JIESBI: Journal of Islamic Economics and Business Ethics
ISSN: 3047-311X (e) & 3064-4585 (p)
Vol. 3, No. 2, 2026, pp. 124-143
Doi: 10.24235/jiesbiv1i3

Machine Learning-Based Sentiment Analysis as a Dynamic Early Warning Sign for Financial Distress in Property and Real Estate Companies

Revina Sofia Septiani[1*] Bayu Tri Cahya[2] Nor ishah Mohd Ali[3]
[1,2] Universitas Islam Negeri Sunan Kudus, Indonesia
[3] UiTM Melaka, 78000, Malaysia

ABSTRACT

This study investigates the integration of sentiment analysis from social media and digital news portals as predictors of financial distress in Indonesian property companies. Intensifying industrial competition in 2024, as reported by KPPU data, has led to shrinking market shares and declining profits. Companies failing to compete face heightened risks of financial distress, necessitating more proactive monitoring tools. Unlike traditional models that rely solely on financial ratios, this research innovates by using investor sentiment as an early warning signal, leveraging advanced crawling techniques such as Tweet Harvest and high-performance analysis in RapidMiner. For the research methods, a total of 1,965 tweets and 602 news articles (2020–2023) were extracted and classified using Naïve Bayes. The relationship between sentiment and financial health (Altman Z-Score) was analyzed using ordinal logistic regression. Results showed that both social media sentiment ($p=0.029$) and news sentiment ($p=0.042$) significantly influence financial health. Positive sentiment on platform X shows the strongest predictive power ($Exp(B)=3.907$) in identifying "healthy" companies. In conclusion, **i**nvestment sentiment is a robust early warning indicator. Positive investor sentiment significantly reduces the likelihood of financial distress, providing a dynamic, real-time monitoring tool for regulators and investors in a competitive market.

ARTICLE INFO

Keywords:
Early Warning Sign
Machine Learning
Naïve Bayes
Property and Real Estate
Sentiment Analysis

Dates:
Received 10 January 2026
Revised 10 March 2026
Accepted 29 May 2026
Published 28 Juny 2026

* Corresponding at the Universitas Islam Negeri Sunan Kudus, Indonesia
E-mail address: revinasofiaa@gmail.com

INTRODUCTION

Amidst technological disruption and global economic volatility, competitive advantage is no longer merely a growth strategy but a critical defense mechanism (Erayanti, 2019). The consistent escalation of the Business Competition Index, which peaked at 4.95 in 2024, signals a structural shift in the Indonesian economy toward a hyper-competitive ecosystem (KPPU, 2024). This environment compels the elimination of conventional business models in favor of innovation-based resilience. Historically, rapid business competition has risked triggering market fragmentation, in which a drastic decline in market share can undermine cash flow stability and drive firms toward financial distress (Arifiana & Khalifaturofi'ah, 2022). As market pressures erode profit margins and threaten operational sustainability, the implementation of financial distress prediction models becomes a crucial early warning system to identify potential financial failure before a company reaches bankruptcy (Pratiwi & Imelda, 2022).

Financial health serves as the fundamental foundation for business entity sustainability; however, a phenomenon of profitability degradation has plagued the property sector (Chen & Chen, 2020). his is evidenced by the decline in net profit of PT Bumi Serpong Damai Tbk and the increase in operating expenses at PT Agung Podomoro Land Tbk. In 2023, PT Bumi Serpong Damai Tbk's profit fell to IDR 2.25 trillion, a 1.33 percent decrease from IDR 2.65 trillion in 2022 (Hidayat, 2024). By the end of December 2022, PT Agung Podomoro Land Tbk's net loss had widened. Based on the company's financial statements, PT Agung Podomoro Land Tbk's gross profit—after deducting direct costs and cost of goods sold—amounted to IDR 2.26 billion, bringing the total to IDR 1.16 trillion, indicating significant structural vulnerability (News, 2023). When profit margins erode and cash flows contract, companies enter a financial distress zone, which, if left unmitigated, leads to insolvency.

As a sector highly sensitive to macroeconomic dynamics, the property industry frequently faces price volatility and policy changes that directly threaten the stability of issuers' financial performance. An entity's inability to mitigate these risks can accelerate financial deterioration into financial distress, which, if undetected early, can lead to bankruptcy (Mahastanti & Utami, 2022). The urgency of early detection is paramount as a protective instrument for stakeholders; for instance, investors can execute timely exit strategies, creditors can secure their receivables through preventive actions, and regulators can intensify supervision. Thus, developing a responsive prediction model is not merely a managerial requirement but an essential mechanism to maintain the stability of the investment ecosystem in the property market (Shiller, 2017).

The limitations of traditional mathematical models, which rely solely on retrospective accounting data, create a predictive gap in detecting financial distress in real time. Amid volatile market dynamics, integrating investor sentiment analysis via the social media platform X and news article narratives is a crucial tool that provides prospective information dimensions not captured in conventional financial reports (Dunham & Garcia, 2021). The synergy between fundamental indicators and public sentiment enables the development of a more responsive early warning model, in which market perception fluctuations can serve as leading indicators of potential financial health deterioration and future bankruptcy risks (Al Qamar Ze & Juanda, 2023).

