoning (Rizinski et al., 2024).

Machine-learning pipelines, by contrast, build classifiers on datasets marked with sentiment labels. Popular algorithms—Naïve Bayes, Logistic Regression, Random Forest, Support Vector Machine—are trained after text is converted into numerical form via TF-IDF or Bag-of-Words. Umar (2023) reported that both Naïve Bayes and Logistic Regression gauge sentiment in financial news with impressive precision when features are extracted at the document level. Still, these methods can be hungry for manual feature tuning and vast annotated collections, making them costly in both time and data.

In recent years, deep learning approaches have transformed sentiment analysis by removing the need for manual feature engineering and by exposing subtle meaning that earlier methods often missed. amongst the earlier successes, recurrent neural networks—especially the long short-time period reminiscence units and their slimmer counterpart, the Gated Recurrent Unit—have proven powerful at following the chain of phrases over the years, which makes them a natural fit for tasks in financial sentiment analysis (Paulraj, 2024). Researchers soon began stacking layers and adding convolutional filters, so the bidirectional LSTM-CNN combination could simultaneously pick up short-range phrases and broader sentence structure, producing noticeable gains in classification accuracy (Aluvala et al., 2023).

Lately, transformer architectures have taken the lead, and models such as BERT (Bidirectional Encoder Representations from Transformers) routinely set the bar higher for many benchmarking datasets. For instance, Lin and Wang (2024) paired BERT with a lightweight transformer layer that brings together polarity scores extracted from news headlines and features extracted from transaction logs. Their hybrid system reached an impressive 60% forecasting accuracy for index movements. Despite these breakthroughs, the heavy appetite for processing power and data remains a concern that researchers must manage when scaling such models for real-world applications.

Hybrid modelling approaches are attracting considerable attention as researchers seek to leverage the complementary advantages of distinct analytical techniques. One promising example is the recently proposed MS-SSA-LSTM architecture, which fuses standards from swarm intelligence, a curated sentiment index, and a multi-layer deep learning framework to deliver brief-time period forecasts for inventory charges. preliminary opinions propose that this composite system consistently outperforms traditional single-algorithm baselines (Madhuri et al., 2024). Another noteworthy development pairs knowledge graphs with attention-enhanced long short-term memory (LSTM) networks, thereby refining the model's ability to encode financial news and its intricate ties to subsequent market dynamics (Zhang, 2023).

Yet, progress in these areas is far from straightforward. Financial commentary—whether published articles, press releases, or tweets— frequently employs sarcasm, area of interest terminology, and complex syntactic bureaucracy that prevent dependable sentiment extraction. moreover, even when preprocessing pipelines succeed in cleaning the textual content, the signals themselves remain volatile and, at times, mutually exclusive, especially when the input stream aggregates markedly different platforms such as formal news sites and informal social feed.

2.4.2 Application in Financial News

Analysing the emotional tone of financial news headlines has increasingly proven itself useful for anticipating stock price movements, especially in emerging markets like Malaysia, where prices can react with particular intensity to newly released information. By mining the sentiment woven into news stories, analysts supplement standard quantitative inputs—such as past prices and trading volumes—with richer qualitative data that better captures the market's mood. This combo of qualitative and quantitative evidence tends to strengthen the forecasting accuracy of inventory prediction algorithms, specially in fast-moving sectors wherein investor sentiment drives buying and selling behaviour.

The predictive capacity of information-derived sentiment has been demonstrated across a number of investigations, which have hired more than a few machine-mastering and deep-mastering techniques. for example, Khonde and associates (2024) combined both vector-area models and lexicon-primarily based classifiers with lengthy quick-time period reminiscence (LSTM) networks to categorise Malaysian monetary headlines as tremendous or bad. Their approach yielded an 86-in keeping with-cent accuracy charge for sentiment tagging and 80 three in line with cent at the same time as forecasting the direction of inventory actions that followed, thereby highlighting how efficaciously news sentiment can beautify conventional time-collection prediction frameworks.

latest improvements in forecasting techniques, significantly the Context-aware Bidirectional long brief-term reminiscence version—abbreviated as CAB-LSTM—have achieved noteworthy gains in accuracy via weaving collectively topic modelling with sentiment scores derived from monetary headlines. In a practical take a look at, Xiao (2024) determined that CAB-LSTM always outperformed trendy forecasting frameworks whilst deployed on news tales about Malaysian stocks, correctly forecasting each sectoral tendencies and price shifts. The model's dual capacity to sense contextual relevance while gauging the intensity of sentiment proved pivotal for securing better directional forecasts, especially during turbulent market phases.

Within the Malaysian financial landscape, sentiment has emerged as a powerful predictor both at the sectoral and the broader macroeconomic levels. Chan and colleagues (2024) documented long-lived, asymmetric impacts of sentiment on different stock prices, with especially pronounced effects in the finance and consumer sectors. Paradoxically, the technology sector demonstrated a marked insensitivity to these sentiment swings, implying that investor psychology varies significantly from one industry to the next. helping this broader view, Ho and Ong (n.d.) illustrated how sentiment alerts lifted from newspapers can nowcast Malaysian commercial enterprise self belief and forecast the trajectory of private investment growth with a lead time of two to three quarters. Their findings thus extend the relevance of sentiment analysis well beyond the confines of equity markets.

Transformer architectures have steadily gained popularity not only for their efficacy with text but also for the seamless way they can accommodate numerical inputs. In a recent study, Dong and colleagues (2024) examined how sentiment scores from social media and news comments could be blended with traditional Transformer frameworks to forecast stock price movements. They reported that this dual-stream input—sentiment data running in parallel with standard technical indicators—produced both sharper prediction accuracy and a noticeably shorter latency between data arrival and output. Their results were consistent with earlier work by Chou and Ramachandran (2021), which demonstrated that coupling time-series models with extracted news sentiment substantially cut forecast errors and made markets' responses quicker to unfolding events.

The larger picture of collective data usage has also evolved, thanks to newly minted datasets designed specifically for the financial domain. The FNSPID set conceived by Dong et al. (2024) stands out in this regard; it pairs minute-by-minute stock prices with timestamped sentiment tags, thereby granting researchers a rich, synchronized resource for training sentiment-enhanced predictive networks. Building on this momentum, Kurisinkel and co-authors (2024) underscored the value of event-driven language models that feed in fresh news as it breaks, allowing forecasts to refresh themselves in near-real time. Such an architecture points toward a markedly more flexible blueprint for financial prediction, one that aims to keep pace with the constant ebb and flow of market-relevant information.

While significant headway has been made in applying sentiment analysis to financial markets, obstacles remain. A particular hurdle in the Malaysian context is the limited availability of well-annotated sentiment datasets that capture local nuances. Without a robust pool of quality labels, there is little room to effectively train or rigorously test forecasting models. adding to this project, the predictive strength of sentiment scores frequently hinges on several external factors: the reliability of the news outlet, the specifics of the information occasion itself, and the period among publication and observable market movement. these variables signal a urgent need for studies this is both extra geographically focused and methodologically bendy.

2.5 Deep Learning for Stock Price Prediction

2.5.1 Long Short-Term Memory (LSTM)

Long-Short Term Memory (LSTM) models have emerged as a fave within the finance literature. Not like traditional forecasting methods—be it ARIMA, linear regression, or transferring averages—LSTMs are not shackled with the aid of assumptions of linearity or strict stationarity. Their architecture is designed specially to capture complicated, nonlinear relationships over prolonged time frames, making them well acceptable for predicting inventory fees. The technical edge of LSTMs lies of their gating gadgets, which carefully manage the passage of records at some point of the sequence. by doing so, they in large part overcome the vanishing gradient issue that regularly hinders preferred recurrent neural networks. As a end result, LSTM fashions can keep crucial past indicators alive whilst filtering out noise, aligning them intently with the unpredictable rhythms regular of financial markets (Gaurav et al., 2023; Zhong, 2024).

A developing frame of empirical studies highlights the effectiveness of lengthy quick-time period reminiscence (LSTM) networks for forecasting stock prices. Pan (2024) lately showed that an LSTM structure without a doubt outperformed aid vector regression, choice timber, and linear regression models whilst tested on excessive-profile equities like Apple and Tesla; the LSTM performed each a decrease imply squared errors and a higher coefficient of determination. In a exceedingly different environment, Ku et al. (2023) carried out the same approach to the Malaysian inventory market and enriched the version inputs with investor insights and numerous technical indicators. This integration now not handiest boosted prediction accuracy however additionally led to significantly better cumulative returns than those yielded by means of random choices or conventional choice strategies.

Researchers have additionally superior LSTM frameworks by means of experimenting with feature engineering and hybrid architectures. Ozupek et al. (2024) introduced a model that pairs LSTM with empirical mode decomposition and several broadly used technical indicators, yielding importantly more potent consequences for fashion forecasting. inside the area of commodity expenses, Brown et al. (2024) stated that a similar LSTM configuration produced a lower suggest absolute errors in oil-rate predictions when as compared against traditional time-series methods which include ARIMA and exponential smoothing.

LSTM certainly provide powerful predictive skills, yet they arrive with continual drawbacks. training those architectures commonly needs sizable computational strength together with large, datasets, each of which may be in quick supply, particularly for smaller economic groups. Furthermore, LSTMs behave like so-called “black boxes,” meaning that analysts and stakeholders often struggle to understand precisely how the models arrive at their forecasts—an interpretive gap that is particularly troubling in the risk-sensitive world of finance (Alamu & Siam, 2024). Encouragingly, ongoing research into explainable artificial intelligence and various model-tuning techniques is steadily enhancing both the transparency and the overall effectiveness of LSTM-based systems for financial prediction.

2.5.2 Gated Recurrent Unit (GRU)

To address some of these challenges, the Gated Recurrent Unit (GRU) has emerged as a popular alternative for time-series tasks in the financial domain. Proposed by Chung and colleagues in 2014, the GRU reduces the architecture’s computational burden by folding the separate forget and input gates into a single update gate, which in turn eliminates the need for dedicated memory cells. This streamlined design allows GRUs to be trained faster and to make predictions with lower resource overhead compared to full-scale LSTMs. As a result, the GRU is often the desired choice in scenarios stressful rapid inference, along with actual-time stock fee prediction on the unexpectedly converting Malaysian marketplace.

current proof factors to the developing effectiveness of Gated Recurrent Unit (GRU) fashions inside the field of monetary forecasting. In his 2024 evaluation, Makinde reviews that a GRU architecture, first-rate-tuned with the Adam optimization set of rules, produced a lower Root mean square error (RMSE) even as converging more quick than traditional LSTM networks while carried out to day by day stock-charge predictions in rising markets. Kalbaliyev and Szegedi (2020) corroborate those findings by noting that the GRU framework yielded superior overall performance all through extended prediction horizons, producing large buying and selling returns and smaller mistakes margins, especially whilst marketplace situations have been erratic.

