IS460
Group 3

Final Project Presentation

LLM-Enhanced Stock Price Prediction:

Improving the Synergy of Price Prediction and Sentiment Analysis Models

Brenda Su, Cheryl Yeo, Chong Le Kai, Wong Swee Kiat, Tammy Lee, Trina Tan, Xavier Santimano

Motivation

- **Maximise Returns, Minimise Risks:**
  - Accurate stock price predictions can reduce exposure to large portfolio losses, helping investors make smarter financial decisions

- **LLM's Potential:**
  - Sentiment analysis alone can already improve prediction accuracy by approximately 5% (Wang et al, 2018)
  - LLMs offer a richer, more context-aware understanding of financial news, potentially enhancing prediction models and providing a competitive edge in volatile markets

- Wang, Zhaoxia, Ho, S.-B., & Lin, Z. (2018). Stock market prediction analysis by incorporating social and news opinion and sentiment. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 1375–1380. https://doi.org/10.1109/icdmw.2018.00195

Problem Statement

This project explores **<u>using LLMs for stock price prediction by integrating historical data with financial news sentiment</u>** for better investment decisions.

While sentiment analysis is well-studied, the advantages of LLMs over traditional methods remain under-studied. We evaluate feature extraction and multimodal architectures to assess their impact on prediction accuracy.

IS460
Group 3

Related Work

EXISTING APPROACH

- Price Prediction Models
- Sentiment Analysis Models

ENHANCING MODEL

- Use of LLMs in Stock Prediction
- Multimodal Architecture

Existing Approach

Price Prediction Models

1. Integrating Sentiment Scores of Financial News and MLP-Regressor (Maqbool et al, 2024)

- Multilayer Perceptron (MLP) model combined sentiment scores and historical stock data
- Feature label to indicate if news articles were company-specific improved model accuracy

2. Comparative Study between Traditional Statistical Approach and Machine Learning Approach (Bhattacharjee et al, 2019)

- Predictions from traditional methods had higher Mean Squared Error
- RF, MLP and LSTM models produced the most accurate predictions

- Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., & Ganaie, I. A. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-Regressor: A machine learning approach. Procedia Computer Science, 218, 1067–1078. https://doi.org/10.1016/j.procs.2023.01.086
- Bhattacharjee, I., & Bhattacharja, P. (2019). Stock price prediction: A comparative study between Traditional Statistical Approach and machine learning approach. 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 1–6. https://doi.org/10.1109/eict48899.2019.9068850

Existing Approach

Sentiment Analysis Models

3. Stock market prediction analysis by incorporating social news opinion and sentiment (Wang et al, 2018)

- Addressed non-linear nature of financial data using voting method to aggregate scores calculated by NLTK and Vader
- Optimised model with look-back period and sentiment effect parameter

- Wang, Zhaoxia, Ho, S.-B., & Lin, Z. (2018). Stock market prediction analysis by incorporating social and news opinion and sentiment. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 1375–1380. https://doi.org/10.1109/icdmw.2018.00195

Enhancing Model

Use of LLMs in Stock Prediction

4. Assessing the Performance of LLMs for Target-Level Sentiment Analysis (Muhammad et al, 2025)

- StockTime is a fine-tuned LLM that improves forecasting by embedding both textual and time-series data into the same LLM space.
- StockTime uses frozen LLaMa-3 embeddings and an LSTM-based autoregressive decoder.

5. LLM Factor: Extracting Profitable Factors through Prompts (Wang, 2024)

- Use LLMs to analyse the news content and identify factors that affect stock prices
- Integrates background knowledge, stock-related factors and temporal data

- Muhammad, I., & Rospocher, M. (2025). On assessing the performance of LLMS for target-level sentiment analysis in financial news headlines. Algorithms, 18(1), 46. https://doi.org/10.3390/a18010046
- Wang, M., Izumi, K., & Sakaji, H. (2024). LLMFACTOR: Extracting profitable factors through prompts for explainable stock movement prediction. Findings of the Association for Computational Linguistics ACL 2024, 3120–3131. https://doi.org/10.18653/v1/2024.findings-acl.185

Enhancing Model

Multimodal Architecture

6. Multimodal Stable Fusion via Gated Cross-Attention Mechanism (Zong and Zhou, 2024)

- Mechanism: MLP + pre-trained LLMs + graph attention neural network
- Higher prediction accuracy by filtering out noise

7. Hybrid (multimodal) Neural Network Architecture (Cote, 2022)

- Mechanism: MLP for tabular data, LSTM for textual data and Transfer Learning for image inputs
- Hybrid network (tabular+text) performs better than tabular-only model

- Zong, C., & Zhou, H. (2024, December 2). Stock movement prediction with multimodal stable fusion via gated cross-attention mechanism. arXiv.org. https://doi.org/10.48550/arXiv.2406.06594
- Cote, D. (2022, November 18). Hybrid (multimodal) neural network architecture : Combination of tabular, textual and image inputs to predict house prices. Medium. https://medium.com/@dave.cote.msc/hybrid-multimodal-neural-network-architecture-combination-of-tabular-textual-and-image-inputs-7460a4f82a2e

Summary

Source	Relevance to project
1 & 2	• Select top model (**LSTM**) for baseline stock price predictions
3	• Integrate sentiment analysis with stock price predictions • Use identified **look-back period of 7 days** for relevant news and **decaying sentiment effect of 0.01%**
4 & 5	• Try VADER, transformer models (DistilRoBERTa, deBERTa) and LLMs (Gemini, etc.) for sentiment predictions
6 & 7	• Explore different methods of combining stock price data and sentiment analysis

IS460
Group 3

Datasets

STOCK PRICE DATA

NEWS DATA

For our project, we used data related to the following stocks: AAPL, AMZN, MSFT, CRM, IBM, NVDA

Datasets

Stock Price Data

Method of retrieval:

- yfinance (retrieve Yahoo! Finance* data)

Data:

- 30 companies (DJIA)
- Narrowed down to 6 stocks (AAPL, AMZN, CRM, IBM, MSFT & NVDA)
- 10 years: 2015 - 2024
- 6 columns (Adj Close, Close, High, Low, Open, Volume)

Used for main model development.

Processing

Data Cleaning:

- Check for null values, drop if necessary
- Target variable: **Adj Close**, drop other columns

Preprocessing:

- Drop rows depending on number of lags
- Transform index into Datetime
- MinMaxScaler
- Train Test Split (80:20)

- Yahoo! (n.d.). Yahoo Finance – Stock Market Live, quotes, business and Finance News. Yahoo! Finance. https://sg.finance.yahoo.com/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAI-LWWCbrbS7Jdn9AlafEaG84ZHZ1tNUKTIYXom7FRW1vUtTmZlIUBBL0eA68mXC8u25WVLC6jR75Dbtii4ZJhfltKeboeH4QLW7nUMWPF8Xfdbsbe1-Bi-Y-PSSa5TSLRCd_4raroQfxv2CJSstSwlw9bT8AdPFXDqOtm9hXrXv

Datasets

News Data

Method of retrieval:

- New York Times API: Article Search*
  - Filter by organisation (e.g. “Apple Inc”)

Data:

- NYT articles on the 6 stocks
- 10 years: 2015 - 2024
- 9 columns (pub_date, abstract, lead_para, headline, doc_type, section_name, type_of_material, rank, web_url)

Used for main model development.

Processing

Data Cleaning:

- Check for null values for headlines and abstract, drop if necessary

Preprocessing:

- Remove non-ASCII and special characters
- DistilRoberta and Deberta: AutoTokenizer from pre-trained models

- The New York Times. (n.d.). Article Search. The New York Times. https://developer.nytimes.com/docs/articlesearch-product/1/overview

Datasets

Financial Text Data

Method of retrieval:
- Financial Phrasebank Kaggle Dataset*

Data:
- 4837 rows of Positive, Negative and Neutral financial news headlines
- Perspective of a retail investor
- 2 columns (sentiment and headline)

Used for finetuning of Transformers Models.

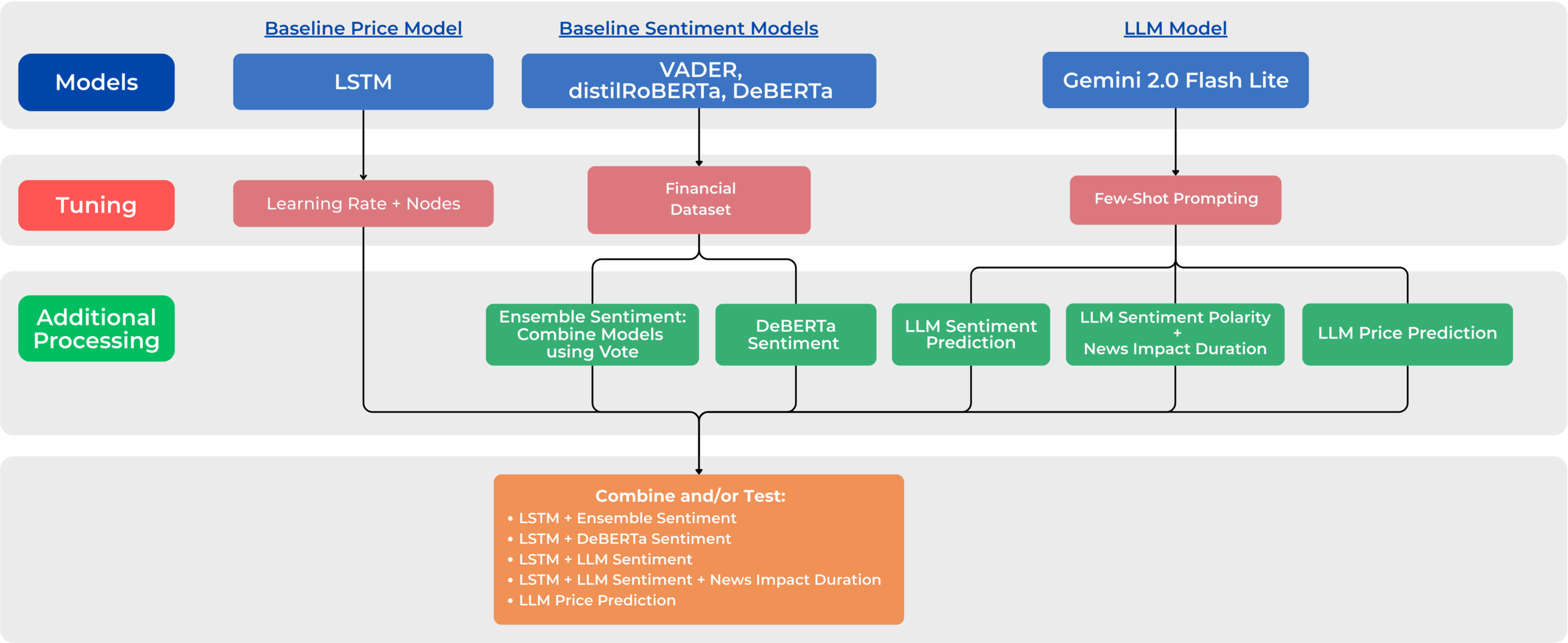
Processing

Preprocessing:
- Remove non-ASCII and special characters
- DistilRoberta and Deberta: AutoTokenizer from pre-trained models