Investor sentiment is a critical psychological factor that directly mitigates fluctuations in issuer performance in the property sector. Fundamentally, this sentiment is shaped by the convergence of internal financial performance, industry prospects, and macroeconomic and political stability (Griffith et al., 2020). In the digital era, sentiment volatility is accelerated by massive information flows through social media platforms like Twitter (X), which provide real-time public opinion data. For market participants, utilizing opinion analysis of current social media is no longer optional but a strategic tool for predicting market reactions and minimizing investment risks amid global and domestic economic dynamics (Sul et al., 2017).

Although liquidity, leverage, and profitability ratios can empirically predict financial distress, conventional accounting models often exhibit limited accuracy (Giovanni et al., 2020). For example, the Altman model has been reported to have an accuracy of only 58%, with a margin of error of 31%. This accuracy gap indicates a need for predictive instruments that are more dynamic and extend beyond historical financial statement data (Sembiring & Sinaga, 2022). Therefore, this research offers novelty through an integrative approach that combines accountancy-based financial data analysis with real-time investor sentiment analysis from Twitter and news articles (Qu, 2023). By synergizing fundamental indicators and public opinion, this model is expected to provide a more responsive and comprehensive early detection system for mitigating corporate financial failure risks.

Building upon this premise, the novelty of this study lies in its systematic integration of real-time digital narratives and fundamental financial data through a robust computational workflow. This research employed an automated data-acquisition process using Tweet Harvest to extract high-velocity social media discourse, which was then processed using RapidMiner's advanced analytics platform. By leveraging RapidMiner's machine learning operators, this study implemented sophisticated text-mining and classification algorithms to transform unstructured investor sentiment into quantifiable risk indicators. This approach enabled a more rigorous integration of behavioral data with accounting ratios, significantly enhancing the model's predictive power and providing a more technologically advanced framework for financial distress detection in the property sector.

This study examines the correlation between investor sentiment and the probability of financial distress by positioning sentiment analysis as a proactive early warning system. Departing from traditional models, this approach uses content from the social media platform Twitter and news articles as data sources believed to have predictive value for a firm's financial health before it reaches a crisis or bankruptcy phase. The primary premise is that public opinion volatility and digital information flows reflect issuers' real conditions more dynamically. Consequently, by analyzing these investor perceptions, companies can identify signs of financial anomalies earlier and implement targeted strategic corrective measures.

METHOD, DATA, AND ANALYSIS

This study employed a quantitative approach that integrates sentiment analysis and econometric modeling to predict corporate financial health (Al Qamar Ze & Juanda, 2023). The methodology was systematically structured, beginning with targeted data acquisition, followed by text preprocessing, sentiment classification using machine learning, and culminating in statistical testing via ordinal logistic regression (Basuki & Prawoto, 2016).

The research population comprises all property and real estate companies listed on the Indonesia Sharia Stock Index (ISSI) during the 2020–2023 observation period. Sample selection was conducted using purposive sampling to ensure data representativeness aligned with the research objectives (Sugiyono, 2013). The inclusion criteria were defined as follows: (1) issuers listed on the ISSI throughout the study period with complete financial reports; (2) issuers exhibiting negative volatility, characterized by significant stock price declines; and (3) issuers with sufficient sentiment information available on the X platform and digital news portals.

Based on these criteria, eight property sector issuers were selected: ASRI, BKSL, BSDE, CTRA, LPKR, LPCK, PWON, and SMRA.

Data Crawling

The data acquisition process involved a dual-stream crawling strategy to capture both informal and formal market discourse. Primary social media data were harvested from the X platform using the Tweet Harvest tool, targeting specific keywords (e.g., $ASRI, $BSDE) and filtering for Indonesian-language content, while excluding retweets to ensure originality of sentiment, yielding 1,965 relevant tweets (Rajan & S.P.Victor, 2014). Simultaneously, secondary data were gathered from the Detik.com news portal using a Python-based web scraper in a Google Colab environment, targeting keywords such as company names and industry terms, including "PT Alam Sutera Realty Tbk" and "ASRI", to retrieve 602 news articles.

Data Implementation Process

Raw data underwent a rigorous preprocessing pipeline in RapidMiner software to transform unstructured text into a machine-readable format (Wongkar & Angdresey, 2019). This stage included:

1. Cleansing involves removing noise such as hashtags, mentions, symbols, URLs, and duplicate entries.
2. Case Folding: standardizing all text characters to lowercase.
3. Tokenization, the segmentation of text into individual words based on non-letter criteria, yielded 2,848 regular attributes.
4. Filtering, applying word length filters (2–25 characters), and Stop-word Removal to eliminate Indonesian conjunctions that lack substantive meaning.

Sentiment Labeling and Classification

This research implemented a hybrid labeling strategy to ensure high accuracy. Manual labeling was initially conducted to establish the ground truth, capturing complex linguistic nuances and sarcasm. The dataset was then split into 60% for training and 40% for testing. The Naïve Bayes Classifier algorithm was utilized for automated labeling (Samsir et al., 2021). This supervised learning model was trained on manual labels and applied to the test set using the Apply Model and Union operators. This process generated a consistent classification of sentiment into three categories: Positive, Negative, and Neutral (Ramadhan & Setiawan, 2019).

