

DATA-DRIVEN INSIGHTS ON POST-COVID MSME RECOVERY: INTEGRATING TEXT AND REVENUE ANALYSIS IN WEST SUMATRA

Ares Albirru Amsal^{1*)}, Berri Brilliant Albar²⁾

^{1,2)} Faculty of Economics and Business, Universitas Andalas

Email: aresalbirruamsal@eb.unand.ac.id, berribrilliantalbar@eb.unand.ac.id

ABSTRACT

This research examines how media narratives frame the condition of MSMEs during the crisis and post-pandemic recovery by utilizing topic and sentiment analysis from 1422 online news stories containing the keyword MSMEs. The main problem raised is the lack of empirical understanding of how media reports reflect changes in MSME performance and whether the dynamics of these narratives move in the same direction as business recovery. This research aims to map topic patterns, measure news sentiment using a special MSME–crisis sentiment dictionary, and link it with MSME income trends as an business indicator. Through the stages of text preprocessing, LDA topic modeling, and lexicon-based sentiment calculations, this research found that in the early phase of the pandemic, the news was dominated by crisis topics and had a negative tone in line with the decline in MSME income. Entering the recovery phase, the proportion of government support topics, economic programs and market activity increased, followed by a shift in sentiment to become more positive. In the post-pandemic period, the media narrative has become increasingly stable and growth-oriented, consistent with the trend of increasing MSME income. These findings show that the dynamics of media narratives move parallel to the economic conditions of MSMEs and can be a supporting indicator for reading the crisis, transition and recovery phases. In conclusion, text mining in local news can provide strategic insight for local governments and MSME actors in understanding public perceptions and directing communication strategies and economic recovery policies.

Keywords: MSMEs, text mining, topic modeling, sentiment analysis, business recovery, media narratives

INTRODUCTION

MSMEs play a strategic role in the regional economy as the main driver of production and employment, which has a significant impact on local economic growth. In Indonesia, MSMEs have become the backbone of the economy with a large contribution to national GDP and economic inclusion. Nugroho (2012) shows that MSMEs in Indonesia make a significant contribution to the growth of employment and national exports. Quantitative studies also illustrate that the growth of MSMEs nationally has a positive impact on regional economic growth (Surya et al., 2021). At the provincial level such as West Sumatra, research shows that MSMEs have adaptation strategies that are crucial in establishing competitive advantages (Aisyah et al., 2022). Thus, an analysis of the dynamics of media narratives and how this interacts with the recovery of MSME incomes in West Sumatra is important to understand how MSMEs not only survive but also recover after shocks such as the COVID-19 pandemic.

At the start of the COVID-19 pandemic, many MSMEs experienced enormous economic pressure. A study conducted by the United Nations Industrial Development Organization (UNIDO) shows that around 60% of respondents reported that at least half of their workforce was unable to return to work due to social networks and business operations (UNIDO, 2020). In Indonesia, a national survey shows that 84% of micro, small and medium businesses experienced a decrease in income as a direct impact of the pandemic (Tempo, 2020). The main factors triggering a decline in performance include reduced demand, supply chain disruption, mobility, and liquidity difficulties (UNIDO, 2020). Although there are currently no publications that specifically focus on the impact of the pandemic on MSME income in West Sumatra province, research examining regional development disparities during the crisis shows that the region experienced worsening disparities during the pandemic period (Puslit DPR RI, 2021). Thus, the economic impact of the pandemic on MSMEs in West Sumatra is expected to follow the national trend: a significant decline in income and operations, with a medium-term recovery process.

The role of the media is increasingly important in shaping public perceptions and providing a framework for understanding regional economic issues. Research shows that the media not only disseminates information but also determines the focus of problems through agenda-setting and framing processes, which then influences how society views economic conditions and government policies (Boumans et al., 2023). In Indonesia, a study of economic reporting states that local media in the Global South region faces challenges in highlighting narratives of business recovery and equitable development, which means that the way MSME issues and post-traumatic recovery from crises such as the pandemic are presented can strengthen or increase public perception of policy effectiveness (Ausat, 2023). Thus, in a regional context such as West Sumatra, how the media reports on the economic recovery of MSMEs can shape the hopes, trust or skepticism of residents and MSME actors towards the recovery process.

In the literature on the MSME sector, there is a glaring void in linking changes in media narratives to MSME financial or operational performance indicators. Most studies focus on the adoption of technologies such as social media and digital marketing, but rarely examine how media coverage or public framing affects income, consumer confidence, or post-shock recovery. For example, although the United Nations Industrial Development Organization (UNIDO) and others collect data on pressures on MSMEs during the crisis, research that directly links media narratives and recovery trends in MSME performance is still limited (Fu et al., 2024). In the MSME and social media domains, research shows that the level of media adoption is still low and empirical research on its impact on performance is limited (Meier & Peters, 2023). This shows the need for further research that combines media reporting analysis, including topics, sentiment, and trends, with MSME performance data at the local level such as the province.

This research aims to map and analyze how local media narratives in West Sumatra regarding MSMEs and post-pandemic recovery are formed, as well as examining their relationship with MSME income trends. Theoretically, this study contributes to the media framing and regional economics literature by combining topic analysis and news sentiment with MSME performance indicators, complementing the void of previous research which focused more on technology adoption or digital marketing (Atanassova & Clark, 2020). From

a local context, studies such as Novitasari (2022) show that the MSME sector has a significant contribution to GDP and employment, while Janah and Tampubolon (2024) emphasize the role of MSMEs as the backbone of the national economy. Practically, it is hoped that the findings of this research will provide guidance for policy makers and MSME actors to design public communication and news strategies that can strengthen the trust of MSME actors and the public in the economic recovery process. Thus, this research offers an empirical and applied framework that is relevant for the development of MSMEs at the regional level.

