Evaluation of News Sentiment Impact on Crude Oil Prices via BERT-Enhanced Deep Learning Classifiers

Ahmad Muhariya[1*], Indrawan Ady Saputro[2], Dziky Ridhwanullah[3]

[1] *Department of Computer Engineering, Universitas Tiga Serangkai, Indonesia*
[2] *Department of Informatics, STMIK AMIKOM Surakarta, Indonesia*
[3] *Department of Informatics, Universitas Tiga Serangkai, Indonesia*
*corr-author: ahmadmuhariya@tsu.ac.id

Abstract – Various factors, including market perception reflected in media information, Influence crude oil price fluctuations. This study aims to analyse the Influence of news sentiment on crude oil price movements using a deep learning–based sentiment analysis approach. The dataset consists of 108 news headlines and daily closing oil prices from January to May 2025. It is important to note that this dataset is relatively small for deep learning models like LSTM, GRU, and BiLSTM, which constitutes a major constraint for this study. The news text was processed with case folding, tokenisation, stopword removal, and lemmatisation (not stemming to preserve semantic integrity for BERT), then automatically labelled using the DistilBERT model. The BERT-based vector representations were used as input for three classification models: LSTM, GRU, and BiLSTM. The evaluation results showed that all three models achieved the same average validation accuracy of 85.27%. However, the GRU model is identified as the optimal performer, achieving the lowest validation loss (0.3324), indicating better generalisation performance than LSTM and BiLSTM. Further analysis reveals that news sentiment tends to align with oil price trends, particularly during significant market shifts.

Keywords: Sentiment analysis, deep learning, DistilBERT, crude oil prices, LSTM.

I. INTRODUCTION

Globalization and advances in information technology have significantly impacted various aspects of human life, including the operation of global energy markets. Crude oil is strategically essential in this regard, as it is one of the most critical commodities in the worldwide economy. Crude oil prices can fluctuate due to numerous factors, including supply and demand, OPEC policies, geopolitical conditions, and widespread market opinion in the media and digital channels. In today's information age, the abundance of news, reports, and opinions widely disseminated through online news portals, social media, and economic forums can rapidly influence market participants' decisions in the short and medium term [1]. This demonstrates that information is not only a means of conveying facts but also a driver of market sentiment, which can directly impact crude oil price volatility [2].

Attention to the influence of news on commodity price movements has increased in recent years, particularly due to advances in data analytics and machine learning [3]. Sentiment analysis is a technique for identifying and categorizing emotional content in text, whether it is positive, negative, or neutral. It is a highly sought-after approach [4]. Sentiment analysis allows us to assess whether circulating news is likely to drive crude oil prices up or down. For example, when news indicates a possible supply crunch caused by geopolitical tensions, market sentiment tends to be positive towards oil prices, which then drives prices up. Conversely, when news indicates a supply glut or a global economic slowdown, the market tends to react negatively, resulting in a decline in oil prices [5]. Therefore, the ability to capture and measure sentiment from such news is crucial in understanding crude oil price dynamics.

Quantitative studies show that automatically detected sentiment helps predict commodity markets. Research finds that sentiment signals from media and Twitter reveal new evidence about oil price volatility [2]. News sentiment also improves the accuracy of oil futures return and volatility forecasts [5]. These results suggest that using sentiment scores regularly improves forecasting. Sentiment analysis provides useful insights and is a strong tool for predicting short-term trends in the crude oil market.

Approaches such as Support Vector Machine (SVM), Naive Bayes, and K-Nearest Neighbor (KNN) have been increasingly used in text-based sentiment analysis, alongside the development of classification methods in the field of artificial intelligence [6]. These approaches offer advantages such as ease of implementation and the ability to perform classification based on the similarity of text features. In other words, these models enable text-based sentiment identification [7]. However, on the other hand, advances in deep neural network architectures have also been encouraged. Conversely, more complex sentiment analysis methods, such as Long Short-Term Memory (LSTM), are an evolution of the Recurrent Neural Network (RNN) [8]. Its ability to overcome the problem of vanishing gradients and understand long-term context in sentences makes LSTM a popular tool for capturing temporal patterns in text data.

Additionally, Bidirectional LSTM (BiLSTM) can process information from both forward and backward directions, providing a richer context for understanding the text's meaning. As a result, it is becoming increasingly popular. As an alternative, the Gated Recurrent Unit (GRU), which is more lightweight than LSTM, offers higher computational speed without compromising sequential processing accuracy [9]. These three methods have proven effective in analyzing sentiment across various domains, particularly in the financial and cryptocurrency sectors, such as Bitcoin, by utilizing a text analysis-based approach to predict price direction [10]. However, the computational complexity and large data requirements mean that these methods have not been widely applied in oil commodity market sentiment analysis, particularly in local research [11]. Considering the need for researchers and investors in Indonesia to understand the dynamics of the global oil market in relation to their investment decisions, this study utilizes English-language global news data obtained through news portals and public datasets.

This phenomenon necessitates research specifically examining how sentiment contained in news about crude oil influences its price movements. Furthermore, it is also essential to assess how well classification techniques such as KNN identify this relationship. This study aims to provide a more comprehensive understanding of the relationship between public opinion and energy price changes by combining historical oil price data and relevant news. Furthermore, this study aims to investigate how classification models categorize news sentiment and relate it to changes in oil prices. The results of this study will aid in the development of data-driven trading strategies, assist market participants in making decisions more aligned with changes in global data, and contribute to the development of evidence-based and adaptive energy policies.

