Discover Artificial Intelligence

Research

Deciphering news sentiment and stock price relationships in Indonesian companies: an AI-based exploration of industry affiliation and news co-occurrence

Andry Alamsyah[1] · Dian Puteri Ramadhani[1] · Farida Titik Kristanti[1] · Arbi Haza Nasution[2] · Mohd Sham bin Mohamad[3] · Rajalingam Sokkalingam[4] · Sri Widiyanesti[1] · Muhammad Apriandito Arya Saputra[5]

Received: 8 January 2025 / Accepted: 26 May 2025
Published online: 03 June 2025

Abstract
The rapid increase of textual data has transformed the way we understand and forecast financial market behavior. Investor sentiments, often swayed by news, are pivotal in determining stock prices. Analyzing a dataset of 192.582 Indonesian financial news articles published between 2018 and 2023. This study investigates the complex connections between news sentiment and stock market behavior of Indonesian companies. We leverage AI-based sentiment analysis and natural language processing techniques, including identity recognition, network analysis, and correlation assessment, to explore how news sentiment affects stock prices at the levels of individuals, industries, and news co-occurrence clusters. While earlier research has addressed the effect of sentiment on stock prices at both the company and industry levels, there is a significant lack of studies focused on media co-occurrence clusters, which is vital for comprehending the collective media portrayal of interconnected firms. Our results show that sentiment-price correlations strengthen hierarchically, with individual companies at 0.26, industry groupings at 0.30, and news co-occurrence clusters at 0.43. This research introduces a unique analytical framework that explores sentiment across various levels, highlighting co-occurrence clusters that reflect business relationships beyond traditional industry lines. It demonstrates that companies frequently mentioned together in the news exhibit stronger and more stable sentiment-price correlations, offering a new analytical perspective for AI-driven investment strategies and underscoring the potential of big data analytics in Indonesia's capital market.

Keywords AI-driven sentiment analysis · Stock market dynamics · Entity recognition · News co-occurrence · Indonesian companies

1 Introduction

A crucial aspect of financial institutions is determining the value of a company's stock [1]. The essential task involves assessing the intrinsic worth of a company's shares. The stock price serves as a representation of a company's inherent value, encapsulating the perception of the company's financial performance [2, 3]. Investors and market participants perceive the stock prices as indicators of the company's worth within the public sphere. Changes in stock prices commonly indicate shifts in investor sentiment [4], shifts in existing market conditions, and overarching economic

✉ Andry Alamsyah, andrya@telkomuniversity.ac.id | [1]School of Economics and Business, Telkom University, Bandung, Indonesia. [2]Department of Informatics Engineering, Universitas Islam Riau, Riau, Indonesia. [3]Centre of Mathematical Sciences, University Malaysia Pahang. Al-Sultan Abdullah (UMPSA), Pahang, Malaysia. [4]Fundamental and Applied Science Department, Universiti Teknologi Petronas, Perak, Malaysia. [5]School of Business and Management, Bandung Institute of Technology, Bandung, Indonesia.

situations [5]. Stock prices hold a fundamental role in influencing investment decisions, corporate strategies, and stakeholders' perceptions.

Financial markets are characterized by complex inter-dependencies arising from the ever-changing interactions among market participants [6, 7, 8], financial instruments [9], and their connections to larger dynamic macroeconomic factors [10]. The instances of financial system failures trigger broader economic impacts. The valuation of a company's shares within the stock market is subject to a multitude of economic factors. It exhibits a profound interconnectedness with the prices of other publicly traded securities, financial policies, the national economy trajectory, and the performance of specific industrial sectors [11].

Academics and market participants employ various techniques to cluster stocks exhibiting similar characteristics [12]. A common method involves classifying securities according to their respective industrial sectors. Many academic studies [12] suggest that a stock's industrial categorization accounts for a notable proportion of the correlations observed in its returns, surpassing the explanatory power of the broader market dynamics. A prior study highlighted the presence of positive correlations in the stock prices of companies belonging to the same industry or categorized within specific sectors. A study discovered that industry-related factors play a more significant role in stock return behaviors than other basic stock attributes [13]. The observation emphasizes that companies within a common industry often experience similar price movements, indicating a tendency for their stock values to align. The trend is influenced by broader economic factors, market trends, and industry-specific drivers, which collectively contribute to the convergence of stock price behaviors among companies sharing industry affiliations [14].

Jing et al. [15] asserts a strong correlation between investment and the Efficient Market Hypothesis. This hypothesis suggests that the prevailing stock market prices accurately reflect all available information. The news emerges as a potential catalyst for shaping market sentiments, which in turn can exert influence over stock prices. Curme [16] underscores the profound significance of the link between stock price fluctuations and the information content within the news. However, the inherent qualitative nature of news [17] poses a formidable challenge as it defies from the classical quantitative measurement.

Leveraging advanced data analytic techniques offers a pivotal role in empowering investors to formulate well-informed decisions [18, 19], such as gaining insight into prevailing public sentiment regarding stocks in the capital market. Big data refers to the exponential growth and availability of data, including structured and unstructured data, that surpasses the processing capabilities of traditional data management systems [20]. This collection encompasses a wide range of sources, such as online transactions, social media interactions, sensor measurements, and digital documents [21, 22–24]. The large volume of data offers great potential for gaining insights, making informed decisions, process optimization [25, 26, 27], and driving progress across various fields through the application of advanced analytical methods and artificial intelligence [22].

As artificial intelligence (AI) continues to evolve and permeate various fields, it has proven particularly adept at managing and analyzing large-scale unstructured data. In the financial sector, advanced AI approaches—such as automated text mining, sentiment analysis, and entity recognition—enable researchers and practitioners to uncover new insights beyond what traditional models offer. This study leverages AI-based sentiment analysis of digital news articles, combined with a network-based framework, to better understand how media narratives shape stock market behavior. By examining Indonesian companies through multiple lenses—individual firms, industry affiliations, and co-occurrence clusters—our research contributes to an emerging body of AI-driven financial analytics. Such methods deepen our comprehension of market dynamics and support more informed and transparent investment decisions, ultimately fostering economic growth and resilience. Moreover, this AI-centric perspective dovetails with the United Nations Sustainable Development Goal 9 (Industry, Innovation, and Infrastructure), underscoring the role of innovative AI methodologies in cultivating robust financial infrastructures and promoting sustainable industrialization.

One of the most significant surges in data is the textual data [28, 29]. The newfound accessibility to vast volumes of textual data has ushered in a transformative shift in the methodologies employed to comprehend, dissect, and forecast the intricate workings of financial markets [30, 31]. Among sources of textual data, news articles have emerged as a reliable and authoritative source of information. The transformation of traditional print newspapers has given rise to a rich tapestry of digital news data [32]. Digital news articles provide a dependable and valuable flow of information by comprehensively covering timely events and trends. Digital news data's inherent potential lies in its ability to unlock insights into the global dynamics of financial markets, heralding a new era of understanding and predicting the complex interplay of global financial entities [33, 34].

Sentiment analysis is a notable method aims to utilize varied information sources to improve stock prediction models [30, 33, 35, 36]. The approach centers on the detection and classification of sentiment expressed within news articles, segregating them into three primary categories: positive, neutral, and negative. These sentiment classes exert a significant influence on the fluctuations and movements of stock prices in the financial markets [37, 38, 39]. The emergence of positive sentiment in news reports strongly correlates with rising stock prices, while a higher frequency of negative news corresponds to declining asset values. This approach establishes a compelling relationship between news sentiment and stock market fluctuations [40].

Wan et al. [41] propose that strong media sentiment targeting a specific company might signal significant changes in media sentiment regarding related companies. Companies are considered related if they are neighboring entities in a financial network constructed by examining news co-occurrences. The financial market functions as a complex system of interdependencies among companies. The co-occurrence of multiple companies in news articles reveals the underlying relationships between them, offering a unique perspective for investigating these interconnections [41].

Numerous studies have explored the intricate relationships among companies and the factors influencing their stock prices. One key finding is that the stock prices of companies within the same industry tend to be correlated, indicating the presence of interconnectivity among firms in a given sector. This interconnectivity can lead to associations between the price movements of different companies' stocks. Other research has highlighted the significant impact of public sentiment on stock prices. The sentiment expressed by the public toward a company can have a direct effect on its stock performance. Interestingly, this sentiment not only influences the company in question but also extends to related companies frequently mentioned in news articles. This phenomenon gives rise to two potential groupings of companies: industry-based clusters and co-occurrence clusters. Industry-based clusters consist of companies operating within the same sector, while co-occurrence clusters comprise companies that are often mentioned together in news reports, regardless of their industry. Within these clusters, the sentiments toward individual companies can mutually influence each other, ultimately impacting the stock prices of the entire sector group.

Previous studies [37–40] have explored how sentiment affects stock prices, but a detailed examination of the relative interactions between industry and news co-occurrence cluster relationships remains less investigated. A comprehensive exploration of the varying extents of sentiment's impact within the possible groupings presents an opportunity for extended scholarly investigation [41]. This research seeks to fill the knowledge gaps and enhance our comprehension of sentiment-driven connections that shape stock prices.

This study explores the comprehensive advantage of digital news data by harnessing the capabilities of natural language processing through advanced text mining techniques. Our primary goal is to gain insight into the underlying sentiments present in relevant news articles. We employ an entity recognition approach to extract the presence of Indonesian main board stock companies within each article and track their occurrences alongside other companies. The network analysis is deployed to visually represent and identify clusters of co-occurrence entities that emerge from these relationships. We aim to uncover whether these relationships are more decipherable at the industry level or within clusters of news co-occurrence.

Following this introduction, Sect. 2 provides a focused review of the literature surrounding sentiment analysis and its influence on stock prices, distinguishing different grouping mechanisms (individual firms, industries, and co-occurrence clusters). Section 3 outlines our methodology, including data collection, entity recognition, and the network analysis approach for detecting co-occurring entities in news data. Section 4 presents the empirical results, contrasting sentiment-price correlations at the individual, industry, and co-occurrence cluster levels. Section 5 concludes by summarizing our findings, discussing the theoretical and practical implications, and suggesting directions for future research.

This research addresses key questions about the complex relationships between news sentiment and stock prices. The fundamental question is, 'How is the news sentiment correlated to the stock prices of individual companies?'. Exploring deeper into the intricacy of inter-company relations, we explore the industry-specific impact of news to ascertain, 'How is the news sentiment correlated to stock prices from the perspective of companies' industries?'. Beyond the industry affiliations, we aim to uncover another layer of complexity. Companies are often referenced concurrently in journalistic reports with peer entities, attributable to partnerships, competitions, or shared stakes in specific global events or challenges. This leads us to compelling questions: 'Might clusters of relevant companies, based on their co-occurrence in similar news articles, exhibit a higher level of sentiment and stock price relationship?' and 'Does the sentiment associated with specific news exhibit a higher influence compared to the prevailing sentiment within the industry? '. By delving into these questions, we aim to decipher a higher level of sentiment and stock price relationships based on the company's industry affiliation and news co-occurrence propensities.

2 Related research

2.1 Stock prices and sentiment analysis

Stock prices play a pivotal role in the financial market. The prices serve as a crucial indicator of a company's overall success and management efficacy. Investors pay close attention to stock prices, as they encapsulate the present value of future earnings that shareholders will receive. The investors' objective for engaging in stock market transactions is to secure a financial return, either in the form of dividend income or capital gain, as well as to acquire ownership of a company. Prior to the investment, individuals weigh the anticipated rate of return alongside the intrinsic value of the company. The stock price serves as an indicator of the company's value for publicly traded companies [42]. An elevated stock price is indicative of greater corporate worth [43]. In simpler terms, when stock prices go up, it typically signifies an increase in the company's value, benefiting all involved stakeholders. Hence, investors place significant emphasis on stock prices because their fluctuations have a direct impact on potential returns.

The closing price is often utilized as the benchmark for assessing a stock's value [44–46]. This is because the closing price is widely considered to be the most accurate representation of a stock's real value at the end of each day's trading. This critical metric is frequently reported in the media and meticulously monitored by investors, making it an essential tool for investment decisions [47].

Predicting stock prices has long been a challenging [48] yet crucial area for a broad range of professionals, financial experts [49], investors [50], and economists [51]. The stock market functions as a venue for trading, distributing, and selling shares, enabling companies to grow by raising capital through Initial Public Offerings (IPOs). Investors aim to maximize their profits by accurately timing their purchase and sale of stocks from diverse companies. Two dominant methodologies for analyzing stock market behavior are technical and sentiment analysis. The former is grounded in examining historical stock prices to identify patterns, while the latter assesses the emotional tone in financial news to forecast price changes. Various external elements, such as financial news, social media, national policies, and regional economies, can have a significant impact on stock values [52, 53].

The Efficient Market Hypothesis states that market financial flows are governed by the availability of various factors, such as information and products, that influence a company's value [54]. Essentially, a market is considered efficient if stock prices promptly and fully reflect all available information. This information goes beyond financial metrics to include political events, social occurrences, and other types of data [55]. According to this theory, the market information alone cannot generate the excess returns investors may seek. Investors often analyze news and social media trends, as these can lead to the formation of abnormal returns on investments [56].

Previous research [57, 58] focused on the relationship between investor sentiment and stock market performance, specifically in the context of the Shanghai stock index. Their research constructs a sentiment index to examine the correlation between variations in investor sentiment and price fluctuations. The findings substantiate that investor sentiment has a notable positive influence on both market return and volatility.

Recent research has leveraged machine learning techniques for sentiment analysis, especially natural language processing (NLP) [53]. Specifically, Bidirectional Encoder Representations from Transformers (BERT) models have proven effective in classifying sentiment in financial news articles [53]. Summarizing the prediction of stock prices remains a complex effort due to the confluence of various factors like historical data, external elements, and human psychology. However, advances in machine learning and sentiment analysis offer promising avenues for more accurate predictions [11, 53].

2.2 The role of industry affiliation in stock price volatility

The Indonesia Stock Exchange (BEI) has instituted the IDX Industrial Classification (IDX-IC) as a mechanism to categorize its listed entities. The allocation into specific sectors, sub-sectors, industries, or sub-industries. The IDX-IC employs a systematic framework to classify listed entities predicated upon the market exposure pertaining to the terminal goods or services they produce. In 2006, JASICA's old classification delineated 9 sectors accompanied by 56 sub-sectors. After advancements in 2021, the classification system has been expanded to incorporate 12 primary sectors. These are further segmented into 35 sub-sectors, 69 distinct industries, and 130 sub-industries, thus offering a more comprehensive classification framework [59]. Each company is subjected to a more precise categorization. The sectors introduced in this revised classification are explained in Table 1.

The industry affiliation of a company is frequently employed to create homogenous groups of stocks to manage portfolio risk, evaluating relative value, and performing analyses among peer groups. Academic studies support the notion that industry-related factors play a major role in the correlations of stock returns, extending beyond general market influences. Companies within the same industry tend to show stronger correlations within their group and clearer distinctions from correlations with other groups. In real-world applications, the research conducted by financial analysts is frequently structured according to industry categories, reflecting the approaches used in esteemed financial publications [12]. In portfolio construction, investment managers place considerable importance on the industry connections of the included stocks. Furthermore, academic research, as cited in [60, 61] articles, highlights that investors frequently observe a positive correlation between the share price of a company making an announcement and those of its industry peers. This emphasizes the concept that a firm's share price in a particular industry may correlate with the share prices of other companies in the same industry sector.