- Sinha, A. (2020, May 27). Sentiment Analysis for Financial News. Kaggle. https://www.kaggle.com/datasets/ankurzing/sentiment-analysis-for-financial-news

Methodology

Methodology



Baseline Models

Baseline Pricing Models

Random Forest (RF)

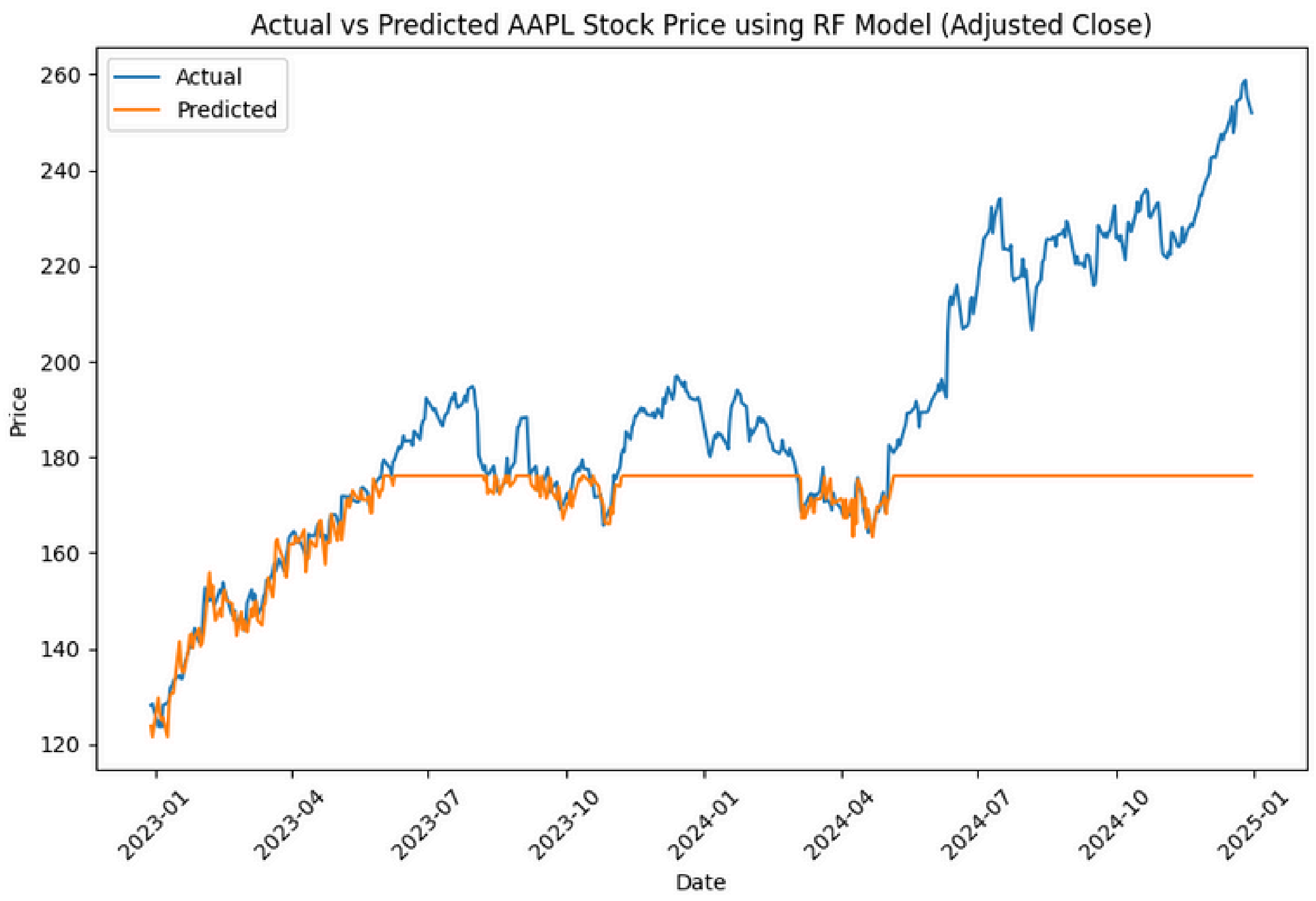


Figure 1: Baseline RF model for AAPL Stock Price

Mean Absolute Error: **16.672**
Mean Squared Error: **743.909**

Multilayer Perceptron (MLP)

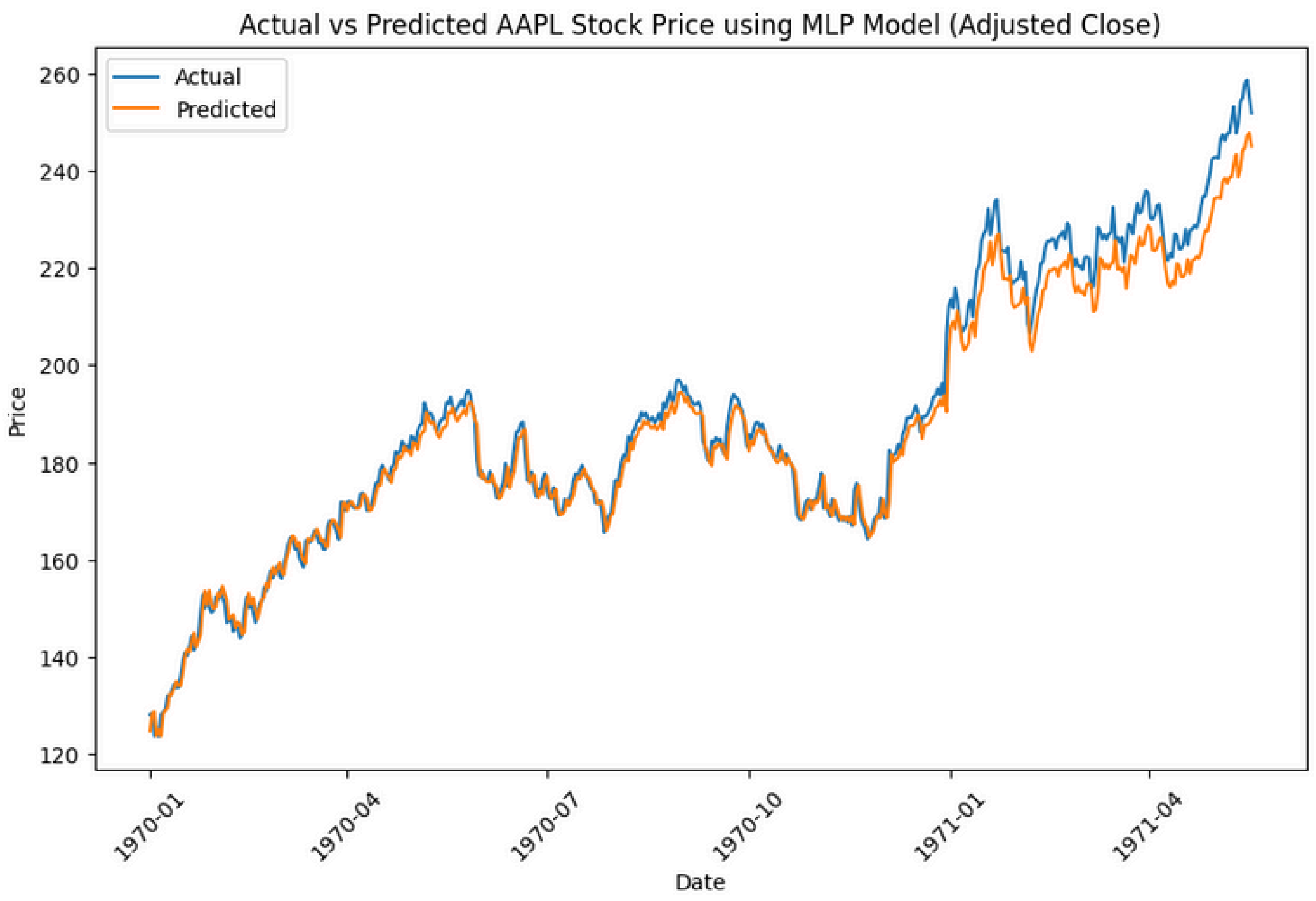


Figure 2: Baseline MLP model for AAPL Stock Price

Mean Absolute Error: **3.235**
Mean Squared Error: **19.320**

Baseline Pricing Models

Long-Short Term Memory (LSTM)