Previous Research has extensively explored the use of end-to-end transformer models for commodity market prediction, which typically require large-scale datasets to achieve convergence. However, there is a significant Research gap in the context of highly specific and limited datasets, where pure transformer models often suffer from overfitting. Unlike previous studies that rely on complex hybrid models or massive social media datasets, this study offers a novel hybrid-lite approach: leveraging the efficiency of DistilBERT for accurate automatic labelling of short texts, then employing RNN layers (LSTM, GRU, BiLSTM) as classifiers.

The use of RNNs over BERT vector representations was chosen because they can capture sequential dependencies in small datasets more reliably than full transformer models with much larger parameter sets. Furthermore, while most of the literature focuses on historical data, this study provides a more up-to-date contribution by analysing crude oil market dynamics from January to May 2025. Thus, this Research fills a methodological gap regarding the effectiveness of deep learning-based sentiment classification under data-limited conditions in the energy sector.

II. METHOD

In this Research, the authors propose a deep learning-based classification framework to assess the impact of news sentiment on crude oil price movements. To transform words into numeric vector representations, BERT (Bidirectional Encoder Representations from Transformers) contextual embeddings are used, which are then fed into LSTM, BiLSTM, and GRU models [12]. This approach aims to identify patterns and relationships between news and changes in oil prices more effectively. This Research utilises a news dataset on energy markets, particularly those involving crude oil [13]. Preprocessing was performed to clean and structure the text, converting it to a numeric format ready for in-depth analysis [14, 15]. After preprocessing, each news item was categorised as positive or negative using a transformer-based labelling approach.

Next, models were developed using the labelled data to compare the performance of each algorithm using evaluation metrics such as the confusion matrix, accuracy, and validation loss to assess generalisation capabilities [16]. The goal of this Research is to develop an optimal model identified as GRU based on validation performance that accurately measures the impact of news sentiment on fluctuations in crude oil prices. The Research stages are illustrated in Fig. 1.

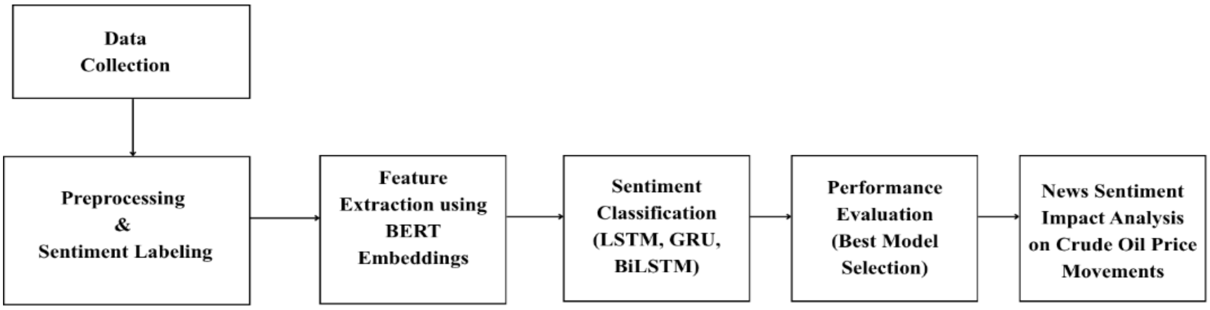

Fig. 1 Systematic research stages

A. Data Collection

The primary data used in this study were obtained from various credible sources, providing a comprehensive dataset for the period of January to May 2025:

- https://www.investing.com Used as the primary source for obtaining historical data on global crude oil prices (WTI and Brent), including daily closing prices and related market news.
- https://finance.yahoo.com Served as a supporting source for validating daily crude oil price data in numeric format (CSV).
- https://newsapi.org & Kaggle: Used to collect the latest news headlines related to the oil market, OPEC policies, and global energy dynamics.

This dataset was developed to support the classification process of news influence on crude oil price movements. The data consists of five primary attributes and one class attribute, which describes the sentiment of each news item. Details of the data structure are shown in Table I.

B. Preprocessing & Sentiment Labeling

Preprocessing and sentiment labeling are crucial stages in the data classification process. Their primary goal is to clean the data of noise and assign sentiment labels (positive or negative), enabling more accurate classification. The preprocessing stages in this study include the following steps [15]:

1) Case Folding: Converting all text to lowercase to eliminate any discrepancies in meaning caused by capitalization [17].

2) Tokenizing: Breaking sentences into words (tokens) and removing punctuation to facilitate data processing [18].

3) Stopword Removal: Removing common words that do not significantly impact the meaning of the text [19].

4) Lemmatization: Converting words to their base form (lemma) using WordNetLemmatizer. This was chosen over stemming to maintain the semantic integrity of tokens required for BERT-based contextual models. [20].

5) Transformer (Labeling): For the labeling process, we used a transformer model from Hugging Face, DistilBERT, which has been specifically trained for sentiment analysis. We utilized a fine-tuned DistilBERT model from https://huggingface.co/distilbert/distilbert-base-uncased-finetuned-sst-2-english, which was trained explicitly for general sentiment tasks. This model is capable of recognizing context and emotions in text with a high level of accuracy, making it suitable for providing sentiment labels to the news data used in this study. Because the news data we used was sourced from global English-language news, the English-based DistilBERT pretrained model has strong initial capabilities in understanding the semantic context of the short texts. We acknowledge that this model was trained on general sentiment data (not specific to the financial domain), so its interpretation of the commodity market context may be limited. Nevertheless, this model was chosen due to its ability to understand the semantic context of short texts and the availability of pretrained models [21].

