

CHAPTER 1

INTRODUCTION

1.1 Rationale

Stock market fluctuations are influenced not only by numerical data but also by investor perception [1] and sentiment shaped by information sources [2]. These movements are further driven by irrational behavior [3], investor emotions [4], and sentiment-based reactions [5], which reflect key aspects of economic behavior. The large volume of information makes it challenging for investors to quickly identify market sentiment, often leading to delayed or irrational decisions [1]. It also complicates the identification of credible sources, affecting sentiment accuracy. Although technology has improved data access, decision-making still depends on human judgment [6], leaving room for sentiment-based forecasting approaches.

Sentiment analysis has emerged as a key technique for extracting qualitative information, making it possible to classify market-related news as positive, negative, or neutral signals that may correlate with stock movements. Sentiment-based stock analysis has been widely explored in countries like the US [5, 7–10] and China [3, 7, 11], showing that public perception and media-driven sentiment can significantly affect market behavior. National developments, such as political events, policy announcements, corporate mergers or splits, and global occurrences, often have a strong impact on market behavior [12]. Similarly, Indonesia is exposed to a range of high-impact national issues, such as the launch of its sovereign wealth fund (Danantara), the development of the new capital city (IKN), and shifting government policies, which receive extensive news coverage and shape investor sentiment. Given these conditions, sentiment analysis holds strong potential in the Indonesian context, as market reactions may also be influenced by the emotional tone and frequency of such information in local media.

While widely explored in English and Chinese languages, its application in Indonesian contexts remains limited due to a lack of linguistic resources and tools [13]. This poses challenges in developing accurate models, particularly for the Indonesian stock market, such as the LQ45 index. The existing studies in the Indonesian stock market have applied various models to forecast LQ45 and JKSE, including ARIMA-LSTM [14], SMO and Random Forest [15], and LSTM on financial and infrastructure sectors [16–18]. Recent works also integrated sentiment features using sources like GDELT, Kompas.com, or financial headlines [19–21], often combining them with deep learning models. However, studies like [22] rely on translated Indonesian news labeled with English sentiment classifiers (e.g., SIEBERT) to fine-tune their aspect sentiment classification IndoBERT model, which risks losing context and its meaning, highlighting the need for native Indonesian sentiment models like IndoBERT. Furthermore, some of these works simplify sentiment classification into

only two labels (positive and negative), which fails to capture the full nuance of real-world sentiment, and none of them apply sentiment probability as a predictive input.

As sentiment gains importance in forecasting approaches, stock market prediction aims to anticipate future price movements using both historical patterns and behavioral signals. Traditional models like Linear Regression and ARIMA rely on numerical indicators but often fall short in capturing the complexity of market dynamics [23]. Deep learning models such as Long Short-Term Memory (LSTM) address this limitation by learning temporal dependencies and nonlinear relationships [24]. IndoBERT [25], based on BERT architecture [26], is a pre-trained model for Indonesian, offering high-precision sentiment classification on IndoLEM. IndoBERT addresses these gaps by providing both hard sentiment labels and class probabilities via softmax output. Its effectiveness in extracting sentiment from financial news makes it ideal for forecasting stock trends. When enhanced with sentiment features extracted using IndoBERT, LSTM models present a promising direction for improving the accuracy of stock index prediction in the Indonesian context.

Based on these gaps, this study explores how integrating sentiment features with historical LQ45 stock index data improves forecasting performance, particularly when using sentiment probabilities, either directly or in aggregated form, rather than features derived from hard labels. To evaluate this, an LSTM model is implemented to forecast stock index close values by combining historical index data with sentiment-driven features. These features are extracted using IndoBERT, which analyzes Indonesian financial news headlines to generate sentiment probabilities and class labels. This study also explores raw sentiment probabilities as direct input features. While IndoBERT is used as the sentiment classification tool, the study primarily focuses on evaluating its outputs within the LSTM-based forecasting framework, rather than assessing the classification model's performance itself.

This study aims to investigate whether the integration of sentiment analysis from Indonesian financial news can improve the performance of stock index forecasting models. Traditional models often rely solely on historical numerical indicators, which may overlook the impact of market sentiment and investor perception. To address this limitation, the study proposes integrating sentiment probabilities, extracted using an IndoBERT-based classifier, into an LSTM forecasting model. The research question is whether sentiment features can enhance prediction accuracy compared to baseline models using only historical price data. This aim is addressed through a series of structured experiments and scenario-based evaluations. The model's performance is evaluated using standard error metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Absolute Percentage Error (MAPE), to determine whether sentiment-enhanced models outperform those relying solely on historical data.

The remainder of this paper is structured as follows. Section 2 reviews related work on stock index prediction and sentiment analysis. Section 3 details the proposed methodology,

including sentiment feature extraction and forecasting model design. Section 4 presents the experimental results along with the discussion. Section 5 provides the conclusion, with a summary of key findings and a discussion of limitations and future work.

