UTM
UNIVERSITI TEKNOLOGI MALAYSIA

PROJECT PROPOSAL PRESENTATION

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

PRESENTED BY : NUR AINA FARRAAIN BINTI ZAHANIZAM
MATRIC NO : MCS241053
LECTURER : ASSOC. PROF. DR MOHD SHAHIZAN BIN OTHMAN

01 INTRODUCTION

02 LITERATURE REVIEW HIGHLIGHTS

03 METHODOLOGY

04 INITIAL FINDINGS

05 CONCLUSION & FUTURE WORKS

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

INTRODUCTION

BACKGROUND

- Stock market trends reflect the economy’s condition, making price prediction crucial.
- Sentiment analysis turns news headlines into data to help forecast stock movements.
- CIMB and Maybank, key banks in Malaysia, are chosen due to their influence on the financial sector.
- Machine learning (LSTM, GRU, ACNN-LSTM) enhances prediction accuracy using both historical prices and sentiment data.

PROBLEM STATEMENT

- Predicting stock prices is hard due to market volatility, unstructured news data, and external influences.
- Traditional models can’t fully capture the impact of news sentiment.
- Integrating sentiment analysis with deep learning offers a new way to improve forecasting accuracy.

RESEARCH QUESTIONS & OBJECTIVES

Research Question

- RQ1: How do news headlines (positive/negative/neutral) relate to the next closing price of CIMB and Maybank?
- RQ2: Can sentiment analysis predict the next-day stock prices of CIMB and Maybank?
- RQ3: Is news sentiment more predictive than historical prices and trading volume?

Research Objectives

- To analyse the sentiment in financial news headlines.
- To examine how sentiment correlates with next-day closing prices of CIMB and Maybank.
- To develop and evaluate predictive models (LSTM, GRU, ACNN-LSTM) based on sentiment

SCOPE OF THE STUDY

Data Scope	• Financial news headlines related to CIMB and Maybank (2019–2025). • News sources: The Star, Malay Mail, The Edge Market, New Straits Times, Business Today. • Stock price data sourced via Yahoo Finance using yfinance Python library.
Methodological Scope	• Perform sentiment analysis using VADER via NLTK. • Align sentiment scores with corresponding stock price data (next-day closing). • Train and evaluate deep learning models: LSTM, GRU, and ACNN-LSTM. • Use evaluation metrics: MSE, RMSE, and MAE.
Temporal Scope	• Data collected and analysed covers January 2019 to June 2025. • Time-aligned analysis for sentiment vs. next-day closing stock price.
Research Depth & Phases	• Current phase: Daily sentiment classification and next-day stock price prediction. • Extended phase: Add macroeconomic variables, test transformer models and expand to other sectors or regional banks.

LITERATURE REVIEW

Overview of Literature Review Focus

Topic Focus

Exploration of sentiment analysis and deep learning models for stock price prediction.

Key Themes Identified

- Importance of investor sentiment and financial news in influencing stock prices.
- Application of machine learning and deep learning models (e.g., Bi-LSTM, LSTM, GRU).
- Use of hybrid models that combine sentiment analysis and technical indicators.

Main Purpose of Reviewed Studies:

- To improve stock market prediction accuracy by leveraging news sentiment, social media data, and AI models.