Media and the role of information in regional economic dynamics

Media plays a crucial role in the social construction and public perception of regional economic issues. The concept of media framing according to Robert Entman is the process of selecting and highlighting certain aspects of reality which then give meaning to the audience (Entman, 1993 in Sodikin & Rini, 2020). For example, research in Indonesia found that online news about halal policies for MSMEs highlighted aspects of the costs and burdens of small business actors, which ultimately shaped public perceptions of these policies (Sodikin & Rini, 2020). At the international level, studies show that media narratives can also influence economic expectations and regional development processes: credible and consistent narratives through the media can shape the commitment of institutions and economic actors and thus have an impact on economic growth (Jiang & Man, 2025). In the regional economic context, the media not only distributes information, but also frames how the economic crisis, recovery and policy interventions are understood by local communities and MSME actors.

Furthermore, regional economic literature states that the dynamics of information and public narratives have implications for investment, business confidence, and ultimately regional economic performance (Kraus et al., 2020). In a regional context such as West Sumatra, local media functions as an important mediator between public policy, institutional support, and the perceptions of micro-small business actors. However, so far there has not been much research that has empirically analyzed how media reporting topics and sentiments correlate with MSME performance indicators at the regional level.

The impact of external shocks on MSMEs and their recovery process

In the micro, small and medium enterprise (MSME/SME) sector, external shocks such as the COVID 19 pandemic have placed great pressure on business continuity and performance. International research shows that SMEs that have good financial strength, operational efficiency and supply chain flexibility tend to be better able to face the risk of disruption due to the pandemic (Brown et al., 2022). For example, an Australian study found that efficiency, financial strength, and sourcing flexibility were negatively correlated with supply chain risk during the pandemic shock, suggesting that a firm's internal capabilities are a key recovery factor (Brown et al., 2022).

In the Indonesian context, preliminary studies show that MSMEs experienced a systematic decline in income during the pandemic, with indications that recovery will depend on institutional support and digital adaptation. However, few have explored the recovery process empirically at the regional level or used medium-term quantitative performance indicators. This shows the need for research that connects the dynamics of MSME recovery research with operational variables such as income, employment, or service innovation in the local environment.

Text mining as an approach to analyzing business issues in the news

Text mining has now become an important tool for extracting insights from large text data such as media reports, online reports and public documents, which were previously difficult to process quantitatively. For example, a study by Kobayashi (2018) states that the text mining approach in organizational research helps researchers recognize hidden themes and context patterns in large texts without first manually assigning categories. In the context of economics and media, a study by Hossain (2021) analyzes Bangladeshi newspaper headlines using word-cloud, clustering, and sentiment analysis techniques to reveal how dominant socio-political issues emerge in economic reporting. These techniques, such as topic extraction, word frequency analysis, and sentiment classification, enable the transformation of text data into quantitative indicators that can be linked to economic variables.

Furthermore, the application of text mining in the context of business and public policy is increasingly widespread. Bholat et al. (2015) say that central banks use text mining to examine various documents, including speeches, reports and mass media, as part of macroeconomic and monetary policy analysis. This method allows real-time monitoring of changes in public narratives and economic expectations that traditionally could only be measured through surveys or lagging aggregate data. Furthermore, Yang et al. (2022) developed a news-based text-mining framework for business sentiment analysis, showing that CNN-based models can achieve better performance in sentiment classification and have direct applications to economic decision making.

Thus, for studies linking media narratives with MSME performance, text mining offers an appropriate methodology: collection and preprocessing of news corpus, identification of topics related to MSME/regional economy, sentiment calculation, and aggregation of scores to be analyzed together with MSME performance data. This approach allows for a more systematic empirical examination of how news dynamics can be linked to trends in local economic performance.

Position of this study in the existing literature

Studies on MSMEs have developed significantly in the last two decades, but research that specifically links media narratives with MSME performance is still limited. Most of the MSME literature focuses on internal factors such as market orientation, entrepreneurial dynamics, and dynamic capabilities in improving performance (Panjaitan et al., 2025). Meanwhile, in the realm of media and communication, studies of framing, agenda-setting, and economic narratives show that the media not only conveys facts but also shapes public priorities and social interpretations (Saari et al., 2022). Therefore, studies that bridge these two fields, namely how media coverage influences MSME performance indicators, occupy a relatively new and interdisciplinary position.

Furthermore, the literature survey shows that research based on big text (big text data) through text mining in the context of MSMEs and economic media is still developing. For example, systematic research on social media adoption by MSMEs found that although many articles discussed increasing visibility or customer interaction, few tested the correlation between media reporting narratives and business performance outcomes empirically (Meier & Peters, 2023). This shows the importance of studies that use methodologies such as text mining to bridge the gap between media content and MSME performance data.

In the Indonesian context, although there is research that examines digital transformation and the readiness of MSMEs to face technological change (Erlanitasari et al., 2019), there is a lack of research that links local news, especially economic media or MSME reporting at the regional level, with micro-enterprise economic data such as income, employment, or level of recovery after external shocks. Thus, this study positions itself as an effort to fill this gap: integrating media report corpus analysis with MSME performance data at the provincial level. This contribution is expected to enrich the literature on media framing, regional economics and MSME management with evidence based on empirical data.

METHODS

Data sources and the basis for selecting keywords

This research uses an online news corpus from local media Langgam.id which contains the term "UMKM" as the main keyword, with a collection period from January 2020 to October 2025 (1422 news stories). The choice of the single keyword "UMKM" is intended to capture all news relevant to micro, small and medium enterprises in the study area, while maintaining data consistency and systematicity. In the context of text-mining, the use of appropriate keywords is essential for building a focused and relevant corpus (Georgieva-Trifonova & Dechev, 2021).

Data characteristics include metadata variables such as publication date, news title, and main paragraph of text. After collection, text preprocessing is carried out, which includes cleaning special characters, normalizing letters, tokenizing, removing stop-words, and stemming/lemmatizing according to research policy. These steps are in line with the general practice of text-mining for media news, where methods such as word-cloud, sentiment analysis, and clustering have proven useful in analyzing reporting patterns (Hossain et al., 2021).