2.3 News co-occurrence network and news co-occurrence company clusters

Network analysis provides a comprehensive framework for a deeper understanding of entities' underlying relationships [61]. Classical economic models often fail to decipher intricate market dynamics [62]. The application of data-driven and computational modeling methods portrays financial markets from the perspective of computational networks. The individual constituents (such as individual stocks or other financial assets) serve as nodes, while the connections among these constituents form the edges [63–66]. This model has been applied to various scenarios, including depicting the spread of systemic risk across global economic systems [67] and evaluating the transmission of default risk within industrial networks [68].

The advancement of Natural Language Processing (NLP) techniques [69] provides support for the computational examination of large-scale unstructured data, which was formerly unsuited for numerical processing (such as textual data found in financial news articles or research reports). Studies have been dedicated to developing networks of companies based on their mutual appearance in news sources. Whenever a company is co-mentioned with other companies, it is treated as a connected edge [41].

In the study by Wan et al. [41], the co-occurrence network of companies was employed to investigate the dynamics of sentiment correlations. The co-occurrence network within news articles is suitable to uncover the underlying relationships between companies, offering a new viewpoint for the inter-company association analysis. The study demonstrates that a significant portion of company industries establish strong connections by gaining frequency with which companies are jointly referenced within a specific temporal span in news content. Moreover, this network can offer more profound insights when distinguishing between news co-occurrence clusters and industries, potentially encompassing perspectives extending beyond industry-based classifications.

Certain nodes exhibit strong interconnections within the associative network, while others possess a lower degree of connection. The intensity of the internal connections exceeds the external node connections denoted as a community. The overarching arrangement of communities in the entire network is termed the community structure. The process of community detection aims to locate these closely linked communities within the network, thereby revealing additional insights into the interconnectedness among financial firms [70]. According to Barabasi [71], network approaches encompass several metrics that support the analysis, one of which is modularity. Modularity quantifies and sizes communities within an identified network. The Louvain Modularity method [16] is employed to uncover the communities among interconnected companies.

2.4 Existing research on news co-occurrence and its impact on stock prices

The study by Wan et al. in 2021 [41] emphasizes the complex interdependencies in modern financial markets. News sentiment of a company can influence the market performance of that company and other related companies, particularly those in related sectors or groups. The research utilizes Natural Language Processing (NLP) to analyze news sentiment for 87 prominent companies reported on Reuters over seven years. By constructing a news co-occurrence network based on how frequently companies are mentioned together in news articles, the study provides insights into the interconnectedness of companies. This network helps identify relationships and understand how companies within the same sector or co-occurrence group are linked through news media.

The study finds that strong media sentiment towards one company can lead to significant changes in media sentiment towards related companies [41]. The research suggests that negative sentiment changes in a company often correlate

Table 1 IDX industrial classification

Code	Sector	Explanation
IDXENERGY	Energy	(1) Oil & Gas, (2) Coal, (3) Oil, Gas & Coal Supports, (4) Alternative Energy Equipment, (5) Alternative Fuels
IDXBASIC	Basic Materials	(1) Chemicals, (2) Construction Materials, (3) Containers & Packaging, (4) Metals & Minerals, (5) Forestry & Paper
IDXINDUST	Industrials	(1) Aerospace & Defense, (2) Building Products & Fixtures, (3) Electrical, (4) Machinery, (5) Diversified Industrial Trading, (6) Commercial Services, (7) Professional Services, (8) Multi-sector Holdings
IDXNONCYC	Consumer Non-Cyclicals	(1) Food & Staples Retailing, (2) Beverages, (3) Processed Foods, (4) Agricultural Products, (5) Tobacco, (6) Household Products, (7) Personal Care Products
IDXCYCLIC	Consumer Cyclicals	(1) Auto Components, (2) Automobiles, (3) Household Goods, (4) Consumer Electronics, (5) Sport Equipment & Hobbies Goods, (6) Apparel & Luxury Goods, (7) Tourism & Recreation, (8) Education & Support, (9) Services, (10) Media, (11) Entertainment & Movie Production, (12) Consumer Distributors, (13) Internet & Homeshop Retail, (14) Department Stores, (15) Specialty Retail
IDXHEALTH	Healthcare	(1) Healthcare, (2) Equipment & Supplies, (3) Healthcare Providers, (4) Pharmaceuticals, (5) Healthcare Research
IDXFINANCE	Financials	(1) Banks, (2) Consumer Financing, (3) Business Financing, (4) Investment Services, (5) Insurance, (6) Holding & Investment Companies
IDXPROPERT	Properties & Real Estate	(1) Real Estate Management & Development
IDXTECHNO	Technology	(1) Online Applications & Services, (2) IT Services & Consulting, (3) Software, (4) Networking Equipment, (5) Computer Hardware, (6) Electronic Equipment, Instruments & Components
IDXINFRA	Infrastructures	(1) Transport Infrastructure Operator, (2) Heavy Constructions & Civil Engineering, (3) Telecommunication Service, (4) Wireless Telecommunication Services, (5) Electric Utilities, (6) Gas Utilities, (7) Water Utilities
IDXTRANS	Transportation & Logistic	(1) Airlines, (2) Passenger Marine, (3) Transportation, (4) Passenger Land Transportation, (5) Logistics & Deliveries
	Listed Investment Product	(1) Investment Trusts, (2) Bonds

with larger movements across the related group, particularly in the financial services sector. Positive sentiment changes can have a more symmetric effect across consumer product-related companies. The findings imply that news sentiment, particularly when it shows extreme changes, can be a predictor of market movements. This predictive value extends beyond individual companies to sectors or groups of companies, highlighting the importance of considering network effects in financial analysis. The findings suggest that understanding the dynamics of news sentiment and its network effects could be crucial for market participants and regulators' financial decision-making. The research advocates for a shift in analysis from individual companies to groups or sectors for a more comprehensive understanding of market dynamics.

3 Methodology

This study integrates sentiment analysis, entity recognition, network analysis, and correlation evaluation to investigate the intricate relationships between news sentiment and stock market dynamics. Figure 1 illustrates the research workflow, which is further elaborated in the following sub-sections.

3.1 Data collection and preparation

The data collection process involves consolidating information from various reputable sources, encompassing stock prices, the list of companies registered on Indonesia's main stock board, and digital news related to the companies. The data collected encompasses 348 companies listed on Indonesia's main stock board, detailing the monthly closing stock prices and the corresponding industry affiliations of each company.

To capture a broad spectrum of economic and business perspectives, we collected 192,582 news articles from CNBC Indonesia spanning 2018 to October 2023, without imposing keyword-based filters. These articles address diverse topics

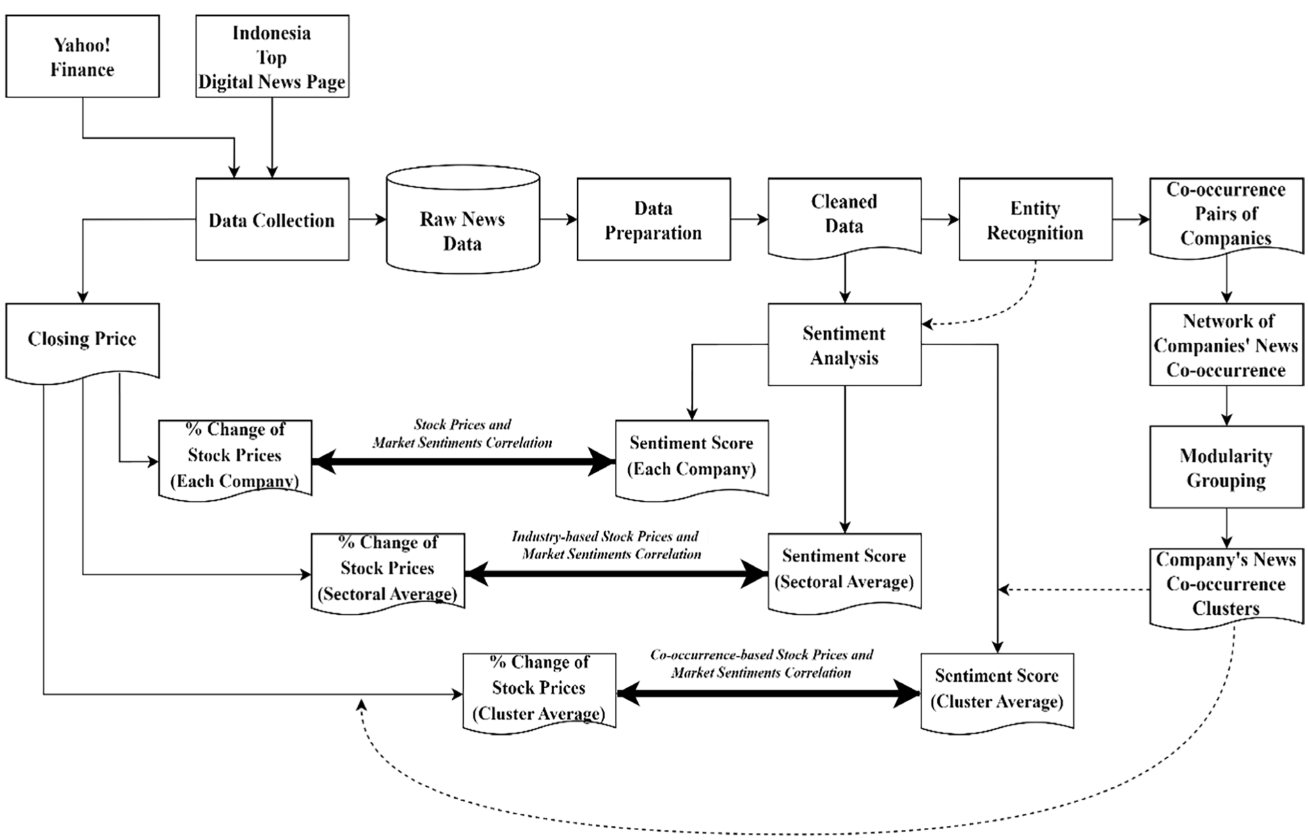


Fig. 1 Research workflow

such as finance, commerce, investment, technology, politics, and entrepreneurship, providing a holistic view of current economic developments.

The news articles were collected through automated web scraping using Python-based tools, with proper adherence to the website's terms of service. During the data preprocessing stage, we removed non-news content (e.g., advertisements, and promotional materials), eliminated duplicate articles, and cleaned irrelevant characters and HTML tags to ensure data consistency. Articles without any mention of listed companies were excluded based on our entity recognition results. In the subsequent data-preparation phase, we applied a named entity recognition (NER) approach to identify and extract 85,908 organizational entities mentioned within the news corpus.

Finally, to align all datasets—stock prices, industry affiliations, and news coverage—we segmented the data into monthly intervals for further analytical procedures. This step enables consistent comparison across companies, industries, and sentiment signals derived from the news, thereby ensuring cohesive and streamlined analytics.

3.2 Entity recognition and sentiment analysis

Manual processing is infeasible for the vast volume of data. The Named Entity Recognition (NER) for computerized text analysis aims to identify and categorize named entities in text into predefined classifications such as people, organizations, and locations. It stands as a crucial research area in natural language processing, serving as a foundational tool for numerous NLP studies, including information extraction and machine translation [72]. NER was dominated by lexicon and rule-based methodologies, which relied on external lexicons for classification and were effective in specific contexts but required frequent updates and struggles with complex texts from diverse fields. NER tasks mark a notable shift since machine learning gained prominence in NLP [73]. Effectively addressing the sequence annotation is become crucial for enhancing the NER. However, this approach demands meticulous feature selection involving determining the specific feature sets that aptly represent the unique attributes of a given named entity and result in limited generalizability.

In recent years, the rapid progress in deep learning techniques has propelled neural network models to the forefront of Named Entity Recognition (NER) tasks. One of the most notable benefits of these models is their ability to operate independently of labor-intensive manual feature extraction processes. Text data is transformed into vector format through learned embedding models and is fed into the neural network, which encodes the text sequence information. The decoding layer interprets this encoded data to produce the final annotated sequence [74]. In 2018, Devlin et al. [75] introduced BERT (standing for Bidirectional Encoder Representation from Transformer), a deep learning-oriented NER technique. BERT is a language representation model developed by the Google Artificial Intelligence Language team in 2018. BERT demonstrated superior performance in pre-trained language models for NLP. It is adept at capturing long text features and dynamically producing word vectors across varied contexts with enhanced computational efficiency [75].

Deep learning techniques have been integrated into Indonesian NER studies. A notable contribution comes from Wilie et al. [76], who developed and refined the Indonesian BERT (IndoBERT) model for twelve distinct NLP tasks, including NER. Earlier research using the identical dataset was undertaken with a focus on quotation identification [77]. Data were specifically labeled for pinpointing quotations. Khairunnisa et al. [78] curated a more reliable Indonesian NER dataset by revising a formerly inconsistent one, generously offering it to the wider research community for future explorations.

In our research, we employ the IndoBERT model and utilize the renowned NERGrit dataset for the Indonesian language, sourced from the Grit-ID repository, comprising 2,090 sentences. This dataset uses the IOB chunking representation for labeling and encompasses three primary named entity tags: PERSON (individual's name), PLACE (location's name), and ORGANIZATION (organization's name) [76]. Since NERGrit isn't explicitly designed to extract names of companies listed on the main board stock and considering the inherent complexity where a single organization may be represented in various naming conventions (for instance, "Apple Inc.", "AAPL", and "Apple" all denote the same entity), we undertook fine-tuning efforts to enhance the entity recognition outcomes.

We conduct entity detection for companies by considering the stock codes, the formal company names, and the commonly used names. As an example, "BBCA", "Bank Central Asia Tbk.", "Bank Central Asia", "Bank BCA", and "BCA" all pertain to a singular entity, namely "BBCA". Out of the 348 companies listed on Indonesia's main board stock companies list, only 342 companies have been referenced in CNBC news during the observed period. Some of the companies that are not mentioned in the news, including SCCO—Supreme Cable Manufacturing, INCI—Intanwijaya Internasional Tbk, ESTI—Ever Shine Tex Tbk, STAR—Buana Artha Anugerah Tbk, PSSI—IMC Pelita Logistik Tbk, and ADCP—Adhi Commuter Properti Tbk.

We selected IndoBERT for sentiment analysis because of its high performance in Indonesian language processing and ability to capture Indonesian financial news's linguistic and contextual characteristics. Compared to conventional machine learning models such as Naïve Bayes and Support Vector Machines (SVM), IndoBERT demonstrates superior accuracy and contextual understanding. In our experiments, the IndoBERT model achieved an accuracy of **94.34%** in classifying sentiment polarity. This high performance, combined with its ability to process large-scale unstructured news data efficiently, made IndoBERT the most appropriate choice for this study.

3.3 Network of news co-occurrence companies and modularity grouping

A network of co-occurrence companies is constructed based on the simultaneous mention of companies within a specific news article. After constructing the list of companies mentioned across various news articles, we identified pairs of companies that were jointly mentioned. In this process, we paired any two companies mentioned in the same article without distinguishing between primary and secondary roles. As a result of this methodology, we created a network of undirected relationships between companies. Over time, we have documented 10.749 co-occurrence pairs among the 335 companies in the news. Examples of company pairs are presented in Table 2.