Such flexibility to house the moving, nonlinear characteristics of monetary time series makes the GRU specifically appropriate for modeling Malaysian stock indices. Chen et al. (2023) recently introduced a GRU variant that integrates multi-stock functions from distinctive sectors, demonstrating that this move-enterprise facts fusion method curbs overfitting and extensively enhances model generalization. This approach is well timed for Malaysia, in which interdependencies amongst sectors can quick adjust basic market sentiment.

Hybrid models that pair GRU with other architectures have additionally started to expose promise. A 2022 take a look at titled “stock fee Prediction the use of Bi-LSTM and GRU-based Hybrid Deep gaining knowledge of technique” located that fusing the bidirectional reminiscence of Bi-LSTM with the computational performance of GRU caused higher lengthy-time period trend forecasts than either model may want to achieve in isolation across numerous stock datasets.

latest traits have demonstrated that gating mechanisms like those in gated recurrent unit (GRU) can paintings distinctly well while nested inside ensemble architectures. of their 2024 examine, He and colleagues reported on a filtering ensemble that harnesses numerous parallel GRU networks; the innovation has already translated into constant profits in forecast precision over medium to lengthy horizons. perhaps most putting, the ensemble exhibits stable performance regardless of the precise time frame concerned, a functionality that investors within the Malaysian market have located beneficial given that volatility frequently ripples outward from worldwide and nearby occasions.

but the generation isn't a common panacea. Velarde and co-authors mentioned in 2022 that GRU profits tend to flatten out whilst carried out to series displaying little temporal range; in such cases, classical approaches, along with easy moving averages, can on occasion identical or surpass state-of-the-art architectures. The researchers additionally cited that bringing in sentiment ratings derived from financial journalism and social media chatter can help close the performance hole that also lingers among the GRU and LSTM community. Their findings underscore a broader lesson: for effective forecasting in finance, depending totally on one sort of statistics or one modeling paradigm is rarely sufficient.

2.5.3 Attention-Based CNN-LSTM (ACNN-LSTM)

ACNN-LSTM architectures have begun to draw interest in the discipline of economic forecasting, mainly for stock rate predictions that draw on both textual and numerical inputs. by pairing the neighbourhood characteristic extraction power of Convolutional Neural Networks with the sequence-mastering prowess of long brief-term reminiscence units, and then augmenting the aggregate with an attention layer that highlights the maximum salient indicators, those hybrid models are capable of expand forecasts that experience extra context-sensitive. The practice of feeding sentiment-rich information headlines into the gadget along uncooked price information is a major reason why the method has shown promise.

In the first stage, CNN blocks process structured time-series information—such as price charts or fixed-length vectors obtained from embedded text—to mine for short-term trends and recurrent seasonal motifs. Next, the LSTM component steps in to unravel longer temporal dependencies, shedding light on how previously detected patterns drift and change. Attention modules, like the Convolutional Block Attention

Module (CBAM), further streamline the workflow by automatically assigning higher weights to features that have proven useful in the past. In a recent study, Li et al. (2023) integrated CBAM within their CNN-LSTM pipeline for BBVA stock data and reported an impressive Mean Absolute Error of just 0.0058 and an R^2 score of 0.9673, marks that considerably eclipse those of more conventional CNN-LSTM configurations.

Recent research on hybrid architectures has turned to combining convolutional long short-term memory (CNN-LSTM) networks with gradient-boosted tree methods and attention layers in order to leverage the character benefits of each approach. Zhu and Chen (2023) constructed a machine wherein function representations extracted through a CNN-LSTM spine have been fed into an severe gradient boosting (XGBoost) regressor. Their effects showed that this -level pipeline completed better accuracy than either the neural network alone or the tree model used separately. In a related study, Shi et al. (2022) embedded an additive attention mechanism directly into their CNN-LSTM framework, reporting gains in directional prediction performance for stock price movements.

Zhang et al. (2023) undertook a comparative analysis involving a convolutional bidirectional LSTM network augmented with attention, pitting it against both standard LSTM and CNN-LSTM models across multiple financial datasets. The attention-enhanced architecture repeatedly outperformed its peers, confirming the utility of more nuanced weightings on past inputs. Wu et al. (2024) added further evidence by evaluating a CNN-LSTM variant outfitted with what they termed an additive attention mechanism (AAM); their findings indicated an 8 to 10 percent reduction in forecast error compared to the baseline CNN-LSTM, thus highlighting the dual benefits of increased accuracy and stability.

Incorporating unstructured text data—such as news headlines—into ACNN-LSTM frameworks substantially expands their potential to gauge investor sentiment. Li and Hua (2024) showed that through embedding attention modules, the structure is capable of highlight phrases wealthy with sentiment. whilst these highlighted cues are integrated with traditional stock indicators and technical metrics, the model's generalization performance on Malaysian forecasts markedly improves. Similarly, Liu and Zhang (2022) argue the attention mechanism augments ACNN-LSTM's resilience to market volatility by allowing the system to pivot toward the most relevant financial signals that mirror prevailing sentiment.

The seamless integration of localized feature extraction, temporal processing, and real-time attention lets ACNN-LSTM exploit diverse data streams, thereby lifting forecasting accuracy. More importantly, this attention layer bolsters the model's interpretability, clearly indicating which features or time steps attract focus. Such transparency is especially critical in finance, where stakeholders demand robust explanations alongside numerical predictions.

2.6 Previous Work on Sentiment Analysis in Financial Forecasting

Sentiment evaluation has steadily received prominence in inventory forecasting partly due to the fact social media and on line news now flood the general public with opinions. Academics and market analysts are testing numerous strategies to funnel the emotional undercurrents from forums, tweets, press releases, and conference calls into algorithmic forecasts, hoping this will yield a sharper snapshot of market mood.

latest research suggests that integrating sentiment analysis with time-series inventory records can tremendously improve predictive accuracy. as an example, neural architectures which include convolutional neural networks and long short-term reminiscence groups, when fed sentiment indicators extracted from on line inventory boards, have outperformed traditional forecasting techniques (Jing et al., 2021). similarly, hybrid fashions that pair the VADER sentiment lexicon with LSTM frameworks using social media streams have yielded superior results (Dutta et al., 2021). Sulistianingsih and Martono (2024) extended these approaches to the Indonesian market, thereby demonstrating that sentiment-driven techniques retain their effectiveness across diverse linguistic and cultural settings.

On a more advanced front, transformer-based architectures like BERT have been employed to mine sentiment from news headlines and corporate earnings transcripts (Fu & Zhang, 2024; Sarkar & Shahid, 2025). The deep contextual embeddings generated with the resource of these models permit for a extra nuanced interpretation of language, translating into giant upgrades in predictive energy. but, practitioners even though deal with limitations such as casual wording, sarcasm, and the general noisiness of short-text data, challenges that underscore the necessity of thorough preprocessing and resilient language models.

Overall, the reviewed literature supports the integration of sentiment analysis in financial forecasting models, reinforcing its relevance to this study's aim of improving stock price prediction through deep learning and news sentiment.

Table 2.1 summarizes key studies that have applied sentiment analysis techniques in financial forecasting.

Author/Year	Title	Research Focus	Techniques Used
Jing et al. (2021)	Stock Forum Sentiment Analysis Using CNN for Price Movement Forecast	Shows domain-specific sentiment improves forecasting accuracy	CNN-based sentiment classification
Dutta et al. (2021)	Sentiment Analysis Using VADER and LSTM for Market Behavior Prediction	Hybrid sentiment + LSTM enhances market movement prediction	VADER + LSTM hybrid model
Khonde et al. (2024)	Forecasting Stock Trends with LSTM and News Sentiment Analysis	Validates value of integrating sentiment with sequential models	Sentiment classification + LSTM
Sulistianingsih & Martono (2024)	LSTM-CNN Hybrid with Lexicon Scoring for Indonesian Stock Forecast	Combines emotional tone and time-series for more stable predictions	Lexicon-based scoring + LSTM-CNN

Zhu et al. (2023)	Attention-Based CNN-LSTM-XGBoost for Stock Price Forecasting	Leverages attention to improve nonlinear sequence learning for price movement	Attention-based CNN-LSTM + XGBoost
Fu & Zhang (2024)	BERT–LLA for Sentiment and Technical Indicator Fusion	Demonstrates strong performance in combining text and numeric trends for forecasting	BERT–LLA (BERT + Lookahead Attention)

Table 2-1 Previous Work Study

2.7 Research Gap

While a lot of research has demonstrated the effectiveness of sentiment-based forecasting models in Western financial markets, there is a notable gap in the application of such methods in Southeast Asian contexts. Specifically, the Malaysian banking sector remains underexplored in this regard. Most existing studies either focus on global indices or rely solely on historical numerical data, overlooking the predictive potential of localized sentiment data.

Moreover, limited research has compared multiple deep learning models (LSTM, GRU, ACNN-LSTM) within a unified framework applied to Malaysian banks. This study aims to bridge this gap by analyzing the predictive relationship between news sentiment and stock prices of CIMB and Maybank, offering a comprehensive and context-specific contribution to the literature.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

This chapter will explain about the methodology and model used for this research. The goal is to predict the stock prices using sentiment analysis of financial news headlines. The developed model will incorporate both the stock price movement and text sentiments based on financial news articles. By integrating deep learning and sentiment analysis, the model aims to predict future price movement of stocks for CIMB and MAYBANK. This chapter will discover the model architecture, data processing techniques and the steps involved in training the model.

3.2 The Framework

This research consists of seven phases which are:
Phase 1: Problem Identification
Phase 2: Data Collection
Phase 3: Data Pre-processing
Phase 4: Exploratory Data Analysis (EDA)
Phase 5: Modelling
Phase 6: Model Evaluation
Phase 7: Deployment
The details of the research framework for this study are shown in the Figure 1 below