1.2 Theoretical Framework

This section explains the theoretical framework that underlies the research of building an LQ45 stock index prediction with LSTM, integrated with sentiment features derived from IndoBERT.

1. Behavioral Finance: Suggests that market movements are not solely driven by fundamentals but also by psychological factors, including investor sentiment, emotion, and perception.
2. Sentiment Analysis: Refers to the extraction of subjective information (positive, negative, neutral) from financial text data, which may influence investor behavior and market movements. In this study, the targeted outputs of sentiment analysis are both the sentiment class label and the probability scores for each class, enabling richer feature integration into the forecasting model.
3. Time Series Forecasting with LSTM: LSTM models are designed to capture long-term dependencies in sequential data. In this study, LSTM is used to learn patterns from both historical index movements and sentiment dynamics.
4. Language Representation Models (IndoBERT): A pre-trained transformer-based language model optimized for Bahasa Indonesia, capable of generating sentiment probabilities that serve as behavioral indicators in forecasting.

1.3 Conceptual Framework/Paradigm

This study integrates textual sentiment information with time-series data to enhance stock index forecasting. Financial news titles are obtained individually and processed using IndoBERT to extract sentiment probabilities for positive, negative, and neutral classes. These sentiment features are then combined with historical LQ45 closing values to form the input for an LSTM-based forecasting model. Two sentiment integration strategies are explored: direct use of sentiment probabilities and aggregated sentiment scores derived from either probabilities or hard sentiment labels. The goal is to evaluate whether the integration of sentiment-related features can improve model performance compared to relying solely on historical index values, evaluated with MAE, RMSE, and MAPE. The model is trained and tested through a time series cross-validation approach to ensure robust evaluation across different time periods.

1.4 Statement of the Problem

Investor sentiment, shaped by financial news and public perception, significantly influences stock index movements. In recent years, sentiment analysis has been widely applied in financial forecasting, especially in English and Chinese markets. In the Indonesian context, its potential remains underexplored in financial forecasting, partly due to the lack of high-quality language tools tailored to Bahasa Indonesia. With the development of models like IndoBERT, there is an opportunity to extract sentiment signals from local financial news more precisely and robustly. However, the use of sentiment probabilities remains underexplored in forecasting frameworks. This study addresses this gap by leveraging sentiment probabilities extracted from IndoBERT to enhance stock forecasting. The goal is to develop a more robust and precise LQ45 index prediction model by evaluating how sentiment features improve accuracy, as reflected in lower MAE, RMSE, and MAPE scores compared to baseline models.

1.5 Objective and Hypotheses

The main objective of this study is to examine the effectiveness of integrating sentiment analysis from Indonesian financial news into LQ45 stock index forecasting models. Specifically, the research explores how sentiment probabilities, extracted using the IndoBERT language model, can be implemented to enhance the predictive accuracy of Long Short-Term Memory (LSTM) models for the LQ45 stock index. Rather than relying solely on historical price data, which may overlook qualitative aspects influencing investor behavior, this study aims to integrate behavioral signals captured from financial news sentiment.

To achieve this, the study proposes two sentiment integration approaches: first, using raw sentiment probabilities (positive, negative, neutral) directly as additional features; and second, transforming those probabilities into aggregated sentiment scores, both from probability values and from hard classification labels. These approaches are evaluated through a structured set of experiments to determine which integration method yields the best performance in terms of forecasting error metrics. The underlying hypothesis is that sentiment-enhanced models will produce more robust and precise forecasts compared to traditional models, and that probability-based sentiment features will offer better predictive value than hard label-based approach.

1.6 Assumption

This study assumes that financial news headlines reflect investor sentiment and can influence LQ45 stock movements. IndoBERT is assumed to accurately extract sentiment probabilities from Indonesian-language news. It is also assumed that these probabilities, whether used directly or in aggregated form, can enhance LSTM forecasting when combined

with historical LQ45 index data. This news is considered relevant to market conditions, especially during impactful national or economic events.

1.7 Scope and Delimitation

This study focuses on forecasting the LQ45 stock market index daily close integrated with sentiment features extracted from Indonesian financial news headlines published between January 2024 to July 2025. The sentiment model used is IndoBERT, limited to classifying news into positive, negative, and neutral categories. Only news titles are analyzed, without considering full article content. The forecasting model is based solely on LSTM, and the evaluation emphasizes the impact of sentiment integration rather than sentiment classification accuracy.

1.8 Significance of the Study

This research contributes to the field of financial forecasting by demonstrating the value of integrating sentiment analysis into deep learning models. It shows that sentiment probabilities from Indonesian news sources, extracted using IndoBERT, can enhance LSTM-based prediction of the LQ45 index. The findings are expected to benefit investors, analysts, and researchers by providing a sentiment-aware forecasting framework tailored to the Indonesian market context.