Analysis and Predictive Modeling

Classified sentiment results were integrated with financial health indicators measured by the Altman Z-Score model (Resfitasari et al., 2022). An Ordinal Logistic Regression model was applied to test the predictive power of sentiment variables X_1 (Platform X) and X_2 (News Articles). The analysis used the Likelihood Ratio Test for model fit and Nagelkerke R-Square for variance explanation (Algifari, 2000). The use of sentiment as an early warning sign represents an innovation over traditional models relying exclusively on historical financial ratios.

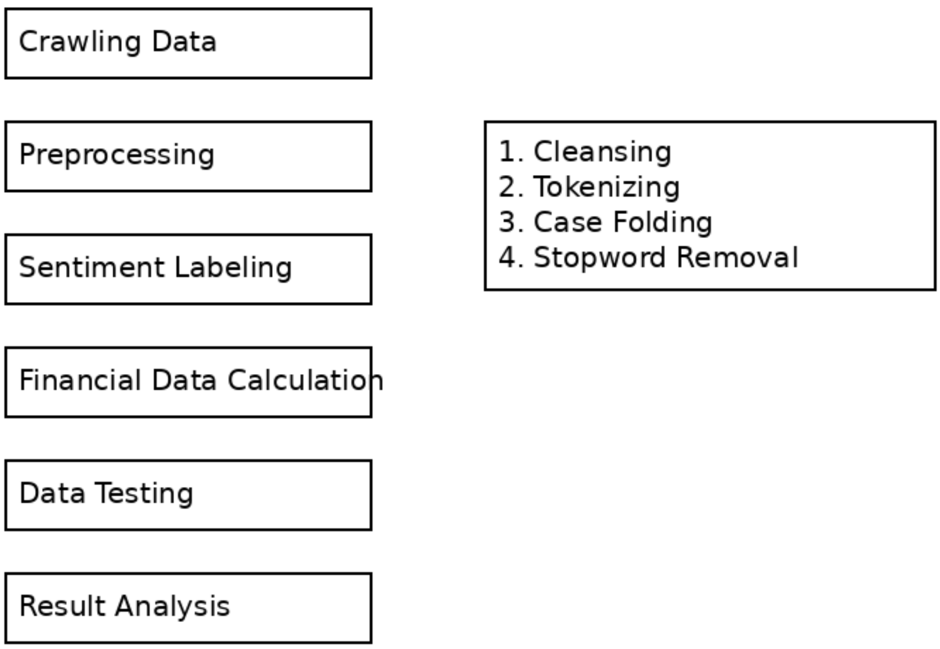

Figure 1. System Design Model

RESULTS AND DISCUSSION

Data Acquisition and Tools

This study employed a hybrid sentiment analysis approach, using Python for data extraction and preprocessing and RapidMiner for advanced sentiment modeling. The dataset comprised primary data in the form of public opinions (tweets) extracted from the social media platform X (formerly Twitter) and secondary data consisting of news articles sourced from detik.com.

Social Media Data Crawling (X Platform)

Primary data collection was conducted through web crawling techniques using Python within the Google Colab environment. The observation period spanned four years, from January 1, 2020, to December 31, 2023. Data extraction focused on specific property-sector issuers using a combination of entity names and stock tickers as keywords, including: ASRI, BKSL, BSDE, CTRA, LPKR, LPCK, PWON, and SMRA.

A total of 1,965 tweets were successfully retrieved, and filtered specifically for domestic localization (Indonesia) to ensure relevance to the local market context. The raw data were archived in .csv format to maintain structural integrity. Subsequently, a systematic labeling process was performed to classify each entry into predefined sentiment categories, resulting in a robust dataset ready for testing within the financial distress prediction model.

News Article Data Extraction

To supplement social media insights, secondary sentiment data were gathered from Detik.com. This source was selected to capture formal narratives and informative public discourse regarding corporate reputation and property market trends. Similar to the social media phase, data were collected via Python-based web crawling on Google Colab, targeting the same sample entities (ASRI, BKSL, BSDE, CTRA, LPKR, LPCK, PWON, and SMRA). The observation remained consistent with the four-year study period (2020–2023), yielding 602 relevant news articles. These textual data were converted into .csv format and integrated with the Twitter dataset. The integration of these two distinct data sources aims to provide a comprehensive multidimensional perspective, enhancing the accuracy of predicting a company's financial condition through sentiment signals.

Data Implementation and Processing Workflow

The data implementation phase follows a structured pipeline that begins with partitioning the crawled dataset into training and testing sets. To facilitate accurate classification, the data undergoes rigorous preprocessing. The training data is manually labeled into positive and negative sentiment classes. Subsequently, both datasets are subjected to feature selection and preprocessing before being integrated into a Naïve Bayes classification model. The final output of this system is the classified sentiment result, which serves as a predictive indicator.