Features:
Lag(-1) value of stock

Parameters:
96 hidden neurons
20 epochs

Test Metrics:
Mean Absolute Error: **2.387**
Mean Squared Error: **10.159**

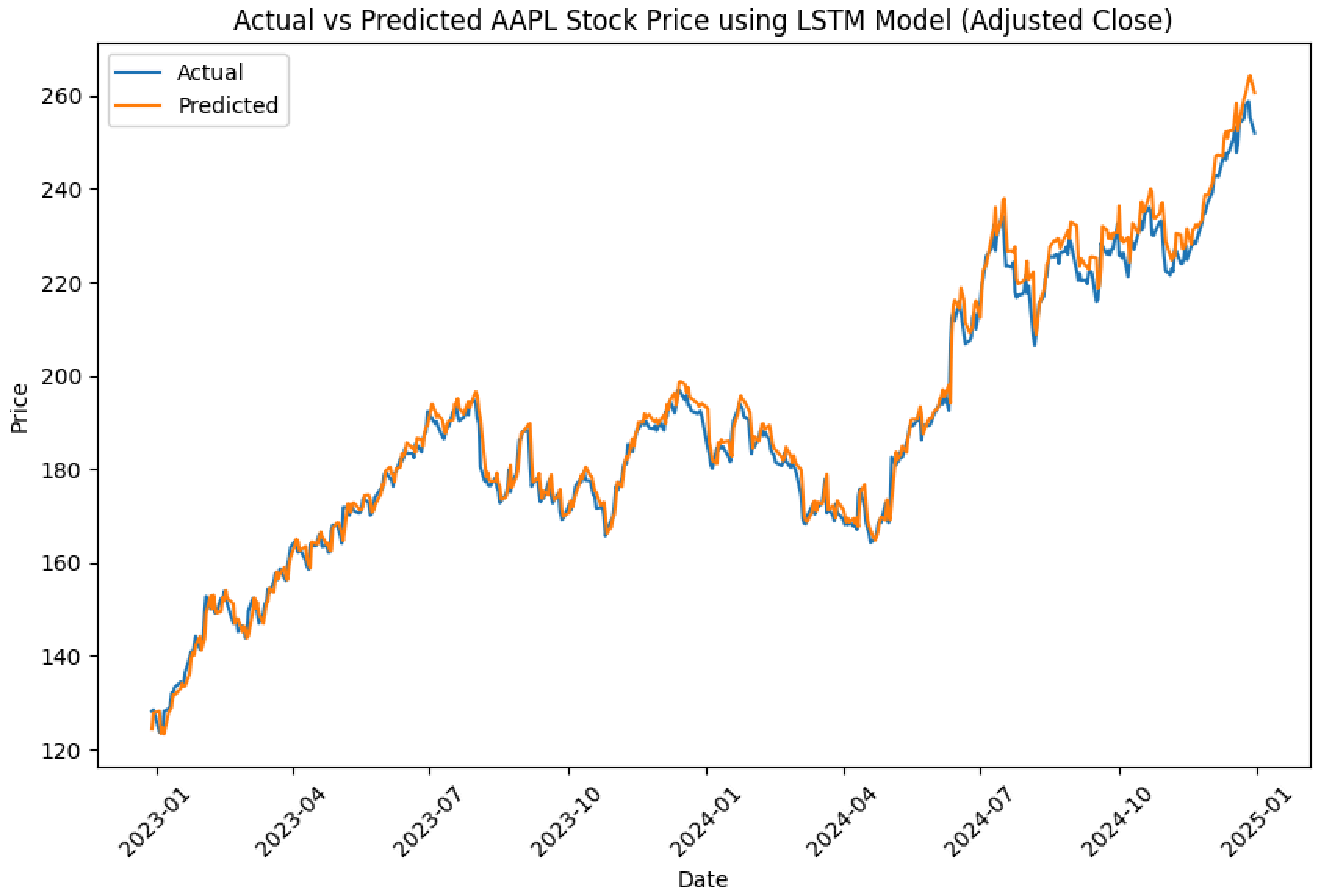


Figure 3: Baseline LSTM model for AAPL Stock Price

Baseline Tuned LSTM Model

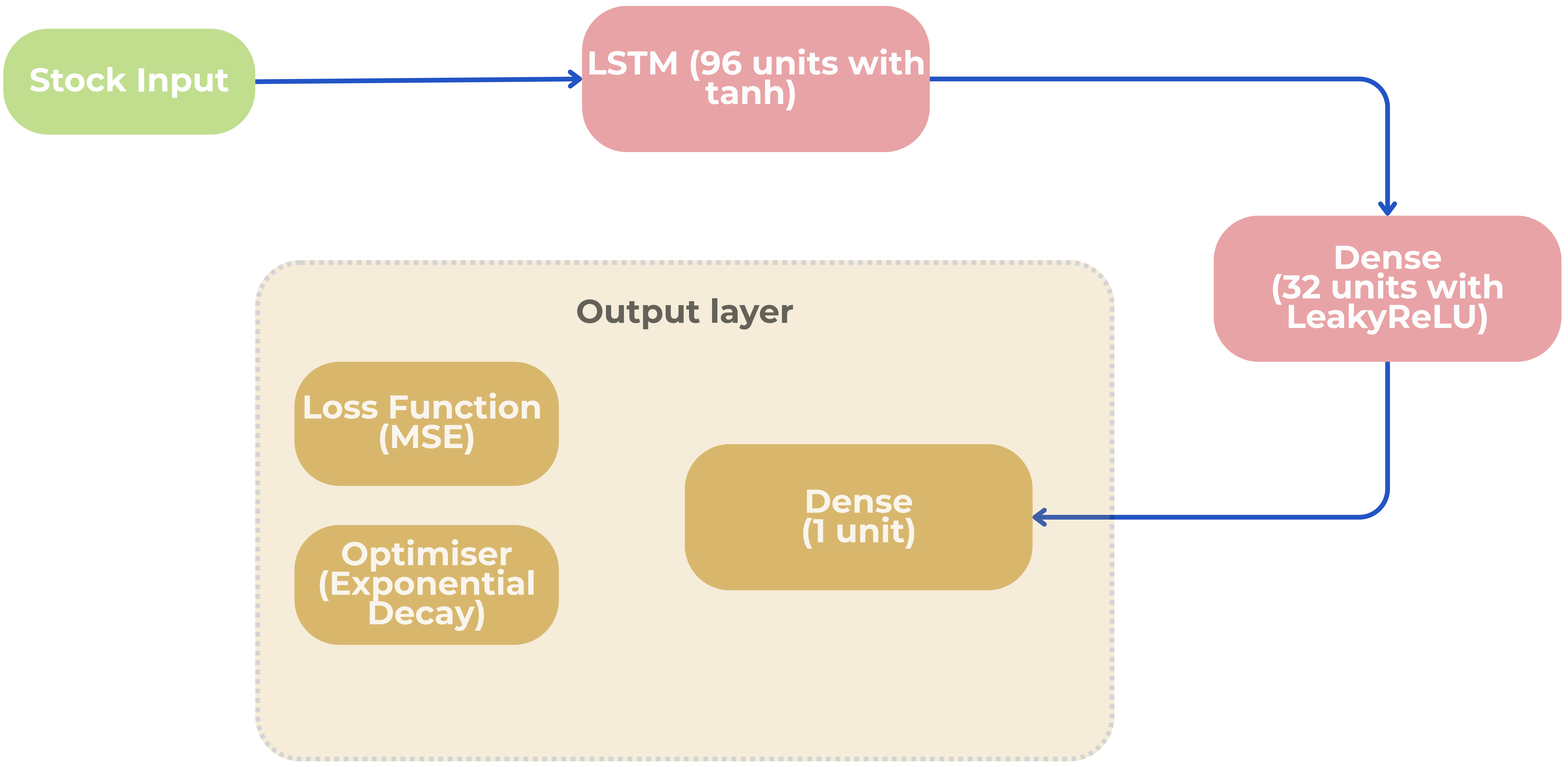


Figure 4: Graph of Baseline Tuned LSTM Architecture

Baseline Tuned LSTM Model

Long-Short Term Memory (LSTM)

- Chosen for ability to capture long-term temporal dependencies in sequential stock data

Architecture & Parameters:

LSTM Layer
- 96 units , tanh (default) Activation

Dense Layer
- 32 units, LeakyReLU Activation

Test Metrics:

