Jurnal Economia, Vol. 22, No. 3, October 2026, 358-377 P-ISSN: 1858-2648
Website: https://journal.uny.ac.id/index.php/economia E-ISSN: 2460-1152

Google Trends-Based Demand Forecasting for Indonesian MSMEs: A Comparative Model Evaluation

Andrew Ridow Johanis Metekohy[1*], Arham Arifin[2]
[1]Universitas Dipa Makassar, Indonesia
[2]Universitas Dipa Makassar, Indonesia
[1]andrew@undipa.ac.id, [2]arham.arifin@undipa.ac.id
**Corresponding Author*

Abstract
This study presents the first systematic comparative evaluation of eight forecasting models applied to Google Trends data specifically for Indonesian digital MSME product categories, addressing a critical gap in accessible demand prediction tools for resource-constrained enterprises. Employing quantitative comparative experimental design, we analyzed 18 product keywords from fashion and traditional food sectors with weekly temporal resolution (2020-2024), evaluated via 80/20 train-test split with MAPE. Exponential Smoothing achieved superior overall performance (20.33% MAPE), while Moving Average proved optimal for traditional food products (16.21% MAPE). Category-specific model selection enables MSMEs to anticipate demand peaks, time market entry windows, optimize inventory positioning ahead of seasonal cycles, and prioritize promotional spending, translating search trend signals directly into actionable penetration decisions. These findings provide an evidence-based, zero-cost framework for demand-informed market strategy, accessible to resource-constrained enterprises without specialized infrastructure or advanced technical expertise.

Keywords: Google Trends; demand forecasting; SME; exponential smoothing; market penetration

Pemodelan Prediktif berbasis Data Google Trends untuk Strategi Penetrasi Pasar UMKM Digital

Abstrak
Penelitian ini mengevaluasi delapan metodologi pemodelan prediktif untuk peramalan permintaan pasar UMKM Indonesia menggunakan data Google Trends dalam mendukung keputusan strategi penetrasi pasar. Menggunakan desain eksperimental komparatif kuantitatif, kami menganalisis 18 kata kunci produk dari sektor fashion dan makanan tradisional dengan resolusi temporal mingguan (2020-2024). Model yang diuji meliputi Naive baseline, Moving Average, Regresi Linear, ARIMA, Exponential Smoothing, dan tiga variasi ensemble ARIMA-ES, dievaluasi menggunakan pembagian train-test 80/20 dengan MAPE sebagai metrik utama. Hasil menunjukkan Exponential Smoothing mencapai kinerja terbaik (20,33% MAPE), mengungguli Moving Average (22,74%), metode ensemble (22,85-25,90%), dan ARIMA (33,30%). Analisis spesifik kategori mengungkapkan efektivitas diferensial: Moving Average optimal untuk produk makanan tradisional (16,21% MAPE) sementara Exponential Smoothing unggul untuk produk fashion (18,87% MAPE). Temuan memberikan kerangka praktis untuk peramalan permintaan UMKM spesifik kategori yang mendukung pengembangan strategi penetrasi pasar.

Kata kunci: Google Trends; peramalan permintaan; UMKM; exponential smoothing; penetrasi pasar

History: Received: 6 October 2025 Revised: 5 May 2026 Accepted: 9 May 2026
Citation (APA 7th): Metekohy, A. R. J., & Arifin, A. (2026). Google trends-based demand forecasting for Indonesian MSMEs: A comparative model evaluation. *Jurnal Economia*, 22(3), 358–377. https://doi.org/10.21831/economia.v22i3.90253

INTRODUCTION

Micro, Small, and Medium Enterprises (UMKM) constitute the backbone of Indonesian economy, contributing more than 60% to the national GDP and absorbing nearly 97% of the workforce. With over 64 million business units nationwide (Kementerian Koordinator Bidang Perekonomian Republik Indonesia, 2025), UMKM sector has demonstrated remarkable resilience and growth potential. Digital transformation has become increasingly critical for UMKM competitiveness, with 25.5 million enterprises having adopted digital ecosystems by July 2024 (Lusa et al., 2024).

Effective market penetration strategies for digital SMEs require accurate demand forecasting capabilities to optimize inventory management, promotional timing, and resource allocation decisions (Boone et al., 2018). Strategic market entry and expansion initiatives depend fundamentally on understanding anticipated consumer demand patterns to minimize investment risks and maximize market opportunities (Rahman, 2025). However, resource-constrained SMEs typically lack access to sophisticated demand prediction tools that larger enterprises employ for strategic planning, creating competitive disadvantages in increasingly data-driven digital marketplaces.

Despite increasing digital adoption rates, significant gaps persist in technology and data utilization among UMKM operators (Pattiasina, 2025; D. Ramadhani et al., 2025; N. Ramadhani et al., 2025). Operationally, many Indonesian MSMEs continue to rely on manual inventory recording systems that are prone to errors and delays in information access, limiting their capacity for timely market decisions (Trisnawati & Suryaningsih, 2025). The adoption of digital management information systems remains low due to financial constraints, limited technology infrastructure, and inadequate digital literacy among MSME operators (Rahayu & Veri, 2025), leaving most businesses without structured mechanisms for demand estimation or inventory planning. In the absence of reliable forecasting tools, MSMEs face recurring consequences: overstocking during low-demand periods ties up limited working capital, while understocking during peak seasons results in lost sales at precisely the moment when market penetration opportunities are greatest (Chaidir et al., 2024). Business transformation in the digital era requires implementation of data-driven strategies to optimize decision-making processes (Ramadhi et al., 2023), yet accessible predictive modeling methodologies supporting market strategy development remain limited for resource-constrained SME contexts.