Furthermore, determining the period January 2020-October 2025 allows temporal mapping of news before, during and after the main phase of the COVID 19 pandemic, thereby enabling analysis of topic and sentiment dynamics based on the context of external shocks and business recovery. Methodology articles from the field of text-mining emphasize the importance of temporal representations that are long enough to detect changes in themes and trends (Pohlmann, 2023). The data will then be analyzed through topic modeling and sentiment analysis, and then combined with MSME performance indicators (annual revenue) to explore the link between media narratives and economic performance.

Text processing procedures

Before topic and sentiment analysis can be carried out, the news corpus must first be processed through pre-processing steps, which include cleaning, normalization, and tokenization. Text cleaning includes removing special characters, numbers, links, excess punctuation, and changing all letters to lowercase to ensure consistency and avoid duplication of tokens that are actually identical. This step is important because online news data often contains noise such as HTML tags, multimodal metadata, or document formats that need to be converted first before text mining can run smoothly (Pohlmann, 2023).

After cleaning, the normalization process includes replacing variant word forms with basic forms through stemming or lemmatization, as well as removing stop-words that do not carry significant meaning in the analysis (Hossain et al., 2021). The use of domain-adjusted

stop-word lists can significantly increase the efficiency and quality of text mining analysis (Alshanik et al., 2020).

Next, tokenization is the stage of breaking down the text into meaningful units (tokens), usually in the form of words or phrases, which become the basis for subsequent analysis. These tokenization decisions can influence the final results of topic or sentiment modeling because variations in token formation can potentially cause analytical bias (Zhang et al., 2025). Therefore, implementing appropriate tokenization procedures (e.g. removal of stop-words, punctuation, numbers, as well as sorting multi-word phrases) is a crucial step in preparing the news corpus for further analysis.

Topic modeling

In applying topic models such as Latent Dirichlet Allocation (LDA), decisions regarding the main parameters and the number of topics (k) become very critical because they directly affect the interpretability and quality of the resulting topics. Empirical studies show that choosing a k that is too small can obscure different themes, while a k that is too large can produce overlapping and less meaningful topics. For example, research concludes that there is a relationship between corpus size and the optimal number of topics following the semi-power-law, where the larger the corpus and the longer the document, the more valid topics tend to increase.

To ensure the validity of the model, several metrics such as topic coherence, topic similarity, and model complexity penalty have been developed; one study introduced the singular Bayesian information criterion (sBIC) for more systematic selection of k . In the context of economic news research with 1422 documents from long-term local media portals like this, the recommended procedure is: (1) test the model with a range of k for example 4 to 12 topics; (2) choose k that provides a balance between topic readability and thematic coherence; (3) verifying the results through manual interpretation and topic consistency with the domain context (Maier et al., 2021).

After the parameters are determined, topic interpretation is carried out by exploring the dominant keywords in each topic, looking at examples of news fragments with high topic weight, and naming the topics according to their thematic content (for example: "Government Support", "Post-Pandemic Recovery", "Financing Challenges"). This allows further analysis of changes in the proportion of topics temporally and their relationship to MSME income indicators.

Lexicon-based sentiment for MSMEs and the crisis

The process of compiling a domain-specific sentiment dictionary, in this case for the context of MSMEs and economic crises such as pandemics, is an important stage for accurate sentiment analysis. Research in the field of finance shows that domain-aware sentiment dictionaries (for example, a lexicon that accommodates phrases such as "profit decreases") can provide more accurate classification results than generic lexicons. For example, Park, Lee & Cho (2021) developed a contextual lexicon that includes «direction-dependent words» to capture changes in word meaning in a financial context.

In the context of analyzing economic and MSME reporting in local media, the steps in compiling a lexicon include: identification of initial keywords related to MSMEs (e.g. "contribution", "growth", "decrease in income", "support", "liquidity"), categorization of words as positive (e.g. "growth", "success"), negative (e.g. "failure", "threatened"), or neutral/crisis

(e.g. "emergency", "emergency policy"). Empirical studies show that SME/MSME sector-specific dictionaries increase relevance because the terms used by the media are different from general terms (Schwaiger et al., 2016).

Then, after preparing keywords and categories, validation is carried out through reliability testing: for example taking news samples and checking whether the sentiment scores match the manual interpretation. This follows the practice recommended by the text mining literature that sentiment dictionaries should be tested for local contexts so as not to introduce meaning bias (Kobayashi et al., 2018). Thus, the resulting sentiment dictionary will be used in calculating the aggregate sentiment score per news, which is then combined with topics and MSME performance indicators for further analysis.

RESULTS AND DISCUSSION

Comparison of topic intensity in Covid and non-Covid news

The news pattern in the Figure 1 shows that the proportion of MSME news related to Covid was very dominant in the early phase of the pandemic, especially throughout 2020 to early 2021. In this period, the share of Covid news often exceeded 40 to 60 percent of the total monthly articles. This dominance reflects the media's focus on the direct impact of the pandemic on MSMEs, especially the issue of activity restrictions, reduced income, and various government emergency policies.