TABLE I
DATA ATTRIBUTES TO CRUDE OIL PRICE MOVEMENTS

No	Attribute	Description
1	Date	Date of oil price news and data release
2	Close Price (USD)	Closing price of crude oil in United States dollars (USD)
3	Headline	Main title of an article or news related to crude oil
4	Sentiment	Sentiment value generated from text analysis (positive = 1, negative/neutral = 0)
5	Label	Class labels as classification targets, following the results of preprocessing and sentiment clustering

To mitigate the risk of mislabeling, a small sample (approximately 20%) of the automatically labelled data was manually examined (human annotation) to validate the correlation of sentiment with price movements. While manual validation was not possible for the entire data set due to resource constraints, this process provided initial confidence. Future work should include more comprehensive human validation or the use of models explicitly trained on financial market sentiment data to ensure sentiment labels accurately reflect commodity market sentiment.

C. *Feature Extraction using BERT Embeddings*

After preprocessing stages such as case folding, tokenization, stopword removal, lemmatization, and labeling with transformers, a clean set of words is obtained, ready to be represented as vectors. In this study, word representations were computed using BERT (Bidirectional Encoder Representations from Transformers) [22]. BERT is a pre-trained transformer-based model that can understand the meaning of a word in a bidirectional manner, namely by considering the words before and after it. This makes BERT superior to traditional word representation methods, such as GloVe [20], as it can produce more nuanced and contextual word meanings [23].

In this study, word representations were generated using BERT. BERT is superior to traditional methods like GloVe because it produces nuanced, bidirectional contextual word meanings. BERT maps each headline into a numeric vector with 768 dimensions. These representations capture the deeper meaning of news content, thereby improving classification performance. BERT maps each word in a document into a numeric vector. If a word is not present in the pre-trained model, it is replaced with a specific token. The resulting representation is then used to construct a representation of the entire document. Some standard methods used to build document representations from BERT output include [24]:

1) *Pooling:* Taking the average or maximum value of all word vectors in the document.

2) *Concatenation:* Combining vectors from all words in the sentence sequence to form a complete document vector.

With BERT-based representation, classification models can more accurately recognize meaning and sentiment in documents.

D. *Sentiment Classification (LSTM, GRU, and BiLSTM)*

After the data has been preprocessed and word-weighted, the next step is classification using a deep learning model. This study employs three classification algorithms: Gated Recurrent Unit (GRU), Bidirectional Long Short-Term Memory (BiLSTM), and Long Short-Term Memory (LSTM) to analyze news sentiment related to crude oil price movements [25, 26]. At this stage, each word in the processed news text is converted into a vector representation using the BERT (Bidirectional Encoder Representations from Transformers) model. BERT enables more contextual word weighting by considering the semantic and syntactic relationships between words within a sentence.

The resulting word representations from BERT are then used as input for the three predetermined classification The resulting word representations from BERT are then used as input for the three predetermined classification algorithms (GRU, BiLSTM, and LSTM). Although all models achieved identical accuracy, the best-performing model is determined based on validation loss. While all models achieved identical accuracy, the best-performing model was selected based on validation loss to ensure superior generalization. The performance of each model is evaluated using metrics such as accuracy to determine which model produces the best classification results [27].

E. *Performance Evaluation (Best Model Selection)*

Model evaluation was conducted by comparing the average training loss, training accuracy, validation loss, and validation accuracy values of each classification model used: LSTM, GRU, and BiLSTM [25, 28]. Particular emphasis was placed on validation loss to assess the model's ability to avoid overfitting. This study aims to determine which model is most effective and efficient in classifying news sentiment regarding crude oil price movements.

Training loss measures the model's error rate during training, while training accuracy indicates how accurately the model classifies the training data. Validation loss measures the model's ability to avoid overfitting and maintain performance on new data, while validation accuracy indicates the model's ability to classify the validation data accurately. The ideal model is one with a low loss value (both on the training and validation data) and a high accuracy value. By comparing these four metrics, this study aims to determine which model is most effective and efficient in classifying news sentiment regarding crude oil price movements.

F. *News Sentiment Impact Analysis on Crude Oil Price Movements*

The sentiment labels generated by the selected classification model are compared to crude oil price movements to examine the relationship between news sentiment and market trends. This analysis was carried out through descriptive observation of sentiment patterns and price fluctuations that occurred during the research period.

III. RESULT AND DISCUSSION

This Research focuses on sentiment analysis to assess the impact of news on fluctuations in crude oil prices. The data used was collected from various news sources and global oil price data from January 2025 to May 2025. The Research process involved several stages, starting with the collection of news and oil price data, followed by text data preprocessing, which included case folding, tokenization, stopword removal, lemmatization, and labeling using transformers. These stages aimed to clean the data from noise and structure the text into a more structured form.

Next, word weighting was performed using the BERT (Bidirectional Encoder Representations from Transformers) model. BERT is used to transform words in news reports into contextual vector representations, enabling them to capture the meaning and relationships between words in sentences more deeply. These representations are then used as input for the classification process. Classification experiments were conducted using three deep learning algorithms: the Gated Recurrent Unit (GRU), the Bidirectional Long Short-Term Memory (BiLSTM), and the Long Short-Term Memory (LSTM). Each model was trained to classify news sentiment as positive or negative regarding crude oil price movements. Model evaluation was performed using a confusion matrix and evaluation metrics, including accuracy, precision, recall, and F1-score, to determine which model was most effective in classifying news sentiment.