The methodology involves transforming company pairings into a network structure. We utilize two types of input data: the node list and the edge list. The node list serves as a tabular representation that outlines the specific characteristics attributed to each individual node existing within the network. In this context, the node list defines the companies' characteristics, including their notations, names, and respective operation industries, as exemplified in Table 3. The edge list represents the connections or relationships between pairs of nodes in the network. Each row in the edge list signifies a unique connection between two companies jointly mentioned in a news article. The edge list includes essential information such as the source company, target company, and the year of their co-occurrence, as illustrated in Table 4.

The network construction follows an undirected approach, where the connections between companies lack directionality. We generate two distinct visualization representations. The first visualization maps companies across industries, offering a sectoral overview of the network's structure. The second visualization focuses on mapping companies based on their co-occurrence clusters within news articles, revealing interconnected groups of companies frequently mentioned together. While the industrial information for each company is readily available from stock market data providers, identifying news co-occurrence clusters demands additional effort. We employ a method based on co-mention frequency within news articles to identify clusters in the network. This method, known as "modularity grouping", groups nodes closely linked by co-mentions. This analysis reveals significant co-occurrence patterns among networked companies.

The Louvain Modularity formula serves as a pivotal tool in identifying distinct communities within complex networks. The formula, which is commonly applied in community detection algorithms, quantifies the extent to which a network can be partitioned into separate modules or communities that are more densely interconnected internally than with the rest of the network [79]. The Louvain Modularity formula mathematically captures the difference between the observed number of edges within communities and the expected number of edges that would occur in a random network. We select the Louvain Modularity algorithm because of its computational efficiency, scalability, and effectiveness in detecting meaningful community structures in large-scale networks. Given the volume of co-occurrence pairs in our dataset and the complexity of the Indonesian capital market, the Louvain method offered a suitable approach to uncover clusters of companies that frequently appear together in news articles. Moreover, its capability to reveal latent community structures aligns with our objective of identifying interconnected companies beyond conventional industry classifications.

The Louvain formula is denoted as below.

$$Q = \frac{1}{2m} \sum_{i,j} \left(e_{i,j} - \frac{k_i k_j}{2m} \delta(C_i, C_j) \right) \tag{1}$$

In the network, $\boldsymbol{e_{i,j}}$ represents the connection strength between nodes $\boldsymbol{i}$ and $\boldsymbol{j}$, which, in this context, is determined by the similarity of company mentions in news articles. For nodes $\boldsymbol{i}$ and $\boldsymbol{j}$, $\boldsymbol{k_i}$ and $\boldsymbol{k_j}$ represent the total weights of their respective connecting edges. In this context, $\boldsymbol{C_i}$ and $\boldsymbol{C_j}$ depict the specific communities (or company groups) to which $\boldsymbol{i}$ and $\boldsymbol{j}$ are affiliated. The Kronecker delta function, represented as δ, equals 1 when $\boldsymbol{x} = \boldsymbol{y}$ and 0 when $\boldsymbol{x} \neq \boldsymbol{y}$. Using the Louvain method, every node starts in its unique community. The modularity change ($\boldsymbol{\Delta Q}$) is determined by tentatively moving node $\boldsymbol{i}$ from its initial community to others it's linked to. If this results in a modularity increase, node $\boldsymbol{i}$ shifts to the community giving the maximum $\boldsymbol{\Delta Q}$. If not, it retains its initial allocation. This stepwise process continues for every node until no further modularity enhancement exists.

Table 2 Example of company pairs

News	Entity Recognized	Company Pairs
Jokowi Ubah Struktur Kepemilikan BMRI & BBRI *Jakarta, CNBC Indonesia—Presiden Joko Widodo atau Jokowi mengubah struktur kepemilikan saham negara di dua perusahaan pelat merah, yakni PT Bank Rakyat Indonesia (Persero) Tbk dan PT Bank Mandiri (Persero) Tbk. Hal itu tertuang dalam Peraturan Pemerintah (PP) Nomor 31 Tahun 2023 dan PP Nomor 32 Tahun 2023. Pasal 1 PP 31/2023 mengatur perubahan dilakukan dengan negara menjual sebagian saham serta menambahkan modal. Kemudian, dalam pasal selanjutnya tertulis sebagian saham Seri B dialihkan kepada Lembaga Pengelola Investasi sebagai tambahan penyertaan modal, sesuai dengan PP 111/2021 sebanyak 5,49 miliar saham. Penjualan sebagian saham milik negara, penambahan modal, serta pengalihan sebagian saham Seri B membuat komposisi saham negara di BRI berubah menjadi 1 saham Seri A dan 80,61 miliar saham Seri B. Dengan demikian struktur kepemilikah saham Pemerintah Republik Indonesia di BRI menjadi 53,19% yang terdiri atas satu saham Seri A dan 80,61 miliar saham Seri B. Adapun PP tersebut ditetapkan pada 16 Juni 2023. Sementara itu, perubahan struktur kepemilikan pemerintah di Bank Mandiri diatur dalam PP 32/2023. Tertulis bahwa 3,73 miliar saham BMRI dialihkan kepada Lembaga Pengelola Investasi sebagai tambahan penyertaan modal. Dengan demikian kepemilikan saham pemerintah di BMRI menjadi 52% yang terdiri dari 1 saham Seri A dan 24,26 miliar saham Seri B. PP 32/2023 juga ditetapkan oleh Presiden Jokowi pada 16 Juni 2023*	BBRI–PT Bank Rakyat Indonesia (Persero) Tbk BMRI–PT Bank Mandiri (Persero) Tbk	BBRI—BMRI

Table 3 Example of the node list

Node	Company's Name	Industry
MAPI	Matahari Putra Prima	Consumer Cyclicals
LPPF	Lippo Malls Indonesia	Properties & Real Estate
BBTN	Bank Tabungan Negara	Financials
AMRT	Alfamart	Consumer Cyclicals
FREN	Smartfren Telecom	Technology
RALS	Ramayana Lestari Sentosa	Consumer Cyclicals

Table 4 Example of the edge list

Source	Target	Period
RALS	LPPF	2018
RALS	MAPI	2018
BSDE	BBTN	2018
BNGA	BBTN	2018
BBNI	AMRT	2018
ISAT	FREN	2018

The identified communities are referred to as "co-occurrence-based clusters of companies" within the scope of this paper. With the clusters of news co-occurrence companies derived through the procedures mentioned above, we conduct a comparison and contrast of these clusters with the established ground-truth company industries.

3.4 Sentiment analysis

Sentiment is defined as a viewpoint expressing positive or negative opinions about a specific issue [80]. Sentiment analysis identifies user-expressed reviews in terms of sentiment or polarity[81] and is primarily used to extract information from unstructured text [82]. This approach utilizes sophisticated Natural Language Processing (NLP) tools to decipher, process, and assess sentiments conveyed in textual content, categorizing them into positive, negative, or neutral designations [83]. In this research, we proposed a sentiment analysis model to uncover the emotional tone of news articles by analyzing the full content on the digital platform. The outcomes are primarily categorized as positive or negative, along with their quantitative polarity values. We use BERT-based sentiment analysis for this purpose.

Leveraging transformer technology, BERT is adept at discerning the contextual relationships between words or subwords in a sentence, the meaning of which may vary based on the sentence context [84]. BERT applies deep learning methodologies to address the intricate nuances present in linguistic data [75]. When contrasted with traditional machine learning techniques such as Naïve Bayes and SVM, which depend on simpler rules or features, BERT offers a superior understanding of word contexts in sentences [85]. The Naïve Bayes algorithm is noted for its efficiency and simplicity [86] yet faces challenges in capturing detailed context and operates under the feature independence assumption. SVM excels in processing high-dimensional data and can achieve commendable results across various classification tasks [87]. SVM's parameter tuning can be complex and may not naturally discern contextual relationships in text. BERT remains a preferred model considering these comparisons.

IndoBERT is a specialized variant pre-trained for the Indonesian language [76]. This dataset encompasses formal language without abbreviations, as well as everyday conversational sentences sourced from various platforms, including blogs, websites, and social media [76]. IndoBERT can categorize text sentiments as positive, negative, or neutral concerning a specific issue [88]. IndoBERT sentiment analysis is suitable since the news articles are presented in the Indonesian language, given that IndoBERT provides intricate sentiment analysis capabilities.

In this research, we employ the IndoBERT pretrained model to determine the sentiment of each news article related to the 342 companies. Our dataset encompasses a total of 85.908 news articles. We fine-tuned the model to refine the precision of our analysis and ensure alignment with our research goals. Employing this method, we obtain the probability of the sentiment polarity for each news article. The scores range from -1 to 1, with values closer to -1 indicating a more negative sentiment and those approaching 1 reflecting a more positive sentiment. A threshold of 0 is established as the neutral midpoint in this scale. The BERT-based sentiment analysis model demonstrates a remarkable level of performance, as evidenced by its impressive 94.34% accuracy rate. The high degree of accuracy indicates the model's reliability and effectiveness in accurately differentiating between the two emotional states.

3.5 Correlation comparison

This study aggregated the sentiment data on a monthly basis and calculated the monthly percentage change in closing stock prices. We observed the limited number of companies featured in news articles during 2018, which resulted in insufficient data for monthly sentiment analysis. We decide to exclude the news from the year and focus our correlation analysis on the dataset spanning from 2019 to October 2023. Our investigation encompassed three distinct types of correlation analyses: the correlation between stock prices and market sentiments (individual company level), the industry-based correlation between stock prices and market sentiments (average sectoral level), and the co-occurrence-based correlation between stock prices and market sentiments (average cluster level, derived from co-occurrence grouping results) as explained below:

1. In the initial phase, we focus on uncovering the intricate correlation between sentiment trends and market dynamics at the granular level of individual companies. We quantify the interplay between fluctuations in sentiment and their impact on individual companies' performance without considering the network effects or interdependencies between companies in a specific industry.
2. Transitioning into the next phase of our investigation, we shifted our attention from individual companies to a more macroscopic perspective by investigating the correlation between sentiment patterns and market movements within distinct industries. We utilized the average monthly sentiment values and the average monthly changes in closing prices for each industry and performed a correlation analysis on the parameters. This transformation in our analytical approach enabled us to uncover broader industrial trends and ascertain whether sentiment played a significant role in shaping market dynamics across different segments of the economy.
3. In the final stage, we leveraged the concept of co-occurrence clusters within our data and probed the correlation between sentiment and market performance from another perspective. We utilized the average monthly sentiment values and the average monthly changes in closing prices for each modularity cluster and performed a correlation analysis on the parameters. By identifying the cluster-based correlation, we discover whether the collective sentiment patterns within these clusters signify any discernible impact on the market.

We compare the three correlation approaches. To statistically validate the differences in correlation strength across the three analytical levels, we applied the Fisher r-to-z transformation test to assess the significance of correlation changes between individual, industry, and cluster levels. This comprehensive analysis allowed us to draw valuable insights into the complex relationships between sentiment trends and market behavior and provides insight into the varying degrees of influence that sentiment exerts at different levels of analysis.

4 Result and discussion

This research analyzes the temporal evolution of news co-occurrence networks for Indonesian main board companies from 2018 to 2023, as shown in Fig. 2. The network structure exhibits progressive expansion in both nodes and edges throughout the observation period. The expansion rate shows an accelerating pattern with each subsequent year. This growth pattern characterizes the increasing complexity of relationships between listed companies in Indonesian business news media narratives.

The quantitative network analysis examines the growth of company relationships in news media through multiple network metrics. In this network structure, nodes represent individual companies listed on Indonesia's main board stock, while edges indicate the co-occurrence of companies within the same news article. The initial network in 2018 contained minimal relationships with only 4 companies (nodes) and 2 co-mentions (edges) in news articles. The network expanded substantially to reach 335 companies and 10,749 co-occurrences by 2023, encompassing 96.26% of all listed companies on Indonesia's main board. This extensive coverage implies increasing media attention on company interconnections and their roles in Indonesia's capital market dynamics. The network's connectivity parameters demonstrate exponential growth through two key measurements: average degree and average weighted degree. The average degree represents the mean number of unique company pairs in news co-occurrences, while the average weighted degree accounts for the

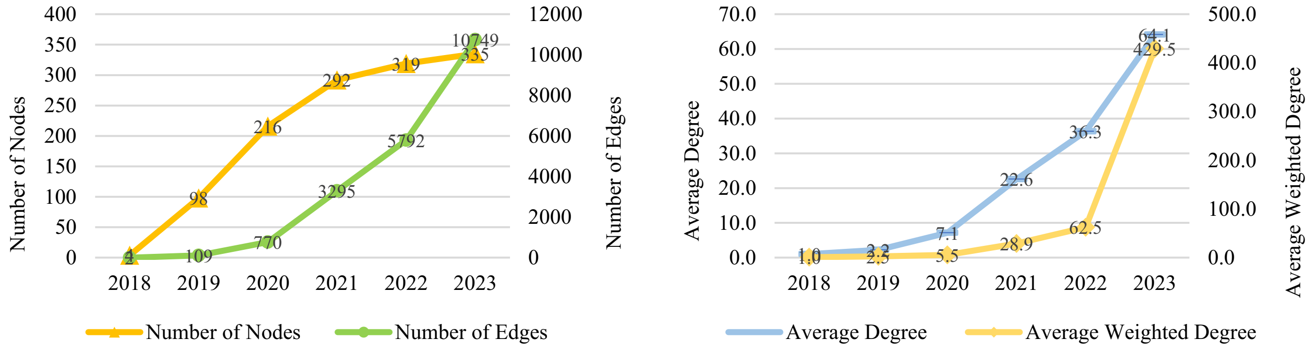


Fig. 2 Growth of news co-occurrence network components from 2018 to 2023

frequency of these co-occurrences between the same company pairs. The average degree metric increased from 1.0 in 2018 to 64.1 in 2023, showing that companies appear in news articles with an expanding network of different companies over time. Until 2022, this metric showed steady growth, reaching 36.3, indicating gradual diversification of company relationships in news coverage. The average weighted degree experienced intensive growth in 2023 (jumping from 62.5 in 2022 to 429.5 in 2023), showing intensified news coverage of existing company relationships. This contrasting pattern between metrics suggests an evolution in media reporting behavior: the initial years (2018–2022) focused on establishing connections between different companies, while 2023 showed concentrated coverage on established company relationships. This shift implies that while the network of unique company connections stabilized, certain company pairs received significantly more repeated coverage, potentially indicating strengthening business relationships or increased market attention on specific company partnerships within Indonesia's corporate ecosystem.

4.1 Industry-based and news co-occurrence company clusters

The network metrics reveal the overall growth in company co-occurrences within news coverage. A deeper understanding of this evolution requires examination of industry-specific co-occurrence patterns. The analysis of company co-occurrence distribution across different industrial sectors provides insights into how companies are mentioned together in news articles, revealing interconnected relationships between business segments in Indonesia's capital market. The analysis of unique company co-occurrences across different industries from 2018 to 2023 is shown in Fig. 3. The Financial sector dominated in co-mentions with 46,421 co-occurrences in 2023, followed by the Energy sector with 15,341 co-occurrences and Infrastructure sector with 13,964 co-occurrences. The Customer Non-Cyclicals sector also maintained a substantial interconnected presence with 13,469 co-occurrences, while Basic Materials recorded 12,999 co-occurrences in the same period. This distribution pattern shows the concentrated interconnection of finance-related companies with other sectors, which aligns with their central role in Indonesia's capital market.