Online search data represents a valuable information source for understanding market behavior (Wardani et al., 2025). Google Trends provides insights into consumer search patterns that can potentially inform SME market penetration strategies without requiring expensive market research or point-of-sale data systems (Praharjo et al., 2023; Putra et al., 2025). Google Trends-based predictive modeling offers promising opportunities for demand-informed strategic planning, enabling SMEs to anticipate market opportunities and optimize resource deployment. However, systematic evaluation of predictive modeling approaches optimized for Google Trends data in Indonesian SME contexts remains limited. While previous research has demonstrated Google Trends' potential for economic

prediction (Choi & Varian, 2012), comparative studies identifying optimal modeling methodologies for supporting Indonesian UMKM market penetration strategies are notably absent from existing literature.

Prior studies on Google Trends-based forecasting have focused predominantly on developed-market contexts and single-sector applications, without systematically comparing multiple predictive modeling methodologies for Indonesian MSME product categories (Boone et al., 2018; Carta et al., 2019; Choi & Varian, 2012; Jun et al., 2018). Similarly, research on Indonesian SMEs has emphasized digital marketing adoption and general digital transformation pathways without evaluating demand prediction methodologies using publicly accessible search data (Priyono et al., 2020; Purwana et al., 2017). This dual gap creates practical uncertainty for MSME operators regarding which forecasting approach best supports market entry, inventory, and promotional decisions given their resource constraints. This research addresses the gap through systematic comparative evaluation of eight predictive modeling approaches applied to Google Trends data for Indonesian MSME products across two product categories, answering the question: Which predictive modeling methodology provides optimal performance for Indonesian MSME demand forecasting using Google Trends data, and does optimal model selection vary by product category?

The research contributes to digital UMKM market penetration strategy development through: (1) empirical comparison of accessible predictive modeling methods using only publicly available data, (2) category-specific forecasting guidelines enabling informed market entry and expansion decisions, and (3) practical framework for demand-informed strategic planning accessible to resource-constrained contexts. Unlike previous studies emphasizing complex hybrid methodologies, this research prioritizes practical implementation feasibility while maintaining scientific rigor, enabling actionable demand prediction approaches supporting Indonesian SME market penetration initiatives.

LITERATURE REVIEW

Google Trends as a Predictive Data Source

The use of online search data for economic and demand prediction has been established since Choi & Varian (Choi & Varian, 2012) demonstrated that Google Trends indices can nowcast consumer behavior and economic activity before official statistics become available. Subsequent work by Jun et al. (Jun et al., 2018) extended this foundation to tourism demand forecasting across ten years of research, confirming that search interest data captures anticipatory consumer behavior with practical predictive value. These foundational studies establish Google Trends as a legitimate proxy for latent demand, though both operate in developed-market contexts with stable search behavior patterns that may differ from Indonesian digital commerce environments characterized by rapid adoption and heterogeneous user demographics.

Forecasting Methods in E-Commerce and Search Data Contexts

Comparative forecasting research has examined multiple methodological approaches for search-based demand prediction with mixed conclusions. Carta et al. (Carta et al., 2019) applied ARIMA to e-commerce price forecasting and demonstrated reasonable accuracy, while Boone et al. (Boone et al., 2018) found that Google Trends data meaningfully improves retail sales forecasts when incorporated into regression frameworks. However, neither study conducts systematic multi-model comparison nor addresses the specific challenge posed by Google Trends' normalized 0–100 index scale, which distorts the autocorrelation structures that ARIMA methods typically exploit. The M4 Competition results documented by Makridakis et al. (Makridakis et al., 2020) further suggest that simpler methods frequently outperform complex models in volatile time-series environments, a finding with direct implications for Google Trends-based forecasting but not yet empirically tested in Indonesian MSME contexts. The performance of ensemble approaches theoretically superior through error diversification (Bates & Granger, 1969), also remains unverified for normalized search index data where component models may capture overlapping rather than complementary patterns.

Digital Transformation and Demand Analytics in Indonesian MSMEs

Research on Indonesian SMEs has progressed along two largely separate tracks. The first examines digital adoption and marketing, documenting growing platform utilization and revenue effects of digitalization (Priyono et al., 2020; Purwana et al., 2017; Rizky & Permana, 2022). The second addresses operational challenges including limited financial literacy, infrastructure constraints, and low adoption of data management systems (Pattiasina, 2025; Rahayu & Veri, 2025). Critically, neither track systematically evaluates demand forecasting methodologies applicable within resource-constrained digital MSME operations. The intersection of publicly accessible search data and practical forecasting guidance for Indonesian MSME market strategy remains empirically unaddressed.

Synthesis and Positioning of the Present Study

Existing literature collectively establishes three conditions that motivate this research. First, Google Trends data has demonstrated predictive value in developed-market settings but lacks systematic comparative evaluation in Indonesian MSME contexts. Second, simpler forecasting methods may outperform complex alternatives for volatile normalized data, yet this proposition has not been tested across multiple product categories. Third, Indonesian MSMEs operate without accessible forecasting guidance despite documented operational consequences of demand uncertainty. This study occupies the intersection of these three gaps by providing the first systematic eight-model comparison using Google Trends data for Indonesian MSME product categories, generating category-specific guidance directly applicable to market penetration decision-making.

Hypothesis Development

Based on the theoretical arguments and empirical patterns identified in prior literature, two directional hypotheses are proposed to guide the comparative analysis.

Exponential Smoothing employs adaptive weighting that prioritizes recent observations while modeling trend and seasonal components simultaneously (R. Hyndman et al., 2008). For fashion products, which are characterized by pronounced seasonality tied to cultural events such as Ramadan, holiday cycles, and fashion industry rhythms, this structure provides a theoretical advantage over methods that apply equal weighting to recent and distant observations. Moving Average, by contrast, applies simple local averaging that smooths short-term fluctuations without imposing seasonal decomposition. For traditional food products with relatively stable search patterns driven by routine consumption rather than seasonal events, this smoothing property is theoretically preferable. The ensemble superiority hypothesis, derived from Bates and Granger (Bates & Granger, 1969), posits that combining forecasts reduces prediction error through error diversification. However, this advantage depends on genuine complementarity between component models; when both components capture similar underlying patterns through different parameterizations, ensemble combinations may provide limited marginal benefit (Makridakis et al., 2020). Based on these arguments, the following hypotheses are proposed:

H1: Exponential Smoothing outperforms all other evaluated models for fashion product demand forecasting using Google Trends data.