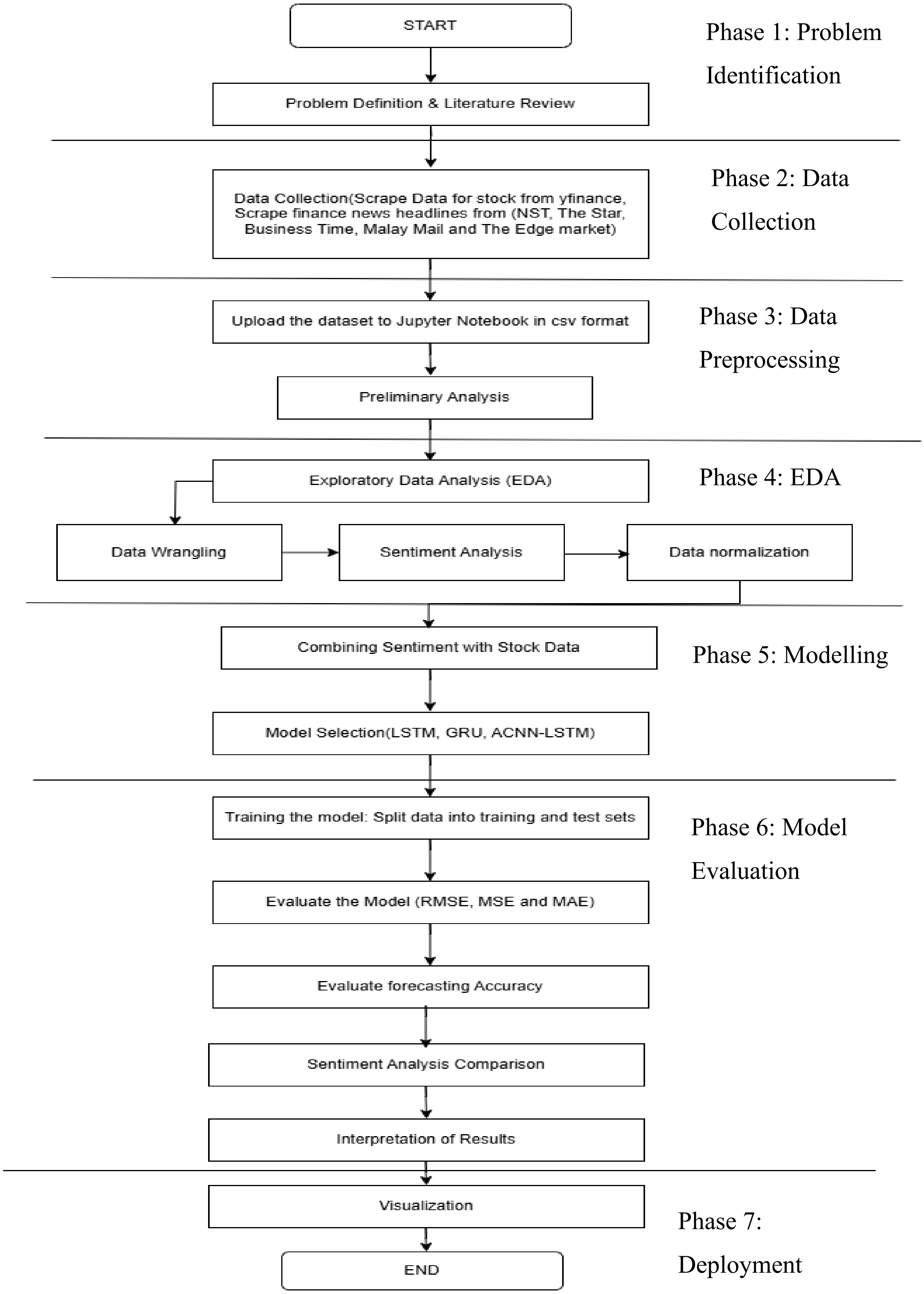


Figure 3-1Research framework of Predicting stock market using sentiment analysis

3.3 Problem Identification

The primary goal of this research is to increase a system studying version that leverages sentiment evaluation of financial news headline to are expecting stock market movement. By using deep learning techniques including LSTM, GRU and ACNN-LSTM, this looks at objectives to capture the relationship among the sentiment in the information and inventory charge trends.

But there are several demanding situations need to be solved to ensure the version accuracy.

1. Making sure the facts great and consistency is the main problem as the dataset want to be complete. For these studies, the dataset will include stock fees and economic information headlines related to Maybank and CIMB from 2019 to 2025. The dataset must not have missing value and must properly handle. Only financial headlines and topic related to it are chosen to ensure the relevancy of the dataset.

2. The model must accurately capture the sentiment (positive, neutral and negatives) based on the financial news headlines. The use of sentiment analysis models such as VADER is important to make sure the financial news is properly interpreted. To train the model, the sentiment scores need to be combine with historical stock price data.

3. It is critical to account for the dynamic nature of the stock market. This because the stock prices movement are influenced by the geopolitical events, economic policies, and global market sentiments even it is not usually address in news headlines. The model designed must be able to adapt to market evolves also must consider the external factors.

3.4 Data Collection and Preprocessing

There are two datasets used in this study which are the stock price datasets and financial news headline. The stock prices datasets were extracted from CIMB and Maybank, two main blue-chip companies in banking sector in Bursa Malaysia. Both are chosen because of the financial performance and reputation. The yfinance, a python library tools that allow to access Yahoo Finance's stock data is used to fetch the stock price data. The datasets include daily stock prices open, high, low, close prices, and volume for the years 2019 to 2025. The data consist of 35384 records for both CIMB and MAYBANK

```
[5]: df=pd.read_csv('MAYBANK_YF.csv')
     df.head()
     print(df.head())
     df.tail()
     print(df.tail())
```

```
                        Date      Open      High       Low     Close   Volume  \
0  2019-01-02 00:00:00+08:00  5.940646  5.940646  5.865606  5.871860  4096900
1  2019-01-03 00:00:00+08:00  5.828086  5.909379  5.828086  5.853099  6983900
2  2019-01-04 00:00:00+08:00  5.821834  5.865607  5.821834  5.846847  4276600
3  2019-01-07 00:00:00+08:00  5.859353  5.946900  5.859353  5.915633  7580400
4  2019-01-08 00:00:00+08:00  5.934393  5.940646  5.903126  5.921886  6692800

   Dividends  Stock Splits
0        0.0           0.0
1        0.0           0.0
2        0.0           0.0
3        0.0           0.0
4        0.0           0.0
                           Date  Open  High   Low  Close    Volume  Dividends  \
1568  2025-05-28 00:00:00+08:00  9.87  9.89  9.82   9.84   8025100        0.0
1569  2025-05-29 00:00:00+08:00  9.85  9.89  9.82   9.87   8959200        0.0
1570  2025-05-30 00:00:00+08:00  9.89  9.90  9.78   9.78  29844200        0.0
1571  2025-06-03 00:00:00+08:00  9.80  9.84  9.70   9.76  10002900        0.0
1572  2025-06-04 00:00:00+08:00  9.76  9.78  9.71   9.72   8806600        0.0

      Stock Splits
1568           0.0
1569           0.0
1570           0.0
1571           0.0
1572           0.0
```

Figure 3-2 The code and information first 5 rows and last 5 rows of Maybank stock data

Same process was conducted for different dataset. Below are summarized of the data

Dataset Name	Source	Format	Number of rows
CIMB_YF	yfinance	.csv	1581
MAYBANK_YF	yfinance	.csv	1581

Table 3-1 Dataset for Stock Market

Second, Selenium is used as scrapping tools for financial news headlines from several finance news websites such as New Straits Times, The Star, Malay Mail, The Edge Market and Business Today. The Selenium can handle web pages that require user interaction like clicking next button or scrolling to load more content. Besides, the use of BeautifulSoup is limited for the websites that use JavaScript to load headlines and content. There is total 32222 headlines scrapped from this website related to CIMB, Maybank, business, and finance.

```
[3]: import pandas as pd
     df=pd.read_csv('MAYBANKFULL.csv')
     df.head()
     print(df.head())
     df.tail()
     print(df.tail())
```

```
                                            headline          date
0  Maybank Upgrades Capital A To "Buy" On Stronge...  June 2, 2025
1  Analysts Raise Concerns Over Maybank's Mid-Ter...  May 27, 2025
2  Today's Shares: Maybank Shares Dip 0.5% Amid M...  May 27, 2025
3  Solid Start For Maybank With Q1 Profit Rising ...  May 26, 2025
4  Maybank Signs LOI To Finance RM2.35 Billion Of...  May 19, 2025
                                                headline         date
16636  Bursa Malaysia ends higher on 11th-hour window...  28 Dec 2018
16637             Bursa Malaysia stays in red at mid-day  28 Dec 2018
16638  Bursa Malaysia stays in negative territory at ...  28 Dec 2018
16639                        Bursa Malaysia opens lower  28 Dec 2018
16640  Bursa Malaysia reacts positively to Wall Stree...  27 Dec 2018
```

Figure 3-3 The code and Information of first 5 rows and last 5 rows of Maybank news headlines

All the data uploaded in Jupyter Notebook for further analysis. Initial analysis dataset is performed to check the data structure, types, and missing value. Any duplicated data are handled properly.

Dataset Name	Source
CIMB_MMORI	https://www.malaymail.com/
CIMB_NSTORI	https://www.nst.com.my/business
CIMB_TSORI	https://www.thestar.com.my/
CIMB_BTORI	https://www.businesstimes.com.sg/
CIMB_TEMORI	https://theedgemalaysia.com/
CIMBALL	Merged all the dataset of CIMB headlines
MAYBANK_MMORI	https://www.malaymail.com/
MAYBANK_NSTORI	https://www.nst.com.my/business
MAYBANK_TS	https://www.thestar.com.my/
MAYBANK_BT	https://www.businesstimes.com.sg/

MAYBANK_TEM	https://theedgemalaysia.com/
MAYBANKFULL	Merged all the dataset of Maybank headlines

Table 3-2 Dataset for news headlines

3.5 Data Cleaning and Sentiment Analysis

The data cleaning process is proceeded after the dataset was uploaded and analysed. This step to ensure the quality of the datasets before used for machine learning training.

The unnecessary columns, news headlines and row are removed during the data wrangling. The headlines are cleaned by removing the special characters, numbers, and punctuation. The stock price data is then aligned with the news headline based on their timestamps, ensuring that each headline corresponds to the correct stock price data. The numbers of frequent words used in news and stock datasets are identify during EDA. Besides, it helps to understand trends, patterns, and common terms in the financial headlines.

News headlines are analysed using natural language toolkit (NLTK) sentiment VADER. It is a prebuilt tools used to extract sentiment from text data. The VADER model calculates the polarity value and categorized into Positive (>0.1), Negative (<-0.1) and Neutral (>-0.1, <0.1). The SentimentIntensityAnalyzer from VADER is applied to calculate the sentiment score for each headline. The score is then used to identify the sentiment type and to predict the stock price movements.

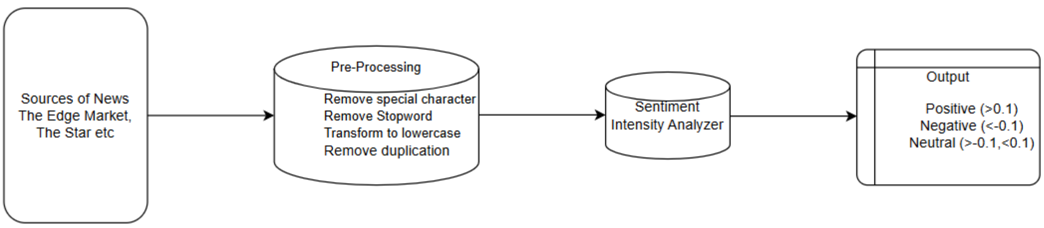


Figure 3-4 Sentiment Analysis Process FlowFigure

The data normalization step begin after the sentiment scores are determined. Stock prices data need to be normalized to ensure the value ranges (between 0 and 1) is not biased before use them for machine learning.