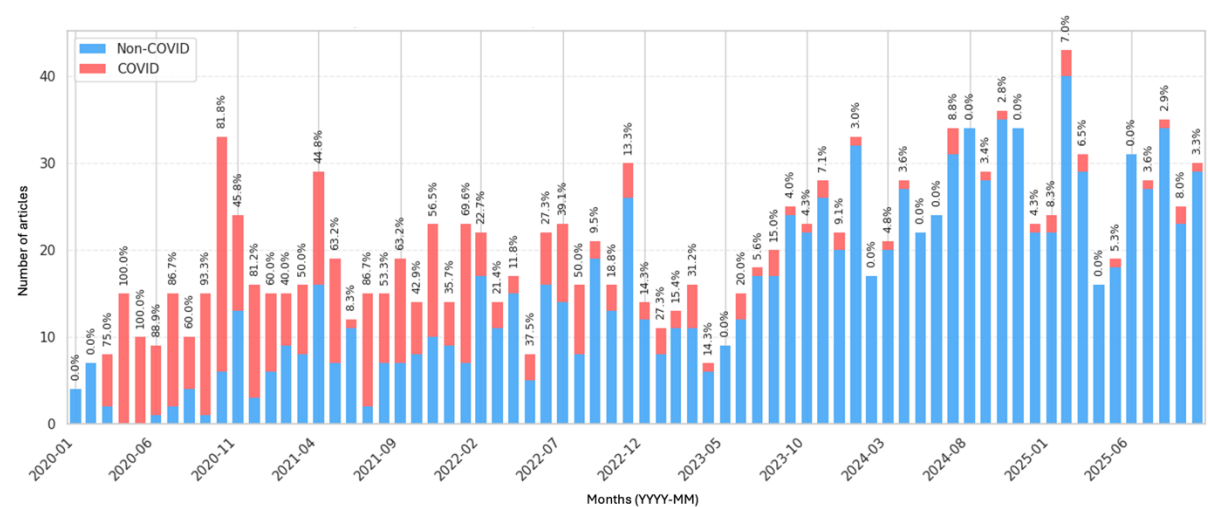


Figure 1. Covid vs Non-Covid News Comparison

Entering mid-2021, the proportion of news related to Covid began to decline and stabilized at a low level. This can be seen from the decreasing intensity of the red color on the graph, coinciding with the increase in the portion of non-Covid news. In this phase, the media began to shift attention to issues of recovery, financing support, digital adoption, and post-pandemic MSME empowerment.

From 2022 to 2025, MSME news will be almost completely dominated by the non-Covid category. The proportion of Covid news tends to be below 5 percent, with some small peaks usually reflecting certain moments, such as changes in health policy or continued aid programs. The dominance of non-Covid news in this period indicates the normalization of

economic conditions and the media's focus returning to structural issues for MSMEs such as growth, markets, innovation and competitiveness.

Topic Interpretation and Comparison of Covid vs Non-Covid Intensity

The six topics that emerged from the modeling provide a picture that is consistent with the reporting characteristics of West Sumatra MSMEs (Table 1). A comparison of the average proportion of topics for Covid and non-Covid news shows that there are differences in the pattern of issue emphasis between the crisis period and the recovery period.

Table 1. LDA Topics Comparison

Topics	COVID	Non-COVID	Overall Score
topic_0: nagari jalan dharmasraya bupati kabupaten bangun tol tanah dukung laksana	0.088	0.137	0.125
topic_1: kembang barat budaya bangun indonesia giat sumatera daya program gubernur	0.123	0.222	0.198
topic_2: padang semen giat halal program pt wali sehat pemko gelar	0.204	0.172	0.180
topic_3: koperasi indonesia andre bantu kerja menteri ketua latih milik kembang	0.192	0.162	0.170
topic_4: payakumbuh pasar pariaman produk wisata rendang dagang pariwisata harga giat	0.153	0.177	0.171
topic_5: bank persen tumbuh uang triliun nagari syariah kredit salur capai	0.240	0.130	0.156

These topics, which were obtained through topic modeling using the Latent Dirichlet Allocation (LDA) algorithm, were applied to 1422 MSME news stories from Langgam.id. All news texts are first processed through cleaning, normalization and tokenization stages to produce a corpus that is ready for analysis. After that, the LDA model was trained by testing several numbers of topics until a theme structure was obtained that was the most stable and easy to interpret. Each topic that appears is represented by a set of words with the highest probability, then labeled based on its dominant word pattern. The proportion of topics in each news story was then calculated, making it possible to compare the contribution of each topic to Covid and non-Covid news.

1. Topic 0 – Infrastructure and local government

The non-Covid average is higher than Covid. This topic contains words such as nagari, road, district, wake, toll, land. These issues are more dominant in non-Covid news because infrastructure coverage usually increases in the recovery period when public attention returns to regional economic development. The lack of focus on this topic in Covid news shows that during the crisis the media was more absorbed by the issue of handling the pandemic and its impact on MSMEs.

2. Topic 1 – Empowerment, government programs, and local economic culture

The non-Covid average is much higher. This shows that development programs and activities such as active, program, culture, power, governor are reported more often in the non-Covid phase. During the pandemic, many field activities and public programs were hampered so that they did not appear as the main narrative.

3. Topic 2 – Activities of major cities and certain industrial sectors

This topic has a slightly higher Covid rate than non-Covid. The appearance of words such as Padang, cement, halal, municipal government, healthy, active indicates a focus on city government efforts, public health and industrial adjustments to pandemic conditions. This is natural because large cities are the center of the initial impact of the pandemic and the location of adaptive policies for the business sector.

4. Topic 3 – Cooperatives, aid and the role of ministries

This topic is also a bit higher on Covid news. The dominance of words such as cooperative, help, minister, work, train shows that during the pandemic, the news more often highlighted government assistance, adaptive training and structural support for MSMEs. The media tends to highlight narratives of policy responses and interventions during times of crisis.

5. Topic 4 – Local trade, markets, culinary, tourism

The non-Covid average is higher than Covid. Keywords such as payakumbuh, market, tourism, rendang, price, trade point to traditional economic activities and tourism which have only recently recovered after mobility restrictions were relaxed. During the pandemic, the tourism sector and traditional markets stopped so that the frequency of news in this category decreased.

6. Topic 5 – Finance, banking, credit, money growth

This topic has the highest proportion of Covid news. The appearance of the words bank, percent, growth, credit, nagari, sharia shows that during the pandemic the media focused on issues of financing, business credit, economic stimulus, as well as financial strategies to maintain MSMEs during the shock. The increase in the proportion of financial topics in the crisis phase is consistent with the dynamics of liquidity pressure and the need for capital assistance.