A. Dataset

This study uses a dataset derived from news stories considered to Influence crude oil price movements, consisting of 108 records from January to May 2025. While the dataset consisting of 108 news headlines is relatively limited for deep learning applications, this study employs several strategies to mitigate the risk of overfitting. First, the use of 10-fold cross-validation ensures that every data point is used for both training and validation, providing a more robust estimate of model performance. Second, the choice of GRU and DistilBERT (a lightweight transformer) is strategic; these models have fewer parameters than full-scale transformers, making them less prone to overfitting on small-scale datasets. The data labeling process was automated using Python and the Hugging Face library, enabling sentiment classification of each news item into positive or negative categories. Although the initial dataset had five attributes, this study used only three primary attributes, as they were most relevant to the analysis objectives: number, date, close price (in USD), and headline. The following metadata used in this study are included in Table II.

B. Implementation of Data Preprocessing and Labeling

In this process, the focus of data transformation is on the Headline attribute, as it is analyzed to determine the news sentiment regarding crude oil price movements. To expedite and simplify the data classification process, this study used 108 news articles on crude oil prices. The dataset includes a collection of news headlines from January to May 2025 that have been sentiment-annotated, resulting in thousands of words analyzed. The results of each preprocessing stage and the final sentiment labels for representative news headlines are presented in Table III. Before the data was entered into the classification model, several preprocessing and labeling stages were carried out, namely:

1) Casefolding: This is an initial preprocessing step that standardizes text formatting by converting all capital letters to lowercase, so that spelling differences such as "Oil" and "oil" are no longer recognized as different terms by the system. For example, the text "Oil Prices Rise Amid Geopolitical Tensions" is changed to "oil prices rise amid geopolitical tensions." This change enables further text processing, including tokenization and sentiment analysis.

TABLE II
CRUDE OIL PRICE NEWS SENTIMENT METADATA

No	Date	Close Price (USD)	Headline
1	2025-01-01	64.75	Oil markets await OPEC decision
2	2025-01-02	60.93	OPEC fails to reach output agreement
3	2025-01-03	74.42	Global demand shows strong recovery
4	2025-01-06	66.55	Refineries increase runs, supporting oil markets
…	…	…	…
108	2025-05-30	72.6	Global demand shows strong recovery

2) Tokenizing: In this stage, the sentences in each headline are broken down into word parts (tokens) using commas as word separators. While casefolded text remains a full sentence in lowercase form, tokenizing separates each headline into individual word sets. This process aims to simplify subsequent stages, such as stopword removal and word frequency analysis. For example, the sentence "oil prices rise amid geopolitical tensions" becomes: ["oil", "prices", "rise", "amid", "geopolitical", "tensions"] after tokenization. By dividing sentences into tokens in this manner, the system can more easily identify essential words and remove irrelevant ones in subsequent stages.

3) Stopword: Common words such as "in," "is," "to," "the," "and," "on," "by," and other auxiliary words that do not have significant informational value are removed from the tokenization results. These words are considered stopwords because they do not contribute significantly to the meaning of the sentence or sentiment analysis. The purpose of this process is to simplify the text and filter only essential and relevant words to be used in the classification stage, enabling the model to focus more on identifying patterns that accurately represent sentiment in news articles regarding crude oil price movements.

4) Lemmatization: This stage was conducted on the collection of news headlines using the WordNetLemmatizer library, transforming affixed words to their lemma (valid base form) while preserving their lexical integrity. For example, words like "markets" become "market," "drops" become "drop," "prices" become "price," "shows" become "show," "supporting" become "support," and "threatens" become "threaten." This process aims to reduce word variation while ensuring the resulting words are semantically valid, which is crucial for maintaining the quality of contextual token representation used by BERT. The lemmatization results are used as input to the word weighting and classification process, which uses GRU, BiLSTM, and LSTM models.

5) Transformers: The result of the transformer process in this study is a sentiment classification into two categories: positive and negative. For automatically labeling crude oil news headlines, a transformer model based on the pretrained language model DistilBERT from Hugging Face, trained explicitly for sentiment analysis tasks, was used. This process automatically categorizes each headline based on its context.

Preprocessing and labeling successfully removed noise, simplified news headlines to a more standardized format, and assigned positive and negative sentiment labels to each news item. Thus, the data, previously a collection of unstructured sentences, has been converted into a cleaner, more consistent format, making it better prepared for analysis. This stage enables a more efficient and accurate classification process, facilitating the identification of sentiment patterns in news reports regarding crude oil price movements from January to May 2025.

TABLE III
TEXT PREPROCESSING PIPELINE EXAMPLE

No	Original Headline	Casefolding	Tokenizing	Stopword Removal	Lemmatization	Label
1	Oil markets await OPEC decision	oil markets await opec decision	[oil, markets, await, opec, decision]	oil markets await opec decision	oil market await opec decision	negative
2	OPEC fails to reach output agreement	opec fails to reach output agreement	[opec, fails, to, reach, output, agreement]	opec fails reach output agreement	opec fail to reach output agreement	negative
3	Global demand shows strong recovery	global demand shows strong recovery	[global, demand, shows, strong, recovery]	global demand shows strong recovery	global demand show strong recovery	positive
4	Refineries increase runs, supporting oil markets	refineries increase runs, supporting oil markets	[refineries, increase, runs, supporting, oil, markets]	refineries increase runs supporting oil markets	refinery increase run support oil market	positive
…	…	…	…	…	…	…
108	Global demand shows strong recovery	global demand shows strong recovery	[global, demand, shows, strong, recovery]	global demand shows strong recovery	global demand show strong recovery	negative

C. Word Weighting using BERT

After preprocessing 108 headlines, each term was transformed into a vector representation using the BERT method After preprocessing 108 headlines, each term was transformed into a vector representation using the BERT method with dimensions (86, 768). This method produces contextual word representations based on the deep learning model's understanding of semantic and syntactic relationships within news texts. Each document (headline) was represented as a vector of dimension (86, 768), obtained from the final layer of the BERT model, accounting for the full context of each word within the sentence.