Mean Absolute Error: **1.881**

Mean Squared Error: **6.477**

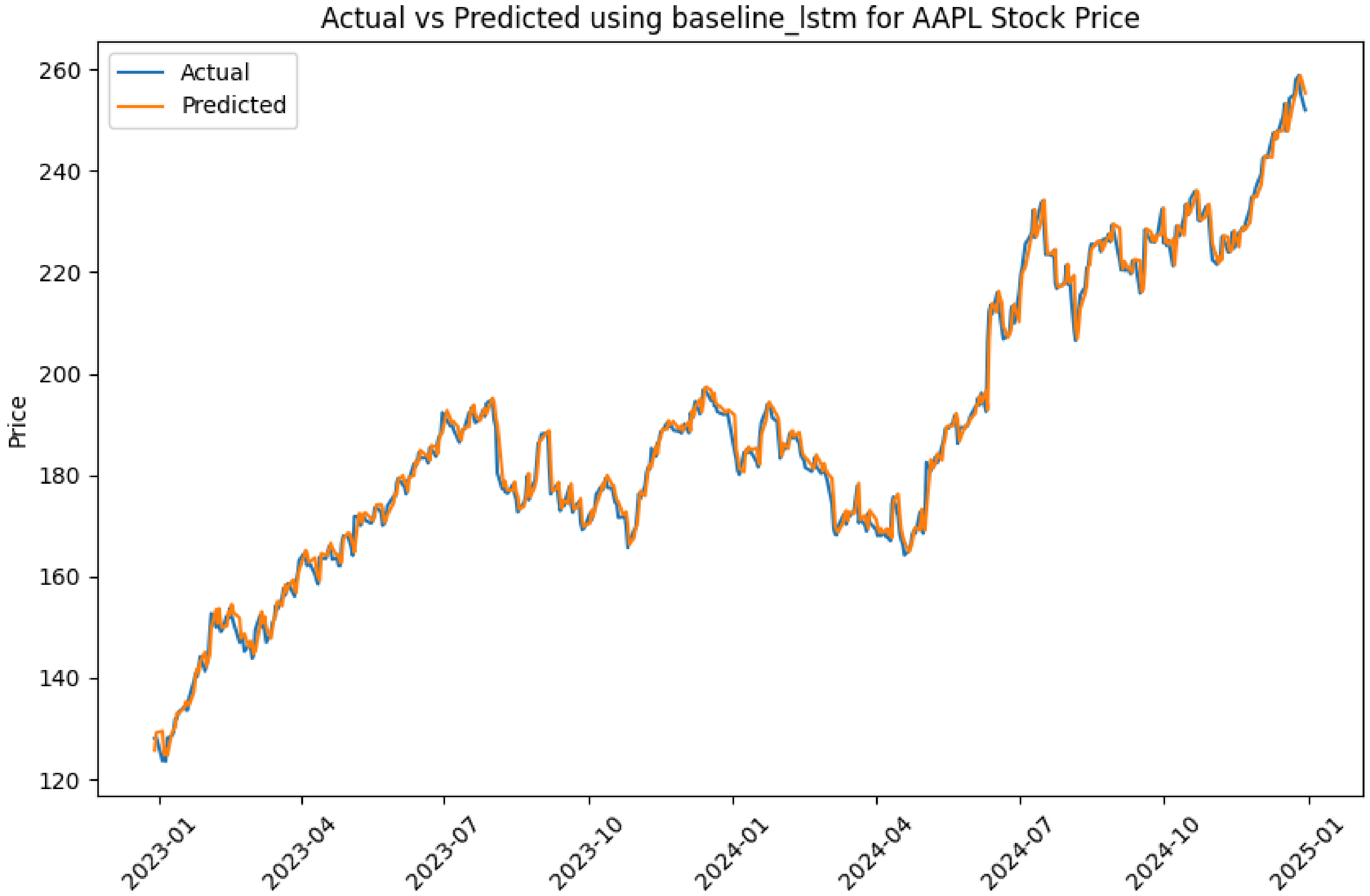


Figure 5: Baseline Tuned LSTM model for AAPL Stock Price

Baseline Sentiment Models

DistilRoBERTa

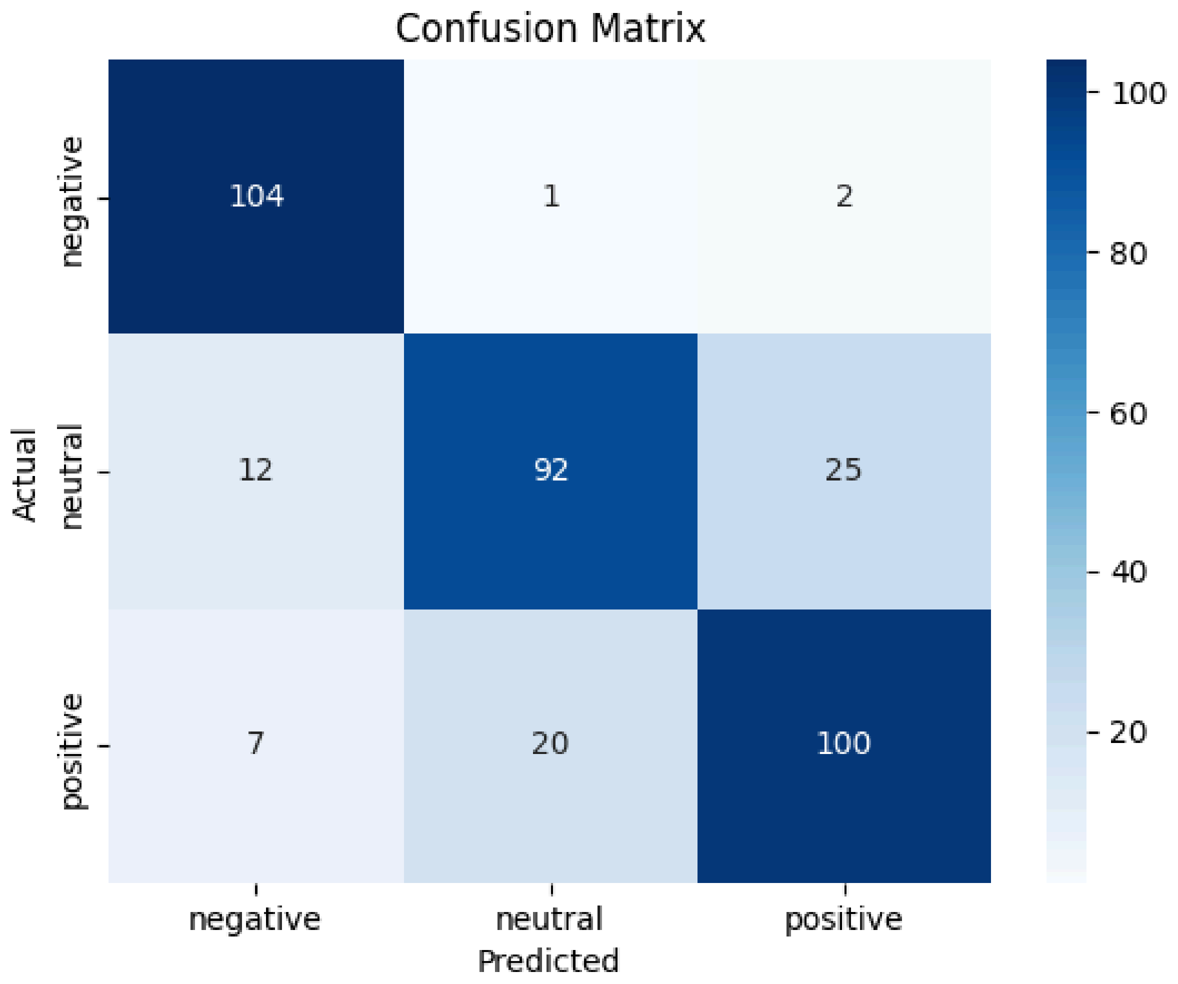


Figure 6: Confusion Matrix for DistilRoBERTa

Accuracy: ***81.5%***
F1 Score: ***81.7%***

DeBERTa

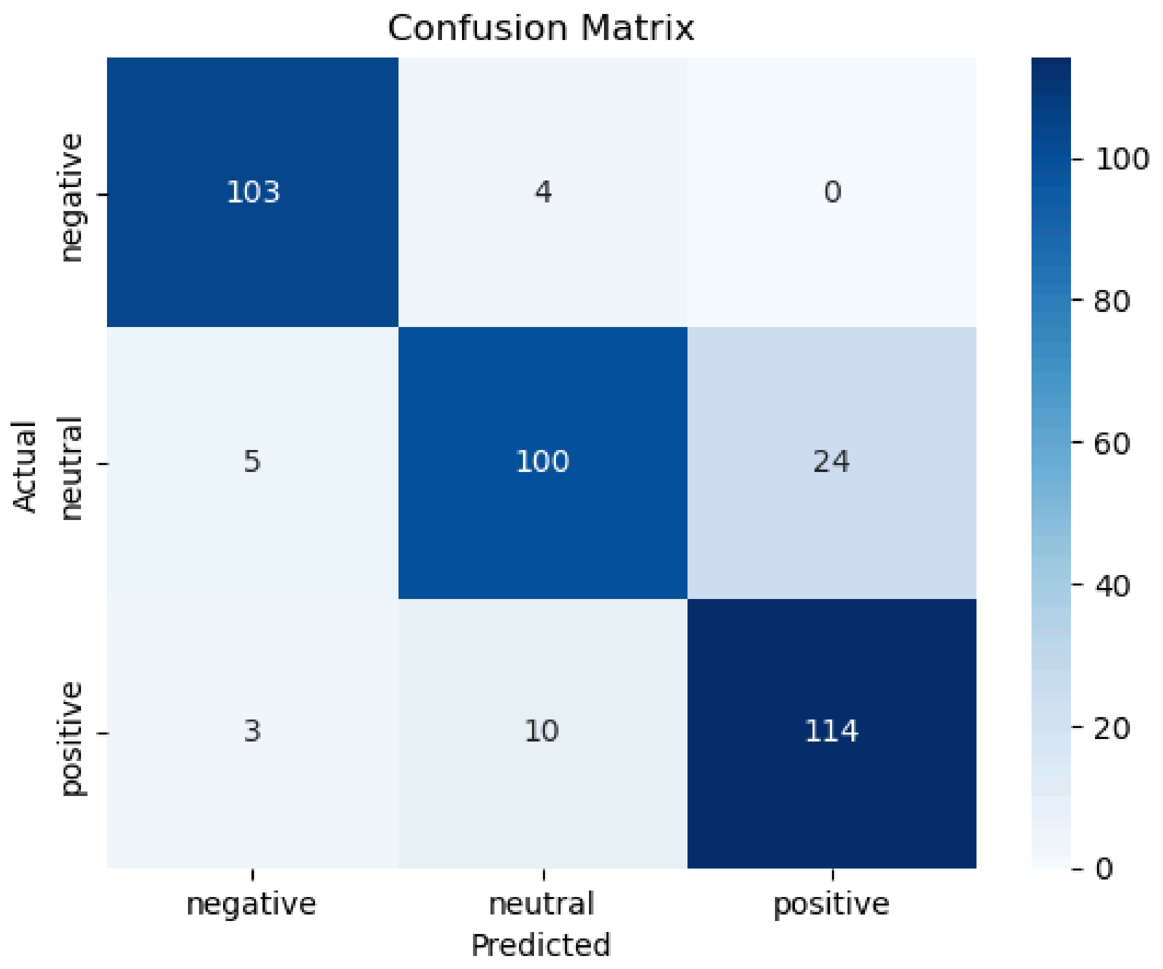


Figure 7: Confusion Matrix for DeBERTa

Accuracy: ***87.3%***
F1 Score: ***87.6%***

Finetuned on Financial Phrasebank from Kaggle

IS460
Group 3

LLM Models

LLM Sentiment

Sentiment Extraction

- Gemini LLM classifies headlines and abstracts into three categories: positive, neutral and negative
- Few-shot prompting for more accurate sentiment predictions

```
def find_sentiment_few_shot(text):
    prompt = f"""
    Classify the sentiment of the financial abstract as 'Positive', 'Negative', or 'Neutral'.
    Use 'Neutral' only if truly unclear. Be decisive—choose 'Positive' or 'Negative' whenever possible.

    Return the result in JSON format:

    Example:
    {{ "Sentiment": "Positive" }}

    Example 1:
    Abstract: "Company Close to Finalizing Its 40 billion dollar funding."
    Sentiment: Positive

    Example 2:
    Abstract: "Trump blocks 10% of funds for key agency in US-China Tech Race."
    Sentiment: Negative

    Example 3:
    Abstracte: "Why Company B could be a key to a Company C's Deal."
    Sentiment: Neutral
```

Figure 8: Examples Provided for Few-Shot Prompting

LLM Sentiment (Dynamic Duration)

Continuous Sentiment Score Extraction

- Gemini LLM evaluates headlines & abstracts
- Sentiment scored from -1 to 1
- Prompt for factors in market dynamics

Estimating Temporal Impact

- Gemini LLM estimates duration impact each financial headline has on Apple's stocks
- Categorised into:
  - 1-3 days (short-lived)
  - 4-7 days (moderate)
  - 8-14 days (significant)
  - 15-30 days (structural)

LLM Sentiment (Dynamic Duration)

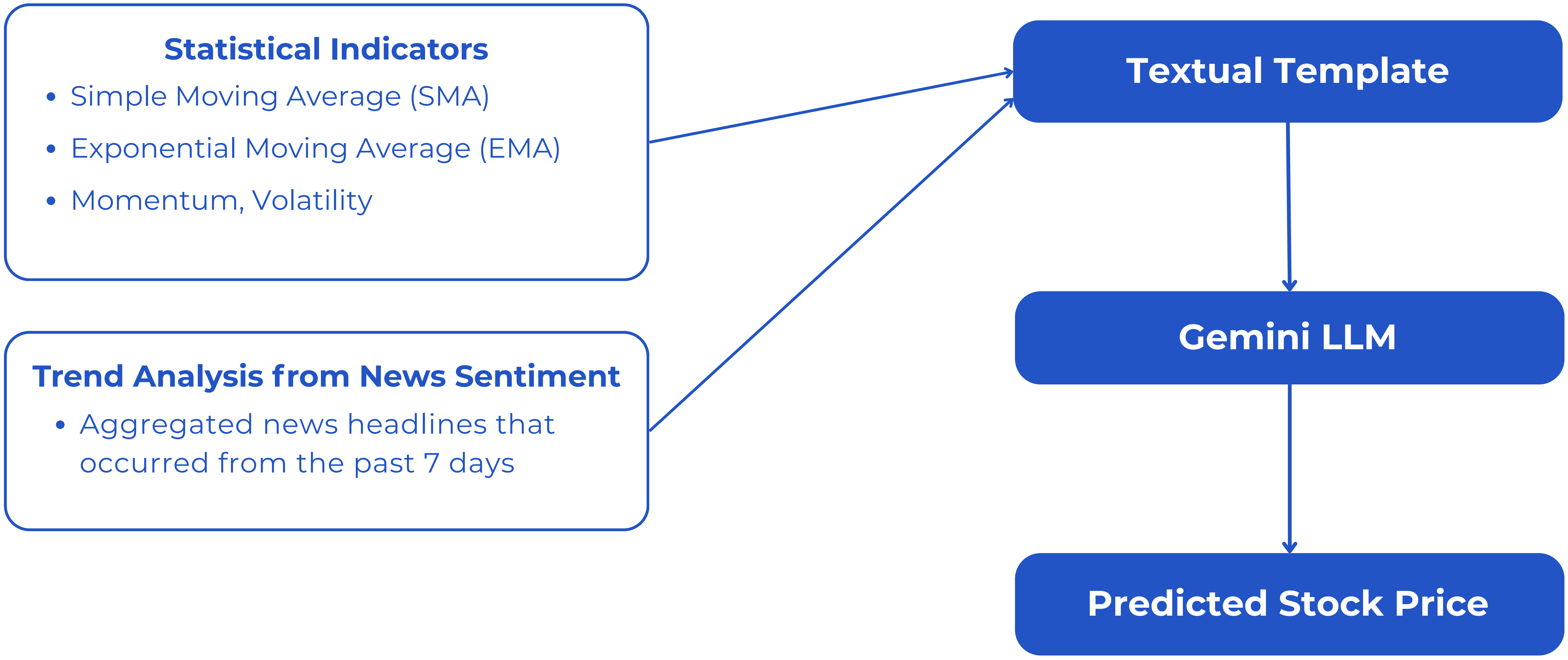
	pub_date	headline	sentiment_score	impact_days
0	2015-04-07	mba programs that get you where you want to go	0.0	3
1	2015-04-14	what were reading	0.0	1
2	2015-04-13	ibm creates watson health to analyze medical data	0.4	8
3	2015-04-22	whats that on beyoncs wrist let me guess an ap...	0.1	1
4	2015-04-01	daily report tech leaders come together to opp...	-0.2	7