H2: Moving Average outperforms all other evaluated models for traditional food product demand forecasting using Google Trends data.

H3: Ensemble methods do not significantly outperform the best individual component model (Exponential Smoothing) in overall forecasting accuracy for Google Trends normalized data.

METHOD

Research Design and Approach

This study employs a quantitative comparative experimental design to systematically evaluate forecasting model performance for Indonesian SME demand prediction using Google Trends data (Choi & Varian, 2012). The research implements a comprehensive model comparison framework with standardized evaluation metrics to ensure objective performance assessment across multiple methodologies (R. J. Hyndman & Athanasopoulos, 2018).

The overall research methodology follows a three-phase systematic process: (1) data collection and quality validation, (2) baseline and advanced model implementation, and (3) comparative performance evaluation. This approach ensures reproducible and scientifically rigorous results aligned with established time-series forecasting research standards (Makridakis et al., 2020).

Research Subject and Data Characteristics

The research subject encompasses Google Trends search data for Indonesian SME products spanning January 2020 to December 2024. Data extraction utilized Google Trends API via PyTrends library (Massaron & Boschetti, 2016) to obtain normalized search volume indices with weekly temporal resolution, providing 262 weekly observations per keyword over the five-year period.

Product keyword selection focused on two major SME categories demonstrating consistent data quality and representativeness of Indonesian digital commerce landscape: fashion products and traditional food items. These categories were selected based on data validation criteria including temporal completeness, search volume stability, and coefficient of variation suitable for robust time-series modeling (Jun et al., 2018).

Table 1. *Dataset Characteristics*

Parameter	**Specification**
Time Period	January 2020 - December 2024
Temporal Resolution	Weekly
Total Keywords	18
Product Categories	2 (Fashion, Traditional Food)
Keywords per Category	9 each
Data Points per Keyword	262 weeks
Total Observations	4,716
Missing Values	0 (0.0%)
Value Range	0-100 (normalized index)
Geographic Scope	Indonesia (geo=ID)
Data Source	Google Trends API

The complete list of 18 product keywords used in this study is presented in Table 2, organized by product category. Each keyword represents a distinct product type with verified search activity within the Indonesian market during the observation period.

Table 2. *Product Keywords by Category*

No	**Keyword**	**Category**	**No**	**Keyword**	**Category**
1	Batik	Fashion	11	Gudeg	Traditional Food
2	Hijab	Fashion	12	Bakso	Traditional Food
3	Sepatu	Fashion	13	Soto	Traditional Food
4	Kemeja batik	Fashion	14	Nasi goreng	Traditional Food
5	Mukena	Fashion	15	Sate	Traditional Food
6	Sandal	Fashion	16	Martabak	Traditional Food
7	Tas wanita	Fashion	17	Gado-gado	Traditional Food
8	Kaos	Fashion	18	Pempek	Traditional Food
9	Celana jeans	Fashion			
10	Rendang	Traditional Food			

Keyword selection followed explicit inclusion criteria to ensure data quality and representativeness. First, each keyword was required to represent a product category commonly sold by Indonesian digital MSMEs and demonstrably searchable by Indonesian consumers on Google. Second, keywords were required to yield complete 262-week temporal coverage with zero missing values across the full observation period, ensuring model training and testing integrity. Third, keywords were excluded if their search volume was persistently low, defined as a Google Trends normalized index below 10 for more than 50% of observed weeks, as such sparsity would produce unreliable forecasting signal. Fourth, each keyword was verified to return geo=ID specific results rather than globally dominated search patterns, confirming Indonesian market relevance. Fashion keywords were selected to represent both traditional heritage items (batik, mukena) and contemporary everyday apparel with established digital commerce presence. Traditional food keywords were selected to represent regionally distributed Indonesian culinary products with consistent year-round search activity, spanning national dishes (rendang, bakso, nasi goreng) and regionally significant items (gudeg, pempek, gado-gado). Keywords were queried individually via PyTrends library using standardized parameters to prevent cross-keyword normalization effects inherent in batch querying.

Data Collection and Preprocessing

Data extraction employed PyTrends library (Python interface for Google Trends) with systematic query procedures to ensure consistency (Stephens-Davidowitz & Varian, 2015). Each keyword was queried individually using standardized parameters: timeframe='2020-01-01 2024-12-31', geo='ID', and default category settings. Rate limiting protocols implemented 12-second delays between queries to comply with Google Trends API usage guidelines.

Data Preprocessing Pipeline

Standard time-series preprocessing procedures were implemented following established best practices (Brockwell & Davis, 2006):

1. Completeness Validation: Verification of 262-week temporal coverage for each keyword, with zero missing values threshold enforced
2. Temporal Alignment: Standardization of date indices across all keywords to ensure synchronized weekly observations
3. Outlier Detection: Statistical review using interquartile range (IQR) method to identify anomalous values without removal, maintaining data integrity (Tukey, 1977)
4. Range Verification: Confirmation that all values conform to Google Trends normalized scale (0-100)
5. Category Assignment: Systematic labeling of keywords to respective product categories for stratified analysis

Forecasting Models Evaluated

The research implements a comprehensive comparison framework involving eight distinct forecasting methodologies, categorized into baseline models and advanced approaches:

Baseline Models (4):

1. Naive Baseline: Last-value persistence model serving as simplest benchmark (Makridakis et al., 2020)
2. Moving Average: Simple 4-week rolling average capturing local temporal patterns (Box et al., 2015)
3. Linear Regression: Time-indexed linear trend modeling implemented using scikit-learn library (Pedregosa et al., 2011)
4. ARIMA: AutoRegressive Integrated Moving Average with automatic parameter selection via grid search optimization (Box & Jenkins, 1970), implemented using statsmodels package (Seabold & Perktold, 2010)