3.6 Model Selection and Evaluation

The stock price data is then combined with the cleaned and processed sentiment score dataset. The datasets are used for model training. The sentiment values are aligned according to the date of events.
Three deep learning model are selected for this study:

- LSTM (Long Short-Term Memory)
  A LSTM model is selected to provide prediction for the stock prices. Its ability to remember long-term sequence data is better compare to other machine learning algorithm thus it is suitable for time-series forecasting. As stock prices data are influenced by the long period of historical data, the application of LSTM algorithm can identify the long-range relationship. The stock prices datasets will be divided to 80% for training and the remaining 20% for testing. The model will be using in the Keras package in Python.

- GRU (Gated Recurrent Unit)
  GRU model will be used for comparison. The model is capable to learn long-term dependencies even with simpler architecture. It also has fast computational speed. The parameter used for this model are fewer compare to LSTM which is more efficient to train a large dataset. GRU is expected to perform like LSTM due to the simpler model result to quicker iterations. The same training data split will be used. However, the hyperparameters will be adjusted for optimization.

- ACNN-LSTM (Attention Convolutional Neural Network-LSTM):
  This hybrid model also will deploy in this project. The model combines the CNN ability in extracting text features from news headlines with LSTM power to learn temporal patterns of stock market movement. The

attention mechanism allows to focus on the important parts of the news. It will improve the prediction accuracy. This approach is useful when working with textual (new headlines) and sequential data (stock prices).

The data is split into training and test set. The training set is used to teach the models the relationship between sentiment and stock prices. The test set is used to evaluate the model capability adapting to new data. There are a lot of external factors to consider to ensure the model optimization.

The Root Mean squared Error (RMSE) is used to evaluate the efficiency of the model to predict the stock prices. This model is commonly used in time series prediction. RMSE will calculate the differences between the observed prices and the predicted values. The smaller the RMSE, the closer the prediction to the actual value. Therefore, the RMSE will be use to assist in predicting prices of stock market. The dropout layer is added to each LSTM layer to avoid overfitting. Mean Squared Error (MSE) and Mean Absolute Error (MAE) are also use to assist the evaluation for model performances for broader view of model performances in predicting stock market

CHAPTER 4

INITIAL FINDINGS

4.1 Introduction

This chapter will discuss about the accuracy of the model to predict stock market prices for Malaysian banking (CIMB and Maybank) based on sentiment analysis of news headline using machine learning. This chapter begin with the step about data collection, preprocessing, exploratory data analysis (EDA), developing model and implementing model using machine learning techniques. The machine learning used in this project are LSTM, GRU and ACNN-LSTM. Based on the results of machine learning, it was found that the ACNN-LSTM model techniques had highest percentage of accuracy compared to others model. The details of the results and analysis are presented in this chapter.

4.2 Data Collection

To collect news article related to CIMB and Maybank, the data collection process was carried out using the multi-threaded Selenium-based scraping framework. The news headlines data was retrieved from five major websites which are Malay Mail, The Edge Market, Business Today, New Straits Time and The Star. Stock price data was obtain using the Yahoo Finance (yfinance) for both CIMB and Maybank. After the data was collected, the data was saved in CSV file format to facilitate for further analysis.

The total data collected are display below:

Bank	Total Articles Scraped	Sources
CIMB	14 369	Malay Mail, The Edge Market, Business Today,
Maybank	17 853	

		New Straits Time, The Star

Table 4-1 Dataset of News Headline

Each record includes the news headline, published date and category/section.

	headline	date	category
1	99 Speedmart's Extended Hours Could Enhance Groups SSSG: CIMB	June 9, 2025	NEWS
2	CIMB Thai Shares Suspended Over Public Float Rules	June 6, 2025	NEWS
3	CIMB Remains Concerned On 7-Eleven's Long Term Outlook	June 6, 2025	NEWS
4	CIMB Has Tools To Navigate Tariff Headwinds	June 3, 2025	NEWS
5	Today's Shares: CIMB Slips 1.15% Despite Analysts' Optimism Post-1QFY25 Results	June 3, 2025	MARKETS

Figure 4-1News Headline data

Yahoo Finance (via yfinace library) was used to scraped historical stock data between 2019-01-01 and 2025-12-31

Bank	Total Daily Records	Ticker
CIMB	1581	1023.KL
Maybank		1155.KL

Table 4-2 Yahoo Finance Dataset

All the data are compress in RNDPM file and will be upload to google colab for further analysis.

4.3 Data Preprocessing

Data preprocessing was conducted before sentiment analysis and correlation with financial stock data.

4.3.1 Date Formatting and Standardization

- News Headline

  The date format for news portal is vary due to the different data structure of news websites. Some entries contained delimiters such as pipes (|) or symbols (@) which are used to separate date and time elements. To standardise this, it was split into separate Date and Time columns. The Date values were then reformatted into ISO format (YYYY-MM-DD) for consistency across all the datasets.

- Stock Market Data

  Yahoo finance date column contains both time and date data. The date and time string were further split into Date and Time column. The Date was formatted but the time column is removed because it was redundant data.

4.3.2 Column Reduction

- News Headline

  The Time column was removed from news articles datasets. It is because the Time column was not relevant in this study.

- Stock Market Data

  Yahoo Finance contain Dividends and Stock Split columns. Both columns are not needed for this analysis thus it was dropped during preprocessing.

4.3.3 Category Filtering

To improve the quality of sentiment analysis, only financial related articles were kept. Other than that, are discard to reduce noise from the data. For example, in The Edge Markets, categories such as 'Market close,' 'Hot stock' and 'Market Open' were retained while 'Aviation' and 'Branded' were removed. The same filter was also applied to others news dataset. Additionally, the section column in Malay Mail was renamed to category. It is to maintain the consistency across sources.

4.3.4 Dataset Merging, Re-Indexing and Null Handling

After all the individual files were cleaned, they were grouped by bank (CIMB and Maybank). The group was then merged using pandas.concat() into two unified datasets. Then, the merge datasets only retained the record between January 1, 2019 and June 13, 2025. This is to ensure the alignment with the available stock data for the comprehensive analysis. Any null values in the Date and Category column were dropped.

<table>
<tr><th>Preprocessing step</th><th>Before preprocess</th><th>After preprocess</th></tr>
<tr><td>Reformat date and drop column</td><td><table><tr><th>headline</th><th>date</th><th>category</th></tr><tr><td>Trading in CIMB Thai suspended for not meeting minimum public spread</td><td>6 June 2025 | 05:29 PM</td><td>MARKETS</td></tr></table></td><td><table><tr><th>headline</th><th>category</th><th>Date</th></tr><tr><td>Trading in CIMB Thai suspended for not meeting minimum public spread</td><td>MARKETS</td><td>2025-06-06</td></tr></table></td></tr>
<tr><td>Reformate date and drop column</td><td><table><tr><th>Date</th><th>Open</th><th>High</th><th>Low</th><th>Close</th><th>Volume</th><th>Dividends</th><th>Stock Splits</th></tr><tr><td>2019-01-02 00:00:00+08:00</td><td>4.320346270049922</td><td>4.335452362037395</td><td>4.244815810114329</td><td>4.267475128173828</td><td>6022200</td><td>0.0</td><td>0.0</td></tr></table></td><td><table><tr><th>Date</th><th>Open</th><th>High</th><th>Low</th><th>Close</th><th>Volume</th></tr><tr><td>2019-01-02</td><td>4.320346270049922</td><td>4.33545236203704</td><td>4.244815810114329</td><td>4.267475128173828</td><td>6022200</td></tr></table></td></tr>
<tr><td>Category Filtering</td><td>1 to 10 of 2825 entries Filter<table><tr><th>headline</th><th>date</th><th>category</th></tr><tr><td>Asian currencies, stock slip on risk-off after Israel strikes Iran</td><td>2025-06-13</td><td>Emerging Markets</td></tr><tr><td>Maybank IB sees upside for CPE Tech from global wafer fab construction boom in first coverage</td><td>2025-06-12</td><td>Stock Focus</td></tr><tr><td>DayOne secures RM15b multicurrency financing for data centres in Johor</td><td>2025-06-11</td><td>Tech</td></tr><tr><td>SST expansion likely to weigh on property developers, not contractors, says Maybank IB</td><td>2025-06-11</td><td>Sector Focus</td></tr><tr><td>Analysts upbeat on M-REITs' dividend yields amid easing rate outlook</td><td>2025-06-10</td><td>Sector Focus</td></tr><tr><td>Bridging Ambition & Achievement Through CFA® Program</td><td>2025-06-10</td><td>Branded</td></tr><tr><td>Paradigm REIT's IPO sees flat close amid broader market dip</td><td>2025-06-10</td><td>New Listing</td></tr><tr><td>Convenience store segment riding high but consumer sentiment could soften as costs rise</td><td>2025-06-09</td><td>Edge Weekly</td></tr><tr><td>A rare high-rental-yield investment</td><td>2025-06-07</td><td>Branded</td></tr><tr><td>Asian currencies poised for weekly gains; rate cut lifts Indian equities</td><td>2025-06-06</td><td>Emerging Markets</td></tr></table></td><td>1 to 10 of 2441 entries Filter<table><tr><th>headline</th><th>date</th><th>category</th></tr><tr><td>Asian currencies, stock slip on risk-off after Israel strikes Iran</td><td>2025-06-13</td><td>Emerging Markets</td></tr><tr><td>Maybank IB sees upside for CPE Tech from global wafer fab construction boom in first coverage</td><td>2025-06-12</td><td>Stock Focus</td></tr><tr><td>DayOne secures RM15b multicurrency financing for data centres in Johor</td><td>2025-06-11</td><td>Tech</td></tr><tr><td>SST expansion likely to weigh on property developers, not contractors, says Maybank IB</td><td>2025-06-11</td><td>Sector Focus</td></tr><tr><td>Analysts upbeat on M-REITs' dividend yields amid easing rate outlook</td><td>2025-06-10</td><td>Sector Focus</td></tr><tr><td>Convenience store segment riding high but consumer sentiment could soften as costs rise</td><td>2025-06-09</td><td>Edge Weekly</td></tr><tr><td>Asian currencies poised for weekly gains; rate cut lifts Indian equities</td><td>2025-06-06</td><td>Emerging Markets</td></tr></table></td></tr>
<tr><td>Merging</td><td>CIMB_BT_ORI.csv CIMB_TEMORI.csv Maybank_BTORI.csv Maybank_TEMORI.csv
CIMB_MMORI.csv CIMB_TSORI.csv Maybank_MMORI.csv Maybank_TSORI.csv
CIMB_NSTORI.csv CIMB_YF.csv Maybank_NSTORI.csv Maybank_YF.csv</td><td>✅ 'CIMB_TEMORI.csv': Columns renamed and filtered
✅ 'cleaned_CIMB_BT_ORI.csv': Columns renamed and filtered
✅ 'cleaned_CIMB_MMORI.csv': Columns renamed and filtered
✅ 'cleaned_CIMB_TSORI.csv': Columns renamed and filtered
✅ 'cleaned_CIMB_NSTORI.csv': Columns renamed and filtered