Figure 9: Sentiment Scores and Estimated Duration of Impact for Headlines

```
def analyze_duration(text):
    prompt = f"""
    Analyze this financial headline and estimate how many days (1-30) this news might impact the market.
    Consider:
    - 1-3 days for short-lived news
    - 4-7 days for moderately impactful news
    - 8-14 days for significant developments
    - 15-30 days for major structural changes or significant corporate events
    Additionally, for {company_name}-specific news (e.g., product launches, earnings results), consider the historical r
    and the potential market sentiment based on previous similar events.

    For market-wide news (e.g., interest rate changes, regulatory updates), estimate how it will ripple through the tech

    Return result in JSON format like:
    {{ "PotentialImpactDays": Z }}

    Examples:

    Example 1:
    Headline: "Company close to finalizing its 40 billion dollar funding."
    {{ "PotentialImpactDays": 14 }}
```

Figure 10: Few-Shot Prompting for Impact Duration

LLM Price Prediction



LLM Price Prediction

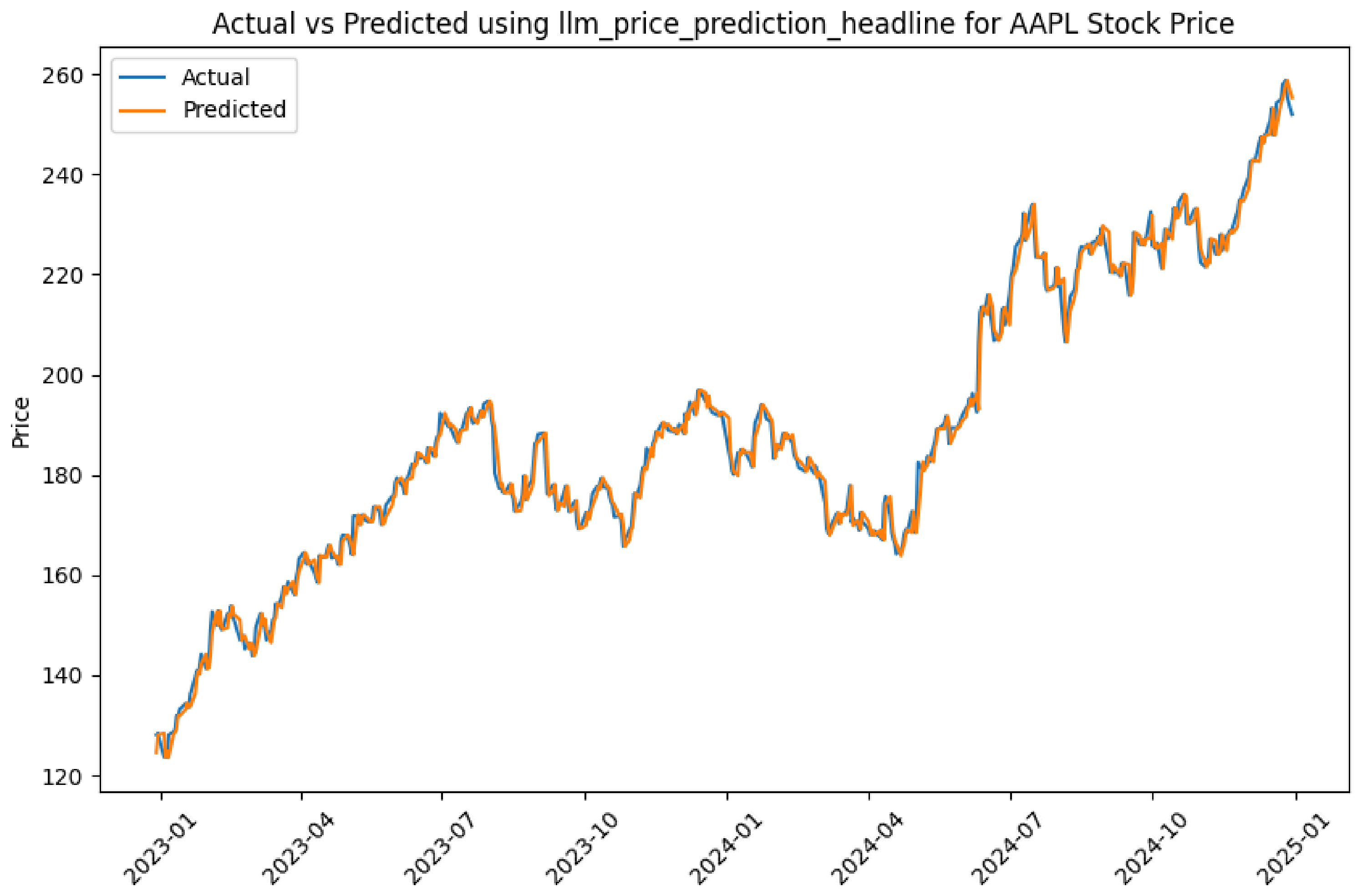


Figure 11: Predicted AAPL Stock Price using LLM

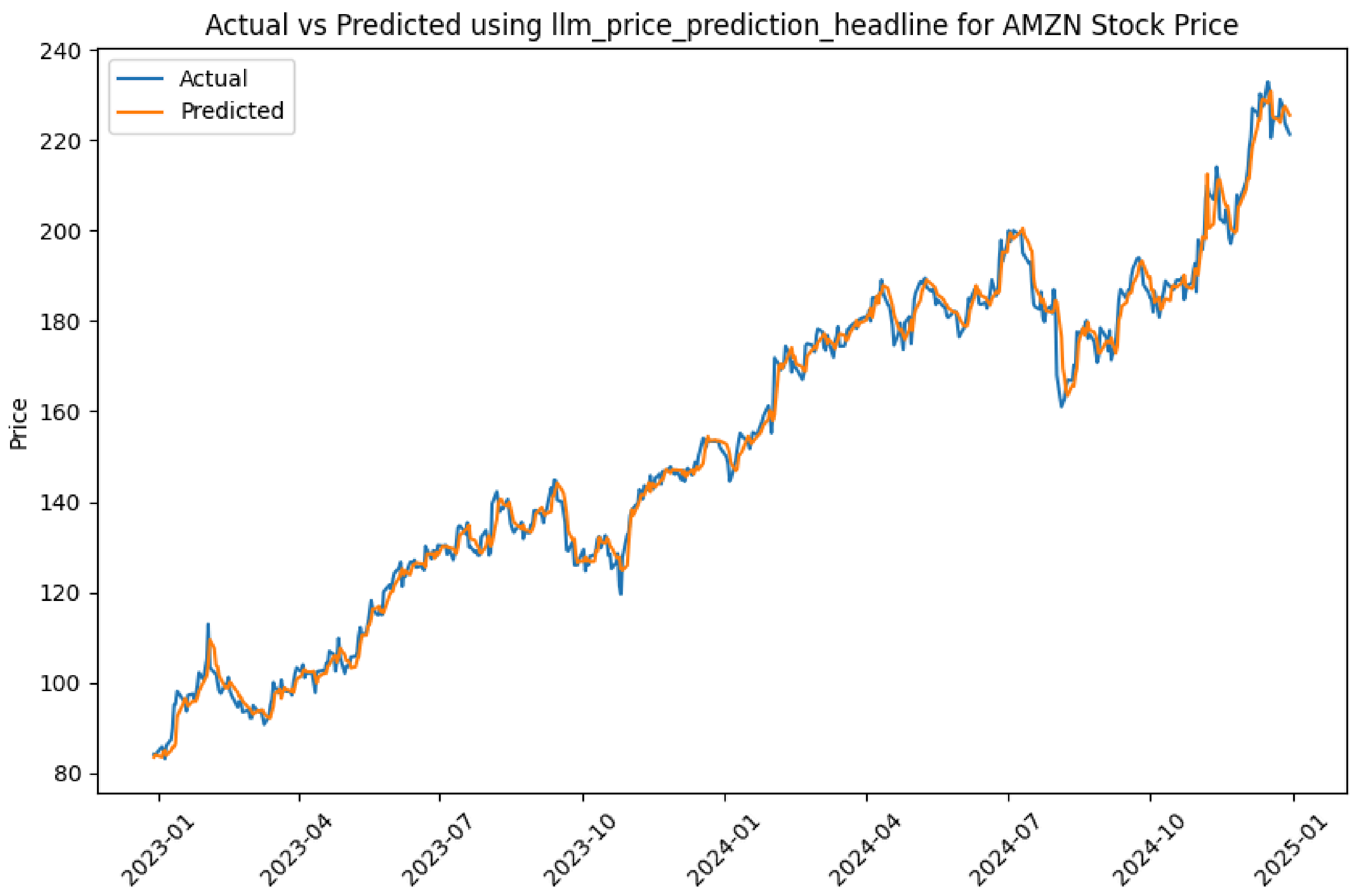


Figure 12: Predicted AMZN Stock Price using LLM

IS460
Group 3

Ensemble Sentiment

Aggregate Sentiment

(1) Voting System (only for ensemble)

- For every article, choose sentiment based on max vote from 3 baseline models
- *Defaults to 0 (Neutral) during a tie*