Existing Model Frameworks and Key Findings

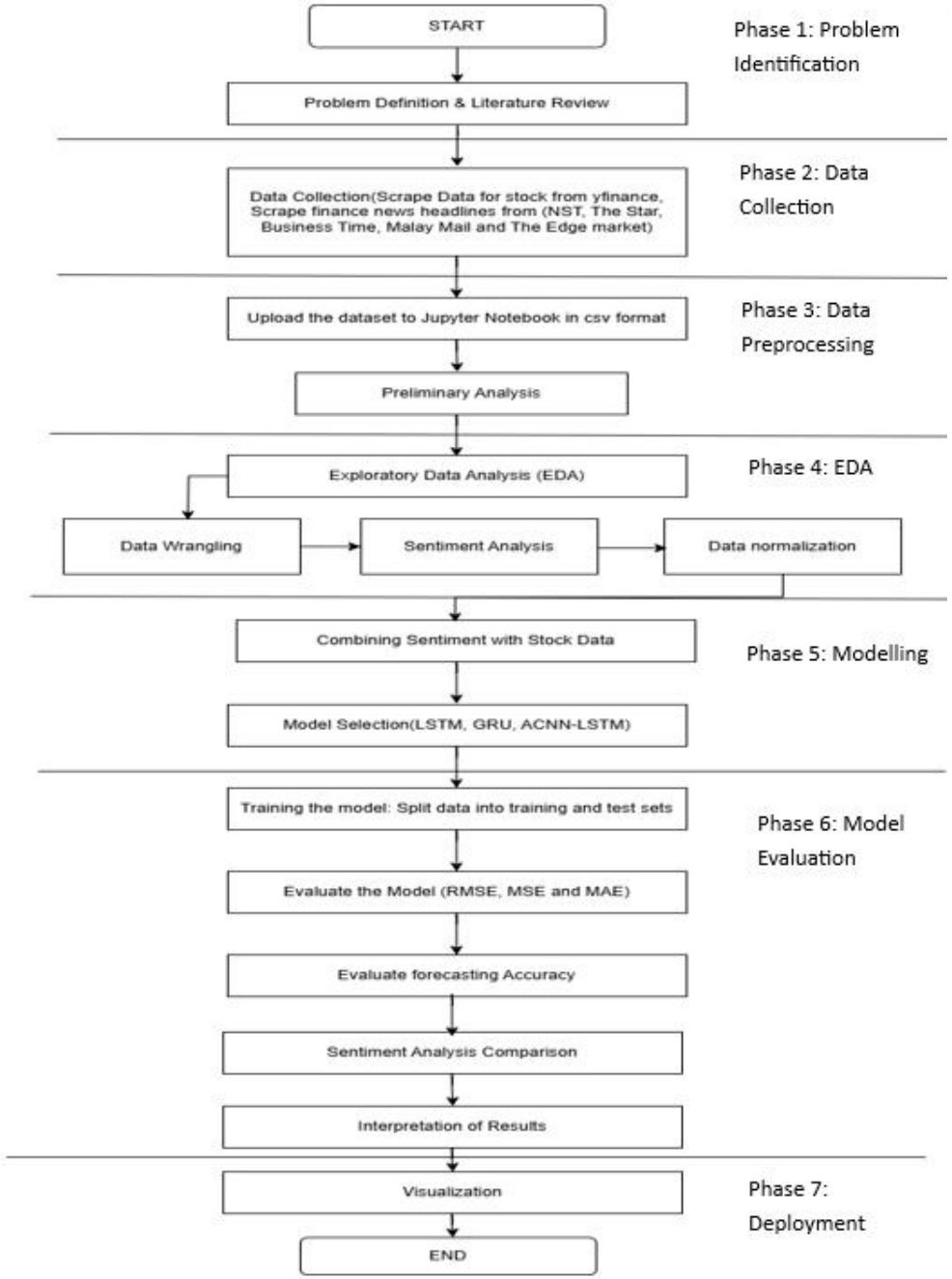
Model Framework	Methodology	Key Finding	Reference
Bi-LSTM	Deep learning with bidirectional sequence processing	Achieved high accuracy in sentiment-based stock prediction.	Aluvala et al., 2023
FinBERT-LSTM Hybrid	FinBERT for sentiment feature extraction + LSTM for time-series data	Effective integration of news sentiment improves price prediction accuracy.	Gu et al., 2024
GRU-Based Model	Gated Recurrent Unit for time-series analysis	GRU enhanced stock price prediction performance with reduced computational cost.	Chen et al., 2023
CNN-BiLSTM-Attention	CNN for feature extraction, BiLSTM for sequence learning, attention mechanism	Provided superior prediction accuracy by focusing on key features in financial data.	Zhang et al., 2023
U-CNNpred	Universal CNN-based architecture	Demonstrated flexibility and high performance across different stock markets.	Hoseinzade et al., 2019