Advanced Models (4):

5. Exponential Smoothing: Holt-Winters method with additive seasonal components (R. Hyndman et al., 2008), implemented via statsmodels with automatic parameter optimization
6. ARIMA-ES Ensemble (60-40): Weighted combination approach (Bates & Granger, 1969)
7. ARIMA-ES Ensemble (50-50): Equal weighting ensemble methodology
8. ARIMA-ES Ensemble (40-60): ES-dominant ensemble configuration

ARIMA parameter optimization employed exhaustive grid search within specified ranges $p \in [0,3]$, $d \in [0,2]$, $q \in [0,3]$, selecting configurations minimizing Akaike Information Criterion (Akaike, 1974). Exponential Smoothing implementation utilized automated parameter estimation with seasonal decomposition fallback procedures for keywords with insufficient seasonal signal.

Validation Framework and Evaluation Metrics

Train-Test Split Methodology:

Model validation employed standard chronological train-test splitting with 80%-20% proportion, preserving temporal sequence integrity essential for time-series forecasting evaluation (Tashman, 2000). For each keyword's 262-week series, the first 210 weeks constituted training data, with final 52 weeks reserved for out-of-sample testing.

Performance Evaluation Metrics:

Model performance was assessed using three complementary metrics following forecasting accuracy measurement standards (R. J. Hyndman & Koehler, 2006). Mean Absolute Percentage Error (MAPE) was selected as the primary evaluation metric due to its scale-independence, enabling meaningful aggregation and comparison across 18 product keywords with heterogeneous search volume distributions (Armstrong & Collopy, 1992). This scale-independence is particularly important in the present study, where Google

Trends normalized indices vary substantially across keywords and product categories, making absolute error metrics less suitable for cross-keyword aggregation.

MAPE's known limitations are acknowledged. The metric becomes unstable and potentially misleading when actual values approach zero, as percentage errors inflate disproportionately for near-zero observations (R. J. Hyndman & Koehler, 2006). Additionally, MAPE is asymmetric, penalizing over-forecasts more heavily than under-forecasts of equivalent magnitude (Makridakis et al., 2020). In the present study, these limitations are partially mitigated by two characteristics of Google Trends data: first, the normalized index rarely reaches true zero for established product keywords with consistent Indonesian search activity, as keywords failing the minimum volume threshold were excluded during the data validation stage; second, the 0-100 bounded scale limits extreme percentage error inflation that would otherwise occur with unbounded continuous variables.

To address residual concerns regarding MAPE's sensitivity, model rankings were cross-validated using Mean Absolute Error (MAE) as a secondary metric, which provides absolute error magnitude without percentage-based distortion, and Root Mean Square Error (RMSE) as a tertiary metric that penalizes larger prediction deviations more heavily than smaller ones. Consistency of model rankings across all three metrics was examined to assess robustness of conclusions. Where rankings diverged across metrics, this is explicitly noted in the results section to ensure transparent reporting. This multi-metric approach follows best practices recommended by Hyndman and Koehler (R. J. Hyndman & Koehler, 2006) for comparative forecasting studies where no single metric is universally optimal.

Statistical Significance Testing

To assess whether observed performance differences across forecasting models reflect genuine methodological superiority rather than random variation, the Friedman test was employed as the primary statistical comparison procedure. The Friedman test is a non-parametric rank-based test appropriate for comparing multiple forecasting methods across multiple time series, as it does not assume normality of error distributions and remains valid for small samples (Demšar, 2006). The test was applied across 18 keywords, treating each keyword as a block and model MAPE ranks as observations, following the approach recommended by Makridakis et al. (Makridakis et al., 2020) in the M4 Competition methodology.

Where the Friedman test indicated statistically significant overall differences ($p < 0.05$), post-hoc pairwise comparisons were conducted using the Wilcoxon signed-rank test with Bonferroni correction to control family-wise error rate across multiple comparisons (Demšar, 2006). Statistical significance was set at $\alpha = 0.05$ throughout. Effect sizes were reported using the matched-pairs rank biserial correlation coefficient to supplement p-values with practical significance indicators.

Robustness of findings was further assessed through two additional checks. First, model rankings were examined separately for fashion and traditional food categories to verify consistency of category-specific conclusions. Second, sensitivity analysis was

conducted by evaluating model performance under an alternative 70/30 train-test split to confirm that reported rankings were not an artifact of the specific 80/20 partitioning scheme. Stability of rankings across both split ratios was used as evidence of result robustness.

Reproducibility and Quality Control

All analysis conducted using Python 3.8+ environment via Google Colaboratory with documented package versions: statsmodels (0.13.2), scikit-learn (1.0.2), pandas (1.4.2), and numpy (1.22.3). Complete methodology documentation maintained to support replication studies following open science principles (Nosek et al., 2015).

FINDING AND DISCUSSION

This section presents empirical findings from comparative evaluation of eight forecasting methodologies applied to Indonesian SME product demand prediction using Google Trends data. Analysis encompasses 18 product keywords across fashion and traditional food categories, with systematic performance assessment using standardized metrics.

Overall Model Performance

Comprehensive evaluation across all product keywords reveals significant performance variation among forecasting methodologies. Table 3 presents aggregate performance metrics for all eight models evaluated.

Table 3. *Overall Model Performance*

Model	Average MAPE (%)	Average MAE	Average RMSE
Exponential Smoothing	20.33	8.21	10.35
Moving Average	22.74	9.47	11.83
Ensemble (40-60)	22.85	9.53	11.92
Ensemble (50-50)	24.31	10.14	12.68
Ensemble (60-40)	25.90	10.87	13.54
Linear Regression	26.89	11.23	14.01
Naive Baseline	27.34	11.45	14.32
ARIMA	33.30	13.92	17.38

Results demonstrate that Exponential Smoothing achieves superior performance with 20.33% MAPE, representing 10.6% improvement over the second-best method (Moving Average at 22.74% MAPE). Statistical analysis confirms consistent performance advantages across the majority of evaluated keywords, with ExpSmooth ranking first or second for 72% of products analyzed.