The temporal evolution of industry co-occurrences exhibits notable growth characteristics. All sectors maintained minimal co-mentions during 2018–2020, with most sectors recording fewer than 200 co-occurrences annually. A moderate increase occurred in 2021, followed by steady growth in 2022. 2023 marked an exceptional surge across all sectors, with the Financial sector experiencing the highest increase from 4,089 to 46,421 co-occurrences. This surge suggests

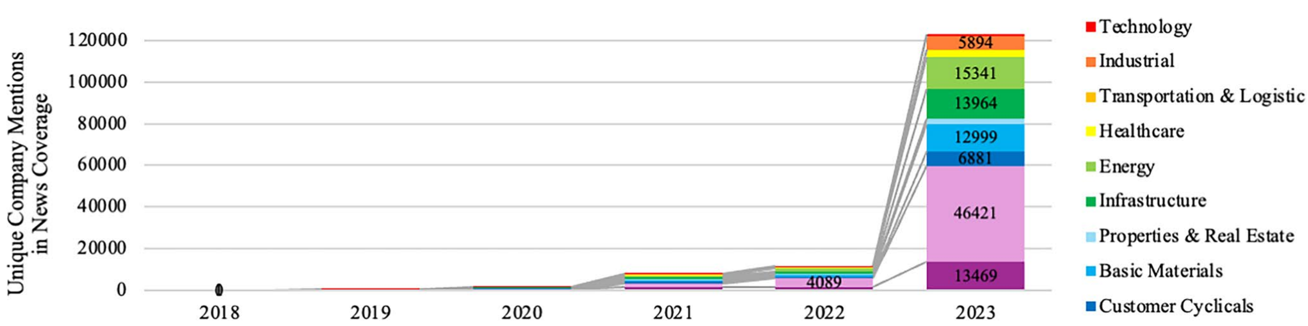


Fig. 3 Sectoral distribution of company co-occurrences in news articles

intensified interconnection between financial companies and other sectors in Indonesia's post-pandemic economy. The disparity in co-occurrences between sectors indicates varying levels of business relationship intensity and interconnectedness. While the Financial, Energy, Infrastructure, and Customer Non-Cyclicals sectors showed extensive co-mentions with other sectors, sectors like Transportation & Logistics (713 co-occurrences) and Technology (1,058 co-occurrences) maintained relatively modest interconnection in 2023. This distribution reflects both the structural interconnectedness of Indonesia's stock market and the relative integration of different sectors in the country's economic network.

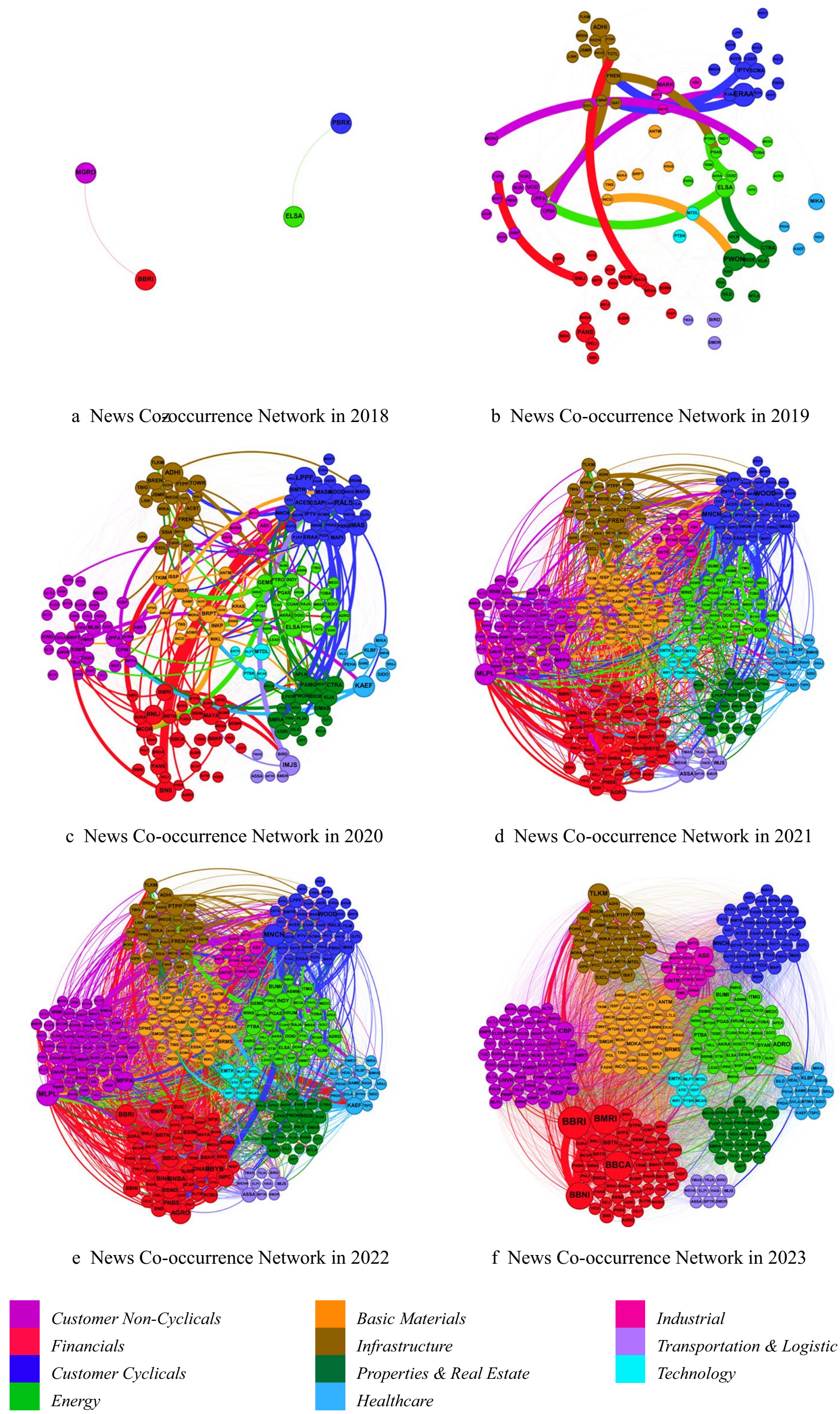
A notable observation emerges from the Healthcare sector's co-occurrence pattern during the COVID-19 pandemic period. Despite the global health crisis in 2020–2021 and the healthcare sector's critical role, news coverage mentioning healthcare companies together with other companies remained relatively modest, with only 76 co-occurrences in 2020 and 375 co-occurrences in 2021. This co-mention pattern continued with 303 co-occurrences in 2022 before increasing to 3,303 co-occurrences in 2023. The limited co-occurrence frequency during the pandemic period shows that the news might focused on individual healthcare companies' specific roles rather than their business relationships with other companies. This trend differs from financial sector co-occurrences, which maintained consistent growth throughout the pandemic period, from 244 co-occurrences in 2020 to 1,652 co-occurrences in 2021, eventually reaching 46,421 co-occurrences in 2023.

The complexity of market dynamics leads us to explore multiple perspectives in understanding the relationship between stock prices and market sentiment. We examine these relationships through industry affiliations and news co-occurrence patterns. We hypothesize that companies' relationships and market behaviors might be better understood through their sectoral connections and how they are collectively represented in media coverage. We employ two distinct approaches to explore these relationships: traditional industry-based clustering and news co-occurrence clustering. Industry-based clustering groups companies by their business sectors, reflecting shared characteristics in business models, market dynamics, and regulatory environments. Meanwhile, news co-occurrence clustering captures relationships based on how companies are mentioned together in media coverage, potentially revealing connections that transcend conventional industry boundaries. This dual approach aims to uncover how market sentiments behave when examined through these collective groupings rather than at the individual company level in the subsequent sections.

To uncover these relationships, we construct annual network representations of company co-occurrences in news articles from 2018 to 2023, as shown in Fig. 4. Companies are represented as nodes, with their sizes proportional to their weighted degree—indicating the frequency of co-mentions with other companies. The connections between nodes (edges) represent co-occurrences in news articles, while different colors denote industry classifications. This visualization approach enables observation of both industry-based clustering and emergent relationships formed through news coverage patterns. The network structure shows progressive transformation from a simple configuration with minimal connections in 2018 to a complex, densely interconnected network by 2023. The node sizes represent weighted degrees, indicating the frequency of companies' co-occurrences with other companies in news coverage. This evolution highlights the increasing media recognition of corporate interconnectedness in Indonesia's market ecosystem. The temporal development of the network exhibits distinct phases of structural evolution. The 2018 network (Fig. 5a) shows minimal interconnection with only four companies, reflecting limited co-mentions in news coverage. A significant expansion emerges in 2019 (Fig. 5b), where many cross-industry co-occurrences begin emerging, particularly visible in the Financial sector (red nodes) and Customer Non-Cyclicals sector (purple nodes). This pattern indicates emerging recognition of sector-based relationships in media coverage, reflecting natural business affiliations and market segments. The network structure undergoes substantial transformation during 2021–2023 (Fig. 5d–f), coinciding with post-pandemic economic recovery. The 2021 visualization shows intensified interconnections across sectors, with the Financial sector maintaining central positioning in the network structure. This centrality of financial institutions signifies their fundamental role in facilitating cross-sector business activities and economic recovery initiatives. By 2022, the network exhibits dense inter-industry connections, reflecting the increased cross-sector collaborations and business integration. The 2023 network reaches peak complexity, characterized by thick edges representing frequent co-occurrences, particularly centered around major financial institutions like BBRI, BMRI, and BBCA. This concentration around major banks suggests their crucial role in Indonesia's corporate financing and business development.

The growing interconnectedness of companies in news articles reflects an intricate network of relationships that might impact market sentiment and stock price movements. As cross-industry interconnectedness becomes more prevalent, companies may not only be grouped based on their sectoral affiliations but also according to the nature of their co-occurrences in news articles. To explore these relationships, we employ two distinct approaches to visualize and analyze company networks: industry-based clustering and news co-occurrence modularity clusters, as shown in Fig. 5. The industry-based network (left) groups companies according to their IDX Industrial Classification, represented by different

Fig. 4 News co-occurrence network in 2018 to 2023



colors. This traditional classification reflects companies' core business activities, market dynamics, and regulatory environments. The news co-occurrence modularity clusters (right) are derived using the Louvain algorithm, which optimizes modularity to identify communities of companies frequently mentioned together in news articles. The modularity score of 0.249 indicates meaningful community structure while acknowledging substantial inter-cluster connections, reflecting the complex nature of corporate relationships in news coverage. The comparison between these two networks

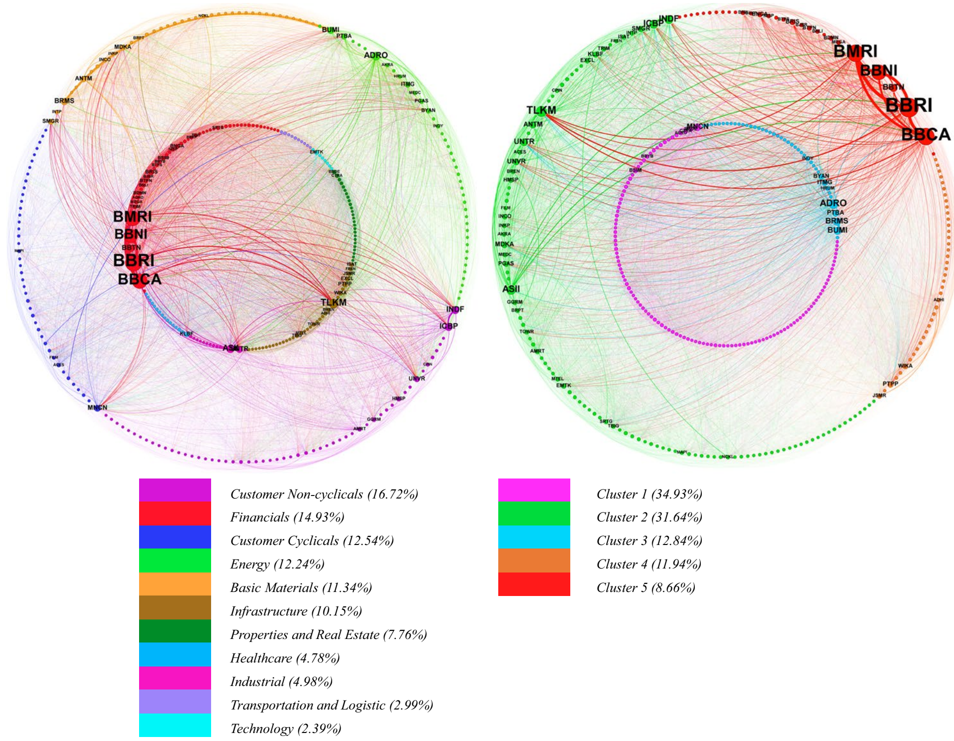


Fig. 5 Industry-based (left) and news co-occurrence company clusters (right)

highlights the divergence between traditional industry boundaries and emerging relationships in news coverage. While the industry-based network shows clear sectoral boundaries, the news-based co-occurrence modularity clusters challenge this traditional classification by demonstrating that companies within the same industry can belong to different clusters when analyzed based on their co-mentions in news articles. Media coverage captures relationships that transcend conventional industry boundaries, potentially reflecting business partnerships, shared market influences, or common regulatory concerns.

To better visualize how companies are positioned within both industry sectors and news co-occurrence clusters, Fig. 6 presents a network visualization where node alignment and colors serve different representational purposes. Companies aligned along the same radial lines belong to the same industry sector, maintaining the traditional IDX Industrial Classification. Meanwhile, nodes sharing the same color belong to the same modularity-based cluster, derived from news co-occurrence patterns. This dual visualization technique shows how companies from the same sector may be distributed across different news-based clusters (shown by different colors along the same line), and conversely, how each cluster (represented by a single color) contains companies from multiple sectors (shown by the distribution of same-colored nodes across different lines). The node sizes represent weighted degrees, with larger nodes indicating companies more frequently co-mentioned in news articles. Notably, major financial institutions like BBCA, BBRI, BBNI, and BMRI appear as prominent nodes, reflecting their central role in news coverage and their frequent co-occurrence with companies across various sectors.

To gain a more detailed understanding of the cluster composition, Table 5 presents the distribution of companies across sectors within each modularity cluster. This cross-tabulation allows us to examine precisely how companies from different sectors are distributed across the five clusters identified by the Louvain algorithm. As shown in Table 5, the Louvain algorithm identified five modularity clusters representing different patterns of company co-occurrences in news coverage discussed as below:

(1) Cluster 1 holds the highest number of companies at 117 (34.93% of total companies), with significant representation across multiple sectors. Properties & Real Estate companies form the largest component with 23 companies (88.46% of the sector), alongside Financial institutions with 19 companies (38% of the sector), and Consumer Non-Cyclicals with 17 companies (30.36% of the sector). This composition captures how media coverage links real estate development activities with financial services and consumer markets, highlighting the interconnected nature of property development, financing, and consumer behavior in Indonesia's economy.