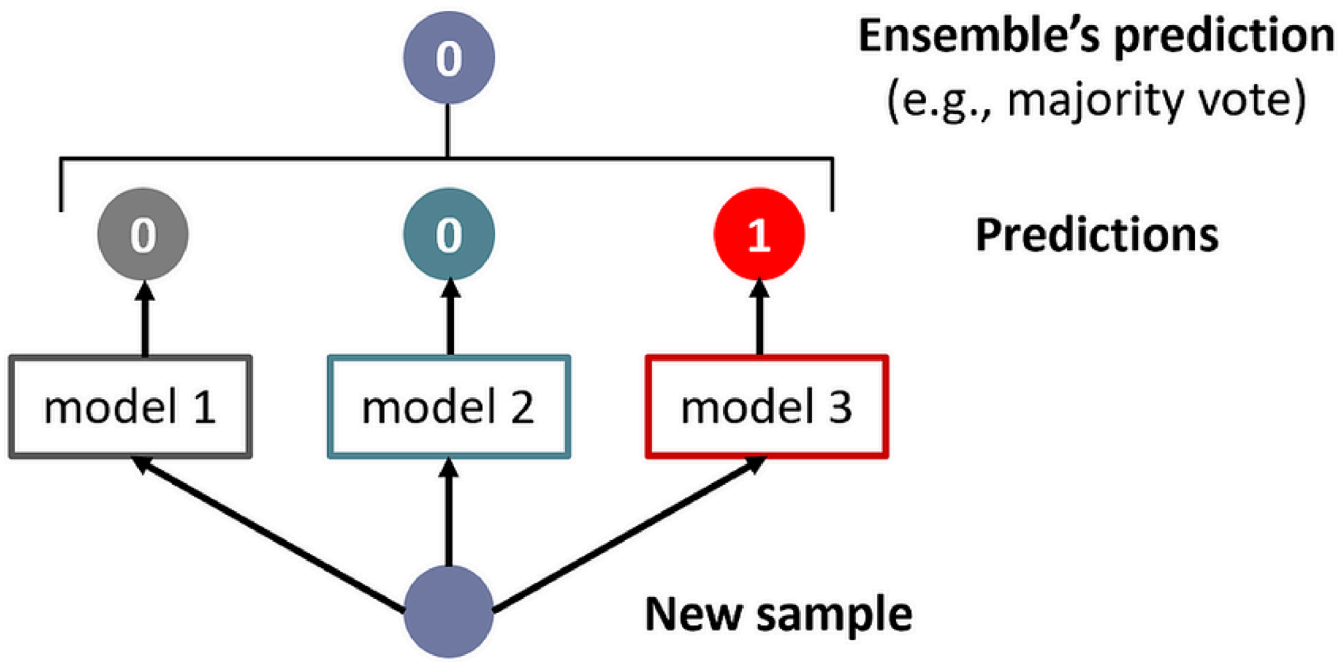


Figure 13: Graph on Ensemble Voting

Yehoshua, R. (2023, August 9). Introduction to ensemble methods - towards AI. Medium. https://pub.towardsai.net/introduction-to-ensemble-methods-226a5a421687?gi=5118569a2ecb

(2) Sentiment Effect Duration (all models)

$$S_t = \sum_{d=t-w}^{t-1} \sum_{i \in A_d} \left(w - (t-d) + 1\right) \cdot \alpha \cdot s_i$$

Figure 14: Sentiment Effect Duration Formula

- ***S_t***: Sentiment score for day (*t*)
- ***w***: Window size (number of days to look back, excluding today)
- ***A_d***: Set of articles published on day (*d*)
- ***s_i***: Sentiment score of article (*i*)
- ***α***: Sentiment effect multiplier

- **(*w*−(*t*−*d*)+1)**: Time decay function — weight decreases the further back the article is

IS460
Group 3

Model Architecture

Multimodal Approach

Early Fusion Model

- Combines stock and text data immediately by concatenating them
- Process through a single LSTM layer, a hidden density layer and outputs price

Late Fusion Model

- Processes stock and text data through separate LSTM networks
- Combines LSTM embeddings, processes through final hidden density layer and outputs price

Attention-Based Model

- Adds Attention-Based mechanism to the Early Fusion Model
- Accounts for importance of features across timesteps*

* Found that adding extra timesteps (lagstep > 1) increased error

Early Fusion Approach

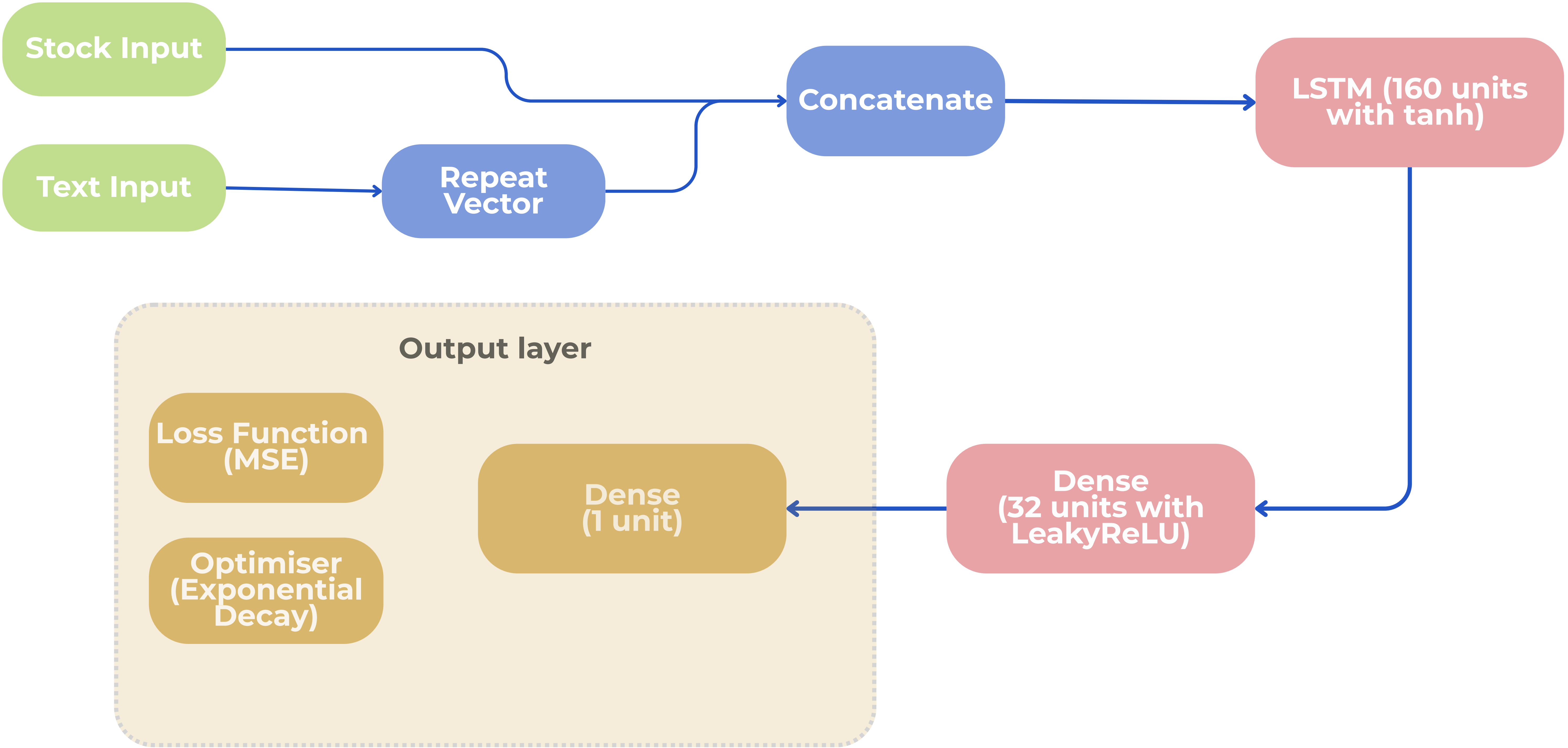


Figure 15: Graph of Multimodal Early Fusion Approach Architecture

Model Training

Details

- **Loss Function:** Mean Squared Error (MSE)
- **Train-Test Split:** 80-20
- **Batchsize:** 32

- To **account for training variability**, we ran each model 30 times with max 50 epochs
- **Early Stopping** based on loss in error with patience of 5 epochs, min delta of 1e-4

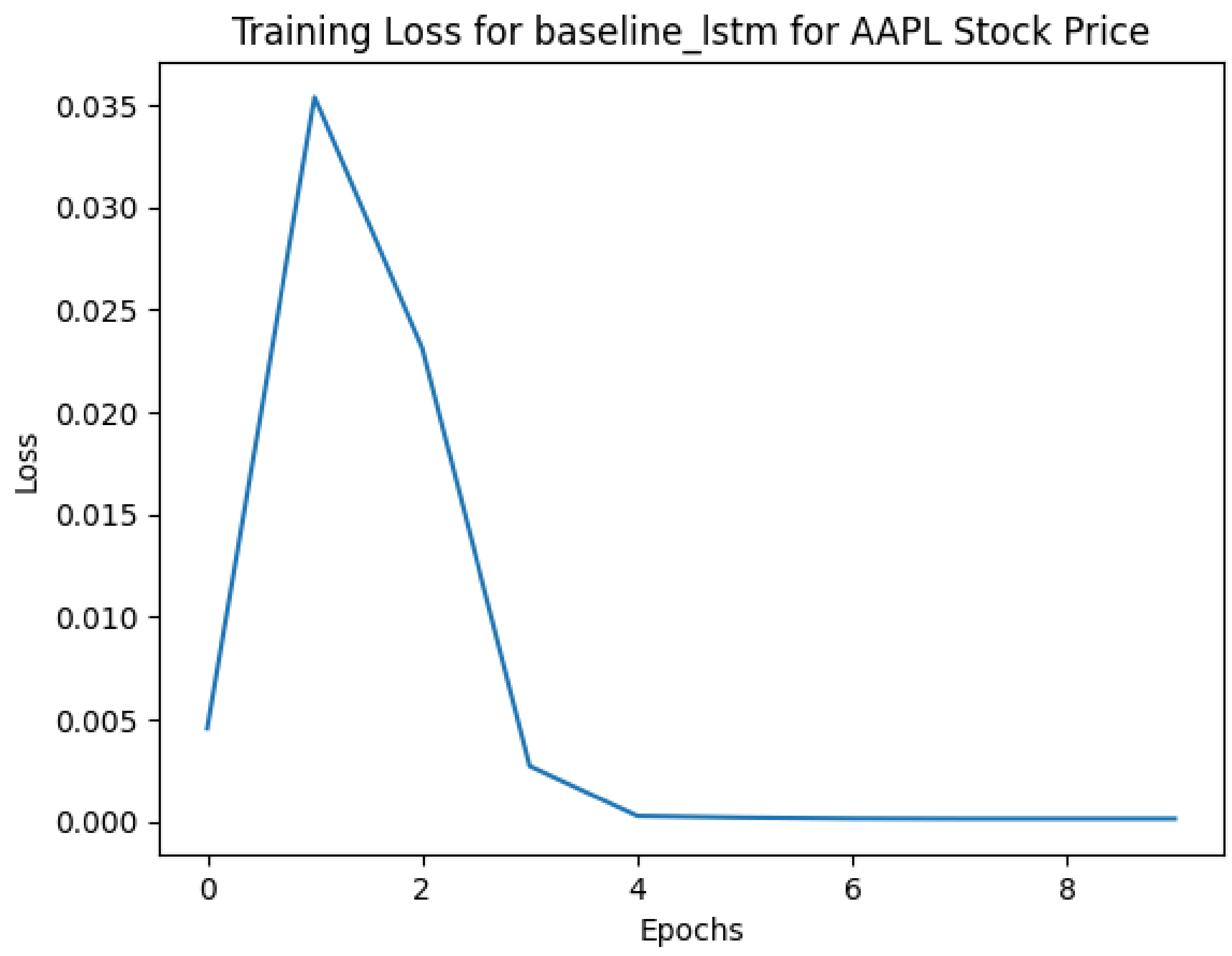


Figure 16: Graph of Training Loss for AAPL stock LSTM model

Results

Discussion of Results

Stock	Best Model	MSE	% Decrease in MSE from Baseline LSTM
AAPL	DeBERTa Sentiment	6.48	**1.04%**
AMZN	Ensemble Sentiment	9.50	**8.50%**
MSFT	DeBERTa Sentiment	24.95	**0.42%**
CRM	Baseline LSTM	27.40	---
IBM	Baseline LSTM	5.68	---
NVDA	Baseline LSTM	6.72	---