The general pattern emerging from the topic comparison shows a clear distinction between the pandemic and non-pandemic phases. During the COVID-19 period, the dominant topics were Topics 2, 3, and especially 5, which reflect a crisis narrative with a focus on government support, financing, and regional responses to the pandemic disruption. Meanwhile, in non-COVID news, the most frequent topics were Topics 0, 1, and 4, reflecting a narrative of development, traditional economic activities, regional government programs, and more structural economic issues. This shift indicates that during the pandemic, the media focused more on issues of aid, financing, health, and business adjustments, while during the recovery period, the narrative shifted toward development, tourism, culture, and regular economic activities, signaling a return to normalcy in the regional economy.

Sentiment Analysis with a Lexicon Approach

The lexicon approach is used to measure the emotional tone of MSME news by utilizing a list of words that have been assigned a sentiment score. Each word in the news that matches the dictionary is assigned a positive, negative, or neutral score according to a predetermined category. The dictionary you use is specifically designed for the context of MSMEs and crises, so it can capture the nuances of economic recovery and business pressures more accurately than a generic lexicon.

This dictionary is divided into four main groups:

1. Positive – general, containing words such as successful, developing, superior, smooth, and productive, which describe progress or good business conditions.

2. Positive – crisis recovery, including words such as recover, rise, and resilient, which reflect efforts to emerge from the pressures of the pandemic.
3. Negative – general, such as loss, difficulty, obstacle, constraint, and decline, which indicate daily operational challenges.
4. Negative – crisis, such as pandemic, slump, stagnant, and shake, which indicate the significant pressures that arise during a crisis.

The scoring process, as shown in Table 2 is carried out by summing all the word values in a single news item, resulting in a sentiment score per article. Positive scores indicate news that is optimistic or supportive of MSME recovery, while negative scores indicate narratives of stress, loss, or crisis. Because the analysis period spans the pandemic and post-pandemic period, this approach allows for the identification of patterns of emotional shifts in news coverage along with changing economic conditions.

Table 2. Lexicon scoring

Category	Words and Scoring
Positive (general)	'positif': 2, 'sukses': 3, 'maju': 2, 'baik': 1, 'bantuan': 1, 'berkembang': 2, 'mudah': 1, 'unggul': 3, 'mendukung': 1, 'naik': 1, 'lancar': 1, 'berhasil': 2, 'kerjasama': 1, 'kuat': 2, 'mandiri': 3, 'komitmen': 1, 'syukur': 3, 'andalan': 2, 'tingkat': 1, 'cipta': 2, 'optimal': 2, 'pesat': 3, 'gizi': 2, 'kolaborasi': 1, 'sinergi': 2, 'terampil': 2, 'luas': 1, 'untung': 2, 'akses': 1, 'saing': 1, 'sehat': 1, 'produktif': 2, 'inovatif': 2, 'apresiasi': 1, 'kontribusi': 1, 'tangguh': 2, 'efektif': 2, 'solusi': 2
Positive (crisis recovery)	'pulih': 3, 'bangkit': 3, 'terjang': 2, 'tahan': 2, 'berantas': 2, 'lepas': 1
Negative (general)	'negatif': -2, 'susah': -2, 'sulit': -2, 'rugi': -3, 'kurang': -1, 'turun': -1, 'hambatan': -2, 'masalah': -2, 'tantangan': -1, 'terkendala': -2, 'buruk': -2, 'sepi': -1, 'minim': -2, 'kendala': -2, 'hadap': -1, 'jerat': -3, 'senjang': -2, 'risiko': -1, 'gagal': -3
Negative (crisis)	'puruk': -3, 'pandemi': -2, 'stagnan': -2, 'goncang': -2, 'paksa': -2, 'miskin': -2, 'rentenir': -3, 'rintang': -2, 'imbas': -1, 'batas': -1

This lexicon approach is then used to provide a quantitative picture of how the media frames the condition of MSMEs from the crisis period to recovery as in the Figure 2.

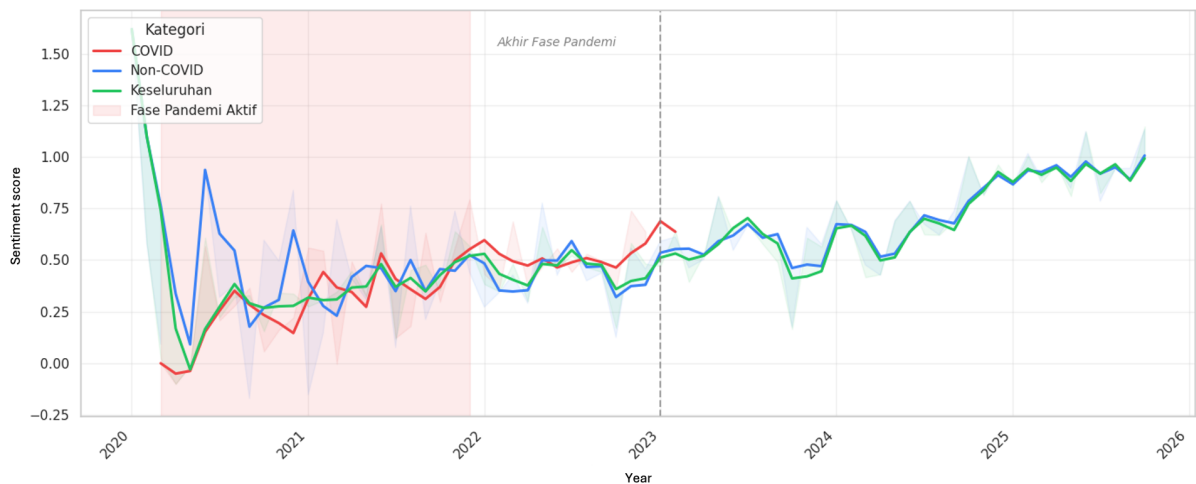


Figure 2. Lexicon-based sentiment trends

Figure 2 shows the development of MSME news sentiment from 2020 to 2025 using previously compiled lexicon scores. In general, the sentiment pattern exhibits three main phases: the initial pandemic shock phase, the gradual recovery phase, and the post-pandemic stable phase.