The results of this transformation process are presented in Table IV, which displays the vector representation of each headline. This representation aims to capture the deeper meaning of news content, thereby improving sentiment classification performance regarding the impact of news on crude oil price movements.

D. Classification using BERT Weighting and LSTM, GRU, and BiLSTM Algorithms

At this stage, the classification process is carried out using vector representations derived from BERT (Bidirectional Encoder Representations from Transformers) feature extraction. Each preprocessed document is converted into a 768-dimensional vector using Huggingface's pretrained model distilbert-base-uncased-finetuned-sst-2-english, which captures both semantic and syntactic context from the news headlines. These vector representations are then used as input for three Recurrent Neural Network (RNN)-based classification models: Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional Long Short-Term Memory (BiLSTM). These algorithms were selected for their ability to learn sequential patterns and capture contextual relationships in text data.

The LSTM model achieves an average validation accuracy of 85.27%, with values ranging from 72.72% to 90.90% across folds. However, its validation loss (0.3702) is higher than that of the GRU and BiLSTM models, indicating suboptimal generalization despite achieving competitive accuracy. Detailed results for the LSTM model are presented in Table V.

TABLE IV
BERT WEIGHTING REPRESENTATIONS

Id/ no	Input_ids	Attention_mask	Predicted_label
1	tensor([[101, 3514, 3006, 26751, 6728, 858	tensor([[1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0,	1
2	tensor([[101, 6728, 8586, 8246, 2000, 3362, 6	tensor([[1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0,	1
3	tensor([[101, 3795, 5157, 2265, 2844, 7233, 1	tensor([[1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,	0
4	tensor([[101, 21034, 3623, 2448, 2490, 351	tensor([[1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,	1
...	....	...	
108	tensor([[101, 3795, 5157, 2265, 2844, 7233, 1	tensor([[1, 1, 1, 1, 1, 1, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0,	0

TABLE V
LSTM MODEL PERFORMANCE DETAILS

Fold	Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
Fold 1	LSTM	0.704808444	0.824742268	0.315160245	0.909090909
Fold 2	LSTM	0.523159258	0.618556701	0.521685541	0.727272727
Fold 3	LSTM	0.430202957	0.865979381	0.517733693	0.727272727
Fold 4	LSTM	0.528250322	0.577319588	0.309604049	0.909090909
Fold 5	LSTM	0.444586307	0.865979381	0.527641237	0.727272727
Fold 6	LSTM	0.727520242	0.659793814	0.324770987	0.909090909
Fold 7	LSTM	0.682730444	0.907216495	0.322162241	0.909090909
Fold 8	LSTM	0.472122282	0.845360825	0.29325828	0.909090909
Fold 9	LSTM	0.450616661	0.846938776	0.288228095	0.9
Fold 10	LSTM	0.455008842	0.846938776	0.281991959	0.9

Furthermore, the BiLSTM model was also used to capture sequential context from both directions (forward and backward). This model achieved an average validation accuracy of 85.27%, with the highest accuracy of 90.90% for Folds 1, 4, 6, 7, and 8, and the lowest accuracy of 72.72% for Folds 2, 3, and 5. The average train loss of 0.5434 and the validation loss of 0.3327 indicate that this model provides consistent, balanced classification results. Details of the BiLSTM model performance are presented in Table VI.

Meanwhile, the GRU model, a lighter, more efficient RNN variant, also demonstrated highly competitive classification performance. This model recorded an average validation accuracy of 85.27%, with a consistent peak accuracy of 90.90% on folds 1, 4, 6, 7, and 8. The average validation loss was 0.3324, and the training loss was 0.4430. Details of the GRU model's performance are presented in Table VII.

E. Comparative Evaluation of the Performance of the Three Classification Methods

Based on the classification experiment using three models, namely LSTM, BiLSTM, and GRU, all models showed the same average validation accuracy, which was 85.27%. However, the difference in performance is seen in the validation loss value of each model. The GRU model achieved the lowest validation loss of 0.3324, followed by BiLSTM (0.3327), while LSTM had the highest validation loss of 0.3702. Lower validation loss values indicate better model generalization capabilities to data that has never been seen before. The evaluation results showed that LSTM, BiLSTM, and GRU achieved an identical average validation accuracy of 85.27%. This phenomenon is primarily attributed to the limited dataset size (108 records), which may constrain the models' ability to differentiate their classification boundaries further. However, the differentiation becomes evident when analyzing the validation loss, where the GRU model achieved the lowest value of 0.3324. From a structural perspective, GRU generalizes better under small-data conditions because it has a simpler architecture with fewer parameters (two gates: reset and update) than LSTM (three gates: input, forget, and output). This reduced complexity effectively lowers the risk of overfitting on niche datasets while maintaining high computational efficiency.

In addition, GRU has the smallest train loss (0.4430), indicating a more efficient learning process than LSTM and BiLSTM. Meanwhile, the LSTM showed considerable fluctuations in validation accuracy across folds, ranging from 72.72% to 90.90%, making its performance less stable than the other two models. A comparison of the three models' performance is shown in Table VIII.

Since all three models achieved the same average validation accuracy, the best model should be selected based on validation loss. Among the three models, GRU achieved the lowest validation loss (0.3324), indicating the best generalization. Therefore, GRU is selected as the optimal model for this study to analyze news sentiment related to crude oil price movements.