1. LLM Models did not outperform Bidirectional Transformers

2. Adding Sentiment Models did not improve predictions for CRM, IBM, NVDA

Figure 17: Table of Best Models for each Stock

Possible Explanations

1. Underperformance of LLM Models: Off-The-Shelf LLM used

- Lack of fine-tuning for LLM sentiment predictions
  - Financial dataset used for tuning: 1.4k rows < 2k could lead to deteriorated performance (Vieira, I. et al 2024)
- Only experimented with the efficacy of Gemini Models

2. No Improvement for Certain Stocks: Sparsity of News Data

- Worse performance in price predictions after sentiment addition could be attributed to sparsity of news
  - CRM, IBM, NVDA (~5% of articles for AMZN)
- Observed that less news increases errors for price prediction across other stocks

Sparsity of News Data

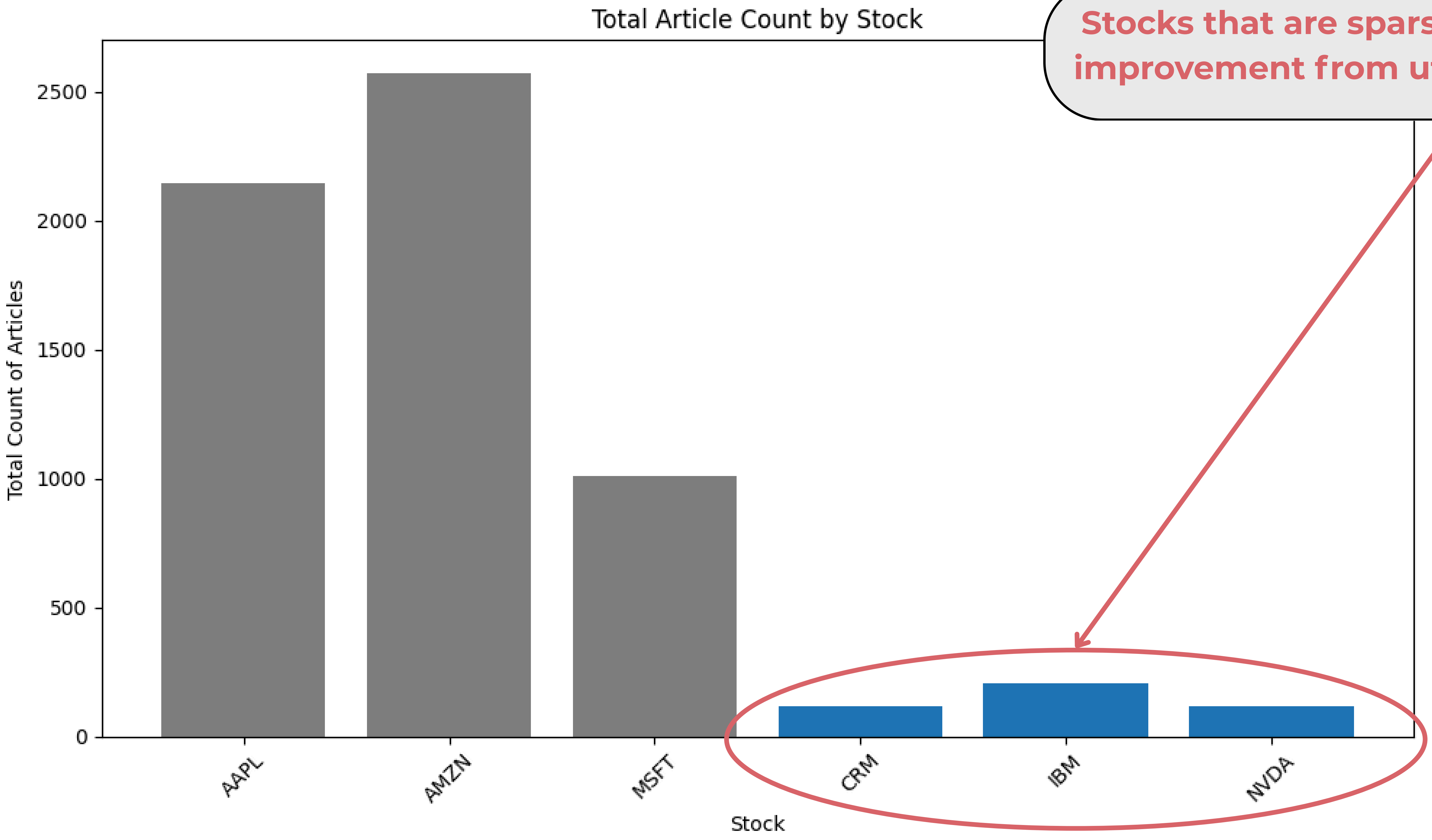


Figure 18: Bar Chart of Article Counts by Day for each Stock

Stocks that are sparse in articles observe no improvement from utilising article sentiment

Stock	Best Model	Baseline MSE
CRM	Baseline	27.40
IBM	Baseline	5.68
NVDA	Baseline	6.72

Excerpt from Fig. 17.

Sparsity of News Data

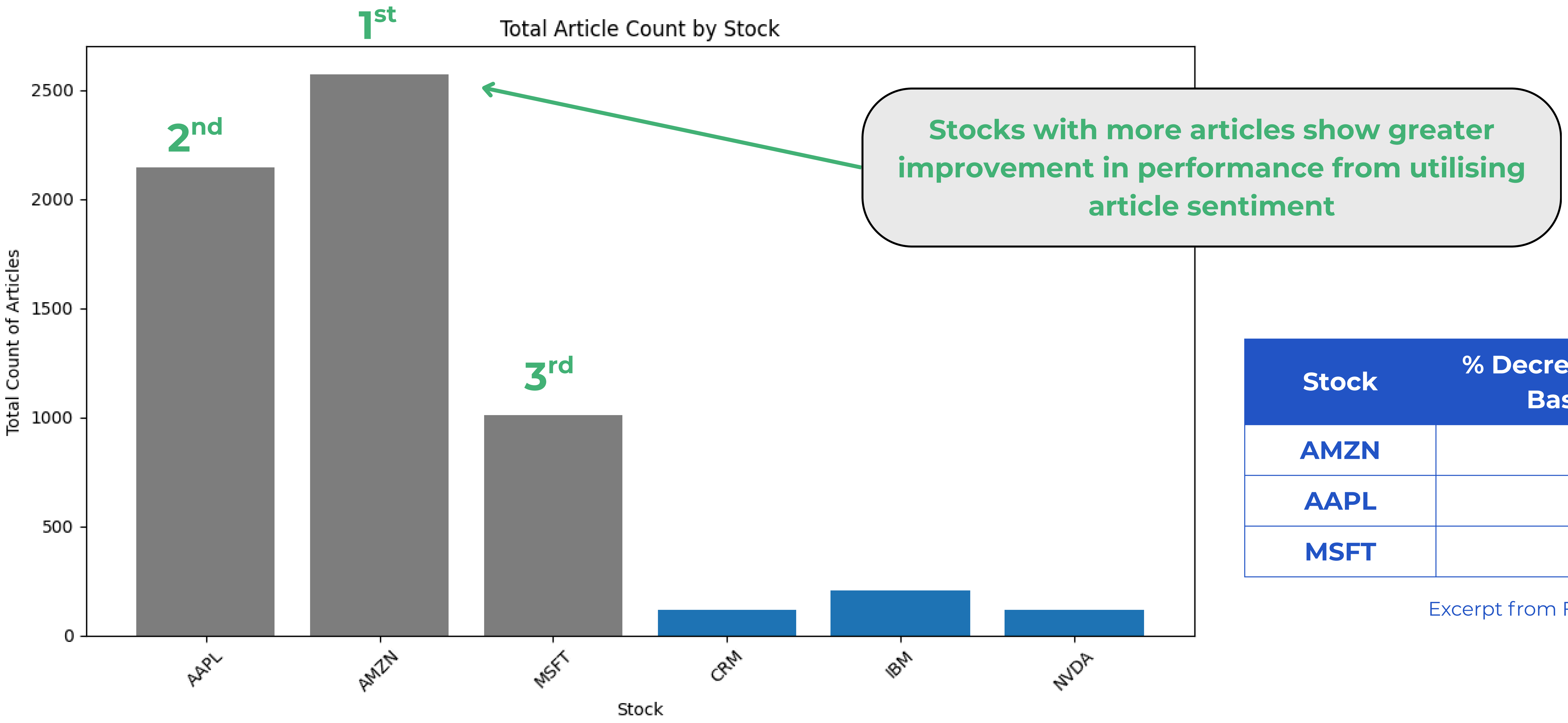


Figure 18: Bar Chart of Article Counts by Day for each Stock

Stock	% Decrease in MSE from Baseline LSTM
AMZN	8.50%
AAPL	1.04%
MSFT	0.42%

Excerpt from Fig. 17.

Summary of Findings

1. Considering news sentiment in stock price prediction may only improve performance when there is sufficient data

2. DeBERTa shows strongest performance in financial sentiment classification: best performing predictor of sentiment across all transformer models, sometimes beating ensemble sentiment models

3. Task-specific Fine-Tuned transformers models may outperform larger, general purpose LLMs

Limitations

Limitations

Data

- Limited Data Volume:
  - News: Few company-specific articles
  - Stock price: Only one data point per time unit (e.g. daily)

- Noise in Data:
  - News data might not always be relevant to the company or market situation
  - Not all articles are as important

LLM Models

- Off-the-shelf LLMs lack domain-specific financial knowledge
- Extensive financial text is needed for fine tuning

Limitations

Temporal Issues

- Stock prices are autocorrelated and models can exploit this short-term inertia without truly learning meaningful features from the text.
- Price reactions may occur before the news is published
- By aligning news and prices by timestamps, there could be a mismatch
    - News breaks after the market closes, but prices move the next day

IS460
Group 3

Future Works

Future Works

Parameters

Investigate Look-back and sentiment effect duration parameters

- Fixed at look-back = 7 and sentiment effect = 0.01% (Wang, 2019), but the ideal parameters could differ across stocks

Custom Loss Function

- Implement a custom loss function that penalises negative changes more heavily to better capture downside risk in market sentiment (Dessain, 2023)

Wang, Zhaoxia, Ho, S.-B., & Lin, Z. (2018). Stock market prediction analysis by incorporating social and news opinion and sentiment. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 1375–1380. https://doi.org/10.1109/icdmw.2018.00195
Dessain, J. (2023). Improving the prediction of asset returns with machine learning by using a custom loss function. Advances in Artificial Intelligence and Machine Learning, 03(04), 1640–1653. https://doi.org/10.54364/aaiml.2023.1193

Future Works

Data

Broaden Text Data Sources

- Expand the text dataset by incorporating additional financial news outlets such as Financial Times and Bloomberg to improve coverage/use of financial jargon and reduce source bias

Abstract-Level Sentiment Analysis

- Explore whether using article abstracts, rather than full texts, improves sentiment clarity and predictive performance