1. Initial Pandemic Shock Phase (2020 – early 2021)

In early 2020, the sentiment line—for both Covid-related and non-Covid-related news, and overall—experienced a sharp decline. This was due to the high occurrence of negative words in the lexicon, such as pandemic, loss, difficulty, stagnation, and constraints, reflecting the significant pressure on MSMEs at the beginning of the crisis.

Although there was a slight positive spike in the first few months (usually related to aid, support, or government programs), the average sentiment remained low due to the dominance of the crisis narrative.

2. Gradual Recovery Phase (2021 – 2023)

Entering mid-2021, the sentiment line began to rise slowly as news coverage became more balanced between negative words and recovery terms such as "recover," "rise," "resilient," and "solutions." During this phase, Covid sentiment began to stabilize, and while still fluctuating, its movements tended to follow non-Covid sentiment patterns. Media also increasingly highlighted issues of MSME adaptation, capital assistance, and various training programs supporting business adjustments. Furthermore, more optimistic words such as "developing," "innovative," and "synergize" emerged, signaling a strengthening of hopes for MSME recovery. This phase lasted until early 2023, indicated by the vertical line on the graph marking the transition to post-pandemic conditions.

3. Post-Pandemic Phase (2023 – 2025)

After 2023, sentiment increased more steadily and strongly, especially in non-Covid news. This trend reflected a shift in the narrative toward rebuilding, growth, and the normalization of business activities. The dominance of positive words such as progress, excellence, smoothness, effectiveness, and contribution led to a consistent increase in the average sentiment score.

From 2024 to 2025, the sentiment line showed a clear upward trend, in line with increasing regional economic optimism, MSME programs, capacity building, and business achievements. During this period, remaining COVID-19 news barely affected the overall trend.

MSME Revenue and News Sentiment (2019–2023)

The visualization shows two parallel patterns since the beginning of the pandemic, the trend in MSME revenue in West Sumatra (BPS) and the trend in average MSME news sentiment. Although both come from different data sources, their directional trends indicate a thematic connection.

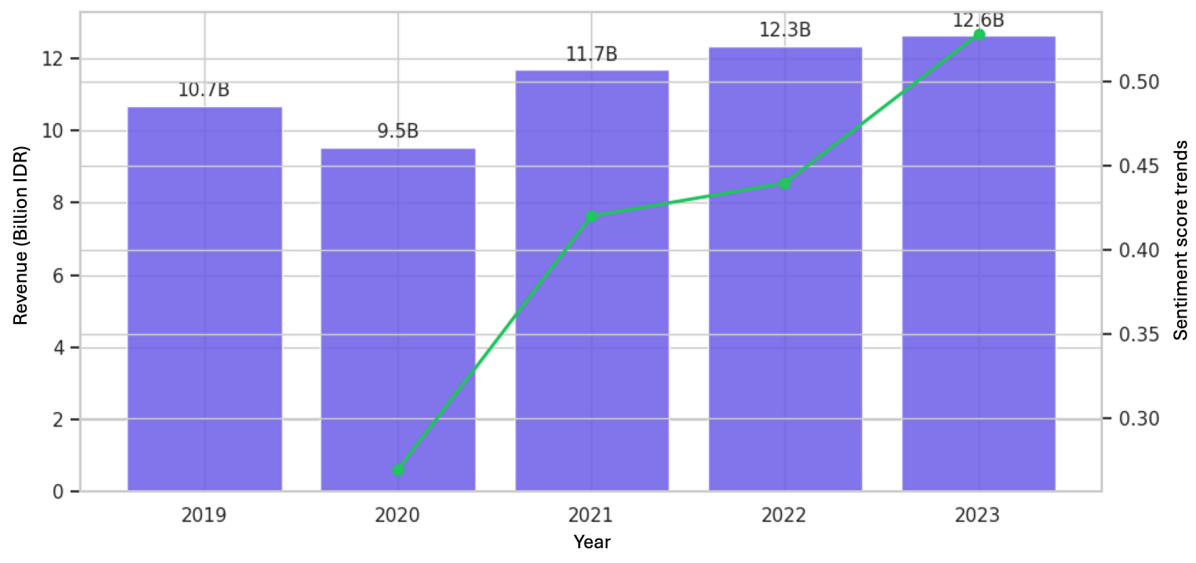


Figure 3. West Sumatra MSME's revenue vs sentiment score

1. Decline in the Early Pandemic Phase (2019–2020)

MSME revenue fell significantly from 10.68 trillion rupiah in 2019 to 9.55 trillion rupiah in 2020. This year was marked by economic contraction due to activity restrictions and demand shocks. During the same period, sentiment scores also reached their lowest point. This reflects the dominance of negative words in news coverage, such as loss, difficulty, obstacle, pandemic, and hurdle, indicating how the media captured the crisis atmosphere on the ground. This directional correlation indicates that the decline in MSME economic performance is reflected in a more pessimistic tone of news.

2. Recovery and Shifting News Tone (2021)

In 2021, revenue increased to 11.69 trillion rupiah, signaling the beginning of economic recovery. Correspondingly, the annual sentiment score rose significantly. This was primarily due to the increased frequency of positive words associated with recovery, such as "recover," "rise," "aid," "synergize," and "resilient." The media began reporting on MSME adaptation efforts, government programs, digital training, and improving domestic demand. This shift indicates that the improving economic conditions of MSMEs are beginning to be clearly reflected in news coverage.