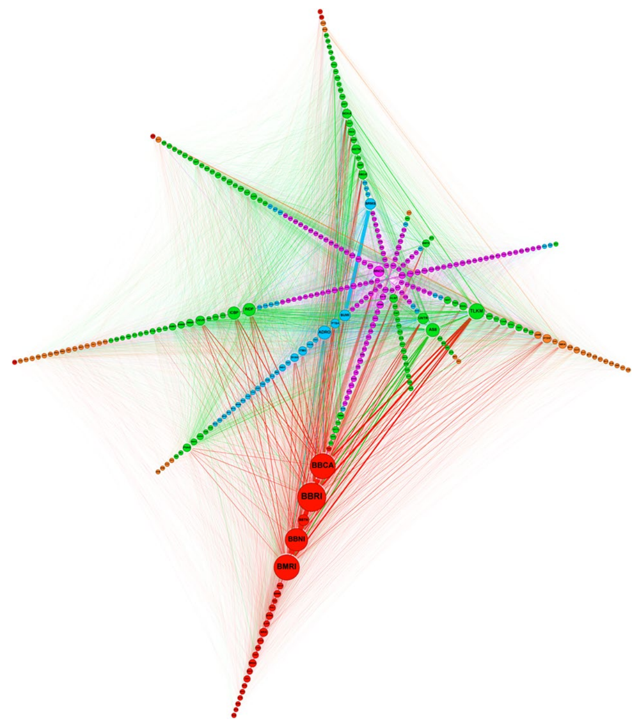


Fig. 6 Network visualization of companies by industry sectors (radial lines) and news co-occurrence clusters (colors)

Table 5 Composition of company sectors within each cluster

Sector \ Cluster	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Total
Customer Non-Cyclicals	17	19	4	15	1	56
Financials	19	5	1	0	25	50
Customer Cyclicals	18	19	3	1	1	42
Basic Materials	10	20	4	2	2	38
Properties & Real Estate	23	1	2	0	0	26
Infrastructure	6	11	2	15	0	34
Energy	7	7	23	4	0	41
Healthcare	1	15	0	0	0	16
Transportation & Logistic	7	1	1	1	0	10
Industrial	4	6	2	2	0	14
Technology	5	2	1	0	0	8
Total	117	106	43	40	29	335

(2) Cluster 2 contains 106 companies (31.64%), with substantial numbers from Basic Materials (20 companies, 52.63% of the sector), Customer Non-Cyclicals (19 companies, 33.93%), and Customer Cyclicals (19 companies, 45.24%). A significant observation is the concentration of Healthcare companies in this cluster, containing 15 of the 16 Healthcare companies (93.75%). This pattern points to frequent media coverage connecting healthcare topics with consumer and materials sectors.

(3) Cluster 3 has a unique composition of 43 companies (12.84%), marked by a high concentration in the Energy sector with 23 companies (56.10% of all energy companies). The presence of Basic Materials (4 companies) and Customer Non-Cyclicals (4 companies) in this cluster aligns with supply chain relationships and market interactions covered in news articles.

(4) Cluster 4 (40 companies, 11.94%) reveals strong representation from Infrastructure (15 companies, 44.12% of the sector) and Customer Non-Cyclicals (15 companies, 26.79%). This co-occurrence reflects media coverage connecting infrastructure development with consumer market dynamics.

(5) The smallest group identified is Cluster 5 (29 companies, 8.66%), which contains a notable concentration of Financial sector companies (25 companies, 50% of all financial companies). This concentration underscores a specific subset of financial institutions that share common news coverage patterns, likely due to their market roles and regulatory environments.

Based on the explanation, Clusters 1 and 2 emerge as the most diversified groups, containing balanced distributions across multiple sectors. These clusters jointly account for 66.57% of all companies, with substantial representation in each sector. In contrast, Clusters 3, 4, and 5 display more focused compositions—Cluster 3 centers on Energy (53.5% of cluster), Cluster 4 concentrates on Infrastructure and Customer Non-Cyclicals (75% of cluster), and Cluster 5 maintains the highest concentration with Financial sector companies (86.2% of cluster). The coexistence of both diversified and focused clusters suggests a dual nature in Indonesia's business media coverage, where some companies operate within broader market narratives. In contrast, others remain within sector-specific news contexts. This contrast between industry classifications and news-based clustering raises an important question about their respective relationships with market sentiment and stock price movements. In the following section, we examine how these different grouping approaches correlate with market behavior, exploring whether industry-based or news co-occurrence relationships better explain the transmission of market sentiments and stock price dynamics than at the individual company level.

4.2 Stock prices and market sentiments correlation

To establish a baseline for comparing different approaches in understanding market sentiment transmission, we first analyze correlations at the individual company level. This baseline analysis examines how stock prices correlate with market sentiments for each company independently before exploring whether industry-based or news co-occurrence groupings reveal stronger patterns of sentiment-price relationships. We calculate the correlation between individual company stock prices and market sentiments without considering their respective industry affiliation or cluster membership. Table 6 presents a matrix of correlation scores between these two variables, where scores range from -1 (perfect negative correlation) to 0 (no correlation) to 1 (perfect positive correlation). These individual-level correlations serve as a reference point for evaluating whether collective groupings—either through traditional industry classifications or news co-occurrence clusters—provide enhanced understanding of sentiment-price relationships.

Table 6 presents the correlation coefficients between stock prices and market sentiment for individual companies from 2019 to 2023. The correlation strength is visualized through color intensity, where darker green shades represent stronger positive correlations and darker red shades indicate stronger negative correlations, enabling quick identification of relationship patterns across companies and time periods. The correlation analysis between individual company stock prices and market sentiments reveals substantial variations that reflect real-world market dynamics. During the pre-pandemic period (2019), correlations show moderate diversity, with financial sector companies like BBCA (0.43) and BBRI (0.64) displaying positive correlations, indicating aligned movement between market sentiment and stock prices. The pandemic onset in 2020 led to notably stronger positive correlations across many companies, particularly in banking (BBNI: 0.88) and construction (PTPP: 0.91), suggesting heightened market sensitivity to the news during this period of uncertainty.

The correlation patterns evolved through 2021–2023, reflecting changing market conditions and investor behavior. The year 2022 marked a significant shift with generally weaker correlations across companies, exemplified by BBRI's decrease from 0.88 (2021) to 0.29 (2022). The decline might reflect market adaptation to post-pandemic conditions and reduced sensitivity to news sentiment. However, 2023 generally shows a resurgence in positive correlations for certain

Table 6 Correlation coefficients between stock prices and market sentiment for individual companies from 2019 to 2013

Year	BBCA	BBNI	JSMR	PTPP	ADHI	ANTM	BBRI	BMRI	PGAS	PTBA	PTSN	ASII	BNGA	BNLI	TLKM	ACES	ADRO
2019	0,43	0,31	0,17	0,71	-0,18	0,60	0,64	0,62	0,34	-0,28	0,45	0,21	0,82	0,67	0,54	0,57	-0,13
2020	0,33	0,88	0,32	0,91	0,53	0,59	0,75	0,76	0,67	0,25	0,69	0,57	0,51	0,90	0,64	0,27	0,34
2021	0,70	0,80	0,04	0,78	0,07	0,32	0,88	0,55	0,58	0,33	0,48	0,39	0,16	0,59	0,48	0,18	0,65
2022	0,28	0,35	-0,02	0,05	0,38	-0,12	0,29	0,74	-0,15	0,02	0,27	-0,16	0,60	0,32	0,35	0,27	0,22
2023	0,15	0,55	-0,21	0,92	0,70	0,25	0,81	0,75	0,23	0,70	-0,16	0,74	0,60	0,34	0,66	0,81	0,90

Year	BBTN	BDMN	INCO	ISAT	KRAS	MEGA	WIKA	CTRA	SMGR	WOOD	ICBP	INDF	MNCN	TRIM	UNVR	BNII	BRMS	BSDE
2019	0,29	0,84	0,57	0,76	0,42	-0,12	0,49	0,34	0,31	0,15	0,84	0,66	0,37	-0,04	-0,27	-0,30	0,19	0,65
2020	0,84	0,66	0,30	0,72	0,70	0,17	0,87	-0,17	0,64	0,79	0,26	0,36	0,51	-0,09	0,41	0,44	0,60	0,27
2021	0,75	0,64	0,46	0,32	0,32	0,52	0,04	0,36	0,35	-0,13	0,14	-0,01	0,41	0,18	0,67	0,43	0,78	0,34
2022	0,22	-0,39	0,46	-0,38	0,15	0,24	0,48	0,39	-0,66	-0,10	-0,18	-0,12	-0,15	0,25	-0,31	0,26	0,13	0,54
2023	0,42	0,36	0,17	0,31	-0,39	0,47	0,44	-0,07	0,65	0,92	0,63	0,16	0,46	0,09	0,57	0,25	0,83	0,49

sectors, particularly evident in commodity-related companies like WOOD (0.92) and construction companies like PTPP (0.92), possibly indicating renewed market responsiveness to sector-specific news in the economic recovery period.

The diverse correlations observed over the five-year span underscore the complex interaction between market sentiment and stock prices, heavily influenced by external economic conditions and sector-specific factors. Notable patterns include stronger positive correlations during the height of market uncertainty (2020–2021), followed by a period of weaker correlations (2022), and selective strong correlations in 2023. These patterns suggest that sentiment-price relationships are not static but evolve with market conditions, investor behavior, and sector-specific dynamics.

The observed inconsistency and high variability in individual company correlations, combined with their sensitivity to external market conditions, highlight the limitations of analyzing sentiment-price relationships at the individual company level. These limitations motivate our investigation of alternative approaches that consider companies' relationships with their peers. In subsequent sections, we examine whether grouping companies through industry affiliations or news co-occurrence clusters provides better insight into sentiment-price relationships. This refined analysis explores how shared characteristics and media coverage patterns might reveal market sentiment dynamics that manifest at broader market levels.

4.3 Industry-based stock prices and market sentiments correlation

Following our examination of individual company correlations, we extend our analysis to investigate sentiment-price relationships at the industry level. This sectoral analysis aims to uncover whether aggregating companies by industrial sectors reveals more systematic patterns in the relationship between market sentiment and stock price movements. Figure 7 presents the temporal evolution of correlations between industry-specific news sentiment and stock prices from 2019 to 2023.

The industry-level correlation analysis reveals fascinating patterns that mirror major economic events and market transformations during 2019–2023. The Financial sector tells a particularly compelling story of market evolution: starting with a near-perfect correlation (0.91) in 2019 when market sentiment strongly drove stock movements, maintaining this tight relationship through the pandemic years (0.84–0.85 in 2020–2021), before dramatically decoupling in 2023 (0.11). This sharp decline suggests a fundamental shift in how financial markets operate post-pandemic, where traditional sentiment-price relationships have given way to more complex market dynamics.

The Energy sector's correlation trajectory captures the sector's turbulent journey through global crises. From a minimal correlation (0.15) in 2019, it strengthened significantly during the pandemic peak (0.86 in 2021) as energy markets became highly sensitive to news. However, a striking shift occurred in 2023, where the correlation turned negative (−0.09), indicating that stock prices began moving in the opposite direction to market sentiment. This negative correlation suggests that positive news sentiment coincided with declining stock prices, or vice versa. This counterintuitive relationship might reflect a disconnect between immediate market reactions and longer-term industry fundamentals, where investors may be acting on factors beyond news sentiment, such as global energy transition policies, long-term demand forecasts, or strategic repositioning of energy companies despite short-term positive news.

Transportation & Logistics sector demonstrates a consistent strengthening in sentiment-price correlation, increasing from 0.54 in 2019 to 0.87 in 2023. This progressive increase in the positive correlation coefficient indicates enhanced synchronization between market sentiment and stock price movements. The strengthening relationship suggests that stock price variations increasingly align with sentiment changes in news coverage, where positive sentiment corresponds with price appreciation and negative sentiment with price decline. This heightened correlation reflects the sector's elevated strategic position in global trade, particularly following supply chain disruptions that emphasized logistics' critical role in economic stability. The strong positive correlation (0.87) in 2023 indicates that approximately 87% of stock price variations align with sentiment changes, demonstrating how market valuations have become highly responsive to sector-specific news regarding operational capacity, freight rates, and supply chain dynamics.

The Infrastructure and Basic Materials sectors exhibit contrasting sentiment-price correlation patterns. The infrastructure sector maintains relatively stable positive correlations ranging from 0.30 to 0.73, indicating consistent alignment between market sentiment and stock performance. This stability likely reflects the sector's defensive characteristics in Indonesia, supported by continued government infrastructure spending through programs like the National Strategic Projects (PSN) and infrastructure development initiatives in the new capital city (IKN) development. The sustained positive correlation suggests that news related to infrastructure development consistently influences stock valuations in predictable ways. Conversely, the Basic Materials sector shows more volatile correlation patterns, with correlation strength peaking during market uncertainty (0.75 in 2020) before inverting to a negative

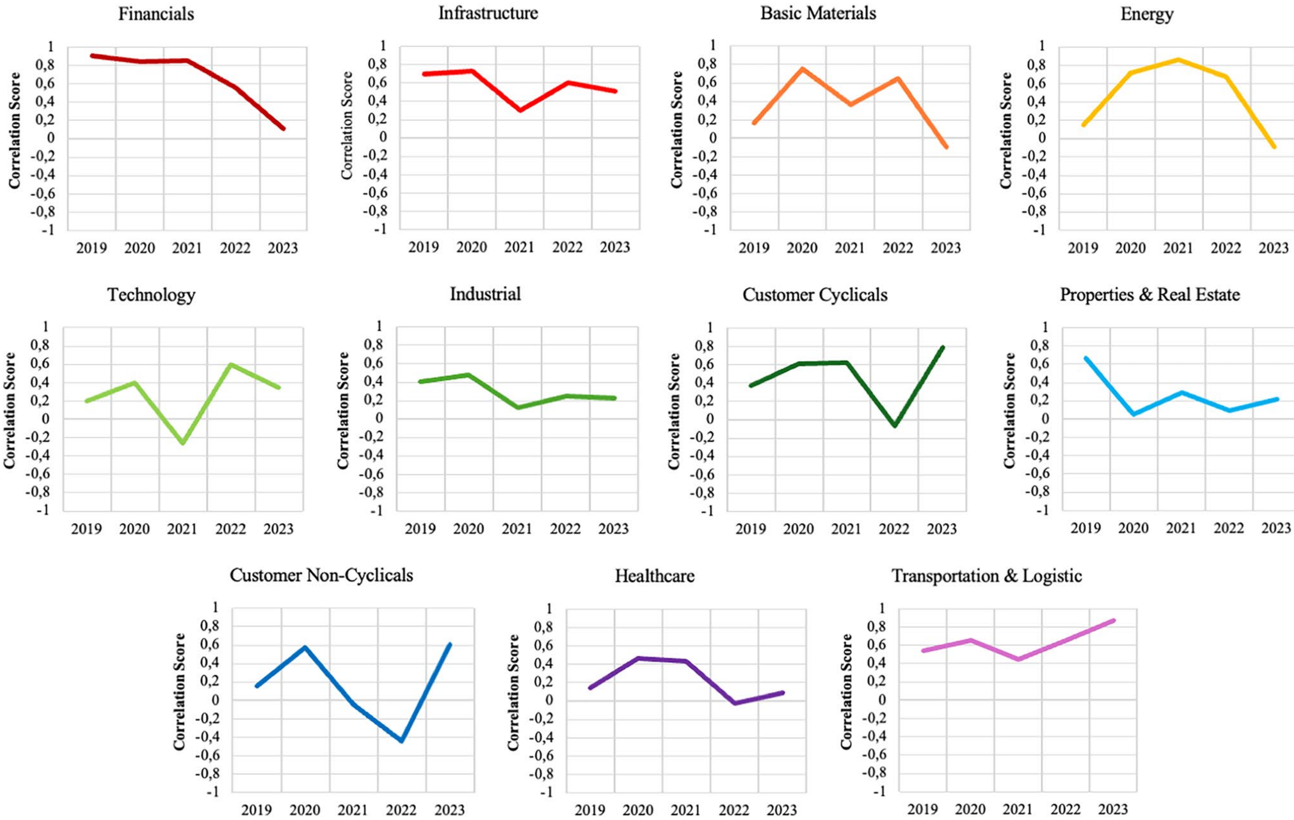


Fig. 7 Temporal evolution of industry-level correlations between stock prices and market sentiment (2019–2023)

correlation (–0.09) in 2023. This volatility reflects the sector's sensitivity to global supply chain disruptions and commodity price fluctuations.