Research Gaps and How This Study Addresses Them

Research Gap	Addressing The Gap
• No combination of sentiment and price data in one model	• This study combines news sentiment with stock price data to give better predictions.
• Most studies focus on foreign stock markets, not Malaysia	• This study looks at CIMB and Maybank, two big banks in Malaysia, to make it more relevant locally.
• Many models don't use powerful deep learning combinations	• This study uses ACNN + LSTM, which can better understand both text meaning and time trends in data.
• Past research uses slow or outdated news sources	• This study uses latest news headlines to help predict stock changes faster and more accurately.

UTM
UNIVERSITI TEKNOLOGI MALAYSIA

METHODOLOGY



A research framework for Stock price forecasting using news Sentiment and Deep Learning

Dataset Used

Dataset Name	Source	Format	Number of Row	Description
CIMB_YF	Yahoo Finance(Yfinance)	CSV	1580	Daily stock prices (2019–2025): Open, High, Low, Close, Volume
CIMBALL	WebScrapping	CSV	9035	Merged and cleaned financial news headlines for CIMB from five news portal
Maybank_YF	Yahoo Finance(Yfinance)	CSV	1580	Daily stock prices (2019–2025): Open, High, Low, Close, Volume
MaybankALL	WebScrapping	CSV	11561	Merged and cleaned financial news headlines for Maybank from five news portal

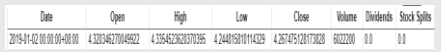
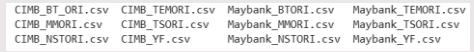
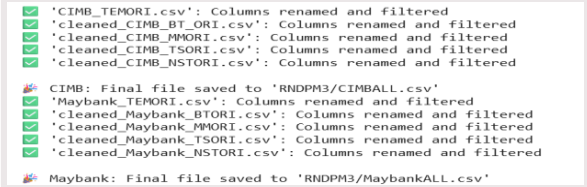
Methodology/Model used

Data collection	Preprocessing	Sentiment Analysis	Model Training	Evaluation Metrics
• Stock prices: yfinance (CIMB & MAYBANK, 2019–2025). • News headlines: Web scraping (Selenium) from 5 news portal websites (The Star, NST, The Edge Market, Business Today, Malay Mail)	• Cleaned stock data (removed duplicates, handled missing values). • Text preprocessing: Removed special characters, numbers, punctuation. • Aligned timestamps for stock prices and news headlines.	• Used VADER (NLTK) to calculate sentiment scores. • Categorized sentiments: Positive (>0.1), Neutral (−0.1 to 0.1), Negative (<−0.1).	• LSTM : Captures long-term dependencies in time-series data. • GRU : Simplified LSTM with faster computation. • ACNN-LSTM : Combines CNN (text feature extraction) + LSTM (temporal patterns) + Attention mechanism.	• RMSE, MSE, MAE to assess prediction accuracy.