F. The Influence of News Sentiment on Oil Price Movements

Table IX illustrates an example of a news ID that represents sentiment toward crude oil price movements, based on the positive and negative tendencies evident in the headlines. Fig. 2, illustrates the movement of crude oil prices from January to June 2025.

TABLE VI
BILSTM MODEL PERFORMANCE DETAILS

Fold	Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
Fold 1	BiLSTM	0.473789833	0.597938144	0.241936579	0.909090909
Fold 2	BiLSTM	0.424532443	0.865979381	0.540643811	0.727272727
Fold 3	BiLSTM	0.735482432	0.649484536	0.458764046	0.727272727
Fold 4	BiLSTM	0.444342684	0.75257732	0.234819517	0.909090909
Fold 5	BiLSTM	0.444201898	0.618556701	0.539542139	0.727272727
Fold 6	BiLSTM	0.444679517	0.762886598	0.255387992	0.909090909
Fold 7	BiLSTM	0.749955855	0.845360825	0.267784268	0.909090909
Fold 8	BiLSTM	0.772213668	0.824742268	0.253119469	0.909090909
Fold 9	BiLSTM	0.46092074	0.714285714	0.268002093	0.9
Fold 10	BiLSTM	0.483994797	0.591836735	0.267083794	0.9

TABLE VII
GRU MODEL PERFORMANCE DETAILS

Fold	Model	Train Loss	Train Accuracy	Validation Loss	Validation Accuracy
Fold 1	GRU	0.50764896	0.659793814	0.222628921	0.909090909
Fold 2	GRU	0.478482973	0.680412371	0.579877317	0.727272727
Fold 3	GRU	0.406860895	0.783505155	0.548561037	0.727272727
Fold 4	GRU	0.399253206	0.731958763	0.237587497	0.909090909
Fold 5	GRU	0.370082065	0.87628866	0.52525574	0.727272727
Fold 6	GRU	0.437579177	0.845360825	0.245883793	0.909090909
Fold 7	GRU	0.435537558	0.659793814	0.248246014	0.909090909
Fold 8	GRU	0.428713936	0.639175258	0.226381391	0.909090909
Fold 9	GRU	0.381517645	0.846938776	0.245680422	0.9
Fold 10	GRU	0.584780879	0.806122449	0.243907899	0.9

TABLE VIII
CLASSIFICATION MODEL PERFORMANCE COMPARISON

Model	Average Train Loss	Average Train Accuracy	Average Val Loss	Average Val Accuracy
LSTM	0.5419	0.7858	0.3702	0.8527
GRU	0.4430	0.7529	0.3324	0.8527
BiLSTM	0.5434	0.7223	0.3327	0.8527

TABLE IX
THE FOLLOWING ID EXAMPLE INDICATES NEWS SENTIMENT TOWARDS CRUDE OIL PRICE MOVEMENTS

ID/No	Date	Close Price (USD)	Headline	Predicted Label
23	2025-01-31	60.14	U.S. inventory drops, oil prices jump	Positive
22	2025-01-30	72.68	OPEC cuts production boosting prices	Positive
36	2025-02-19	60.44	Weak global demand pressures crude prices	Positive
30	2025-02-11	71.71	U.S. inventory drops, oil prices jump	Positive
49	2025-03-10	60.11	OPEC cuts production boosting prices	Positive
50	2025-03-11	70.31	Analysts expect little change in oil prices	Negative
72	2025-04-10	60.04	Oil markets await OPEC decision	Positive
73	2025-04-11	73.52	Global demand shows strong recovery	Negative
87	2025-05-01	74.98	Weak global demand pressures crude prices	Positive
97	2025-05-15	72.82	U.S. inventory drops, oil prices jump	Positive

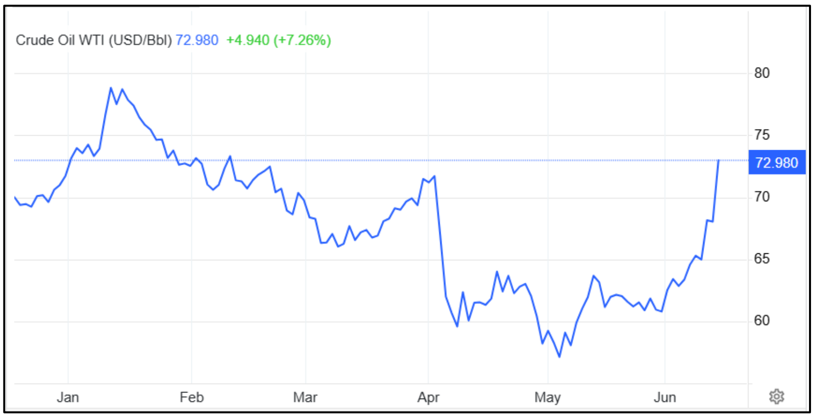

Fig. 2 Graph of crude oil price movements for the period January to June 2025
Source: https://id.tradingeconomics.com/commodity/crude-oil

The classification results are shown in Table IX The table contains five columns: ID/No., news date, crude oil closing price (USD), news headline, and sentiment classification result (positive or negative). For example, the figure shows that news with ID 73, published on 11 April 2025, is classified as having negative sentiment. Referring to Fig. 2, which displays a graph of crude oil price movements, a significant price decline was observed in April, particularly from the beginning to the middle of the month. This indicates a relationship between the emergence of negative news and the downward trend in crude oil prices. The relationship between sentiment and price movements is further validated by observing the alignment between negative sentiment clusters and downward price trends. For instance, a significant price decline in mid-April 2025 directly correlates with a predominance of negative labels generated by the model. Although the current study relies on descriptive correlation through visual mapping of price graphs and sentiment labels, the consistent validation accuracy of 85.27% across all three models suggests meaningful pattern recognition. Future Research should incorporate formal statistical tests, such as the Granger causality test or Pearson correlation coefficients, to quantify the strength of this relationship.