🎉 CIMB: Final file saved to 'RNDPM3/CIMBALL.csv'
✅ 'Maybank_TEMORI.csv': Columns renamed and filtered
✅ 'cleaned_Maybank_BTORI.csv': Columns renamed and filtered
✅ 'cleaned_Maybank_MMORI.csv': Columns renamed and filtered
✅ 'cleaned_Maybank_TSORI.csv': Columns renamed and filtered
✅ 'cleaned_Maybank_NSTORI.csv': Columns renamed and filtered

🎉 Maybank: Final file saved to 'RNDPM3/MaybankALL.csv'</td></tr>
<tr><td>Re-Indexing</td><td></td><td>1 to 10 of 9035 entries Filter<table><tr><th>headline</th><th>Date</th><th>category</th></tr><tr><td>Bursa lower at Friday opening</td><td>2019-01-04</td><td>MONEY</td></tr><tr><td>Best Non-IPO Fundraising (Sukuk/Bond): Yinson's RM950 mil perpetual sukuk a first for the oil and gas industry</td><td>2019-01-07</td><td>Edge Weekly</td></tr><tr><td>Bursa Malaysia higher at opening</td><td>2019-01-07</td><td>MONEY</td></tr></table>9031 to 9035 of 9035 entries Filter<table><tr><th>headline</th><th>Date</th><th>category</th></tr><tr><td>Research houses maintain 4.0% GDP growth forecast for 2025</td><td>2025-06-12</td><td>ECONOMY</td></tr><tr><td>Bursa Malaysia declines at midday as Washington's tariff plan weighs on sentiment</td><td>2025-06-13</td><td>CORPORATE</td></tr><tr><td>Today's Shares: Bermaz Auto Falls 9.7% To RM0.84 Amid Analyst Downgrades, Competitive Pressures</td><td>2025-06-13</td><td>MARKETS</td></tr></table></td></tr>
</table>

Table 4-3 Preprocessing Input and Output of Dataset

4.4 Exploratory Data Analysis (EDA)

Data preprocessing was conducted before sentiment analysis and correlation with financial stock data.

4.4.1 Data Cleaning

Data cleaning is one of the most important steps when dealing with web-scraped data. It is due to the raw data of news headlines are usually have inconsistent formatting, contain missing values, noise, and a redundant field. By cleaning the data, the quality of sentiment analysis and statical modelling are expected to improved.

All character were converted to lowercase to ensure the consistency. It also will help in reducing the lexical duplication. The digits, punctuation and other non-alphabetic characters are eliminated by using the regular expressions. Extra spaces and tabs were removed from the text. NLTK was deployed to removed the English stop words and retain only the meaningful words. Any missing headline values were replaced with empty string. This will prevent the processing errors during analysis. This process was applied to CIMB and Maybank dataset. To reduce noise and redundancy, only three columns are kept ensuring the efficiency of analysis and modelling

1 to 10 of 9035 entries Filter

headline	Date	category
bursa lower friday opening	2019-01-04	MONEY
best nonipo fundraising sukukbond yinsons rm mil perpetual sukuk first oil gas industry	2019-01-07	Edge Weekly
bursa malaysia higher opening	2019-01-07	MONEY
bursa malaysia remains higher midmorning	2019-01-07	MONEY
bursa malaysia remains positive territory midday	2019-01-07	MONEY
bursa malaysia remains higher midafternoon	2019-01-07	MONEY
best privatisation timely sweet exit oldtown shareholders	2019-01-08	Edge Weekly
best share placement khazanah sells rm mil cimb block tight discount	2019-01-08	Edge Weekly
disappointments year year full disappointments bursa	2019-01-08	Edge Weekly
bursa marginally higher opening	2019-01-08	MONEY

Figure 4-2 Cleaned dataset of News headline

4.1.1 Sentiment Analysis

Sentiment analysis is performed using VADER (Valence Aware Dictionary and Sentiment Reasoner) tool. The tool is capable to quantify the emotional tone of the news headline and it effective for domain-specific analysis like financial news. Each headline was passed through VADER's sentiment analyser. The analyser will categorise the score into positive, neutral and negatives. Table below show the compound score that used in the classification of the sentiment label

Compound Score	Sentiment label
>0.05	Positive
<-0.05	Negative
Else	Neutral

Table 4-4 Sentiment Analysis Compound score

The analysis is applied to CIMB and Maybank headlines dataset.

1 to 10 of 9035 entries Filter

headline	Date	category	Sentiment
bursa lower friday opening	2019-01-04	MONEY	-0.296
best nonipo fundraising sukukbond yinsons rm mil perpetual sukuk first oil gas industry	2019-01-07	Edge Weekly	0.6369
bursa malaysia higher opening	2019-01-07	MONEY	0.0
bursa malaysia remains higher midmorning	2019-01-07	MONEY	0.0
bursa malaysia remains positive territory midday	2019-01-07	MONEY	0.5574
bursa malaysia remains higher midafternoon	2019-01-07	MONEY	0.0
best privatisation timely sweet exit oldtown shareholders	2019-01-08	Edge Weekly	0.802
best share placement khazanah sells rm mil cimb block tight discount	2019-01-08	Edge Weekly	0.5423
disappointments year year full disappointments bursa	2019-01-08	Edge Weekly	-0.7184
bursa marginally higher opening	2019-01-08	MONEY	0.0

Figure 4-3 Sentiment Analysis of News Headline

Daily sentiment scores are calculated by averaging all news headlines for each day before combining it with stock prices data. The total of positive, negative, and neutral headlines is also calculated.

1 to 10 of 1871 entries Filter

Date	Open	High	Low	Close	Volume	Dividends	Stock Splits	Avg_Sentiment	Max_Sentiment	Min_Sentiment	Headline_Count	Positive_Headlines	Negative_Headlines	Price_Up
2019-01-02	4.320346270049922	4.33545236203704	4.244815810114329	4.267475128173828	6022200.0	0.0	0.0							0
2019-01-03	4.2901340860756845	4.320346270049922	4.244815810114329	4.267475128173828	5198100.0	0.0	0.0							0
2019-01-04	4.252368927001953	4.282581110200455	4.244815701123511	4.252368927001953	3506600.0	0.0	0.0	-0.296	-0.296	-0.296	1.0	0.0	1.0	1
2019-01-07	4.282581346314875	4.312793531179086	4.259922027587891	4.259922027587891	10358800.0	0.0	0.0	0.23886	0.6369	0.0	5.0	2.0	0.0	1
2019-01-08	4.282581220160947	4.312793404135183	4.259921902101448	4.267475128173828	11516600.0	0.0	0.0	0.008655555555555555	0.802	-0.7184	9.0	3.0	4.0	0
2019-01-09	4.282581220160947	4.297687312148065	4.259921902101448	4.267475128173828	12079200.0	0.0	0.0	0.2361833333333333	0.802	0.0	6.0	3.0	0.0	0
2019-01-10	4.275027994088567	4.297687312148065	4.244815810114329	4.267475128173828	8114200.0	0.0	0.0	0.2553	0.5106	0.0	2.0	1.0	0.0	0
2019-01-11	4.275027994088567	4.2901340860756845	4.244815810114329	4.267475128173828	6385100.0	0.0	0.0	0.0	0.0	0.0	2.0	0.0	0.0	0
2019-01-14	4.267475128173828	4.2901340860756845	4.244815810114329	4.267475128173828	4324900.0	0.0	0.0	0.0638	0.4215	-0.4404	6.0	3.0	2.0	1
2019-01-15	4.267474286663064	4.305239329105084	4.252368197654742	4.282580375671387	9463200.0	0.0	0.0	0.23091	0.6369	-0.296	10.0	5.0	1.0	0

Figure 4-4 Daily Sentiment Dataset

The daily sentiment analysis data was then merged with daily stock prices data for CIMB and Maybank using the date as a key. The missing stock value were forward filled while missing sentiment value are set to zero. The clean dataset for each bank then containing both sentiment and financial information and can be used for further analysis.

11 to 20 of 1871 entries Filter

Date	Open	High	Low	Close	Volume	Avg_Sentiment	Max_Sentiment	Min_Sentiment	Headline_Count	Positive_Headlines	Negative_Headlines	Price_Up	No_News
2019-01-16	4.282581220160947	4.282581220160947	4.259921902101448	4.267475128173828	5718700.0	-0.008125	0.6486	-0.4404	8.0	2.0	3.0	0	0
2019-01-17	4.267475128173828	4.267475128173828	4.244815810114329	4.267475128173828	6605800.0	0.0683	0.2732	0.0	4.0	1.0	0.0	1	0
2019-01-18	4.267474286663064	4.290133240096762	4.252368197654742	4.282580375671387	5980300.0	0.1307	0.5574	-0.296	2.0	1.0	1.0	0	0
2019-01-20	4.267474286663064	4.290133240096762	4.252368197654742	4.282580375671387	5980300.0	0.2732	0.2732	0.2732	1.0	1.0	0.0	0	0
2019-01-21	4.267474286663064	4.290133240096762	4.252368197654742	4.282580375671387	5980300.0	0.2263	0.2263	0.2263	1.0	1.0	0.0	0	0
2019-01-22	4.320346446776484	4.35811185836792	4.282581395342706	4.35811185836792	16396400.0	0.1103857142857142	0.2732	0.0	7.0	3.0	0.0	0	0
2019-01-23	4.320346025316835	4.343005342092759	4.275027751922607	4.275027751922607	12430100.0	0.0793799999999999	0.7003	-0.5994	5.0	2.0	1.0	0	0
2019-01-24	4.2901340860756845	4.312793404135183	4.244815810114329	4.267475128173828	13410700.0	0.1646333333333333	0.4939	0.0	3.0	1.0	0.0	1	0
2019-01-25	4.297687178932736	4.297687178932736	4.267474995894985	4.290133953094482	11523500.0	0.0	0.0	0.0	1.0	0.0	0.0	0	0
2019-01-28	4.305239934185425	4.305239934185425	4.259921660791198	4.275027751922607	10608400.0	-0.2047333333333333	0.0	-0.3182	3.0	0.0	2.0	1	0

Figure 4-5 Sentiment Analysis with Stock Market

4.1.2 Descriptive Statistics

After preprocessing, each dataset contains 14 columns and 1871 rows for CIMB and 1876 rows for Maybank. The dataset ranges from 2 January 2019 to 13 June 2025 and has no null value. The date is also aligned for both sentiment and stock data.