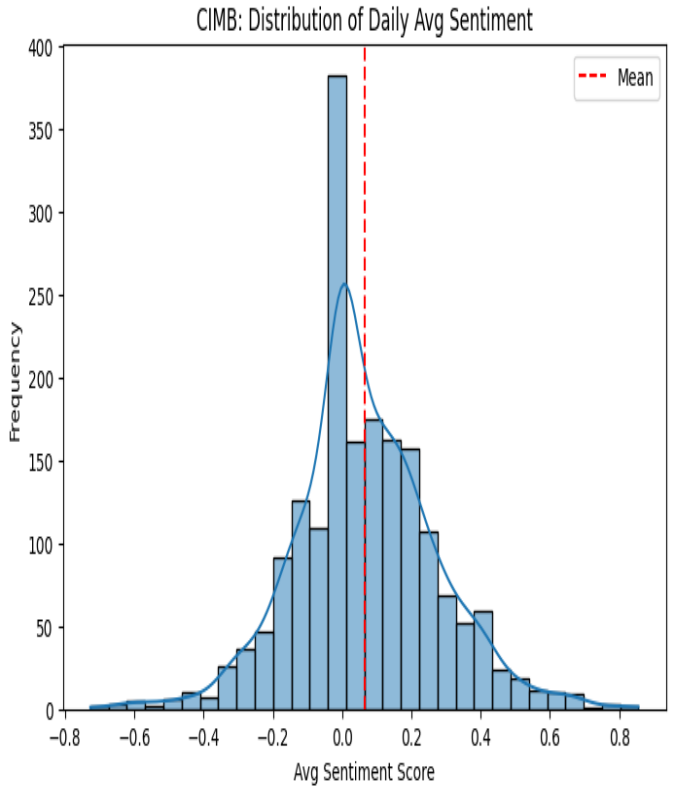
UTM
UNIVERSITI TEKNOLOGI MALAYSIA

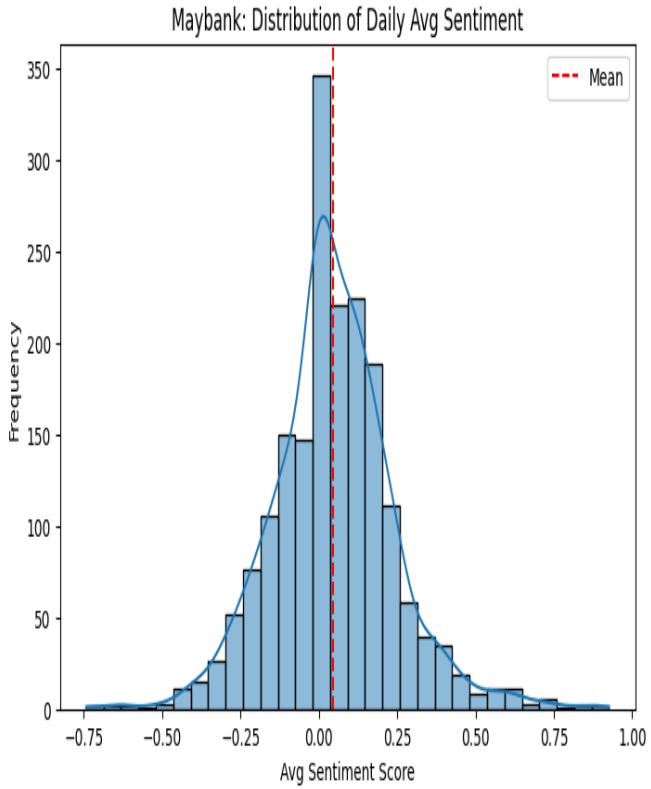
INITIAL FINDINGS

Cleaning and Standardization

Preprocessing Step	Description	Before Process	After Process
Date Formatting	Unified date format (YYYY-MM-DD) for both stock prices and news headlines.		
Column Reduction	Removed irrelevant columns (e.g., "Time," "Dividends," "Stock Splits").		
Category Filtering	Retained only financial/business-related headlines (e.g., filtered out "Aviation" or "Branded" categories).		
Merging	Combined datasets for CIMB and Maybank separately, aligning dates (Jan 2019–June 2025).		