This information can serve as a reference for observing market dynamics at any given time and as a basis for analysis for investors or policymakers to monitor crude oil price movements based on circulating news sentiment. The price movement graph in Fig. 2 shows quite dynamic price fluctuations from January to May 2025. The periodic price increases and decreases are reflected in the graph's clearly visible up-and-down pattern.

Sentiment analysis of news headlines reveals that market perceptions of circulating information significantly Influence crude oil price movements. There is a correlation between sentiment dynamics and price changes, with prices tending to rise after positive sentiment appears and decline when negative sentiment dominates. When positive sentiment prevails, crude oil prices tend to increase, offering opportunities for investors. Conversely, a predominance of negative sentiment can trigger price declines, raising concerns in the energy market.

By understanding this pattern and continuously monitoring news sentiment, investors and market analysts can make more informed decisions in response to oil market dynamics. This highlights the importance of incorporating sentiment analysis into a decision-making strategy in the highly volatile world of energy commodity trading and investment.

In comparison with prior Research, our finding that media sentiment signals provide significant evidence of oil price volatility [2]. Similarly, the observation that incorporating sentiment scores improves forecasting accuracy supports the findings in [5] regarding oil futures. While previous large-scale studies often emphasize the superiority of complex hybrid models, this study demonstrates that even with a constrained dataset, a BERT-enhanced GRU model can achieve a competitive accuracy of 85.27%. This highlights the effectiveness of contextual embedding (BERT) for extracting high-quality features from short news headlines, even when historical records are sparse.

This Research also faced several challenges, particularly in processing English text. One major obstacle was the use of informal language and slang in news reports, which the classification model did not always readily understand. Furthermore, despite the availability of a large English corpus, variations in writing style across news sources complicate classification. As a result, the classification accuracy is not yet optimal and can still be improved in subsequent model development stages.

IV. CONCLUSION

This study demonstrates that sentiment in news headlines is significantly related to crude oil price movements. In the sentiment classification process, all three models (LSTM, BiLSTM, and GRU) produced an identical average validation accuracy of 85.27%. However, based on the validation loss metric for generalization performance, the GRU model is identified as the optimal architecture, with a validation loss of 0.3324. These results indicate that, despite similar accuracy, the GRU exhibits better generalization on relatively small datasets. Consequently, GRU is highly recommended for news-based sentiment analysis in the crude oil market. This study has major constraints, including a very limited dataset (108 records) and the use of a general-domain sentiment model for automated labelling, which may not fully capture the nuances of financial and energy commodity markets. Subsequent studies should utilize larger datasets over longer periods, incorporate domain-specific financial sentiment models, and apply modern transformer architectures to obtain more representative and stable results.

REFERENCES

[1] M. Monge, M. F. Romero Rojo, and L. A. Gil-Alana, "The impact of geopolitical risk on the behavior of oil prices and freight rates," *Energy*, vol. 269, p. 126779, 2023, doi: 10.1016/j.energy.2023.126779.

[2] H. Abdollahi, "Oil price volatility and new evidence from news and Twitter," *Energy Econ.*, vol. 122, p. 106711, Jun. 2023, doi: 10.1016/j.eneco.2023.106711.

[3] S. K. Bharti, P. Tratiya, and R. K. Gupta, "Stock Market Price Prediction through News Sentiment Analysis & Ensemble Learning," in *2022 IEEE 2nd International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, 2022, pp. 1–5. doi: 10.1109/iSSSC56467.2022.10051623.

[4] J. Maqbool, P. Aggarwal, R. Kaur, A. Mittal, and I. A. Ganaie, "Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach," *Procedia Comput. Sci.*, vol. 218, pp. 1067–1078, 2023, doi: 10.1016/j.procs.2023.01.086.

[5] Y. Li, S. Jiang, X. Li, and S. Wang, "The role of news sentiment in oil futures returns and volatility forecasting: Data-decomposition based deep learning approach," *Energy Econ.*, vol. 95, p. 105140, 2021, doi: 10.1016/j.eneco.2021.105140.

[6] M. F. Saleh and R. Imanda, "Public Sentiment Analysis of the Free Meal Program: A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform," *J. Appl. Informatics Comput.*, vol. 9, no. 1, pp. 131–139, 2025, doi: 10.30871/jaic.v9i1.8895.

[7] A. W. Sari, T. I. Hermanto, and M. Defriani, "Sentiment Analysis Of Tourist Reviews Using K-Nearest Neighbors Algorithm And Support Vector Machine," *Sinkron*, vol. 8, no. 3, pp. 1366–1378, 2023, doi: 10.33395/sinkron.v8i3.12447.

[8] S. J. Hegde, H. M. Madhunandana, and Mohana, "Sentiment Analysis with LSTM Recurrent Neural Network Approach for Movie Reviews using Deep Learning," in *2023 3rd International Conference on Innovative Mechanisms for Industry Applications (ICIMIA)*, Dec. 2023, pp. 463–468. doi: 10.1109/ICIMIA60377.2023.10426266.

[9] I. D. Mienye, T. G. Swart, and G. Obaido, "Recurrent Neural Networks: A Comprehensive Review of Architectures, Variants, and Applications," *Information*, vol. 15, no. 9, p. 517, Aug. 2024, doi: 10.3390/info15090517.