Data Preprocessing

The raw dataset was imported into RapidMiner for systematic refinement. This stage is crucial for reducing noise and improving the algorithm's computational efficiency. The preprocessing consists of the following four steps:

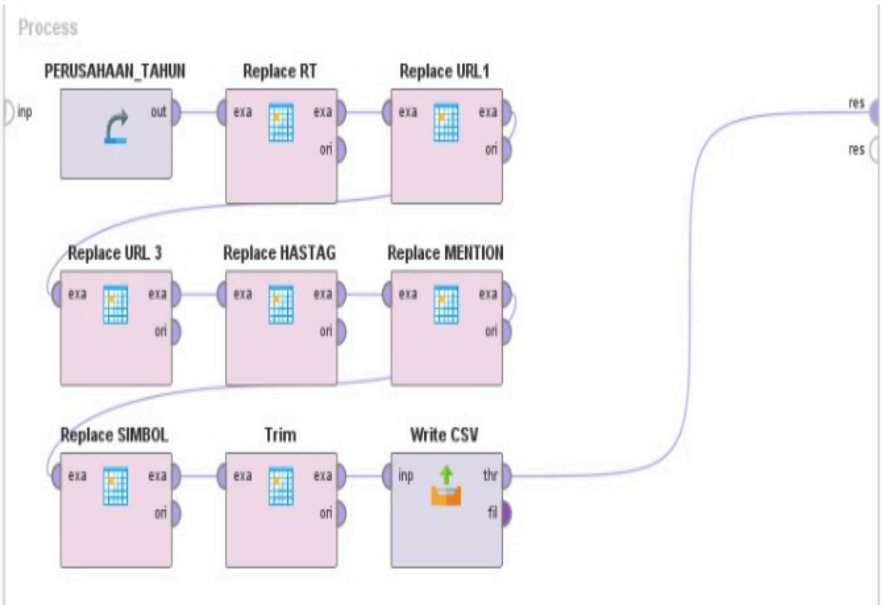

Figure 2. Data Cleaning Proses

Source: Created by Author, 2026

1. Cleansing, the initial step, involves removing redundant or irrelevant information. Specific elements targeted for deletion include hashtags, mentions (@), special symbols, emoticons, excessive whitespace, and duplicate entries. This ensures that the model focuses solely on the textual content.

2. Tokenization: In this phase, the text strings are decomposed into smaller, meaningful units known as tokens (words or phrases). This process identifies the linguistic origins of the data, resulting in a structured set of 2,848 regular attributes.
3. Case Folding: To ensure uniformity, all characters in the text documents are converted into lowercase. This prevents the algorithm from treating the same word as two different entities due to capitalization variances (e.g., "Pasar" vs "pasar").
4. StopWord Removal: This step filters out frequently occurring words that lack significant semantic value, such as conjunctions and prepositions (e.g., "dan," "atau," "yang"). Removing these terms narrows the focus to "feature-heavy" words that carry sentiment weight.

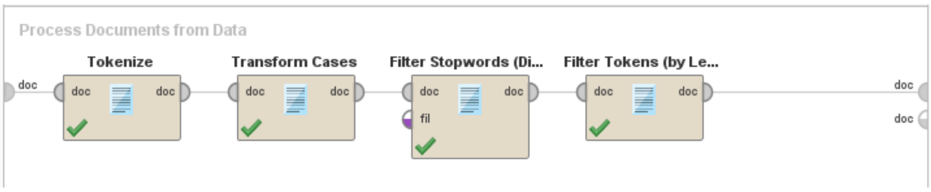

Figure 3. Data Tokenization, Case Folding, StopWord Removal Proses

Source: Created by Author, 2026

Sentiment Labeling Process

The labeling process used a hybrid approach to classify 1,965 tweets and 602 news articles into positive, negative, and neutral categories. Manual labeling was first used as ground truth to address semantic ambiguity and sarcasm, which are difficult to detect automatically. The dataset was then partitioned (data splitting) into 60% for training to build the model using the Naive Bayes Classifier algorithm and 40% for performance validation. This procedure ensures the dataset's reliability before it is integrated into the logistic regression analysis. The automated labeling stage was performed using the Naïve Bayes Classifier algorithm in RapidMiner. This procedure began by training the model (training model) using manual labeling results as a machine learning reference.

Figure 4. Sentiment Labeling Process

Source: Created by Author, 2026

Once the model was established, the testing data was processed using the Read CSV operator, filtered through missing values parameters to ensure that only unlabeled data were processed.

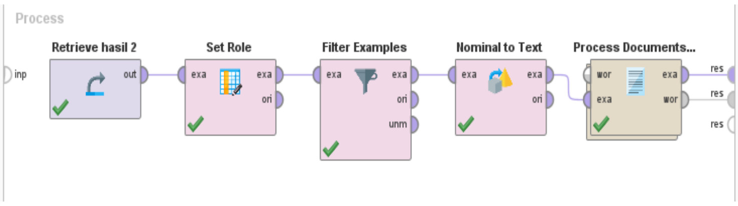

Figure 5. Training Data Modeling Proses

Source: Created by Author, 2026

The testing data modeling stage aims to automatically assign sentiment labels to unclassified data. This procedure began by using the Read CSV operator in RapidMiner, followed by filtering using the parameters "sentiment is missing" and "description is not missing" to ensure input validity. The data was then converted through the Nominal to Text operator before entering the feature extraction stage within the Process Document From Data operator. Inside this operator, four primary sub-processes are performed: (1) Tokenize based on non-letter criteria, (2) Transform Cases for lowercase uniformity, and (3) Filter Tokens (by Length) with a range of 2–25 characters to reduce noise. Subsequently, the Union operator was used to integrate the testing dataset with the previously constructed training model. To maintain data consistency, empty values are addressed using the Replace Missing Values operator with a constant value of 0. The process concluded with the application of the Apply Model operator, which used the Naïve Bayes algorithm to automatically label the data based on the training patterns, producing consistent positive, negative, and neutral sentiment classifications.