- Stock price distribution

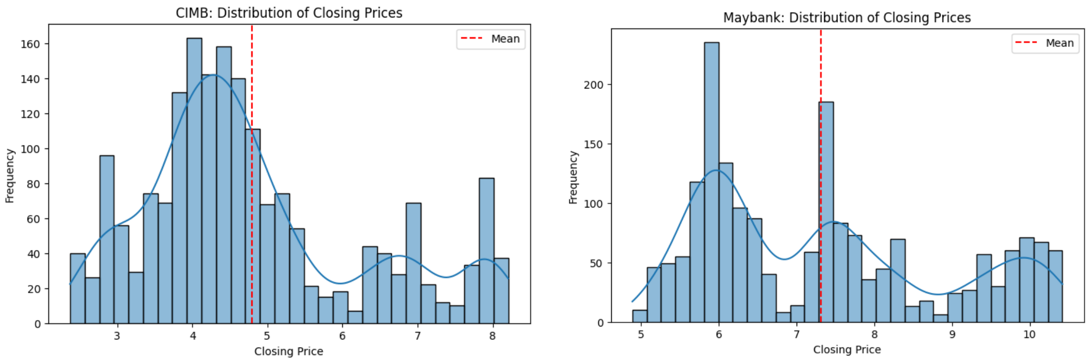


Figure 4-6 Stock price distribution chart for CIMB and Maybank

Based on the graph, CIMB showed a unimodal distribution centered around RM 4.25 with lower volatility. Maybank on the other hand displayed a wider spread and it peaked around RM 5.85-RM 6.20 reflecting a broader trading range

- Correlation heatmaps

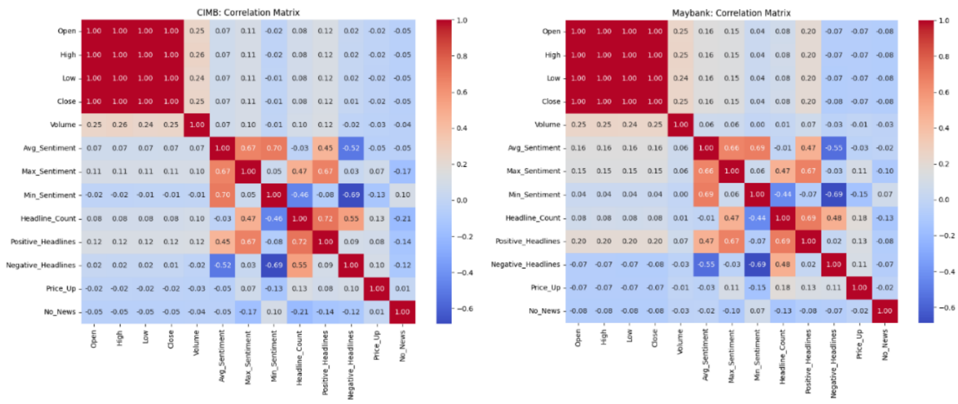


Figure 4-7 Correlation heatmaps for CIMB and Maybank

The heatmap indicated that sentiment indicators such as Avg_Sentiment, Positive_Headlines, Negatives_Headlines had weak positive correlations with price variables. This indicated that sentiment of the news headlines is not directly influences the price stock market.

- Time Series for Price vs Sentimet

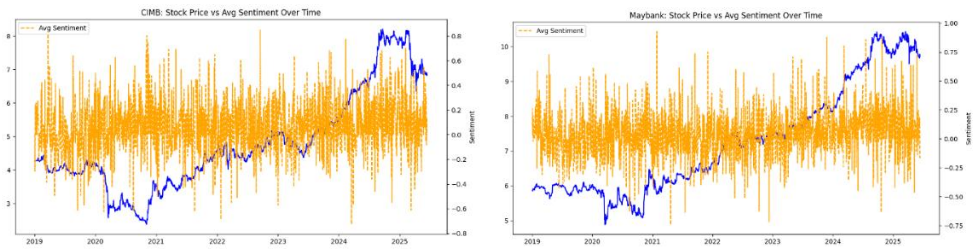


Figure 4-8 Time Series Chart for Price vs Sentimet for CIMB and Maybank

Based on the time series plot, CIMB showed no strong pattern between sentiment and Price. Maybank on the other hand has increases in average sentiment with the upward trends. Maybank showed strong evidence of the relation between sentiment and prices.

- Sentiment Score Distribution

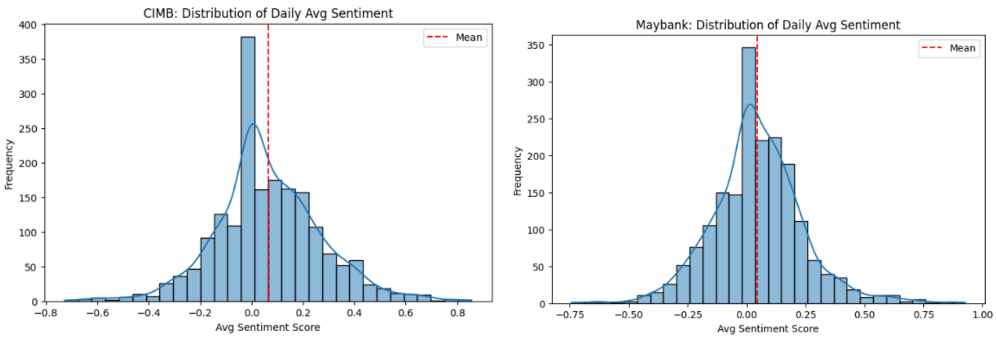


Figure 4-9 Sentiment Score Distribution for CIMB and Maybank

The distribution graph showed that CIMB sentiment distribution was mostly around 0 indicated most of the news are natural. Maybank slightly skewed to the right proving there are more positive sentiment in its news coverage.

- Headline Volume vs Stock Price

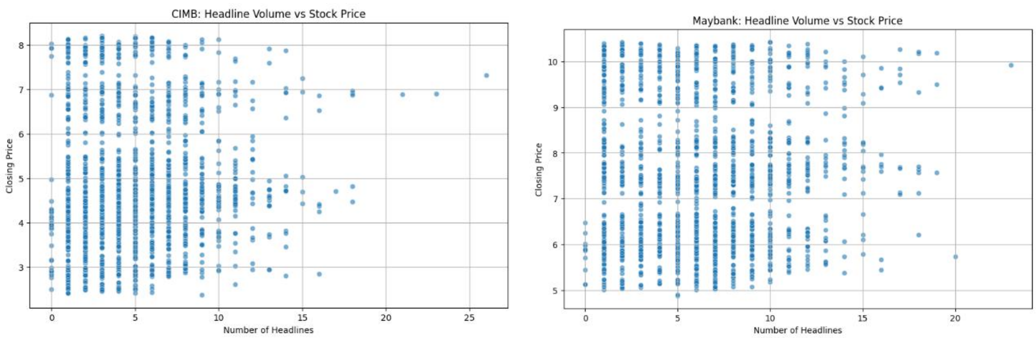


Figure 4-10 Headline Volume vs Stock Price for CIMB and Maybank

Scatterplots between headline count and close price showed that there is no linear trend between these two variables. However major headline volume spikes in Maybank are followed with stock movement activity. This validate that headline count is one of the contributors to market relevant events.

- Sentiment and the closing price

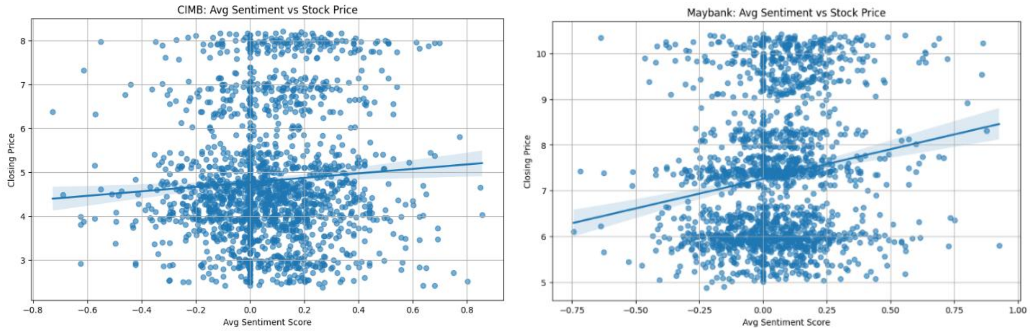


Figure 4-11 Sentiment and the closing price for CIMB and Maybank

The graph revealed that the relationship between Avg_Sentiment and Close price are flat and weak for CIMB. For Maybank, the slope was inclined a bit indicated that higher sentiment related with stock prices market.

- Feature Importance Summary for Stock Price Prediction

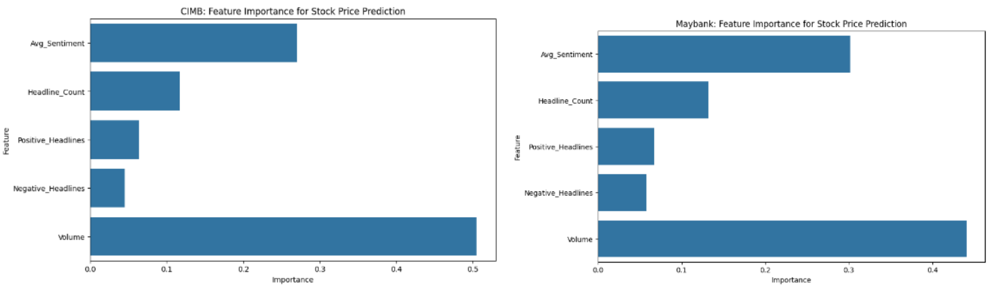


Figure 4-12 Feature Importance Summary for Stock Price Prediction

The bar charts above summarize which features were most important in predicting the stock prices of CIMB and Maybank. Volume had the highest importance indicating that trading activity for both banks affects price movements. The Average Sentiment Score is the second. It means that news sentiment has a noticeable but smaller impact.

4.5 Model Selection, Training & Evaluation

This study compared three deep-learning model to predict the next-day closing price prediction of stock market. A sliding-window of 5 days formed input sequences. 80% of each series used for training while another 20% for testing. To prevent overfitting, all model is then trained using Adam optimiser, MSE loss, mini batch=32 and early-stopping=10

Model	Key idea
LSTM	To capture lone-range temporal patterns
GRU	Fewer parameter
ACNN-LSTM	Extract short-term patterns with 1-D convolutions, then modelling the long-range context

Table 4-5 Core Idea of Model Selection

Evaluation matric used in this project are Mean Squared Error (MSE), square-root (RMSE) and Mean Absolute Error (MAE). Table below showed the result of each model.

Bank	Model	MSE	RMSE	MAE
CIMB	LSTM	0.441	0.664	0.601
	GRU	0.547	0.740	0.665
	ACNN-LSTM	0.060	0.244	0.187
Maybank	LSTM	1.009	1.005	0.969
	GRU	0.522	0.723	0.689
	ACNN-LSTM	0.054	0.232	0.177

Table 4-6 Evaluation of Model

From the table, it can be concluded that ACNN-LSTM model produces the lowest value compare to another model.