3. Recovery Consolidation (2022–2023)

MSME revenue continued to rise to 12.34 trillion rupiah in 2022 and 12.63 trillion rupiah in 2023. During the same period, the annual sentiment score continued to improve and stabilize. Media narratives began to focus more on growth, innovation, collaboration, increased market access, and the return to normal regional economic activity. The tone of the news during this phase consistently reflected growing optimism, in line with the continued recovery of MSME revenues.

4. Medium-Term Trend: Sentiment as a Reflection of Field Conditions

From 2019 to 2023, the two curves move in parallel; when revenues decline, sentiment decreases, and when revenues rise, sentiment increases. Although sentiment is not an economic indicator, this pattern suggests that the media tends to follow the real dynamics of MSMEs: negative narratives intensify during times of stress, while positive narratives increase when recovery begins to emerge. This consistent pattern reinforces the interpretation that sentiment analysis can be a complementary indicator for assessing the condition of regional MSME business activities.

Discussion

The results of this study demonstrate that the dynamics of media narratives regarding MSMEs in West Sumatra can be interpreted through three main elements: topic patterns, shifts in sentiment, and changes in economic performance. All three demonstrate a fairly consistent linear relationship throughout the 2020–2025 period. These findings provide several important insights for the development of theory and practice.

1. Media as a Reflector of Regional Economic Conditions

The emerging topic patterns indicate that local media actively framed the conditions of MSMEs according to the development phase of the pandemic. During the crisis phase, the most dominant topics focused on financing issues, government assistance, business disruptions, and institutional responses. During the recovery phase, the focus shifted to program activities, strengthening regional institutions, tourism, and infrastructure development as a support for economic recovery. This shift confirms that the media does not operate in a vacuum but rather responds to the structural dynamics occurring in the regional economy. Thus, the media can be understood as a fairly accurate reflector of MSME conditions and can be used to monitor the crisis, transition, and recovery phases of the economy.

2. Media Sentiment Represents the MSME Recovery Cycle

Sentiment analysis reveals a very clear pattern: negative sentiment dominated in 2020, began to rise in 2021, and has remained stable and positive since 2022. This shift parallels the increase in MSME revenue, particularly as revenue recovered from 9.55 trillion rupiah (2020) to 11.69 trillion rupiah (2021), and continued to rise to 12.63 trillion rupiah (2023). From an interpretive perspective, media sentiment can be seen as a reflection of collective confidence in MSMEs' ability to survive and recover. The increasingly positive tone of news indicates growing public and institutional optimism about improving economic conditions.

3. The Relationship Between Topics – Sentiment – Revenue: Evidence of Consistency in Economic Narratives

The integration of these three dimensions reveals a mutually reinforcing pattern: (1) when crisis topics dominate, negative sentiment increases and income decreases; (2) when recovery topics and government programs increase, sentiment gradually becomes positive and income begins to recover; (3) When growth and development topics dominate, positive sentiment remains stable, and incomes consistently increase.

This consistency provides the basis for the understanding that media narrative dynamics not only reflect conditions but can also provide signals about the current economic phase. This opens up the opportunity to utilize text mining as a tool to monitor regional economic health.

CONCLUSION

This study shows that the dynamics of media narratives regarding MSMEs in West Sumatra move in line with regional business conditions, where changes in topic proportions and shifts in sentiment reflect the crisis, transition, and recovery phases experienced by businesses from 2020 to 2025. An integrative analysis of topics, sentiment, and revenue confirms that local media not only captures developments but also reflects shifts in collective perceptions of MSME resilience and recovery. These findings provide a novel contribution by demonstrating that text mining in regional news coverage can be a supporting indicator in assessing the temporal health of MSME economies. Furthermore, the innovation of a specific MSME-crisis lexicon and topic-based narrative mapping provides an analytical framework that can be applied to monitor business sector dynamics in other regions. In practice, the results of this study can be used by local governments to assess public perception of MSME policies and develop more responsive communication strategies. They can also help MSMEs understand prominent issues in the media and adjust their business adaptation and promotion strategies.

ACKNOWLEDGEMENT

This research was funded by **Universitas Andalas** under the **Riset Dosen Pemula (RDP)** research grant, Contract No. **T/29/UN.16.17/PT.01.03/Soshum-RDP/2020**, Fiscal Year 2020.

REFERENCE

- Aisyah, H., Puspita, S., & Elizamiharti, E. (2022). Resource-Based View: Strategi Umkm Di Sumatera Barat Untuk Mencapai Keunggulan Kompetitif. *JMD: Jurnal Riset Manajemen & Bisnis Dewantara*, 5(2), 109–120.
- Alshantik, F., Apon, A., Herzog, A., Safro, I., & Sybrandt, J. (2020). Accelerating text mining using domain-specific stop word lists. *2020 IEEE International Conference on Big Data (Big Data)*, 2639–2648. <https://ieeexplore.ieee.org/abstract/document/9378226/>
- Atanassova, I., & Clark, L. (2020). Social media use as an enabler of marketing evolution in knowledge-intensive SMEs. In *Handbook of Research on Innovations in Technology and Marketing for the Connected Consumer* (pp. 39–61). IGI Global Scientific Publishing. <https://www.igi-global.com/chapter/social-media-use-as-an-enabler-of-marketing-evolution-in-knowledge-intensive-smes/239496>
- Ausat, A. M. A. (2023). The role of social media in shaping public opinion and its influence on economic decisions. *Technology and Society Perspectives (TACIT)*, 1(1), 35–44.