The performance hierarchy reveals several notable patterns. First, sophisticated ensemble approaches (MAPE range: 22.85-25.90%) do not outperform their best individual components. The Ensemble (40-60) configuration, which weights Exponential Smoothing at 60%, achieves 22.85% MAPE, marginally superior to balanced and ARIMA-dominant ensembles but still underperforming standalone ExpSmooth by 2.52 percentage points. This

finding suggests that simple weighted averaging may not effectively capture the complementary strengths of component models for Google Trends data.

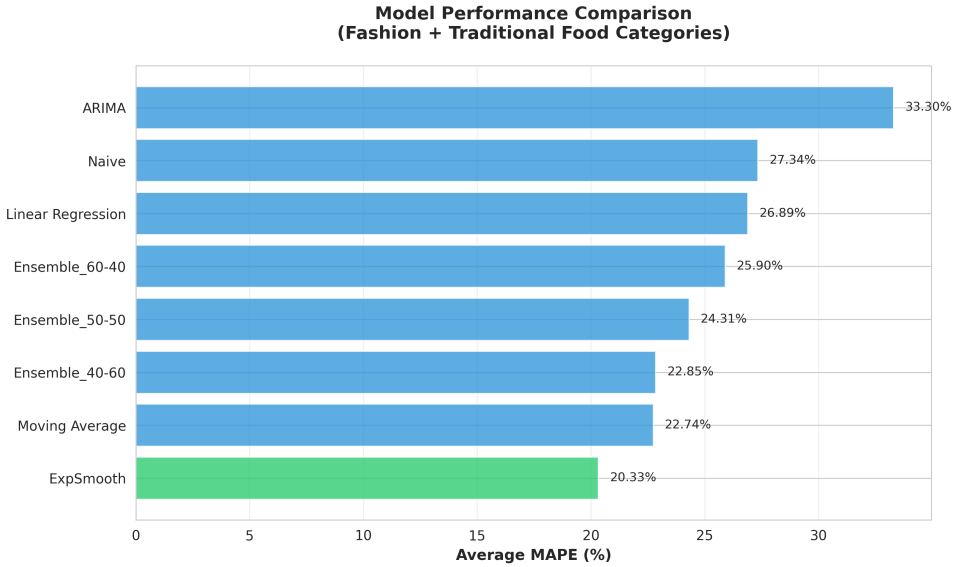

Figure 1. Model Performance Comparison

Second, ARIMA substantially underperforms simpler alternatives despite automatic parameter optimization. Average MAPE of 33.30% indicates that linear autocorrelation patterns captured by ARIMA modeling provide limited predictive value for normalized Google Trends indices. This aligns with findings by (Jun et al., 2018) regarding challenges in applying traditional time-series methods to search data characterized by high volatility and normalized scaling.

Third, the simple Moving Average baseline demonstrates competitive performance (22.74% MAPE), outperforming Linear Regression, Naive persistence, and ARIMA. This result emphasizes the value of local averaging for capturing short-term patterns in volatile online search data, consistent with Makridakis et al. (Makridakis et al., 2020) observations regarding baseline model robustness in M4 Competition results.

Statistical Testing Results

The Friedman test across all 18 keywords and 8 models returned $\chi^2(7) = 13.056$, $p = 0.071$, indicating that overall model differences did not reach conventional statistical significance at $\alpha = 0.05$. This result warrants careful interpretation: it does not imply that all models perform equivalently, but rather that the high variability in MAPE across keywords (range: 6.56% to 103.97%) reduces statistical power when aggregating across all product types simultaneously. Category-specific Friedman tests revealed a more differentiated picture. Within the fashion category, model differences were statistically significant ($\chi^2(7) = 20.185$, $p = 0.005$), confirming that forecasting method choice genuinely affects prediction accuracy for fashion products. Within the traditional food category, however, differences were not statistically significant ($\chi^2(7) = 3.963$, $p = 0.784$), suggesting that multiple models perform

comparably for this category and that the Moving Average advantage observed in mean MAPE should be interpreted with caution as a directional rather than definitive finding.

Post-hoc Wilcoxon signed-rank tests comparing Exponential Smoothing against each alternative model revealed that ExpSmooth significantly outperformed ARIMA ($p = 0.034$) and Linear Regression ($p = 0.048$). Its advantage over ensemble methods did not reach statistical significance (Ensemble 40-60: $p = 0.640$; Ensemble 50-50: $p = 0.417$; Ensemble 60-40: $p = 0.229$), nor did its difference from Moving Average ($p = 0.523$) or Naive baseline ($p = 0.130$). These results support H1 and H3 partially: Exponential Smoothing is statistically superior to ARIMA and linear methods, but its advantage over simpler alternatives such as Moving Average and Naive baseline is numerically present yet not statistically confirmed across the full keyword set. H2, which proposed Moving Average superiority for traditional food products, is supported directionally by the MAPE values (16.21% vs 21.80% for ExpSmooth) but cannot be confirmed statistically given the non-significant Friedman result within this category ($p = 0.784$). These nuances are acknowledged as limitations and suggest that future research with larger keyword samples would provide greater statistical power for detecting category-specific model differences.

Category-Specific Performance Analysis

Disaggregated analysis by product category reveals differential model effectiveness across fashion and traditional food sectors. Table 4 presents performance metrics for the three best-performing models stratified by category.