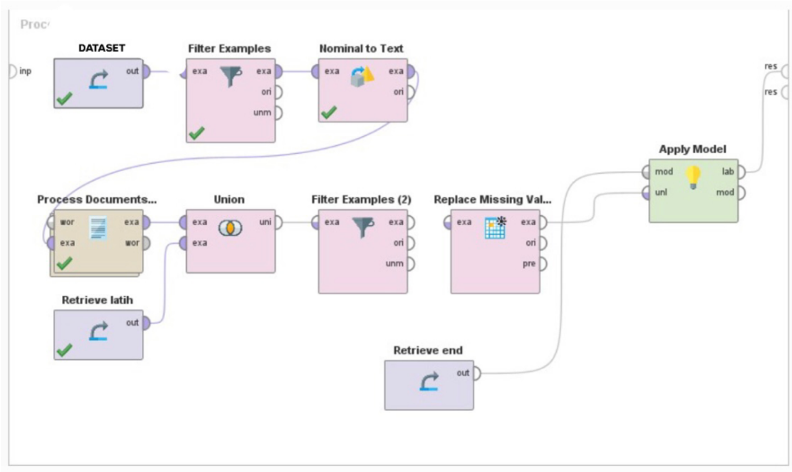

Figure 6. Automated Labeling Process Using the Naive Bayes Algorithm
Source: Created by Author, 2026

Figure 6 shows the automated labeling process flow for sentiment analysis using the Naive Bayes algorithm. This process began with the data-crawling stage, during which text data was collected from various sources. Next, the data undergoes preprocessing, including cleansing, tokenization, case folding, and stopword removal, to ensure it is clean and ready for analysis. After preprocessing, the processed data were used in the sentiment labeling stage using the Naive Bayes algorithm. This algorithm calculates the probability that a text belongs to a specific sentiment category (e.g., positive, negative, or neutral) based on word frequencies and the probability distributions from the training data. The results of this labeling were then used in the financial data calculation stage to link sentiment to the economic or financial variables being analyzed. Data testing was then conducted to verify the accuracy of the classification model used. The final stage was result analysis, in which the test results were evaluated to draw conclusions about model performance and the implications of the sentiment labeling.

Financial Data Calculation

The company’s financial health was measured using the Altman Z-Score model as the dependent variable. Data were sourced from the annual reports of issuers listed on the Indonesia Stock Exchange (Putri & Maulana, 2023), analyzed through the following formulation:

$$\boxed{Z = 1{,}2\ X_1 + 1{,}4X_2 + 3{,}3\ X_3 + 0{,}6\ X_4 + 1{,}0\ X_5}$$

Corporate health classification is determined based on the following score thresholds:

Healthy Zone : Z > 2, 99

Grey Zone : 1,81 < Z < 2,99

Distress Zone : Z < 1,81

Tabel 1. Altman Z Score Result

Company Code	Altman Z-score	Information	Codee
ASRI	0,079334605	Terindikasi	-1
	0,341790962	Terindikasi	-1
	0,720359743	Terindikasi	-1
	0,687867128	Terindikasi	-1
BKSL	0,065139063	Terindikasi	-1
	0,393190144	Terindikasi	-1
	0,288987533	Terindikasi	-1
	0.589805505	Terindikasi	-1
BSDE	0,99002755	Terindikasi	-1
	1,125741164	Zona Grey	0
	1,244841305	Zona Grey	0
	1,317675221	Zona Grey	0
CTRA	0,830927754	Terindikasi	-1
	1,037810544	Zona Grey	0
	1,119236678	Zona Grey	0
	1,173615916	Zona Grey	0
LPCK	-0,779153108	Terindikasi	-1
	0,545900352	Terindikasi	-1
	1,362770563	Zona Grey	0
	1,336571399	Zona Grey	0
LPKR	1,617989202	Zona Grey	1
	2,26593974	Zona Grey	1
	0,459895029	Zona Grey	-1
	0,451735543	Zona Grey	-1
PWON	1,548033014	Zona Grey	0
	1,711416279	Zona Grey	1
	1,61444712	Zona Grey	1
	1,28908211	Zona Grey	0
SMRA	0,584195664	Terindikasi	-1
	0,785676935	Terindikasi	-1
	0,767141885	Terindikasi	-1
	0,768143106	Terindikasi	-1

From the table 1 of results of the Altman Z score model calculation above, it can be concluded that from the entire sample of 8 companies in the period 2020 to 2021, there were a total of 18 data points that indicated financial distress/unhealthy, 10 data points that were in the gray area, and 4 data that did not indicate financial distress/healthy.