To visually compare prediction performance, six-line graphs were generated, each showing the actual closing price vs. predicted price from the test set

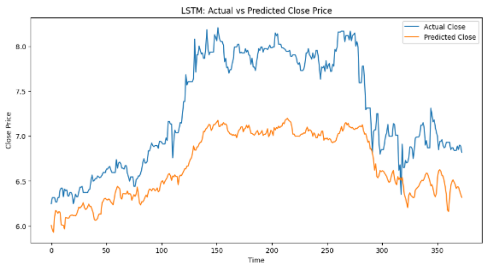
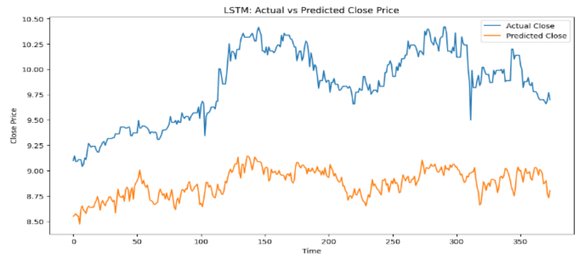
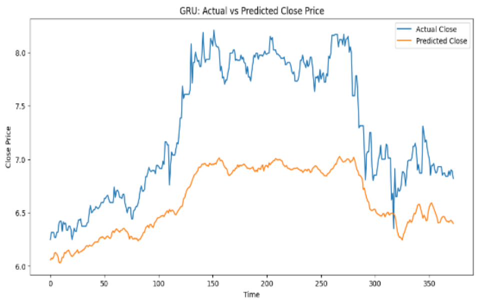
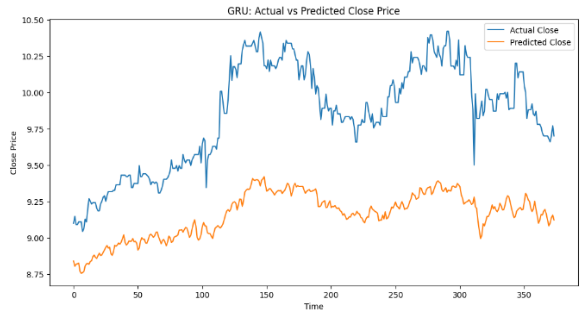
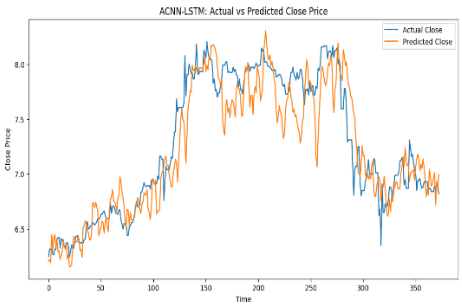
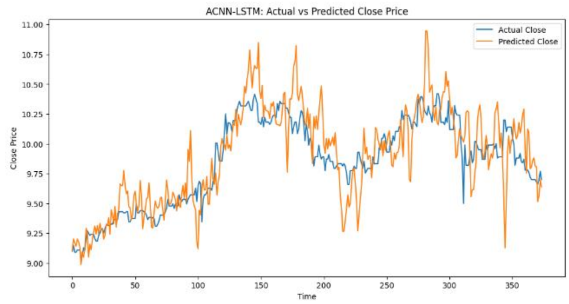
<table>
<tr><th></th><th>CIMB</th><th>MAYBANK</th></tr>
<tr><td>LSTM</td><td>
</td><td>
</td></tr>
<tr><td>GRU</td><td>
</td><td>
</td></tr>
<tr><td>ACNN-LSTM</td><td>
</td><td>
</td></tr>
</table>

Table 4-7 Line Graph for Deep Learning Model

From the table, the convolutional front-end in ACNN LSTM proved to capture short-horizontal patterns. After that, the dense layers of the model will learn longer-term dependencies thus contribute to the lowest prediction error for both CIMB and Maybank. For Maybank GRU outperformed LSTM but for CIMB, LSTM is superior compare to GRU. This show that parameter efficiency is not enough without feature extraction. Overall, ACNN-LSTM is the best model for forecasting and trading-strategy simulations for both banks

CHAPTER 5

CONCLUSION AND RECOMMENDATION

5.1 Conclusion

This study aimed to predict the stock prices of CIMB and Maybank based on sentiment analysis of news headlines using machine learning techniques. The study begins with data collection and preprocessing, followed by the sentiment analysis of the news headlines and lastly evaluate the machine learning models. The data was prepared by scraping the headlines news related to financial while the stock price data are extract from the yfinance. The data was then cleaned, standardized, and merged into one dataset. VADER tool was used for the sentiment analysis of news headlines. The sentiment was classified into three category which are positive, negative, and neutral. These sentiment scores then combined with stock price data and the dataset is used to train the machine learning model.

To predict the next day closing price, three deep learning model were evaluated. It was found that ACNN-LSTM model demonstrated the best performances for both banks compared to the LSTM and GRU. For CIMB, it achieved the lowest MSE (0.06), RMSE (0.244) and MAE (0.187). Like CIMB, Maybank also achieved the lowest MSE (0.054), RMSE (0.232) and MAE (0.177) for the model. Based on the result, it can be concluded that ACNN-LSTM model is capable to predict stock price accurately and sentiment analysis help to improve the result.

Additionally, the study examines the relationship between the number of news headline and the stock price. It was found that there is no direct linear relationship between both. However, it was observed that there is stock price movement when the headline volume is increasing for Maybank. This suggest that the volume of relevant news indicate there is market activity. Combining sentiment analysis with headline volume could result to more accurate prediction of the prices.

In conclusion, sentiment analysis can be used to improve accuracy in predicting stock market prices when combine with ACNN-LSTM machine learning techniques.

5.2 Future Works

There are few gaps that need to be addressed in the future work from this study. Therefore, this study proposes a few suggestions that will be useful for future research to improve the effectiveness of sentiment-based stock price prediction models.

This study only focusses on the three deep learning model. The future study could explore other deep learning architecture especially transformer based to improve the accuracy of the prediction. The study also could focus in the volatile for financial markets. Additionally, the current model could integrate more variables that influencing stock prices. For example, inflation rate, interest rate, and exchange rate could provide a comprehensive analysis for stock price prediction.

As this research was focused solely on CIMB and Maybank in Malaysia, future research can focus to expand the scope of study. It because it will help in assessing the generalizability of the findings. Beside that, a comparative analysis across market could offer insight about the performance of sentiment-based prediction in various financial environments. This will help to identify either the model is applicable globally and show the different market movement of the regions.

This study only used historical data and sentiment analysis. Future research can focus on real-time data processing. By integrating live sentiment analysis with stock market data, the model will become more responsive to real-time market changes. This will enable investors to make decision based on the latest news, proving the model ability to adapt into real-world trading market.

Lastly, future study could explore about the longer-term prediction such as weekly or monthly for stock prices forecasting. A longer prediction time would help to provide valuable insight to investor for investment strategies.

In conclusion, to develop accurate, adaptable, and comprehensive stock prediction models, several areas of improvement could be address based on this research for future study. The insight from the new model about the market behaviour could be used for investor and policymaker in making decision

LIST OF PUBLICATIONS

1. Agarwal, P., & Gupta, A. K. (2024). *Strategic Business Insights through Enhanced Financial Sentiment Analysis: A Fine-Tuned Llama 2 Approach*. https://doi.org/10.1109/icict60155.2024.10544544
2. Alamu, O. S., & Siam, M. K. (2024). Stock Price Prediction and Traditional Models: An Approach to Achieve Short-, Medium- and Long-Term Goals. Journal of Intelligent Learning Systems and Applications, 16(4), 363. https://doi.org/10.4236/jilsa.2024.164018
3. Aluvala, S., Balamurugan, V., Jeyasudha, J., Kumar, R. D., & S, P. K. (2023). *Decoding Financial Markets: Unleashing the Power of Bi-LSTM in Sentiment Analysis for Cutting-Edge Stock Price Prediction*. https://doi.org/10.1109/easct59475.2023.10392444
4. Ansah, K., Denwar, I. W., & Appati, J. K. (2022). Intelligent Models for Stock Price Prediction: A Comprehensive Review. *Journal of Information Technology Research*. https://doi.org/10.4018/jitr.298616
5. Al-Alawi, A. I., & Alshakhoori, N. (2024). *Stock Price Prediction Using Artificial Intelligence: A Literature Review*. https://doi.org/10.1109/icetsis61505.2024.10459442
6. Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. The Journal of Economic Perspectives, 21(2), 129. https://doi.org/10.1257/jep.21.2.129
7. Binti Alyasa Gan, S. S., Mohamed Sayuti, D. S. N., Mokhtar, I., & Tajuddin, A. H. (2022). THE ROLE OF IPO PROCEEDS DISCLOSURE ON SURVIVAL OF PUBLIC LISTED COMPANIES IN MALAYSIA.
8. Brown, M. S., Chen, E. M., Lee, S., Taylor, J., Kim, A., & Johnson, R. C. (2024). *Comparative Analysis of LSTM and Traditional Time Series Models on Oil Price Data*. https://doi.org/10.20944/preprints202410.2355.v1
9. Chan, T.-H., Saqib, A., & Lean, H. H. (2024). Asymmetric Effects of Investor Sentiment on Malaysian Sectoral Stocks: A Nonlinear Autoregressive Distributed Lag Approach. *Asian Academy of Management Journal of Accounting and Finance*. https://doi.org/10.21315/aamjaf2024.20.1.6
10. Chen, C., Xue, L., & Xing, W. (2023). Research on Improved GRU-Based Stock Price Prediction Method. *Applied Sciences*. https://doi.org/10.3390/app13158813
11. Chou, H.-C., & Ramachandran, K. M. (2021). *Combining Time Series and Sentiment Analysis for Stock MarketForecasting*. https://doi.org/10.11159/ICSTA21.132
12. Dong, Z., Fan, X., & Peng, Z. (2024). *FNSPID: A Comprehensive Financial News Dataset in Time Series*. https://doi.org/10.48550/arxiv.2402.06698