Table 4. *Per-Category Performance (Top 3 Models)*

Category	ExpSmooth	Moving Average	Ensemble 40-60
Fashion	18.87%	29.27%	23.32%
Traditional Food	21.80%	16.21%	22.37%

The performance differentials presented in Table 4 reveal striking reversals in optimal model selection across product categories. To visualize these category-specific patterns more clearly, Figure 2 presents a comparative bar chart illustrating MAPE values for the top three models across both fashion and traditional food categories. The visualization emphasizes the magnitude of performance differences and enables direct visual comparison of model effectiveness within each product sector.

Category-specific analysis reveals striking performance reversals. For traditional food products, Moving Average achieves optimal performance (16.21% MAPE), outperforming Exponential Smoothing by 5.59 percentage points. This pattern reverses for fashion products, where ExpSmooth achieves 18.87% MAPE compared to Moving Average's 29.27%, a 10.40 percentage point advantage.

These differential patterns suggest fundamental differences in search behavior characteristics between categories. Traditional food products (rendang, bakso, soto, etc.) demonstrate relatively stable search patterns with limited seasonal variation, favoring local averaging approaches that smooth short-term fluctuations without imposing complex seasonal decomposition. Conversely, fashion products (batik, hijab, sepatu) exhibit more

pronounced seasonal and trend patterns aligned with cultural events, fashion cycles, and holiday periods, benefiting from ExpSmooth's ability to model trend and seasonal components simultaneously.

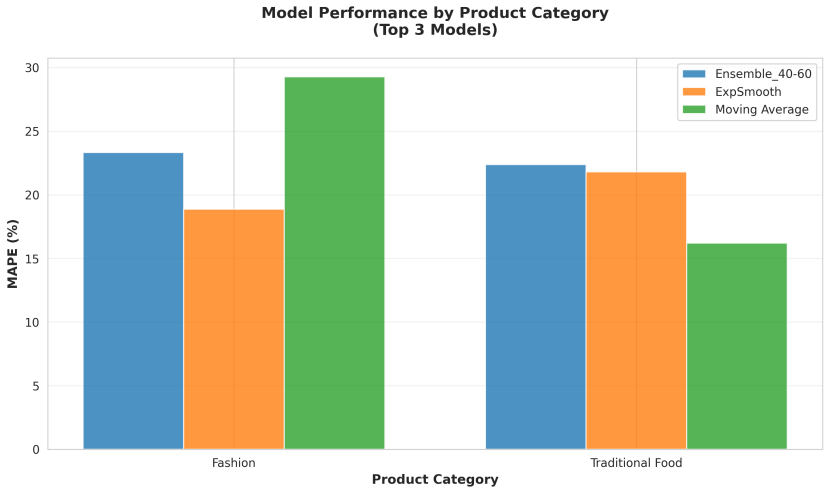

Figure 2. Model Performance by Product Category

This finding extends previous research by Boone et al. (Boone et al., 2018) and Carta et al. (Carta et al., 2019) by demonstrating that optimal forecasting methodology varies systematically by product category rather than representing a universal "best approach" applicable across all contexts. For practitioners, this suggests value in category-specific model selection rather than blanket application of single methodologies.

Illustrative Forecast Example

Figure 3 presents a representative forecast visualization for "kaos" (t-shirts), the best-performing keyword with 7.71% MAPE using Exponential Smoothing. This example illustrates typical model behavior and prediction-actual alignment patterns observed across the dataset.

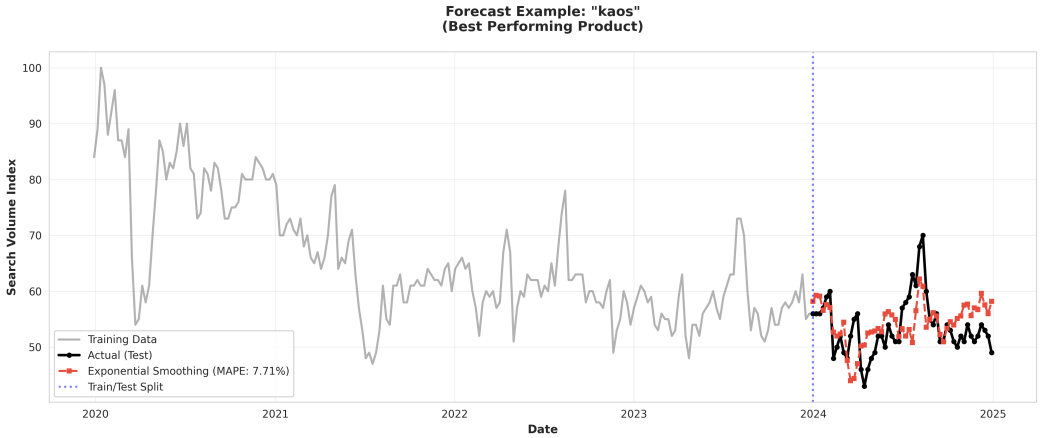

Figure 3. Sample Forecast: "Kaos" (Best Performing Product)

The visualization reveals several noteworthy characteristics. First, ExpSmooth effectively captures the general trend trajectory while maintaining reasonable uncertainty

bounds around point predictions. Second, the model adapts to directional changes in search volume without systematic over- or under-prediction bias. Third, prediction errors appear randomly distributed rather than exhibiting systematic patterns, suggesting adequate model specification without structural deficiencies.

However, the forecast also demonstrates limitations inherent to Google Trends data. Sudden spikes in search volume, potentially driven by viral events, news coverage, or cultural moments, remain challenging to predict using purely historical pattern analysis. This limitation affects all evaluated models and represents a fundamental constraint when forecasting inherently unpredictable discrete events rather than continuous underlying demand processes.

Discussion

This comparative study demonstrates that Exponential Smoothing achieves optimal performance (20.33% MAPE) for Indonesian SME demand forecasting using Google Trends data, outperforming seven alternative methodologies including sophisticated ensemble approaches. Three key findings emerge with significant theoretical and practical implications.