Consumer sectors show distinct sentiment-price correlation patterns. Customer Non-Cyclicals, comprising essential consumer goods producers like food (ICBP, INDF), beverages, and household products, shows highly volatile correlations swinging from positive (0.57 in 2020) to negative (–0.44 in 2022). This volatility shows complex market dynamics where positive news might coincide with stock price declines, possibly reflecting how investors view these defensive stocks during economic uncertainty—often moving against market sentiment as investors seek stability regardless of news coverage. In contrast, Customer Cyclicals, which includes discretionary spending sectors such as retail (MAPI, LPPF), automotive, and consumer electronics, maintains more consistent positive correlations except for a brief negative period in 2022. This stability indicates a closer alignment between market sentiment and stock performance, which is logical for sectors dependent on consumer confidence and discretionary spending power. The positive correlation suggests that news about improving consumer confidence or retail spending directly translates to stock price appreciation, while negative news about declining purchasing power corresponds with stock price declines. This pattern reflects these companies' greater sensitivity to economic cycles and consumer sentiment, unlike their Non-Cyclical counterparts that provide essential goods with relatively stable demand regardless of economic conditions.

Healthcare sector exhibits an intriguing correlation pattern that aligns with its pivotal role during the COVID-19 pandemic. The correlation strengthened from 0.14 in 2019 to peak at 0.47 in 2020–2021 during the height of the pandemic, reflecting heightened market sensitivity to healthcare-related news. This period saw increased attention on healthcare companies like Kalbe Farma (KLBF) and Kimia Farma (KAEF), particularly regarding their involvement in vaccine distribution, medical supplies, and pharmaceutical products essential for pandemic response. However, the correlation declined significantly to -0.02 in 2022 and remained low (0.09) in 2023, suggesting a normalization phase where stock valuations became more influenced by fundamental business metrics and long-term growth strategies rather than pandemic-related news sentiment.

The Industrial sector, which encompasses manufacturing and heavy equipment companies, demonstrates a relatively stable but modest positive correlation. This stability reflects the sector's nature in Indonesia, where industrial companies often operate on long-term contracts and government projects, making them less susceptible to short-term sentiment swings. Companies like Astra International (ASII) and United Tractors (UNTR) exemplify this pattern, where their stock performance appears more tied to fundamental business metrics. The Technology sector's correlation volatility, shifting from moderate positive (0.40 in 2020) to negative (−0.26 in 2021) before recovering (0.34 in 2023), parallels the sector's dramatic journey—from pandemic-driven digital acceleration benefiting companies like Bukalapak (BUKA) and GoTo (GOTO), to post-pandemic concerns about profitability and business sustainability. Meanwhile, Real Estate sector shows declining correlation strength from a strong positive (0.67 in 2019) to consistently low correlations (0.05–0.29) in subsequent years, mirroring the sector's structural challenges.

We evaluated the correlation between monthly news sentiment across various sectors and corresponding stock price movements over a five-year period, as depicted in Fig. 8. These correlation coefficients, ranging from −1 to 1, measure the strength and direction of the relationship between sentiment and price movements, where higher positive values indicate a stronger alignment between positive sentiment and price increases.

The analysis shows considerable variation in correlation strengths across sectors during the 2019–2023 period. At the upper end, the Financial sector leads with the strongest positive correlation (0.62), indicating that approximately 62% of stock price movements align with market sentiment changes. For example, when news sentiment turns positive regarding banking policies, interest rates, or financial market conditions, stock prices in this sector tend to increase proportionally. Conversely, negative news about financial regulations or market stability typically corresponds with price declines. Infrastructure (0.60) and Energy (0.57) sectors follow closely, forming a group of highly sentiment-sensitive industries. These high correlations likely reflect these sectors' strategic importance in Indonesia's economy and their sensitivity to government policies, global market conditions, and macroeconomic news.

At the lower end, Customer Non-Cyclicals (0.09) and Properties & Real Estate (0.13) exhibit minimal correlation with market sentiment. These low correlations indicate that only about 9–13% of price movements align with sentiment changes. For Customer Non-Cyclicals, which includes essential consumer goods companies, this weak correlation shows stock valuations depend more on stable consumer demand and operational performance than market sentiment. Similarly, in the Properties & Real Estate sector, factors such as interest rates, property demand, and development project progress appear to influence stock prices more than general market sentiment. These sectors demonstrate resilience to sentiment fluctuations, potentially offering stability during periods of market volatility.

The cross-industry average correlation of 0.30 (shown as "All across") provides a benchmark for overall market sentiment sensitivity. This moderate positive correlation indicates that, on average across all sectors, there exists a consistent but modest relationship between market sentiment and stock price movements. The distribution of correlations around these average reveals important insights: six sectors (Financials, Infrastructure, Energy, Basic Materials, Transportation and Logistics, and Technology) exceed the market average, while five sectors fall below it. This variation suggests that sentiment analysis might be more valuable for investment decisions in sectors showing above-average correlations, while fundamental analysis might be more crucial for sectors with lower sentiment sensitivity. The fact that this average exceeds some individual sector correlations while falling well below others highlights how different industries respond uniquely to market sentiment, emphasizing the importance of sector-specific approaches in investment strategies.

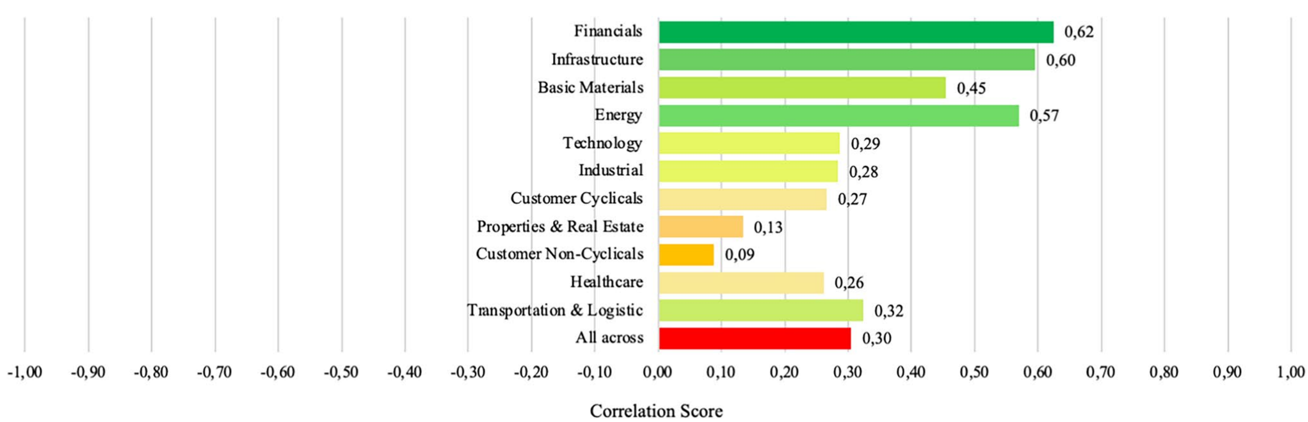


Fig. 8 Average correlation scores between market sentiment and stock price movements by industry (2019–2023)

4.4 Co-occurrence-based stock prices and market sentiments correlation

Following our industry-based analysis, we examine how companies grouped by their co-occurrence in news coverage respond collectively to market sentiment. This approach offers a different perspective from traditional sector-based analysis, focusing instead on companies that frequently appear together in news articles, potentially capturing business relationships and market dynamics that transcend conventional industry boundaries. As shown in Table 5, these co-occurrence clusters contain diverse combinations of companies from different sectors, suggesting complex inter-industry relationships in media coverage.

The correlation analysis, as shown in Fig. 9, reveals varying patterns across news co-occurrence clusters. Clusters 1, 2, and 4, which together comprise 78.5% of the sample (263 companies), maintain relatively stable positive correlations throughout the observation period, ranging from 0.4 to 0.8. Cluster 1 (117 companies) combines diverse sectors including Properties and Real Estate (23), Financials (19), and Consumer sectors (35). This co-occurrence pattern suggests intricate market relationships: financial institutions providing property development funding, real estate projects driving retail space development, and consumer behavior influencing both property demand and financial services. Similarly, Cluster 2 (106 companies) shows interconnected relationships between Basic Materials (20), Consumer sectors (38), and Healthcare (15), where raw material pricing affects consumer product costs, while healthcare services represent a stable demand sector that often appears in broader economic news coverage.

Cluster 5 is heavily concentrated in the Financial sector (25 of 29 companies—25 companies out of 50 companies in the Financial sector), displays a steady decline in correlation strength from 0.70 in 2019 to 0.20 in 2023. This decline is particularly interesting when compared to the Financial sector's industry-level correlation of 0.62 (Fig. 8), suggesting that 50% of Financial companies frequently co-mentioned in the news might experience different sentiment dynamics than the broader financial sector. This could reflect the cluster's specific composition of major institutions that often appear together in specific news.

The most distinctive pattern appears in Cluster 3, dominated by Energy companies (23 of 43 companies), which shows a dramatic shift from strong positive correlation (0.80 in 2021) to negative territory (−0.20 in 2023). This pattern extends beyond the Energy sector's individual correlation trend seen in Fig. 8 (0.57), suggesting that when energy companies are frequently co-mentioned with their supply chain partners (Basic Materials) and major customers (Customer Non-Cyclicals), their collective response to market sentiment becomes more pronounced.

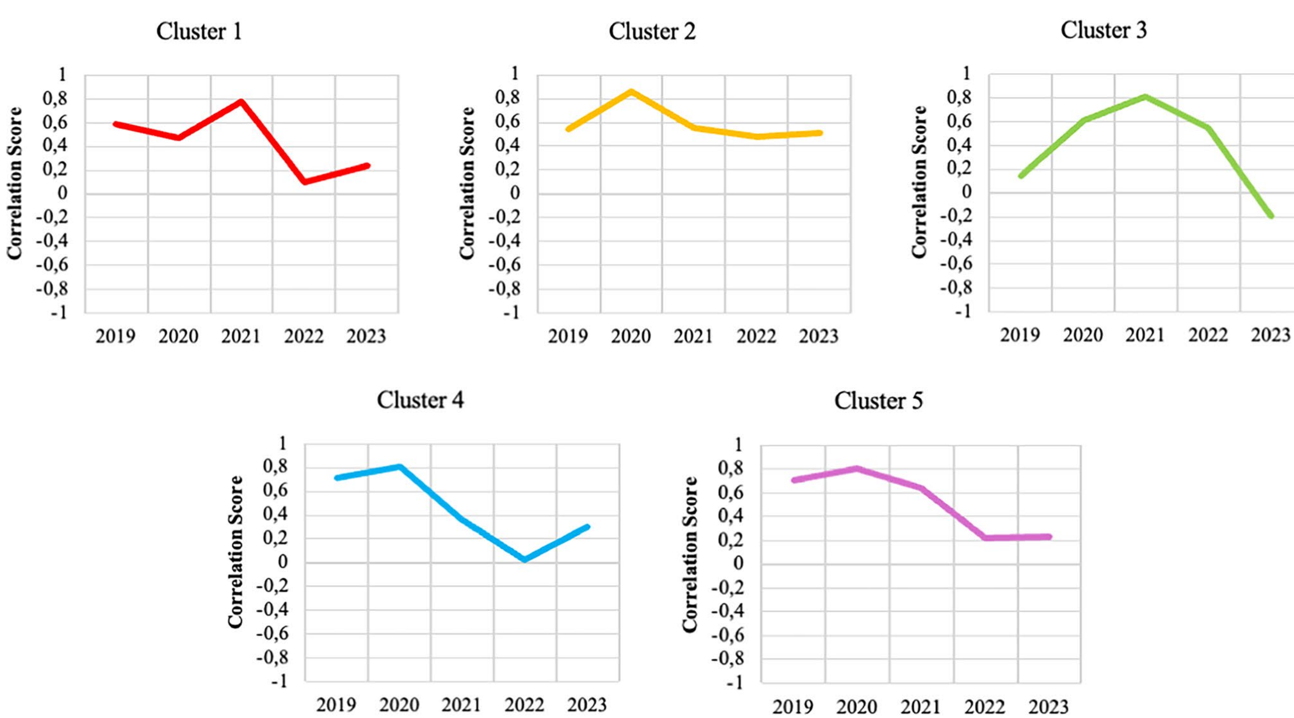


Fig. 9 Temporal evolution of sentiment-price correlations across news co-occurrence clusters (2019–2023)

Shifting our focus to a broader analytical perspective, we assessed the correlation between market sentiment and stock prices for each modularity cluster over a five-year span, as depicted in Fig. 10. The analysis reveals a hierarchy of sentiment sensitivity across clusters, with correlations ranging from moderate to strong positive. Clusters 5 and 4 show the strongest correlations (0.59 and 0.58 respectively), despite their different compositions. Cluster 5's high correlation, driven by its concentration of financial institutions (25 of 29 companies), suggests that companies frequently co-mentioned in financial news share strong sentiment-price relationships. Cluster 4's similar correlation strength, with its mix of Infrastructure (15 companies) and Customer Non-Cyclicals (15 companies), indicates that co-coverage of infrastructure development and consumer markets creates significant sentiment alignment.

A comparative analysis between sectoral and cluster-based correlations reveals an interesting pattern in sentiment-price relationships. While sectoral correlations show wide dispersion (ranging from 0.09 to 0.62), cluster-based correlations demonstrate a more uniform distribution (ranging from 0.36 to 0.59). The overall cross-cluster average correlation (shown as "All across") of 0.43 provides an important benchmark for understanding sentiment-price relationships in co-occurrence networks. This moderate positive correlation exceeds the individual company-level correlation (0.30) and is comparable to the industry-level average, suggesting that news co-occurrence relationships capture sentiment transmission mechanisms at least as effectively as traditional industry groupings. The fact that this average sits comfortably within the range of individual cluster correlations (0.36–0.59) indicates that co-occurrence-based relationships consistently influence how companies respond to market sentiment, regardless of their specific industry. These findings suggest that companies frequently appearing together in news coverage share stronger sentiment-price relationships than those grouped by traditional industry classifications, highlighting the importance of media-based relationships in understanding market dynamics.