[10] D. Nuryadi, D. H. F. Manongga, and I. Sembiring, "Crypto Narratives Sentiment Analysis on Bitcoin Price Prediction Using the Naive Bayes Method," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 10, no. 2, pp. 1087–1098, Mar. 2025, doi: 10.29100/jipi.v10i2.6116.

[11] C. Haas, C. Budin, and A. D'Arcy, "How to select oil price prediction models — The effect of statistical and financial performance metrics and sentiment scores," *Energy Econ.*, vol. 133, p. 107466, 2024, doi: 10.1016/j.eneco.2024.107466.

[12] G. Zhao, G. Liu, L. Fang, B. Tu, and P. Ghamisi, "Multiple convolutional layers fusion framework for hyperspectral image classification," *Neurocomputing*, vol. 339, pp. 149–160, Apr. 2019, doi: 10.1016/j.neucom.2019.02.019.

[13] W. K. Sari, D. P. Rini, and R. F. Malik, "Text Classification Using Long Short-Term Memory With GloVe Features," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 5, no. 2, p. 85, Feb. 2020, doi: 10.26555/jiteki.v5i2.15021.

[14] A. Muhariya, I. Riadi, and Y. Prayudi, "Cyberbullying Analysis on Instagram Using K-Means Clustering," *JUITA J. Inform.*, vol. 10, no. 2, p. 261, 2022, doi: 10.30595/juita.v10i2.14490.

[15] A. Muhariya, I. Riadi, Y. Prayudi, and I. A. Saputro, "Utilizing K-Means Clustering for the Detection of Cyberbullying Within Instagram Comments," *Ing. des Syst. d'Information*, vol. 28, no. 4, pp. 939–949, 2023, doi: 10.18280/isi.280414.

[16] I. A. Saputro, I. A. Prabowo, and R. A. Aziz, "Web-Based Application Development for Mapping Zakat Recipients Using the K-Means Clustering Method," *JEKIN - J. Tek. Inform.*, vol. 4, no. 2, pp. 254–261, May 2024, doi: 10.58794/jekin.v4i2.721.

[17] I. A. Saputro, L. Sugiarto, and F. S. Nugraha, "Analysis of Public Awareness of the Dangers of Internet Phishing Using K-Means Clustering," *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 9, no. 2, 2024, doi: 10.30998/string.v9i2.22404.

[18] T. Agustin, I. A. Saputro, M. L. Rahmadi, F. Patria, A. P. Kartiningtyas, and D. Kurniawan, "Effectiveness of Random Search in Enhancing CNN Performance for Rice Plant Disease Classification," in *2024 6th International Conference on Cybernetics and Intelligent System (ICORIS)*, Nov. 2024, pp. 1–6. doi: 10.1109/ICORIS63540.2024.10903973.

[19] S. K. Sahu, S. Sarangi, and S. K. Jena, "A detail analysis on intrusion detection datasets," *Souvenir 2014 IEEE Int. Adv. Comput. Conf. IACC 2014*, pp. 1348–1353, 2014, doi: 10.1109/IAdCC.2014.6779523.

[20] C. Wang, P. Nulty, and D. Lillis, "A Comparative Study on Word Embeddings in Deep Learning for Text Classification," in *Proceedings of the 4th International Conference on Natural Language Processing and Information Retrieval*, Dec. 2020, pp. 37–46. doi: 10.1145/3443279.3443304.

[21] Y. Wu, Z. Jin, C. Shi, P. Liang, and T. Zhan, "Research on the application of deep learning-based BERT model in sentiment analysis," *Appl. Comput. Eng.*, vol. 71, no. 1, pp. 14–20, May 2024, doi: 10.54254/2755-2721/71/2024MA.

[22] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in *Proceedings of the 2019 Conference of the North*, Jun. 2019, pp. 4171–4186. doi: 10.18653/v1/N19-1423.

[23] R. T. Dewangga and B. Prasetiyo, "Comparative

Analysis of BERT Model and DistilBERT Model for Enhanced Clickbait Headline Structure Detection in Indonesian Online News," *JUITA J. Inform.*, vol. 13, no. 3, pp. 235–244, Nov. 2025, doi: 10.30595/juita.v13i3.26479.

[24] L. B. Cesar, M.-Á. Manso-Callejo, and C.-I. Cira, "BERT (Bidirectional Encoder Representations from Transformers) for Missing Data Imputation in Solar Irradiance Time Series," in *ITISE 2023*, Jun. 2023, vol. 39, no. 1, p. 26. doi: 10.3390/engproc2023039026.

[25] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Sentiment-driven cryptocurrency forecasting: analyzing LSTM, GRU, Bi-LSTM, and temporal attention model (TAM)," *Soc. Netw. Anal. Min.*, vol. 15, no. 1, p. 52, 2025, doi: 10.1007/s13278-025-01463-6.

[26] P. L. Seabe, C. R. B. Moutsinga, and E. Pindza, "Forecasting Cryptocurrency Prices Using LSTM, GRU, and Bi-Directional LSTM: A Deep Learning Approach," *Fractal Fract.*, vol. 7, no. 2, p. 203, Feb. 2023, doi: 10.3390/fractalfract7020203.

[27] M. J. Hamayel and A. Y. Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," *AI*, vol. 2, no. 4, pp. 477–496, Oct. 2021, doi: 10.3390/ai2040030.

[28] A. Nanyonga, H. Wasswa, and G. Wild, "Phase of Flight Classification in Aviation Safety Using LSTM, GRU, and BiLSTM: A Case Study with ASN Dataset," in *2023 International Conference on High Performance Big Data and Intelligent Systems (HDIS)*, Dec. 2023, no. December, pp. 24–28. doi: 10.1109/HDIS60872.2023.10499521.