Key Observation From EDA

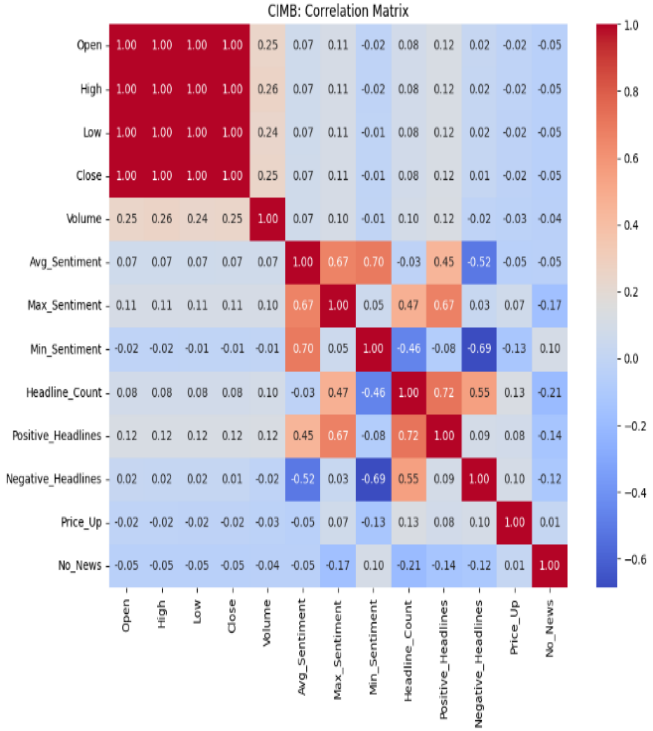


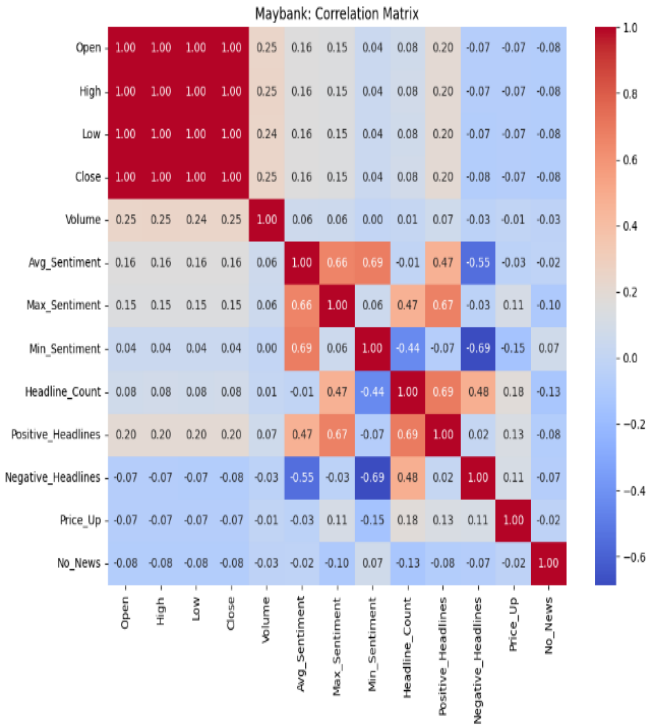


Sentiment Distribution :

- Maybank: More positive headlines (right-skewed sentiment scores).
- CIMB: Mostly neutral sentiment (clustered around 0).