Table 2. Ordinal Logistic Regression Analys

Analys Component	Indicator	Statistic Value	Sig. (p-value)
Model Fitting Information	X^2 (Chi-Square)	14,719	0,000***
Goodness-of-Fit	Pearson	6,305	0,789
	Deviance	8,191	0,789
Pseudo R-Square	Cox and Snell	0,369	
	Nagelkerke	0,435	
	McFadden	0,245	
Parallel Lines	Chi-Square	14,719	0,000
Parameter Estimates	**Variabel**	**Estimate**	**Sig.**
Threshold	[y = -1,00]	0,079	0,877
	[y = 0,00]	3,266	0,000***
Location	X_1	1,363	0,029**
	X_2	0,980	0,042**

Data source: Author's data processing results, 2026

The table 2 overall regression model was deemed fit or feasible, as it yields a p-value of 0.001 ($p < 0.05$) in the Likelihood Ratio test. This is further supported by the significance values for Pearson (0.789) and Deviance (0.610), both of which exceed 0.05, indicating that the model is consistent with the empirical data. The Nagelkerke R-Square value of 0.435 indicates that the independent variables (X_1 and X_2) explain 43.5% of the variance in the dependent variable. Variable X_1 has a positive and significant effect on the probability of higher dependent variable categories (\beta = 1.363; p = 0.029). Similarly, variable X_2 has a positive and significant effect on the probability of higher levels of the dependent variable (\beta = 0.980; p = 0.042).

Statistical tests show a p-value of $p = 0.000$ ($p < 0.05$), leading to rejection of the null hypothesis. This proves that both the X sentiment variable and the news article sentiment variable simultaneously have a significant effect on the company's financial distress level. Partially, the X sentiment variable ($p = 0.029$) and the news article sentiment variable ($p = 0.042$) significantly influence the company's financial distress, as the significance values for both variables are lower than \alpha = 0.05. The regression coefficients for X sentiment (1.363) and news article sentiment (0.980) indicate a positive relationship with the company's financial condition. The odds ratios (Exp(B)) of 3.907 for X sentiment and 2.664 for news sentiment indicate that a one-unit increase in sentiment increases the probability of the company being in the 'healthy' category by 3.9 and 2.6 times, respectively. This improvement in sentiment significantly reduces financial distress risk and enhances the company's chances of moving into a more stable category.

DISCUSSION

This study aims to analyze the influence of social media sentiment and online news on the prediction of financial distress in the property sector in Indonesia. Based on the ordinal logistic regression results, these two data sources have a significant influence, both simultaneously and partially, on the company's financial health, as measured by the Altman Z-Score. The regression analysis demonstrates that social media sentiment on platform X has a significant positive influence on the financial distress of companies listed on the Indonesia Sharia Stock Index (ISSI) during the 2020-2023 period, as evidenced by a p-value of 0.029 ($p<0.05$).

These findings suggest that the escalation of negative digital sentiment acts as a catalyst for financial instability through two primary transmission channels: the erosion of consumer trust, which triggers sales volatility, and reputational degradation, which prompts conservative investment behavior and aggressive stock sell-offs. This phenomenon aligns with behavioral finance theory (Sisbintari, 2018) and corroborates the empirical evidence provided by Mengelkamp et al. (2016), reinforcing the premise that unstructured social media data serves as a potent predictor of corporate credit risk and insolvency. Consequently, this study emphasizes the critical necessity for Sharia-compliant firms to integrate digital reputation management into their risk mitigation strategies to safeguard liquidity and ensure long-term financial resilience against perception-based external shocks.

This study found that news article sentiment significantly influences the financial distress of companies listed on the Indonesia Sharia Stock Index (ISSI) during the 2020-2023 period, as evidenced by a partial test with a p-value of 0.042 ($p<0.05$). The empirical evidence suggests that news sentiment serves as a critical information signal; while positive coverage bolsters investor confidence and share prices, negative news acts as a precursor to financial instability by triggering liquidity constraints and reputational damage. This is exemplified by the Meikarta case, where legal and regulatory scandals led to a massive sell-off and hindered the firm's ability to access capital markets (Ahdiat, 2023). These findings align with the research of Nikkinen & Peltomäki (2020) regarding the impact of information dissemination on crash fears, as well as the assertation of Zhao et al. (2022) that integrating sentiment tone features from news publications is highly effective for financial distress prediction. Ultimately, the results underscore that for Sharia-compliant firms, news sentiment is not merely a reflection of public opinion but a tangible risk factor that directly correlates with corporate solvency.

The main findings of this study indicate that the X sentiment variable (X_1) has a regression coefficient of 1.363, while the news sentiment variable (X_2) has a coefficient of 0.980. The Odds Ratio (Exp(B)) values of 3.907 and 2.664, respectively, indicate that a one-unit increase in positive sentiment on the X platform increases the likelihood of a company being in the "healthy" category by nearly fourfold. This shows that public opinion on social media tends to be a more sensitive leading indicator of changes in financial conditions than formal news. Social media captures investor expectations and concerns in real time, while news on portals like Detik.com is often informative and confirms events that have already occurred.

The results of this study reinforce the Investor Sentiment theory, which holds that public perception could influence market value and corporate access to funding, ultimately impacting financial stability. The use of a hybrid method with the Naive Bayes algorithm for sentiment classification proved effective, as seen from the regression model's ability to explain 43.5% of the data variance (based on the Nagelkerke R-Square value). Although Twitter data is often considered "noisy," systematic preprocessing, including cleansing, tokenization, and stopword removal, has successfully extracted relevant sentiment signals to support technical predictions, such as the Altman Z-Score. Specifically for property issuers such as ASRI, BKSL, and SMRA, which, during the research period (2020-2023), showed indications of financial distress (low Z-Score), a correlation with the developing public narrative was observed (Pradana et al., 2020). These findings have managerial implications: companies must focus not only on fundamental performance but also on proactively managing corporate communications and digital reputation. For investors, this model can serve as an additional Early Warning System (EWS) instrument beyond traditional financial statement analysis.