Future Works

Architecture

Advanced LSTM Feature Engineering

- Enrich LSTM input features using seasonal-trend decomposition and explore optimisation strategies like adaptive learning and niching-based backtracking search algorithm for time series forecasting (Wu et al., 2023)

Ensemble Large Language Models

- Investigate the performance gains from ensembling multiple LLMs for more robust and diverse sentiment interpretation

Wu, Y., Meng, X., Zhang, J., He, Y., Romo, J. A., Dong, Y., & Lu, D. (2023). Effective LSTMs with seasonal-trend decomposition and adaptive learning and niching-based backtracking search algorithm for time series forecasting. Expert Systems With Applications, 236, 121202. https://doi.org/10.1016/j.eswa.2023.121202

References

- Bhattacharjee, I., & Bhattacharja, P. (2019). Stock price prediction: A comparative study between Traditional Statistical Approach and machine learning approach. 2019 4th International Conference on Electrical Information and Communication Technology (EICT), 1–6. https://doi.org/10.1109/eict48899.2019.9068850
- Cote, D. (2022, November 18). Hybrid (multimodal) neural network architecture: Combination of tabular, textual and image inputs to predict house prices. Medium. https://medium.com/@dave.cote.msc/hybrid-multimodal-neural-network-architecture-combination-of-tabular-textual-and-image-inputs-7460a4f82a2e
- Maqbool, J., Aggarwal, P., Kaur, R., Mittal, A., & Ganaie, I. A. (2023). Stock prediction by integrating sentiment scores of financial news and MLP-Regressor: A machine learning approach. Procedia Computer Science, 218, 1067–1078. https://doi.org/10.1016/j.procs.2023.01.086
- Muhammad, I., & Rospocher, M. (2025). On assessing the performance of LLMS for target-level sentiment analysis in financial news headlines. Algorithms, 18(1), 46. https://doi.org/10.3390/a18010046
- Vieira, I., Allred, W., Lankford, S., Castilho, S., & Way, A. (2024, September 5). How Much Data is Enough Data? Fine-Tuning Large Language Models for In-House Translation: Performance Evaluation Across Multiple Dataset Sizes. arXiv.org. https://arxiv.org/abs/2409.03454
- Wang, M., Izumi, K., & Sakaji, H. (2024). LLMFACTOR: Extracting profitable factors through prompts for explainable stock movement prediction. Findings of the Association for Computational Linguistics ACL 2024, 3120–3131. https://doi.org/10.18653/v1/2024.findings-acl.185
- Wang, Zhaoxia, Ho, S.-B., & Lin, Z. (2018). Stock market prediction analysis by incorporating social and news opinion and sentiment. 2018 IEEE International Conference on Data Mining Workshops (ICDMW), 1375–1380. https://doi.org/10.1109/icdmw.2018.00195
- Zong, C., & Zhou, H. (2024, December 2). Stock movement prediction with multimodal stable fusion via gated cross-attention mechanism. arXiv.org. https://doi.org/10.48550/arXiv.2406.06594
- Vieira, I., Allred, W., Lankford, S., Castilho, S., & Way, A. (2024). How Much Data is Enough Data? Fine-Tuning Large Language Models for In-House Translation: Performance Evaluation Across Multiple Dataset Sizes (Version 2). arXiv. https://doi.org/10.48550/ARXIV.2409.03454

References

- *Hugging face – the AI community building the future*. Hugging Face –. (n.d.). https://huggingface.co/
- Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. Proceedings of the International AAAI Conference on Web and Social Media, 8(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550
- The New York Times. (n.d.). The New York Times. https://developer.nytimes.com/docs/articlesearch-product/1/overview
- Yahoo! (2025, February 15). *Apple Inc. (AAPL) stock price, news, Quote & History*. Yahoo! Finance. https://finance.yahoo.com/quote/AAPL/

IS460
Group 3

Final Project Presentation

Thank You

LLM-Enhanced
Stock Price Prediction:
Improving the Synergy of
Price Prediction and Sentiment Analysis Models

Brenda Su Cheryl Yeo Chong Le Kai Wong Swee Kiat
Tammy Lee Trina Tan Xavier Santimano

IS460
Group 3

Appendix

Late Fusion Approach

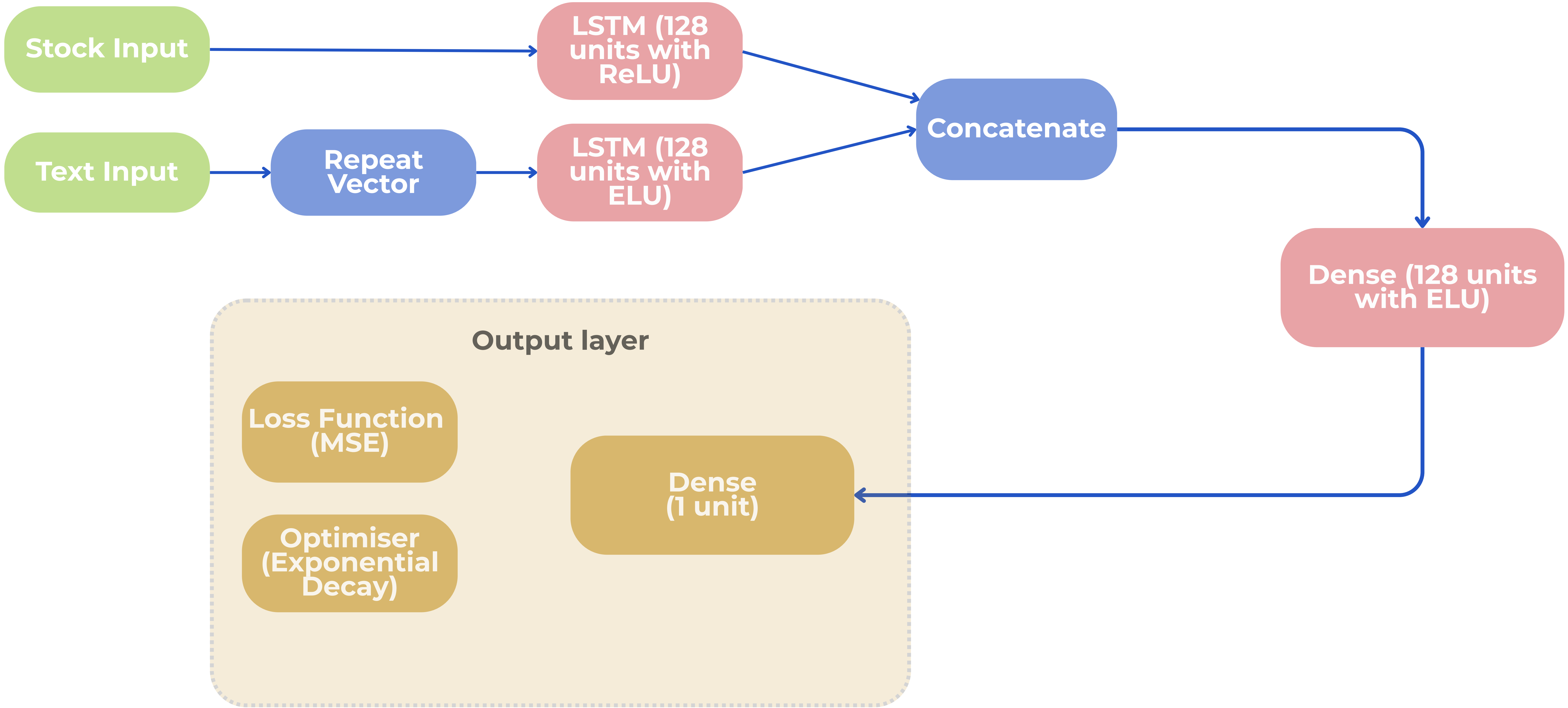


Figure 19: Graph of Multimodal Late Fusion Approach Architecture

Attention-Based Model

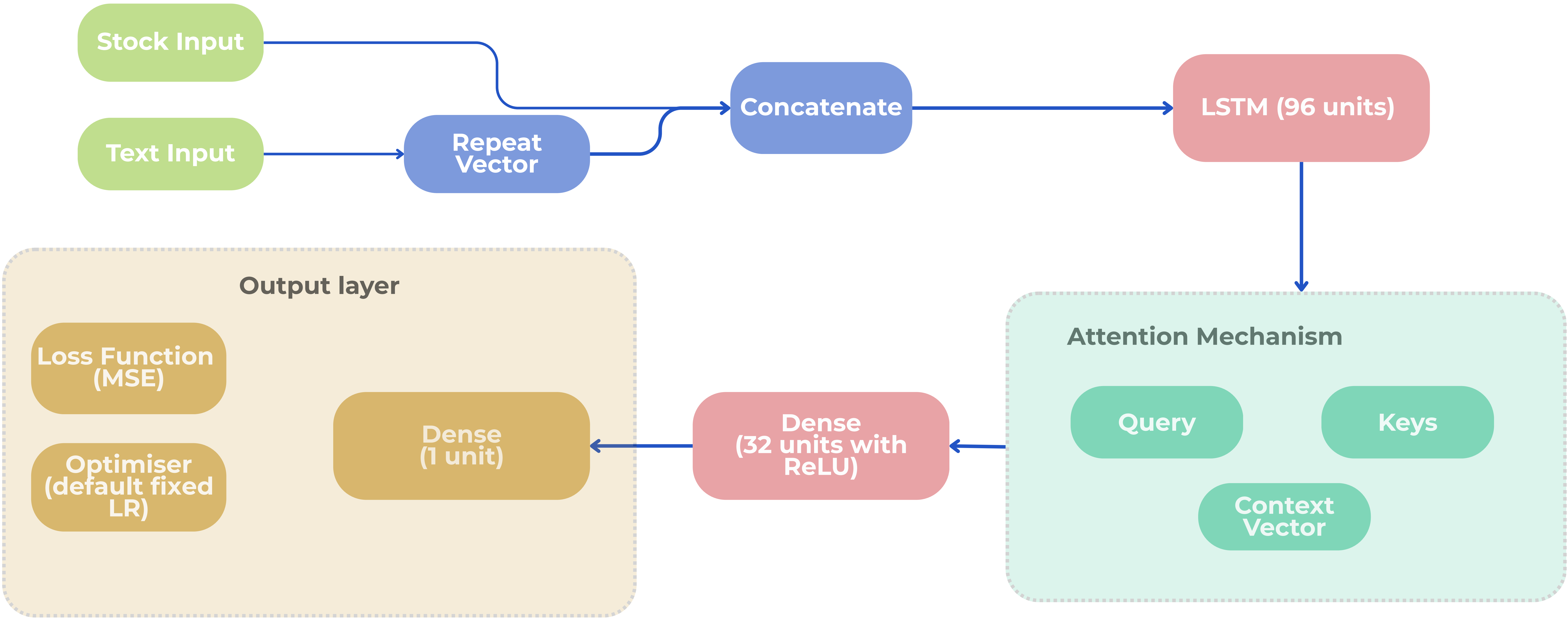


Figure 20: Graph of Multimodal Attention-Based Architecture