- Bholat, D., Hansen, S., Santos, P., & Schonhardt-Bailey, C. (2015). Text mining for central banks: Handbook. *Centre for Central Banking Studies Handbook*, 33, 1–19.
- Boumans, D., Müller, H., & Sauer, S. (2023). How media content influences economic expectations: Evidence from a global expert survey. *Journal of Forecasting*, 42(6), 1295–1308.
- Brown, K., Jie, F., Le, T., Sharafizad, J., Sharafizad, F., & Parida, S. (2022). Factors impacting SME business resilience post-COVID-19. *Sustainability*, 14(22), 14850.
- Erlanitasari, Y., Rahmanto, A., & Wijaya, M. (2019). Digital economic literacy micro, small and medium enterprises (SMES) go online. *Informasi*, 49(2), 145–156.
- Fu, C.-J., Silalahi, A. D. K., Yang, L.-W., & Eunike, I. J. (2024). Advancing SME performance: A novel application of the technological-organizational-environment framework in social media marketing adoption. *Cogent Business & Management*, 11(1), 2360509. <https://doi.org/10.1080/23311975.2024.2360509>
- Georgieva-Trifonova, T., & Dechev, M. (2021). Applying text mining methods to extracting information from news articles. *IOP Conference Series: Materials Science and Engineering*, 1031(1), 012054. <https://iopscience.iop.org/article/10.1088/1757-899X/1031/1/012054/meta>
- Hossain, A., Karimuzzaman, M., Hossain, M. M., & Rahman, A. (2021). Text mining and sentiment analysis of newspaper headlines. *Information*, 12(10), 414.
- Janah, U. R. N., & Tampubolon, F. R. S. (2024). Peran usaha mikro, kecil, dan menengah dalam pertumbuhan ekonomi: Analisis kontribusi sektor umkm terhadap pendapatan nasional di indonesia. *PENG: Jurnal Ekonomi Dan Manajemen*, 1(2), 739–746.
- Jiang, H., & Man, Y. (2025). *Construct to Commitment: The Effect of Narratives on Economic Growth* (No. arXiv:2504.21060). arXiv. <https://doi.org/10.48550/arXiv.2504.21060>
- Kobayashi, V. B., Mol, S. T., Berkers, H. A., Kismihók, G., & Den Hartog, D. N. (2018). Text mining in organizational research. *Organizational Research Methods*, 21(3), 733–765.
- Kraus, P., Stokes, P., Cooper, S. C., Liu, Y., Moore, N., Britzelmaier, B., & Tarba, S. (2020). Cultural Antecedents of Sustainability and Regional Economic Development—A Study of SME ‘Mittelstand’ Firms in Baden-Württemberg (Germany). *Entrepreneurship & Regional Development*, 32(7–8), 629–653. <https://doi.org/10.1080/08985626.2020.1713223>
- Maier, D., Waldherr, A., Miltner, P., Wiedemann, G., Niekler, A., Keinert, A., Pfetsch, B., Heyer, G., Reber, U., & Häussler, T. (2021). Applying LDA topic modeling in communication research: Toward a valid and reliable methodology. In *Computational methods for communication science* (pp. 13–38). Routledge.
- Meier, A., & Peters, M. (2023). Limited engagement of SMEs with social media: A structuration and sensemaking perspective. *Information & Management*, 60(7), 103853.
- Novitasari, A. T. (2022). Kontribusi umkm terhadap pertumbuhan ekonomi era digitalisasi melalui peran pemerintah. *JABE (Journal of Applied Business and Economic)*, 9(2), 184–204.
- Nugroho, S. A. (2012). The economic development and the growth of small-medium enterprises in Indonesia: A hometown investment trust fund approach. *Integration*, 15, 171–193.
- Panjaitan, J. M. P., Darwin, M., Sumiyana, S., Wiva, W., & Setyowati, S. M. (2025). Measuring the Indonesian SMEs’ capabilities for sociodynamic and disruptive innovation: The case for gaining resilience and sustainability. *Journal of Innovation and Entrepreneurship*, 14(1), 130. <https://doi.org/10.1186/s13731-025-00557-x>

- Park, J., Lee, H. J., & Cho, S. (2021). Automatic construction of context-aware sentiment lexicon in the financial domain using direction-dependent words. *arXiv Preprint arXiv:2106.05723*.
- Pohlmann, J. (2023). Text data mining on current newspaper articles from the United States with ProQuest TDM Studio. *Publizistik*, 68(2–3), 305–324. <https://doi.org/10.1007/s11616-023-00805-8>
- Puslit DPR RI. (2021). *Impacts of the COVID-19 pandemic on the micro, small and medium-sized enterprises (MSMEs)*. Research Center DPR RI.
- Saari, M., Haapanen, L., & Hurmelinna-Laukkanen, P. (2022). Social media and international business: Views and conceptual framing. *International Marketing Review*, 39(7), 25–45.
- Schwaiger, J. M., Lang, M., Ritter, C., & Johannsen, F. (2016). *Assessing the accuracy of sentiment analysis of social media posts at small and medium-sized enterprises in Southern Germany*.
- Sodikin, A., & Rini, N. (2020). Framing Analysis of the Indonesian Government's Halal Policy Reports in the Bbc Mass Media. *Policy*, 6(6), 282–296.
- Surya, B., Menne, F., Sabhan, H., Suriani, S., Abubakar, H., & Idris, M. (2021). Economic growth, increasing productivity of SMEs, and open innovation. *Journal of Open Innovation: Technology, Market, and Complexity*, 7(1), 20.
- Tempo. (2020). *BPS: 84% Micro Businesses Suffer Revenue Decline due to Covid-19—News En.tempo.co*. Tempo. <https://en.tempo.co/read/1386624/bps-84-micro-businesses-suffer-revenue-decline-due-to-covid-19>
- UNIDO. (2020). *Summary of COVID-19 Impact Assessment on SME in Indonesia*. United Nations Industrial Development Organization.
- Yang, M., Jiang, B., Wang, Y., Hao, T., & Liu, Y. (2022). News text mining-based business sentiment analysis and its significance in economy. *Frontiers in Psychology*, 13, 918447.
- Zhang, Y., Lin, Z., Tong, C. C., & Ho, S. W. (2025). Enhancing tokenization accuracy with dynamic patterns: Cumulative logic for segmenting user-generated content in logographic languages. *Journal of Computational Social Science*, 8(3), 80. <https://doi.org/10.1007/s42001-025-00406-7>