Key Observation From EDA

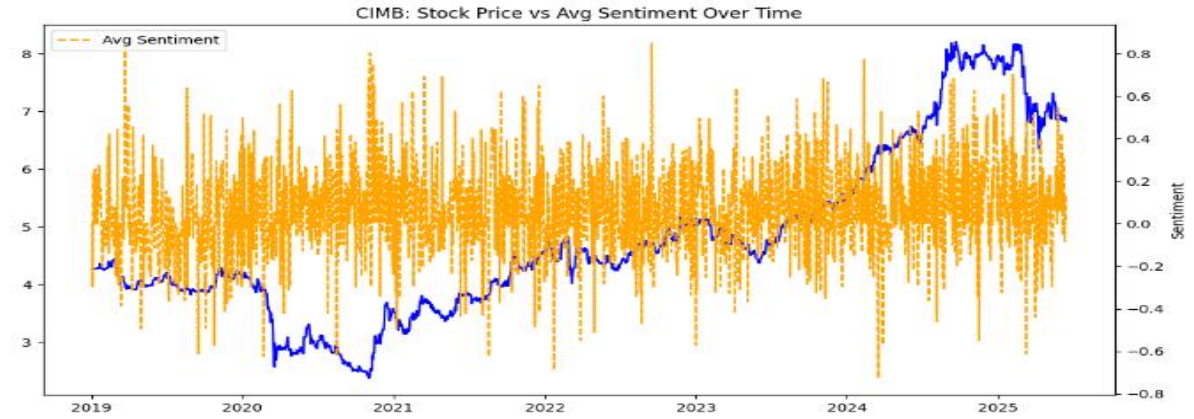


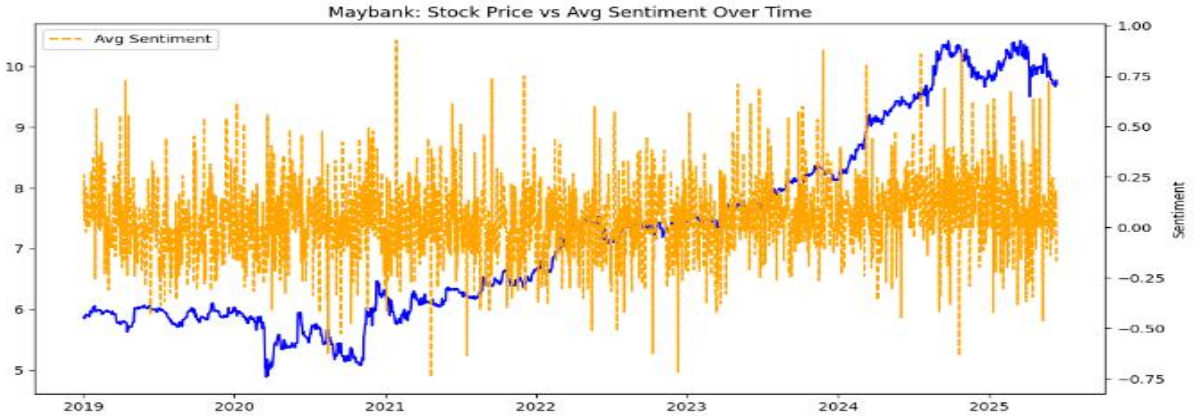


Correlation Insights :

- Weak correlation between sentiment scores and stock prices (heatmap visualization).