4.5 Discussion

This study investigates the complex relationships between news sentiment and stock prices across multiple analytical levels—individual companies, industries, and news co-occurrence clusters—in Indonesia's capital market. By employing advanced AI-driven text analytics, specifically IndoBERT for sentiment classification and Louvain Modularity for clustering, our analysis reveals critical insights about sentiment transmission mechanisms and their influence on market dynamics, showcasing how artificial intelligence can enrich both academic research and practical applications.

1. **Sentiment-price correlation hierarchy**

The results reveal a hierarchical strengthening of correlations from individual companies (0.26) to industries (0.30) and co-occurrence clusters (0.43). This progressive strengthening suggests that AI-enhanced, collective representations in news media amplify sentiment influence through the synergistic effects of interconnected companies' co-mentions. At the individual level, company-specific factors like management changes or business events often overshadow broader sentiment effects. However, when analyzed collectively through AI-based industry or co-occurrence groupings, shared market contexts create more consistent sentiment-price relationships.

2. **Industry-based dynamics**

The study uncovers significant sectoral variations in sentiment-price relationships. The Financial and Energy sectors show strong correlations, reflecting their strategic roles in Indonesia's economy and sensitivity to policy changes.

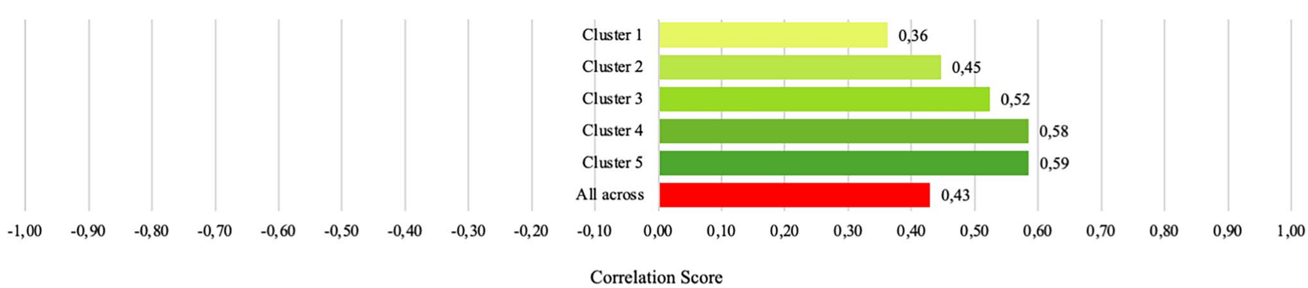


Fig. 10 Average correlation scores between market sentiment and stock price movements by co-occurrence cluster (2019–2023)

Infrastructure sector demonstrates similarly strong correlation, indicating its responsiveness to government initiatives and economic indicators. In contrast, Consumer Non-Cyclicals and Real Estate exhibit weaker relationships, suggesting these sectors' valuations depend more on fundamental factors like consumer demand and property market conditions than market sentiment. These sectoral disparities highlight how different industries process and respond to market sentiment based on their unique characteristics and market roles.

3. **Insights from news co-occurrence clusters**

Our AI-enabled news co-occurrence analysis reveals relationships beyond traditional industry boundaries, capturing intricate market dynamics such as the interplay between financial institutions, real estate developments, and consumer markets. Notably, co-occurrence clusters demonstrate more uniform correlation patterns compared to industry-based groupings, suggesting that media co-coverage patterns better reflect actual market relationships. For instance, Cluster 2's balanced composition of Basic Materials, Consumer sectors, and Healthcare maintains stable positive correlations, showing how AI-supported news coverage captures supply chain relationships and market interdependencies.

4. **Temporal Evolution and Market Sensitivity**

The temporal analysis identifies distinct shifts in sentiment-price correlations, particularly during the pandemic. While early pandemic years showed heightened correlations across sectors, the post-pandemic period revealed selective patterns. Energy sector shows shifts from positive to negative correlation, showing changing market dynamics and policy environments. Financial sector correlations declined from 2019 to 2023, indicating changing relationships between AI-derived financial news sentiment and actual market behavior. These temporal patterns reveal how external events and changing market conditions influence sentiment-price relationships.

5. **Methodological contributions**

The integration of IndoBERT (a transformer-based AI approach) for sentiment analysis, Louvain Modularity for clustering, and correlation evaluations provides methodological rigor. IndoBERT's application to Indonesian financial news achieved 94.34% accuracy in sentiment classification, while Louvain Modularity effectively identified meaningful company clusters with a modularity score of 0.249. These AI-driven techniques enabled granular analysis of Indonesia's market dynamics, demonstrating the value of data-centric and intelligence-driven in understanding sentiment transmission through business networks.

These findings contribute to both academic understanding and practical market analysis by showing how AI-empowered sentiment effects strengthen through business relationship networks rather than traditional industry boundaries. Investors and analysts might benefit from considering AI-identified news co-occurrence patterns alongside conventional sector-based analysis, particularly given the more consistent sentiment-price relationships observed in co-occurrence clusters. The research highlights the importance of considering temporal changes in sentiment relationships, as demonstrated by the varying patterns observed during and after the pandemic period.

4.5.1 Statistical validation of correlation differences

Additionally, to statistically validate the observed differences in correlation strength across the three analytical levels (individual, industry, and co-occurrence cluster), we conducted a Fisher r-to-z transformation test. The observed Pearson correlation coefficients were $r = 0.26$ at the individual level, $r = 0.30$ at the industry level, and $r = 0.43$ at the co-occurrence cluster level.

The Fisher r-to-z test results indicate that the difference between the individual and industry levels yields a $Z = -0.649$ ($p = 0.516$), suggesting that the increase in correlation strength from the individual to industry level is not statistically significant. In contrast, the difference between the industry and cluster levels yields a $Z = -2.248$ ($p = 0.025$), indicating a statistically significant difference at the 0.05 level.

This finding suggests that while aggregation at the industry level only modestly increases correlation strength, aggregation at the co-occurrence cluster level significantly enhances the explanatory power of sentiment-price relationships. Specifically, clustering companies based on news co-occurrence captures broader systemic sentiment signals more effectively than sectoral or individual-level analysis alone.

Furthermore, the results imply that relying solely on conventional industrial classifications may not efficiently capture sentiment-driven stock price dynamics. Since sectoral grouping yields only marginal improvement, whereas co-occurrence clustering reveals a statistically significant enhancement, it highlights the necessity for more dynamic, data-driven grouping methods beyond static industry categories. This observation underscores the value of considering real-time sentiment linkages among companies to better understand and predict market behavior.

4.5.2 Sectoral sentiment sensitivity and structural explanation

The heightened sentiment sensitivity observed in the financial and energy sectors can be attributed to several structural and market-specific factors. The financial sector, which includes banks, securities firms, and other financial institutions, is inherently sensitive to public sentiment due to its pivotal role in the economy and its exposure to market expectations, regulatory policies, and investor confidence. Negative or uncertain news in this sector often triggers immediate market reactions, affecting the financial companies themselves and broader market conditions. Similarly, the energy sector is highly susceptible to external factors such as global commodity price fluctuations, government regulations, and geopolitical dynamics. News related to energy policies, fuel prices, and environmental issues tends to influence market sentiment and stock valuations in this sector significantly. These structural characteristics may explain why news sentiment has a more pronounced impact on stock prices in the financial and energy sectors compared to other industries.

4.5.3 Connection to existing literature

These findings contribute to both academic understanding and practical market analysis by showing how AI-empowered sentiment effects strengthen through business relationship networks rather than traditional industry boundaries. This is consistent with more recent studies that have highlighted the role of investor sentiment in influencing stock volatility and market behavior [36, 57]. In particular, our results regarding the financial and energy sectors corroborate findings on the heightened sensitivity of these sectors to sentiment-driven and policy-related news [40, 81]. Furthermore, by introducing a network-based co-occurrence analysis, our study extends the existing literature beyond industry-level classifications, suggesting that media co-coverage patterns provide additional explanatory power in capturing sentiment transmission dynamics. Investors and analysts might benefit from considering AI-identified news co-occurrence patterns alongside conventional sector-based analysis, particularly given the more consistent sentiment-price relationships observed in co-occurrence clusters. The research also highlights the importance of considering temporal changes in sentiment relationships, as demonstrated by the varying patterns observed during and after the pandemic period.

5 Conclusions

This study provides a comprehensive analysis of the relationships between news sentiment and stock prices in Indonesian companies by employing a multi-layered artificial intelligence framework. The core empirical findings can be summarized as follows:

1. Hierarchical Sentiment-Price Relationship:
The results reveal a hierarchical strengthening of sentiment-price correlations across three analytical levels—individual companies (0.26), industry sectors (0.30), and co-occurrence clusters (0.43). This progressive pattern demonstrates that sentiment effects become more pronounced when companies are grouped based on their industry affiliation or co-mention patterns in news media.
2. Sector-Specific Sensitivity:
Sectoral analysis shows that the Financial and Energy sectors exhibit higher sentiment sensitivity, reflecting their strategic roles in Indonesia's economy and their responsiveness to policy changes and market news.
3. Market Dynamics Beyond Industry Boundaries:
The co-occurrence cluster analysis identifies inter-company relationships that transcend formal industry classifications, revealing latent market dynamics captured by media coverage patterns.

These findings contribute to the existing literature by advancing the application of AI-based techniques in financial market analysis. Unlike prior studies that typically analyze sentiment effects at the individual or sector level, this research introduces a multi-layered analytical framework combining IndoBERT-based sentiment analysis and Louvain Modularity clustering. This approach demonstrates how AI-enhanced methods can uncover complex sentiment transmission mechanisms embedded in financial news data, particularly in emerging markets.

In addition to its academic contribution, this study provides actionable recommendations for market participants and policymakers. Investors and analysts can integrate real-time sentiment indicators derived from IndoBERT-based analysis into their decision-support systems to enhance risk assessment and portfolio management. Regulatory bodies can leverage the co-occurrence cluster analysis to improve market surveillance and develop early warning systems that detect sentiment-driven systemic risks. These applications highlight the potential of AI-driven analytics to enhance market transparency and responsiveness.

However, this research acknowledges certain limitations. The analysis is limited to main board companies in the Indonesian capital market and relies solely on news articles from CNBC Indonesia, potentially excluding broader market influences. Moreover, the monthly aggregation of sentiment and stock price data may overlook finer temporal variations. Future research could expand the dataset to include smaller companies and global news sources, apply real-time sentiment monitoring, or utilize next-generation transformer-based models to explore causal relationships between news sentiment and stock price dynamics.

Author contributions Conceptualization, A.A., D.P.R., and M.A.A.S.; methodology: D.P.R., and M.A.A.S.; data curation: A.A., D.P.R., and M.A.A.S.; writing—original draft preparation: A.A. and D.P.R.; writing—review and editing: A.A., F.T.K., A.H.N., M.S.M., and R.S.; resources: F.T.K. and S.W.; project administration: S.W.; supervision: F.T.K., A.A., and S.W. All authors have read and agreed to the published version of the manuscript. Andry Alamsyah (A.A.), Dian Puteri Ramadhani(D.P.R.), Farida Titik Kristanti (F.T.K.), Arbi Haza Nasution (A.H.N.), Mohd Sham bin Mohamad (M.S.M.), Rajalingam Sokkalingam (R.S.), Sri Widiyanesti (S.W.), Muhammad Apriandito Arya Saputra (M.A.A.S.)

Funding This research was funded by the International Research Collaboration Program between Telkom University, Universiti Teknologi Petronas, Universiti Malaysia Pahang, and Universitas Islam Riau, 2022–2024. Under contract number KWR4.072/PNLT3/PPM-LIT/2022.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate Not Applicable.

Consent for publication Not Applicable.

Competing interests The authors declare no competing interests.

References

1. Sharafoddin S, Emsia E. The effect of stock valuation on the company's management. Proc Econ Finance. 2016;36:128–36.
2. Ali K, Chisti KA, Malik IA. Impact of earnings per share on stock prices and price to earnings ratio. Econ Bus Q Rev. 2022;5(2):147–52. https://doi.org/10.31014/aior.1992.05.02.421.
3. Puspitaningtyas Z. Is financial performance reflected in stock prices?. In: Proceedings of the 2nd international conference on accounting, management, and economics (ICAME 2017), advances in economics, business and management research, vol. 40, pp. 17–28, 2017. https://doi.org/10.2991/icame-17.2017.2
4. Liu Q, Lee W-S, Huang M, Wu Q. Synergy between stock prices and investor sentiment in social media. Borsa Istanbul Rev. 2023;23(1):76–92.
5. Nasarudin IY, Suhendra S, Anggraini LF. The determinant of stock prices: evidence on food and beverage companies in Indonesia. Etikonomi. 2019;18(1):143–54.