CONCLUSION AND SUGGESTIONS

This study concludes that integrating sentiment analysis from social media and digital news portals is a significant predictor of the financial health of property companies in Indonesia. Based on the ordinal logistic regression analysis, public sentiment variables from the X platform (Twitter) and news article sentiment both exert a significant positive influence on the level of financial distress, as measured by the Altman Z-Score model.

The primary findings indicate that public sentiment on the X platform exerts a stronger influence (Exp(B)=3.907) than sentiment from news articles (Exp(B)=2.664). This suggests that every one-unit increase in positive sentiment increases a company's odds of being in the "financially healthy" category by 3.9 times. These results reinforce the Efficient Market Hypothesis (EMH), demonstrating that non-financial information and public opinion evolving in digital spaces effectively reflect a company’s fundamental condition. Furthermore, the Naïve Bayes algorithm proved effective for automated labeling, achieving adequate accuracy to support predictive analysis of financial distress risks. Based on these findings and the limitations identified, several directions for future research are proposed. First, regarding methodology, future researchers are encouraged to explore the use of Deep Learning algorithms, such as Long Short-Term Memory (LSTM) or BERT (Bidirectional Encoder Representations from Transformers), to enhance sentiment classification accuracy, particularly in capturing linguistic nuances, sarcasm, and complex local dialects on social media. Second, as this study is limited to the property sector, the generalizability of the results could be enhanced by expanding the sample scope to other industrial sectors and extending the observation period to encompass various phases of the economic cycle. Third, macroeconomic variables such as interest rate fluctuations, inflation rates, and property-sector regulatory policies should be included as control variables to strengthen the model's predictive power. Finally, from a practical standpoint, regulators and investors are advised to integrate unstructured text data analysis into an early warning system to monitor the financial stability of issuers more dynamically and in real time in the digital era.

ACKNOWLEDGEMENTS

The authors would like to express their sincere gratitude to State Islamic University Sunan Kudusor for providing the academic facilities and support necessary to conduct this research. Special thanks are extended to Dr. Bayu Tri Cahya, S.E., M.Si., for their invaluable guidance, insightful suggestions, and critical discussions throughout the development of the methodology and data analysis. The authors also acknowledge the data provided by Detik.com and X (formerly Twitter), which enabled the extraction of extensive sentiment data for this study. Lastly, we are grateful to our colleagues and family for their continuous moral support and technical assistance in completing this manuscript.

REFERENCE

Ahdiat, A. (2023). Terseret kasus Meikarta, ini keuangan Lippo Cikarang 5 tahun terakhir. *Databoks.*https://databoks.katadata.co.id/properti/statistik/cdd26c632c81aa7/terseret-kasus-meikarta-ini-keuangan-lippo-cikarang-5-tahun-terakhir

Al Qamar Ze, A., & Juanda, J. (2023). Pengaruh herding behavior, media, interaksi sosial, dan advocate recommendation terhadap keputusan investasi dengan sentimen investor sebagai variabel intervening. *International Journal of Science and Society, 1*(2), 1–9.

Algifari. (2000). *Analisis Regresi, Teori, Kasus dan Solusi*. BPFE-YOGYA.

Arifiana, R., & Khalifaturofi'ah, S. O. (2022). The effect of financial ratios in predicting financial distress in manufacturing companies. *Jurnal Riset Bisnis dan Manajemen, 15*(2), 103–108. https://doi.org/10.23969/jrbm.v15i2.5838

Basuki, & Prawoto. (2016). *Analisis regresi dalam penelitian ekonomi dan bisnis (Dilengkapi dengan aplikasi SPSS & EViews)*. RajaGrafindo Persada.

Chen, C., & Chen, C.-D. (2020). Financial distress prediction model: The effects of corporate governance indicators. *Journal of Forecasting, 39*(8), 1238–1252. https://doi.org/https://doi.org/10.1002/for.2684

Dunham, L. M., & Garcia, J. (2021). Measuring the effect of investor sentiment on financial distress. *Managerial Finance*, *50*(12), 1834–1852. https://doi.org/10.1108/MF-02-2021-0056

Pratiwi, E.I, & Imelda, E. (2022). Hubungan kausalitas antara sentimen investor dan pasar keuangan Indonesia. *Jurnal Ekonomi, 27*(3), 182–201. https://doi.org/10.24912/je.v27i03.872

Erayanti, R. (2019). Pengaruh likuiditas, profitabilitas, dan leverage terhadap prediksi financial distress. *Jurnal Riset Akuntansi & Perpajakan (JRAP), 6*(1), 38–51. https://doi.org/10.35838/jrap.v6i01.393

Giovanni, A., Utami, D. W., & Yuzevin, T. (2020). Leverage dan profitabilitas dalam memprediksi financial distress perusahaan pertambangan periode 2016–2018. *Journal of Business and Banking, 10*(1), 151. https://doi.org/10.14414/jbb.v10i1.2292

Griffith, J., Najand, M., & Shen, J. (2020). Emotions in the stock market. *Journal of Behavioral Finance, 21*(1), 42–56. https://doi.org/10.1080/15427560.2019.1588275