Key Observation From EDA

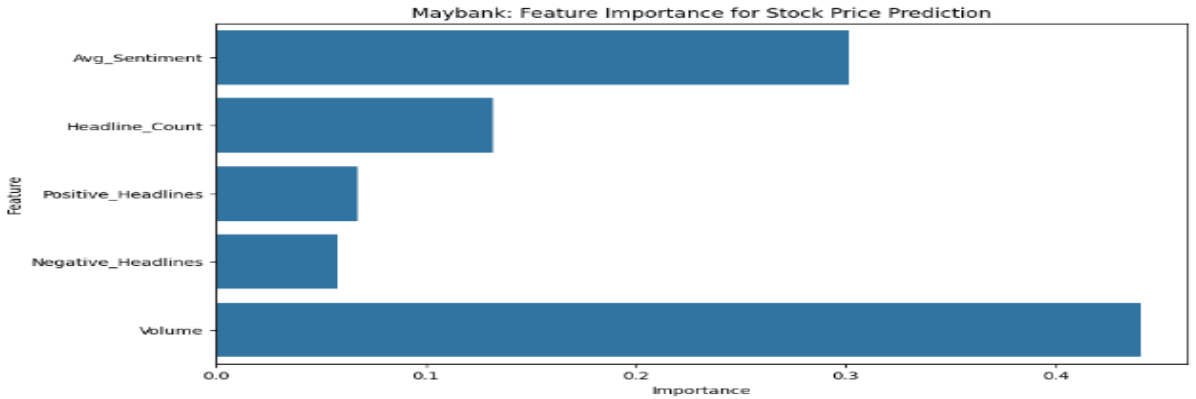


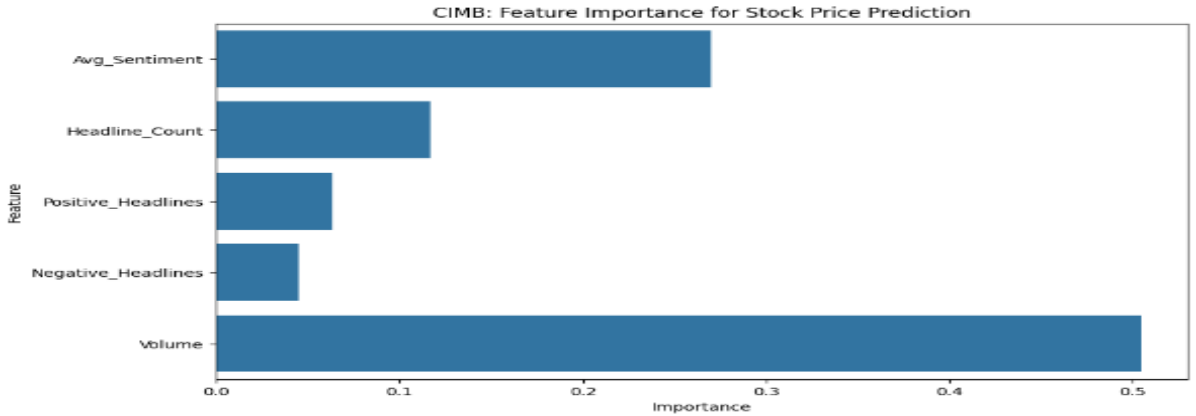


Time Series Trends :

- Maybank showed stronger alignment between sentiment spikes and price movements.

Key Observation From EDA





Feature Importance :

- Trading Volume had the highest impact (~50% for CIMB, ~43% for Maybank).
- Average Sentiment Score ranked second (~27% for CIMB, ~30% for Maybank).

Deep Learning Model Compared

Deep Learning Model	Advantages of Model	Input	Split	Optimization
LSTM	Captures long-term dependencies.	Sliding window of 5 days → predict next-day closing price.	80% training, 20% testing.	Adam optimizer, MSE loss, batch size=32, early stopping (patience=10).
GRU	Faster training with fewer parameters.			
ACNN-LSTM	Hybrid model (CNN for text features + LSTM for temporal patterns + Attention).			

Performance Metrics

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

CIMB Results :

- ACNN-LSTM outperformed others (MSE=0.060, RMSE=0.244, MAE=0.187).
- LSTM slightly better than GRU.

Maybank Results :

- ACNN-LSTM again best (MSE=0.054, RMSE=0.232, MAE=0.177).
- GRU slightly better than LSTM

CONCLUSION AND FUTURE WORKS

Key Findings and Recommendations

Conclusion

ACNN-LSTM Best Model
- Achieved lowest prediction errors for both banks

Sentiment Impact :
- Sentiment scores (via VADER) improved prediction accuracy.
- Maybank showed stronger alignment between sentiment spikes and price trends.

Headline Volume :
- No linear correlation with stock prices, but spikes in relevant news correlated with market activity.

Future works

Advanced Architectures :
- Explore Transformer-based models (e.g., BERT, GPT) for better contextual understanding.

External Factors :
- Integrate macroeconomic variables (inflation, interest rates, exchange rates).

Global Applicability :
- Test the model on banks in other regions for cross-market validation.

Real-Time Systems :
- Develop live sentiment analysis pipelines for real-time trading decisions.

Long-Term Predictions :
- Extend forecasting to weekly/monthly trends for strategic investment planning.

THANK YOU

univteknologimalaysia utm.my utmofficial