13. Dutta, A., Pooja, G., Jain, N., Panda, R. R., & Nagwani, N. K. (2021). *A Hybrid Deep Learning Approach for Stock Price Prediction*. https://doi.org/10.1007/978-981-15-7106-0_1
14. Fu, K., & Zhang, Y. (2024). Incorporating Multi-Source Market Sentiment and Price Data for Stock Price Prediction. *Mathematics*. https://doi.org/10.3390/math12101572
15. Gaurav, A., Arya, V., Chui, K. T., Gupta, B. B., Choi, C., & Lee, O.-J. (2023). *Long Short-Term Memory Network (LSTM) based Stock Price Prediction*. https://doi.org/10.1145/3599957.3606240
16. Geetha, S., Suwetha, C., Sangavi, A., & Nazrin, S. S. (2024). *Time Series Financial Market Forecasting Based on Machine Learning*. https://doi.org/10.1109/aimla59606.2024.10531346
17. Gu, W. Y., Zhong, Y., Li, S., Wei, C. song, Dong, L., Wang, Z., & Yan, C. (2024). Predicting Stock Prices with FinBERT-LSTM: Integrating News Sentiment Analysis. 67. https://doi.org/10.1145/3694860.3694870
18. Gude, V., & Hsiao, D. (2025). Incorporating Media Coverage and the Impact of Geopolitical Events for Stock Market Predictions with Machine Learning. Journal of Risk and Financial Management, 18(6), 288. https://doi.org/10.3390/jrfm18060288
19. He, Z., Zhang, H., & Por, L. Y. (2024). A Comparative Study on Deep Learning Models for Stock Price Prediction. *Advances in Transdisciplinary Engineering*. https://doi.org/10.3233/atde240502
20. Ho, C. C., & Ong, H. H. (n.d.). *Using news sentiment for economic forecasting: a Malaysian case study 1*.
21. Jiang, L., Yue, S., Zhang, M., & Zhang, F. (2023). *AEformer: Asymmetric Embedding Transformer for Stock Market Prediction based on Investor Sentiment*. https://doi.org/10.21203/rs.3.rs-3344960/v1
22. Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems With Applications*. https://doi.org/10.1016/J.ESWA.2021.115019
23. Joshi, A. W., & Younes, R. (2023). Predicting Stocks Changes Using News Sentiment Analysis. https://doi.org/10.2139/ssrn.4605587
24. Joshi, A., Koda, J., & Hadimlioglu, A. (2024). *A Multi-Modal Transformer Architecture Combining Sentiment Dynamics, Temporal Market Data, and Macroeconomic Indicators for Sturdy Stock Return Forecasting*. https://doi.org/10.1109/bigdata62323.2024.10825219
25. Kabir, M. R., Bhadra, D., Ridoy, M., & Milanova, M. (2025). LSTM–Transformer-Based Robust Hybrid Deep Learning Model for Financial Time Series Forecasting. *Sci*. https://doi.org/10.3390/sci7010007
26. Kalbaliyev, E., & Szegedi, G. (2020). *Long Term Stock Market Prediction with Recurrent Neural Network Architectures*. https://doi.org/10.1007/978-3-030-71187-0_100

27. Kapgate, D., & Chaturvedi, Dr. S. (2025). *Exploring Machine Learning Methods for Stock Market Prediction: A Review*. https://doi.org/10.35629/5252-0701276279
28. Kaveri., G., Tejashwini, L., & Manjunath, K. V. (2025). Role of Sentiment in Stock Forecasting. *International Journal For Science Technology And Engineering*. https://doi.org/10.22214/ijraset.2025.66656
29. Kedar, S. V. (2021). Stock Market Increase and Decrease using Twitter Sentiment Analysis and ARIMA Model. Türk Bilgisayar ve Matematik Eğitimi Dergisi, 12, 146. https://doi.org/10.17762/turcomat.v12i1s.1596
30. Khonde, S. R., Virnodkar, S. S., Nemade, S. B., Dudhedia, M. A., Kanawade, B., & Gawande, S. H. (2024). Sentiment Analysis and Stock Data Prediction Using Financial News Headlines Approach. *Revue Dintelligence Artificielle*. https://doi.org/10.18280/ria.380325
31. Ku, C. S., Xiong, J., Chen, Y.-L., Soong, H.-C., & Por, L. Y. (2023). Improving Stock Market Predictions: An Equity Forecasting Scanner Using Long Short-Term Memory Method with Dynamic Indicators for Malaysia Stock Market. *Mathematics*. https://doi.org/10.3390/math11112470
32. Kurisinkel, L. J., Mishra, P., & Zhang, Y. (2024). *Text2TimeSeries: Enhancing Financial Forecasting through Time Series Prediction Updates with Event-Driven Insights from Large Language Models*. https://doi.org/10.48550/arxiv.2407.03689
33. Li, H., & Hu, J. (2024). *A Hybrid Deep Learning Framework for Stock Price Prediction Considering the Investor Sentiment of Online Forum Enhanced by Popularity*. https://doi.org/10.48550/arxiv.2405.10584
34. Li, X., Xie, H., Chen, L., Wang, J., & Deng, X. (2014). News impact on stock price return via sentiment analysis. Knowledge-Based Systems, 69, 14. https://doi.org/10.1016/j.knosys.2014.04.022
35. Lin, Y., & Wang, S. (2024). Enhancing Stock Market Prediction with Sentiment Analysis Using a BERT-based Model. *Transactions on Computer Science and Intelligent Systems Research*. https://doi.org/10.62051/tq61jb84
36. Li, Y., Zhang, X., & Zhu, X. (2023). *Application of LSTM and Attention Mechanism for Stock Price Prediction and Analysis*. https://doi.org/10.2991/978-94-6463-222-4_60
37. Liu, R., Liu, H., Huang, H., Song, B., & Wu, Q. (2023). Multimodal multiscale dynamic graph convolution networks for stock price prediction. *Pattern Recognition*. https://doi.org/10.1016/j.patcog.2023.110211
38. Liu, W., & Zhang, Y. (2022). Attention-based BiLSTM model for stock price prediction. *International Conference on Artificial Intelligence and Pattern Recognition*. https://doi.org/10.1145/3573942.3574019
39. Madhuri, N., Tamilkodi, R., Teja, M. S., Kumar, M., Swaroop, N., & Jayaprakash, L. S. (2024). *Investor Sentiment based Predictive Stock Market*

Forecasting using Deep Learning. https://doi.org/10.1109/icdici62993.2024.10810985
40. Makinde, A. S. (2024). *Optimizing Time Series Forecasting: A Comparative Study of Adam and Nesterov Accelerated Gradient on LSTM and GRU networks Using Stock Market data*. https://doi.org/10.48550/arxiv.2410.01843
41. Melveetil, V. C. (2024). An Optimal Multi-Modal Approach for Stock Market Price Forecasting with Fused Sentiment Analysis for Real Time Data. *International Journal For Science Technology And Engineering*. https://doi.org/10.22214/ijraset.2024.64417
42. Mokhtari, M., Seraj, A., Saeedi, N., & Karshenas, A. (2023). The Impact of Twitter Sentiments on Stock Market Trends. arXiv (Cornell University). https://doi.org/10.48550/arxiv.2302.07244
43. Nagendra, B. V., Chandar, S. K., Simha, J. B., & Bazil, J. A. J. (2024). *Financial Lexicon based Sentiment Prediction for Earnings Call Transcripts for Market Intelligence*. https://doi.org/10.1109/icipcn63822.2024.00103
44. Nyakurukwa, K., & Seetharam, Y. (2024). Sentimental Showdown: News Media vs. Social Media in Stock Markets. *Heliyon*. https://doi.org/10.1016/j.heliyon.2024.e30211
45. Osman, H. A. (2023). *Stock Market Prediction from Sentiment and Financial Stock Data Using Machine Learning*. https://doi.org/10.21428/594757db.40c1a462
46. Ozupek, O., Yılmaz, R., Ghasemkhani, B., Birant, D., & Kut, A. (2024). A Novel Hybrid Model (EMD-TI-LSTM) for Enhanced Financial Forecasting with Machine Learning. *Mathematics*. https://doi.org/10.3390/math12172794
47. Pan, X. (2024). Stock Price Prediction System LSTM Based on Deep Learning. *Journal of Economics and Public Finance*. https://doi.org/10.22158/jepf.v10n4p235
48. Rahim, A. I. A., Ibrahim, M. I., Chua, S.-L., & Musa, K. I. (2021). Hospital Facebook Reviews Analysis Using a Machine Learning Sentiment Analyzer and Quality Classifier. Healthcare, 9(12), 1679. https://doi.org/10.3390/healthcare9121679
49. Rizinski, M., Peshov, H., Mishev, K., Jovanovik, M., & Trajanov, D. (2024). Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex). *IEEE Access*. https://doi.org/10.1109/access.2024.3349970
50. Rizinski, M., Peshov, H., Mishev, K., Jovanovik, M., & Trajanov, D. (2023). Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex). *arXiv.Org*. https://doi.org/10.48550/arXiv.2306.03997
51. Shahi, T. B., Shrestha, A., Neupane, A., & Guo, W. (2020). *Stock Price Forecasting with Deep Learning: A Comparative Study*. https://doi.org/10.3390/MATH8091441

52. Shi, Z., Hu, Y., Mo, G., & Wu, J. (2022). Attention-based CNN-LSTM and XGBoost hybrid model for stock prediction. *arXiv.Org*. https://doi.org/10.48550/arXiv.2204.02623
53. *Stock Price Prediction Using Bi-LSTM and GRU-Based Hybrid Deep Learning Approach*. (2022). https://doi.org/10.1007/978-981-19-3148-2_60
54. Sulistianingsih, N., & Martono, G. H. (2024). Comparative Study on Stock Movement Prediction Using Hybrid Deep Learning Model. *ECTI Transactions on Computer and Information Technology*. https://doi.org/10.37936/ecti-cit.2024184.256303
55. Umar, S. U. (2023). Sentiment Analysis of Financial Textual data Using Machine Learning and Deep Learning Models. *Informatica*. https://doi.org/10.31449/inf.v47i5.4673
56. Velarde, G., Brañez, P., Bueno, A., Heredia, R., & Lopez-Ledezma, M. (2022). *An Open Source and Reproducible Implementation of LSTM and GRU Networks for Time Series Forecasting*. https://doi.org/10.3390/engproc2022018030
57. Wang, Z. (2023). Investor Sentiment Analysis Based on Comment Text for Stock Price Prediction. *BCP Business & Management*. https://doi.org/10.54691/bcpbm.v38i.4178
58. Wu, B., Li, Y., & Fu, G.-S. (2024). Optimizing Convolutional Neural Network-Long Short-Term Memory Architecture with Additive Attention Mechanism for Stock Price Prediction. *Revista Internacional De Metodos Numericos Para Calculo Y Diseno En Ingenieria*. https://doi.org/10.23967/j.rimni.2024.10.56790
59. Wu, J., Xu, K., Chen, X., Li, S., & Zhao, J. (2021). Price graphs: Utilizing the structural information of financial time series for stock prediction. Information Sciences, 588, 405. https://doi.org/10.1016/j.ins.2021.12.089
60. Xiao, P. (2024). Stock Market Prediction Based on Financial News, Text Data Mining, and Investor Sentiment Analysis. *International Journal of Information System Modeling and Design*. https://doi.org/10.4018/ijismd.361593
61. Yu, Y. (2024). Research on the Feasibility of Machine Learning Methods in Stock Price Prediction. *Advances in Economics, Management and Political Sciences*. https://doi.org/10.54254/2754-1169/118/20242000
62. Zhang, Z. (2023). *Integrating Sentiment Analysis and Knowledge Graphs for Enhanced Stock Trend Prediction-A Deep Learning Approach*. https://doi.org/10.1109/ispcem60569.2023.00009
63. Zhang, J., Ye, L., & Lai, Y. (2023). Stock Price Prediction Using CNN-BiLSTM-Attention Model. *Mathematics*. https://doi.org/10.3390/math11091985
64. Zhong, Z. (2024). *Intelligent Financial Modeling Data Analysis Based on MATLAB*. https://doi.org/10.1109/nmitcon62075.2024.10699130

65. Zhu, R., & Chen, J. (2023). *XGBoost and CNN-LSTM hybrid model with Attention-based stock prediction*. https://doi.org/10.1109/ICETCI57876.2023.10176988