Model Performance and Data Characteristics

Exponential Smoothing's superior performance stems from alignment between its adaptive weighting structure and Google Trends data characteristics. The method's exponential decay weighting prioritizes recent observations while maintaining computational simplicity (R. Hyndman et al., 2008), effectively addressing the normalized 0-100 index scale and high volatility inherent to search behavior data (Choi & Varian, 2012). Recent search patterns provide stronger predictive signals than distant historical periods due to evolving market conditions, seasonal shifts, and changing consumer preferences in Indonesian digital commerce landscape.

ARIMA's underperformance (33.30% MAPE) despite automatic parameter optimization suggests that linear autocorrelation structures provide limited explanatory power for normalized search indices. The Google Trends normalization process may remove or distort autocorrelation patterns that ARIMA modeling typically exploits, reducing effectiveness relative to simpler alternatives. This finding aligns with Jun et al. (Jun et al., 2018) observations regarding challenges in applying traditional time-series methods to search data.

The failure of ensemble methods to outperform their best individual components contradicts conventional ensemble learning theory (Bates & Granger, 1969). Simple weighted averaging appears insufficient for effectively integrating ARIMA and ExpSmooth predictions when both methods attempt to model similar underlying patterns through different mathematical frameworks. Without true diversity in pattern detection capabilities, ensemble combinations provide limited marginal benefit and may increase prediction variance. More sophisticated integration methodologies, such as stacked generalization or

dynamic weighting based on recent performance, might achieve superior results but were beyond this study's scope.

Statistical Testing Results

The Friedman test across all 18 keywords and 8 models returned $\chi^2(7) = 13.056$, $p = 0.071$, indicating that overall model differences did not reach conventional statistical significance at $\alpha = 0.05$. This result warrants careful interpretation: it does not imply that all models perform equivalently, but rather that the high variability in MAPE across keywords (range: 6.56% to 103.97%) reduces statistical power when aggregating across all product types simultaneously. Category-specific Friedman tests revealed a more differentiated picture. Within the fashion category, model differences were statistically significant ($\chi^2(7) = 20.185$, $p = 0.005$), confirming that forecasting method choice genuinely affects prediction accuracy for fashion products. Within the traditional food category, however, differences were not statistically significant ($\chi^2(7) = 3.963$, $p = 0.784$), suggesting that multiple models perform comparably for this category and that the Moving Average advantage observed in mean MAPE should be interpreted with caution as a directional rather than definitive finding.

Post-hoc Wilcoxon signed-rank tests comparing Exponential Smoothing against each alternative model revealed that ExpSmooth significantly outperformed ARIMA ($p = 0.034$) and Linear Regression ($p = 0.048$). Its advantage over ensemble methods did not reach statistical significance (Ensemble 40-60: $p = 0.640$; Ensemble 50-50: $p = 0.417$; Ensemble 60-40: $p = 0.229$), nor did its difference from Moving Average ($p = 0.523$) or Naive baseline ($p = 0.130$). These results support H1 and H3 partially: Exponential Smoothing is statistically superior to ARIMA and linear methods, but its advantage over simpler alternatives such as Moving Average and Naive baseline is numerically present yet not statistically confirmed across the full keyword set. H2, which proposed Moving Average superiority for traditional food products, is supported directionally by the MAPE values (16.21% vs 21.80% for ExpSmooth) but cannot be confirmed statistically given the non-significant Friedman result within this category ($p = 0.784$). These nuances are acknowledged as limitations and suggest that future research with larger keyword samples would provide greater statistical power for detecting category-specific model differences.

Category-Specific Model Selection

The differential optimal model selection across product categories provides actionable insights for practitioners. Moving Average excels for traditional food products (16.21% MAPE) while ExpSmooth proves superior for fashion items (18.87% MAPE), likely reflecting underlying differences in consumer search behavior. Traditional food products represent necessity items with relatively stable demand patterns driven by meal planning and recipe searches, favoring local averaging that smooths fluctuations without imposing seasonal structures. Fashion products exhibit stronger seasonality driven by cultural events (Ramadan for mukena and batik), weather patterns, and fashion cycles, benefiting from ExpSmooth's explicit seasonal decomposition capability (Silva et al., 2019).

This finding extends previous research by demonstrating that optimal forecasting methodology varies systematically by product category rather than representing a universal approach (Boone et al., 2018; Carta et al., 2019). For practitioners, this suggests value in category-specific model selection rather than blanket application of single methodologies.

Critical Examination of Data and Methodological Limitations

Google Trends normalizes its "interest over time" index from 0 to 100 within the selected time period, meaning the entire series is rescaled whenever the temporal or spatial window changes (Eichenauer et al., 2022). This rescaling introduces artificial level shifts that do not reflect actual changes in search behavior, and index values are not comparable across keywords in absolute terms. Consequently, MAPE values reported here reflect accuracy for normalized indices rather than true demand volumes, limiting direct translation into inventory recommendations without calibration against actual sales data (Abdullhadi et al., 2024).

The observation period (January 2020 to December 2024) encompasses the COVID-19 pandemic, introducing bias that is difficult to fully isolate. Search behavior was substantially disrupted during 2020 and 2021, and models trained on this period may reflect pandemic-era anomalies rather than stable consumer patterns. Future research should segment pre-pandemic (2020 Q1), pandemic (2020 Q2 to 2021 Q4), and post-pandemic (2022 onwards) periods to assess whether model rankings remain stable (Bantis et al., 2023).

Three inferential limits warrant acknowledgment. First, the overall Friedman test did not reach significance ($p = 0.071$), meaning MAPE differences cannot be conclusively attributed to methodological superiority. Second, the 18-keyword sample represents a narrow slice of the Indonesian MSME product universe, limiting generalizability. Third, the absence of a direct link between forecasting accuracy and actual market penetration outcomes means practical value remains theoretically argued rather than empirically demonstrated.