6. Gallegati M, Gallegati S. Why does economics need complexity? Soft Comput. 2025. https://doi.org/10.1007/s00500-025-10548-5.
7. Johnson NF, Jefferies P, Hui PM. Financial market complexity: what physics can tell us about market behaviour. Oxford Finance Series: Oxford University Press; 2003.
8. International Monetary Fund. Global financial stability report: safeguarding financial stability amid high inflation and geopolitical risks, 2023. https://www.imf.org/en/Publications/GFSR/Issues/2023/04/11/global-financial-stability-report-april-2023. Accessed 25 Apr 2025.
9. Financial Stability Board. Guidance to Assess the Systemic Importance of Financial Institutions, Markets and Instruments: Initial Considerations, Bank for International Settlements, 2009.
10. Humpe A, McMillan DG, Schöttl A. Macroeconomic determinants of the stock market: a comparative study of anglosphere and BRICS. Finance Res Lett. 2025;75:106869. https://doi.org/10.1016/j.frl.2025.106869.
11. Dahal KR, Pokhrel NR, Gaire S, Mahatara S, Joshi RP, Gupta A, et al. A comparative study on effect of news sentiment on stock price prediction with deep learning architecture. PLoS ONE. 2023. https://doi.org/10.1371/journal.pone.0284695.
12. Blau BM, Griffith TG, Whitby RJ. Industry regulation and the comovement of stock returns. J Empir Financ. 2023;73:206–19. https://doi.org/10.1016/j.jempfin.2023.06.005.
13. Ahmed S, Bu Z, Tsvetanov D. Best of the best: a comparison of factor models. J Financ Quant Anal. 2019;54(4):1713–58. https://doi.org/10.1017/S0022109018000947.
14. Massaad A, Moawad R, Fares ON, Vairungroj S. Refining and robust backtesting of a century of profitable industry trends. arXiv preprint arXiv:2412.14361, 2024. https://arxiv.org/abs/2412.14361. Accessed 25 Apr 2025.
15. Jing N, Wu Z, Wang H. A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. Expert Syst Appl. 2021;178: 115019. https://doi.org/10.1016/j.eswa.2021.115019.
16. Curme C, Daisy Zhuo Y, Susannah Moat H, Preis T. Quantifying the diversity of news around stock market moves. J Netw Theory Finance. 2017;3(1):1–20.
17. Nandwani P, Verma R. A review on sentiment analysis and emotion detection from text,". Soc Netw Anal Min. 2021. https://doi.org/10.1007/s13278-021-00776-6.
18. Hasan MM, Popp J, Olah J. Current landscape and influence of big data on finance. J Big Data. 2020. https://doi.org/10.1186/s40537-020-00291-z.
19. Yadegaridehkordi E, Nilashi M, Shuib L, Nasir MHNBM, Asadi S, Samad S, Awang NF. The impact of big data on firm performance in hotel industry. Electron Commer Res Appl. 2020;40:100921. https://doi.org/10.1016/j.elerap.2019.100921.
20. Favaretto M, Clercq ED, Schneble CO, Elger BS. What is your definition of Big Data? Researchers' understanding of the phenomenon of the decade. PLoS ONE. 2020. https://doi.org/10.1371/journal.pone.0228987.
21. Al-Sai Z, Husin M, Syed-Mohamad S, Abdin R, Damer N, Abualigah L, Gandomi A. Explore big data analytics applications and opportunities: a review. Big Data Cogn Comput. 2022;6:157.
22. Cesario E. Big data analytics and smart cities: applications, challenges, and opportunities. Front Big Data. 2023;6:1149402. https://doi.org/10.3389/fdata.2023.1149402.
23. Lutfi A, Alrawad M, Alsyouf A, Almaiah M, Al-Khasawneh A, Al-Khasawneh A, et al. Drivers and impact of big data analytic adoption in the retail industry: a quantitative investigation applying structural equation modeling. J Retail Consum Serv. 2023;70:103129.
24. IBM, "Big Data Analytics," IBM, https://www.ibm.com/analytics/big-data-analytics. Accessed 20 Sept 2023.
25. Badshah A, Daud A, Alharbey R, Banjar A, Bukhari A, Alshemaimri B. Big data applications: overview, challenges and future. Artif Intell Rev. 2024;57(11):290. https://doi.org/10.1007/s10462-024-10938-5.
26. Zhou Z, Liu D, Chen Z, Pancho M. Government adoption of generative artificial intelligence and ambidextrous innovation. Int Rev Econ Finance. 2025;98:103953. https://doi.org/10.1016/j.iref.2025.103953.
27. Batko K, Ślęzak A. The use of big data analytics in healthcare. J Big Data. 2022. https://doi.org/10.1186/s40537-021-00553-4.
28. Li Y, Wang S, Ding H, Chen H. Large language models in finance: a survey," arXiv preprint arXiv:2311.10723, 2023. https://arxiv.org/abs/2311.10723.
29. Sun T, Vasarhelyi MA. Embracing textual data analytics in auditing with deep learning. Int J Digit Account Res. 2018;18:49–67.
30. Farimani SA, Jahan MV, Fard AM. From text representation to financial market prediction: a literature review. Information. 2022;13(10):466.
31. Ashtiani MN, Raahemi B. News-based intelligent prediction of financial markets using text mining and machine learning: a systematic literature review. Expert Syst Appl. 2023. https://doi.org/10.1016/j.eswa.2023.119509.
32. Jiang Y. The transition and countermeasures of traditional media after being impacted by the advent of digital media. In: Advances in social science, education and humanities research, vol. 664, 2022.
33. Yang M, Jiang B, Wang Y, Hao T, Liu Y. News text mining-based business sentiment analysis and its significance in economy. Front Psychol. 2022. https://doi.org/10.3389/fpsyg.2022.918447.
34. Provoost S, Ruwaard J, van Breda W, Riper H, Bosse T. Validating automated sentiment analysis of online cognitive behavioral therapy patient texts: an exploratory study. Front Psychol. 2019. https://doi.org/10.3389/fpsyg.2019.01065.
35. Khan MA, Khan MA, Khan MA. News-based intelligent prediction of financial markets using text mining and machine learning: a systematic literature review. Expert Syst Appl. 2023;217:119563. https://doi.org/10.1016/j.eswa.2022.119563.
36. Cicekyurt E, Bakal G. Enhancing sentiment analysis in stock market tweets through BERT-based knowledge transfer. Comput Econ. 2025. https://doi.org/10.1007/s10614-025-10901-8.
37. Joshi A, Prabhu A, Shrivastava M, Varma V. Towards sub-word level compositions for sentiment analysis of hindi-english code mixed text. In: Proceedings of the 26th international conference on computational linguistics: technical Papers, Osaka, Japan, pp. 2482–2491, 2016.
38. Wang H, Lu S, Zhao J. Aggregating multiple types of complex data in stock market prediction: a model-independent framework. Knowl-Based Syst. 2019;164:193–204.
39. Zhang X, Zhang Y, Wang S, Yao Y, Fang B, Philip S. Improving stock market prediction via heterogeneous information fusion. Knowl Based Syst. 2018;143:236–47.
40. Maqbool J, Aggarwal P, Kaur R, Mittal A, Ganaie IA. Stock prediction by integrating sentiment scores of financial news and MLP-Regressor: a machine learning approach. Procedia Comput Sci. 2023;218:1067–78. https://doi.org/10.1016/j.procs.2023.01.086.

41. Wan X, Yang J, Marinov S, Calliess J-P, Zohren S, Dong X. Sentiment correlation in financial news networks and associated market movements. Sci Rep. 2021. https://doi.org/10.1038/s41598-021-82338-6.
42. Suhadak S, Kurniaty K, Handayani SR, Rahayu SM. Stock return and financial performance as moderation variable in influence of good corporate governance towards corporate value. Asian J Account Res. 2018. https://doi.org/10.1108/AJAR-07-2018-0021.
43. Alamsyah A, Pambudi VS, Ramadhani DP. Sentiment correlation in news network and associated market movements based on co-occurrence network. In: Proceedings of the 2022 international conference on information technology research and innovation (ICITRI), Jakarta, Indonesia, 2022, pp. 71–76. https://doi.org/10.1109/ICITRI56423.2022.9970239.
44. Wei J, Xu Q, He C. Deep learning of predicting closing price through historical adjustment closing price. Procedia Comput Sci. 2022;202:379–84.
45. Phuoc T, Anh PTK, Tam PH, Nguyen CV. Applying machine learning algorithms to predict the stock price trend in the stock market: the case of Vietnam. Hum Soc Sci Commun. 2024;11(1):393. https://doi.org/10.1057/s41599-024-02807-x.
46. Vijh M, Chandola D, Tikkiwal VA, Kumar A. Stock closing price prediction using machine learning techniques. Procedia Comput Sci. 2020;167:599–606.
47. Waskito M, Faizah S. Pengaruh return on equity dan pertumbuhan penjualan terhadap harga saham perusahaan. Jurnal Akuntansi dan Keuangan. 2021;3(1):30–43.
48. Zhao C, Hu P, Liu X, Lan X, Zhang H. Stock market analysis using time series relational models for stock price prediction. Mathematics. 2023. https://doi.org/10.3390/math11051130.
49. Kumari J, Sharma V, Chauhan S. Prediction of stock price using machine learning techniques: a survey. In: 3rd international conference on advances in computing, communication control and networking (ICAC3N), Greater Noida, 2021.
50. Grandhmal DP, Kumar K. Systematic analysis and review of stock market prediction techniques. Comput Sci Rev. 2019. https://doi.org/10.1016/j.cosrev.2019.08.001.
51. Guo Y. Stock price prediction using machine learning. Economics Spring; 2022.
52. Awasthi I, Pushpalatha M. Prediction of stock market behavior using financial news and sentiment analysis. Eur Chem Bull. 2023;12(4):9439–50.
53. Fazlija B, Harder P. Using financial news sentiment for stock price direction. Mathematics. 2022. https://doi.org/10.3390/math10132156.
54. Alamsyah A, Ayu SP, Rikumahu B. Exploring relationship between headline news sentiment and stock return. In: 2019 7th international conference on information and communication technology (ICoICT), Kuala Lumpur, Malaysia, 2019, pp. 1–6, https://doi.org/10.1109/ICoICT.2019.8835298.
55. Wibowo A, Darmanto S. Empirical test of the capital asset pricing model (CAPM): evidence from Indonesia capital market. SSRG Int J Econ Manag Stud. 2020;7(5):172–7. https://doi.org/10.14445/23939125/IJEMS-V7I5P126.
56. Ranco G, Aleksovski D, Caldarelli G, Grčar M, Mozetič I. The effects of twitter sentiment on stock price returns. PLoS ONE. 2015;10(9): e0138441. https://doi.org/10.1371/journal.pone.0138441.
57. Gao Y, Zhao C, Sun B, Zhao W. Effects of investor sentiment on stock volatility: new evidences from multi-source data in China's green stock markets. Financial Innov. 2022. https://doi.org/10.1186/s40854-022-00381-2.
58. Liu J, Leu J, Holst S. Stock price movement prediction based on Stocktwits investor sentiment using FinBERT and ensemble SVM. PeerJ Comput Sci. 2023;9:e1403. https://doi.org/10.7717/peerj-cs.1403.
59. Indonesia Stock Exchange, IDX Industrial Classification Perusahaan Tercatat di Bursa Efek Indonesia. https://gopublic.idx.co.id/media/1401/daftar-sektor_web-go-public_id.pdf. Accessed 15 Nov 2023.
60. Fonseka MM, Rajapakse RLTN, Tian G-L. Competitors' stock price reactions in response to private equity placements: evidence from a transitional economy. Econom Res-Ekonomska Istraživanja. 2018. https://doi.org/10.1080/1331677X.2018.1429293.
61. Ramadhani DP, Alamsyah A, Febrianta MY, Fajriananda MN, Nada MS, Hasanah F. Large-scale cross-cultural tourism analytics: Integrating transformer-based text mining and network analysis. Computers. 2025;14(1):27. https://doi.org/10.3390/computers14010027.
62. Battiston S, Farmer JD, Flache A, Garlaschelli D, Haldane AG, Heesterbeek H, Hommes C, Jaeger C, May R, Scheffer M. Complexity theory and financial regulation. Science. 2016;351(6275):818–9.
63. Almog A, Shmueli E. Structural entropy: monitoring correlation-based networks over time with application to financial markets. Sci Rep. 2019;9:1–13.
64. Carvalho L, Araújo T. The dynamics of exchange traded funds: a geometrical and topological approach. Appl Netw Sci. 2024. https://doi.org/10.1007/s41109-024-00674-8.
65. Alamsyah A, Ramadhani DP, Kristanti FT, Khairunnisa K. Transaction network structural shift under crisis: macro and micro perspectives. Economies. 2022;10(3):56. https://doi.org/10.3390/economies10030056.
66. Alamsyah A, Ramadhani DP, Kristanti FT, Khairunnisa K. Transaction network structural shift under crisis: macro and micro perspectives. Economies. 2022. https://doi.org/10.3390/economies10030056.
67. Starnini M, Boguñá M, Serrano M. The interconnected wealth of nations: shock propagation on global trade-investment multiplex networks. Sci Rep. 2019;9(1):1–10.
68. Barja A, Martinez A, Arenas A, Fleurquin P, Nin J, Ramasco JJ, Tomas E. Assessing the risk of default propagation in interconnected sectoral financial networks. EPJ Data Sci. 2019. https://doi.org/10.1140/epjds/s13688-019-0211-y.
69. Khurana D, Koli A, Khatter K, Singh S. Natural language processing: state of the art, current trends and challenges. Multimed Tools Appl. 2022;82:3713–44.
70. Yang X, Liu Z, Li J, Xie Q. Communities of co-occurrence network of financial firms in news. Inf Technol Quant Manag. 2023;221:821–5.
71. Barabási A-L. Network science. Cambridge: Cambridge University Press; 2016.
72. Peng N, Dredze M. Improving named entity recognition for chinese social media with word segmentation representation learning, In: The 54th annual meeting of the association for computational linguistics (Volume 2: Short Papers), Berlin, 2016.
73. Alves D, Thakkar G, Tadić M. UNER: universal named-entity recognition framework. In: Marenzi I, Gottschalk S, Müller-Budack E, Tadić M, Winters J, editors. Event analytics across languages and communities. Cham: Springer; 2025. p. 3–15. https://doi.org/10.1007/978-3-031-64451-1_1.

74. Cao P, Chen Y, Liu K, Zhao J, Liu S. Adversarial transfer learning for chinese named entity recognition with self-attention mechanism, In: The 2018 conference on empirical methods in natural language processing, Brussels, 2018.
75. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding, arXiv, 2019.
76. Wilie B, Vincentio K, Winata GI, Cahyawijaya S, Li X, Lim ZY, Soleman S, Mahendra R, Fung P, Bahar S, Purwarianti A. IndoNLU: benchmark and resources for evaluating Indonesian natural language understanding. In: The 1st conference of the Asia-Pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing, 2020.
77. Syaifudin Y, Nurwidyantoro A. Quotations identification from Indonesian online news using rule-based method. In: The 2016 international seminar on intelligent technology and its applications (ISITIA), 2016.
78. Khairunnisa SO, Imankulova A, Komachi M. Towards a standardized dataset on Indonesian named entity recognition, In: 1st conference of the Asia-Pacific chapter of the association for computational linguistics and the 10th international joint conference on natural language processing: student research workshop, Tokyo, 2020.
79. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Statist Mech Theory Exper. 2008. https://doi.org/10.1088/1742-5468/2008/10/P10008.
80. Vidanagama D, Silva A, Karunananda A. Ontology based sentiment analysis for fake review detection. Expert Syst Appl. 2022. https://doi.org/10.1016/j.eswa.2022.117869.
81. Zhao Z, Liu W, Wang K. Research on sentiment analysis method of opinion mining based on multi-model fusion transfer learning. J Big Data. 2023. https://doi.org/10.1186/s40537-023-00837-x.
82. Ibrahim NF, Wang X. A text analytics approach for online retailing service improvement: evidence from Twitter. Decis Support Syst. 2019;121:37–50.
83. Mao Y, Liu Q, Zhang Y. Enhancing implicit sentiment analysis via knowledge enhancement and context information. Compl Intell Syst. 2025. https://doi.org/10.1007/s40747-025-01840-w.
84. Topal MO, Bas A, Heerden IV, Exploring transformers in natural language generation: GPT, BERT, and XLNet. In: International conference on interdisciplinary applications of artificial intelligence (ICIDAAI), 2021.
85. Nugroho KS, Sukmadewa AY, Wuswilahaken HD, Bachtiar FA, Yudistira N. BERT fine-tuning for sentiment analysis on Indonesian mobile apps reviews. In: 6th international conference on sustainable information engineering and technology, Malang, 2021.
86. Wongkar M, Angdresey A. Sentiment analysis using naive Bayes algorithm of the data crawler: Twitter. In: Fourth international conference on informatics and computing (ICIC), Semarang, 2019.
87. Alamsyah A, Jannah MNZ, Ramadhani DP, Uncovering customer issues in e-commerce: sentiment analysis and topic modeling approach, In: International conference on information and communications technology (6th ICOIACT), Yogyakarta, 2023.
88. Jayadianti H, Kaswidjanti W, Utomo AT, Saifullah S, Dwiyanto FA, Drezewski R. Sentiment analysis of Indonesian reviews using fine-tuning IndoBERT and R-CNN. ILKOM Jurnal Ilmiah. 2022. https://doi.org/10.33096/ilkom.v14i3.1505.348-354.