Hidayat, K. (2024). Bumi Serpong (BSDE) profit drops 20.04% in 2023: Check out its stock recommendations. *Kontan*. https://english.kontan.co.id/news/bumi-serpong-bsde-profit-drops-2004-in-2023-check-out-its-stock-recommendations#:~:text=Bumi Serpong (BSDE) Profit Drops,Check Out Its Stock Recommendations

Komisi Pengawas Persaingan Usaha. (2024). CEDS Unpad : Tingkat persaingan usaha Indonesia tahun 2023 meningkat. https://kppu.go.id/blog/2024/01/ceds-unpad-tingkat-persaingan-usaha-indonesia-tahun-2023-meningkat/

Mahastanti, L. A., & Utami, A. D. (2022). Perbandingan tingkat akurasi model prediksi financial distress pada perusahaan sektor property dan real estate. *Accounting and Financial Review, 5*(1), 50–63. https://doi.org/10.26905/afr.v5i1.7526

Mengelkamp, A., Wolf, S., & Schumann, M. (2016). Data-driven creation of sentiment dictionaries for corporate credit risk analysis. In *Proceedings of the Americas Conference on Information Systems (AMCIS 2016)* (pp. 1–9).

News, Indo Premier Securities. (2023). Financial statements full year 2022 of APLN.

https://www.indopremier.com/ipotnews/newsDetail.php?jdl=Financial_Statements_Full_Year_2022_of_APLN&news_id=417051&group_news=RESEARCHNEWS&news_date=&taging_subtype=APLN&name=&search=y_general&q=Financial Statements&halaman=1

Nikkinen, J., & Peltomäki, J. (2020). Crash fears and stock market effects: Evidence from web searches and printed news articles. *Journal of Behavioral Finance, 21*(2), 117–127. https://doi.org/10.1080/15427560.2019.1630125

Pradana, M. G., Nurcahyo, A. C., & Saputro, P. H. (2020). Pengaruh sentimen di sosial media terhadap harga saham perusahaan. *Edutic: Scientific Journal of Informatics Education, 6*(2), 67–77. https://doi.org/10.21107/edutic.v6i2.6992

Putri, M. A., & Maulana, H. Z. (2023). Memprediksi kebangkrutan menggunakan Altman Z-score pada perusahaan manufaktur sektor makanan dan minuman. *Jurnal Cendekia Keuangan, 2*(2), 55. https://doi.org/10.32503/jck.v2i2.3431

Qu, S. (2023). A study on the impact of investor sentiment on corporate financial stress based on data mining. In *Proceedings of the International Conference on Big Data Economy and Information Management* (pp. 845–850). https://doi.org/10.1145/3659211.3659357

Rajan, A. P., & S.P.Victor. (2014). Crawling Twitter data. *Journal of Research in Computer Application & Management, 4*(9).

Ramadhan, D., & Setiawan, E. B. (2019). Analisis sentimen program acara di SCTV pada Twitter menggunakan metode naive Bayes dan support vector machine. *E-Proceeding of Engineering, 6*(2), 9736–9743. https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/10708

Resfitasari, E., Taofik Muhammad Gumelar, Andini Ulhaq, & Nina Rusmayanti. (2022). Analisis prediksi financial distress dengan metode Altman Z-score pada PT Waskita Karya Tbk. *Jurnal Aktiva: Riset Akuntansi dan Keuangan, 3*(3), 131–140. 131–140. https://doi.org/10.52005/aktiva.v3i3.123

Samsir, Irmayani, D., Edi, F., Harahap, J. M., Jupriaman, Rangkuti, R. K., Ulya, B., & Watrianthos, R. (2021). Naive Bayes algorithm for Twitter sentiment analysis. *Journal of Physics: Conference Series, 1933*(1), 012019. https://doi.org/10.1088/1742-6596/1933/1/012019

Sembiring, S., & Sinaga, H. W. C. (2022). Analisis akurasi model Altman, Grover, Springate, Zmijewski dalam memprediksi kondisi financial distress. *Jurnal Riset Akuntansi & Keuangan, 8*(2), 299–311. https://doi.org/10.54367/jrak.v8i2.1662

Shiller, R. J. (2017). Narrative economics. *American Economic Review*, *107*(4), 967–1004. https://doi.org/10.1257/aer.107.4.967

Sisbintari, I. (2018). Sekilas tentang behavioral finance. *Jurnal Ilmiah Administrasi Bisnis dan Inovasi, 1*(2), 88–101. https://doi.org/10.25139/jai.v1i2.814

Sugiyono. (2013). *Metode penelitian kuantitatif, kualitatif, dan R&D*. Alfabeta.

Sul, H. K., Dennis, A. R., & Yuan, L. I. (2017). Trading on Twitter: Using social media sentiment to predict stock returns. *Decision Sciences, 48*(3), 454–488. https://doi.org/10.1111/deci.12229

Wongkar, M., & Angdresey, A. (2019). Sentiment analysis using naive Bayes algorithm of the data crawler: Twitter. In *Proceedings of the 4th International Conference on Informatics and Computing (ICIC 2019)* (pp. 1–5). https://doi.org/10.1109/ICIC47613.2019.8985884

Zhao, S., Xu, K., Wang, Z., Liang, C., Lu, W., & Chen, B. (2022). Financial distress prediction by combining sentiment tone features. *Economic Modelling, 106*, 105709. https://doi.org/10.1016/j.econmod.2021.105709