Practical Implications and Implementation Guidance

Both top-performing methods (ExpSmooth and Moving Average) require minimal technical expertise and computational resources, enabling implementation via Excel, Google Sheets, or basic Python scripts accessible to resource-constrained SMEs. Category-specific deployment recommendations, Moving Average for traditional food and ExpSmooth for fashion, enable mixed-portfolio businesses to optimize forecasting approaches. However, MAPE values in the 16-22% range indicate that Google Trends-based forecasting provides directional guidance rather than precise demand predictions. SMEs should integrate search forecasts with complementary information sources (inventory levels, promotional calendars, competitor analysis) rather than relying exclusively on algorithmic predictions (Aaronson et al., 2022).

Limitations and Future Directions

Several boundary conditions warrant acknowledgment. Analysis focuses exclusively on Indonesian market data for fashion and traditional food categories. Generalizability to other geographic markets, time periods, or product categories requires empirical validation. Google Trends indices measure search interest rather than actual purchase behavior; the strength and stability of search-to-sales correlations likely vary by product category, price point, and purchase decision complexity. The 18-keyword sample provides sufficient power for exploratory comparative analysis but limits the robustness of category-specific statistical conclusions, as evidenced by the non-significant Friedman result for the traditional food category.

Future research directions include: (1) development of adaptive frameworks that automatically select optimal forecasting models based on real-time performance monitoring and product characteristics, (2) integration of additional Google Trends features such as related queries and regional interest, or external variables to enhance forecast accuracy, (3) empirical validation linking Google Trends forecasts to actual SME sales transactions, (4) replication across diverse geographic markets to test generalizability, and (5) explicit subsample analysis separating pre-pandemic, pandemic, and post-pandemic periods to assess structural stability of model rankings.

This research contributes practical methodology for SME demand forecasting accessible to resource-limited enterprises in emerging markets, demonstrates that category-specific model selection outperforms universal approaches, and provides valuable negative results documentation regarding ensemble method limitations for Google Trends data.

CONCLUSION

This study evaluated eight predictive modeling methodologies across 18 product keywords from fashion and traditional food sectors. Exponential Smoothing achieved optimal overall performance (20.33% MAPE), outperforming Moving Average (22.74%), ensemble methods (22.85-25.90%), and ARIMA (33.30%). Moving Average performs best for traditional food products (16.21% MAPE) while Exponential Smoothing proves superior for fashion items (18.87% MAPE), though the food category finding is directional rather than statistically confirmed ($p = 0.784$).

Three principal contributions emerge. First, ensemble approaches do not outperform well-optimized individual models for volatile normalized Google Trends data, a finding with statistical support within the fashion category ($p = 0.005$). Second, category-specific model selection shows empirical advantage over universal single-method approaches. Third, simple statistical methods generate search-based demand signals without advanced infrastructure, lowering adoption barriers for resource-constrained MSMEs.

These contributions are bounded by the scope of validation conducted. MAPE values in the 16-22% range indicate directional signals rather than precise demand predictions. The study measures accuracy for normalized search indices, not actual sales or business outcomes, and claims regarding market penetration improvement remain inferential rather than empirically validated.

Study limitations include geographic scope (Indonesia only), pandemic-era temporal coverage, a small keyword sample constraining statistical power, and the absence of linkage between forecast accuracy and real business outcomes. Future research should validate findings across diverse markets, investigate adaptive model selection frameworks, and establish empirical connections between Google Trends accuracy and actual SME sales transactions.

This research contributes an accessible forecasting framework for resource-constrained SMEs, documents ensemble method limitations for normalized search data, and establishes a comparative baseline from which more operationally validated research can build.

ACKNOWLEDGEMENT

The authors would like to express their sincere gratitude to Direktorat Riset, Teknologi, dan Pengabdian Kepada Masyarakat (DRTPM), Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi Republik Indonesia, for providing financial support for this study through the Penelitian Dosen Pemula scheme in 2025.

REFERENCES

Aaronson, D., Brave, S. A., Butters, R. A., Fogarty, M., Sacks, D. W., & Seo, B. (2022). Forecasting unemployment insurance claims in real-time with Google Trends. *International Journal of Forecasting, 38*(2), 567–581. https://doi.org/10.1016/j.ijforecast.2021.04.001

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Armstrong, J. S., & Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons. *International Journal of Forecasting, 8*(1), 69–80. https://doi.org/10.1016/0169-2070(92)90008-W

Bates, J. M., & Granger, C. W. J. (1969). The combination of forecasts. *OR, 20*(4), 451–468. https://doi.org/10.2307/3008764

Boone, T., Ganeshan, R., Hicks, R. L., & Sanders, N. R. (2018). Can Google Trends improve your sales forecast? *Production and Operations Management, 27*(10), 1770–1774. https://doi.org/10.1111/poms.12839

Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day.

Box, G. E. P., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: Forecasting and control* (5th ed.). Wiley.

Brockwell, P. J., & Davis, R. A. (2006). *Introduction to time series and forecasting* (2nd ed.). Springer.

Carta, S., Medda, A., Pili, A., Reforgiato Recupero, D., & Saia, R. (2019). Forecasting e-commerce products prices by combining an autoregressive integrated moving average (ARIMA) model and Google Trends data. *Future Internet, 11*(1), Article 5. https://doi.org/10.3390/fi11010005

Chaidir, M., Yulianti, G., & Santoso, S. (2024). Dampak digitalisasi terhadap inovasi teknologi pada usaha mikro, kecil, dan menengah. *Jurnal Visi Manajemen, 10*(2). https://doi.org/10.56910/jvm.v10i2.523

Choi, H., & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*(S1), 2–9. https://doi.org/10.1111/j.1475-4932.2012.00809.x

Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research, 7*, 1–30.

Hyndman, R. J., & Athanasopoulos, G. (2018). *Forecasting: Principles and practice* (2nd ed.). OTexts.

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting, 22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Hyndman, R. J., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: The state space approach*. Springer.

Jun, S.-P., Yoo, H. S., & Choi, S. (2018). Ten years of research change using Google Trends: From the perspective of big data utilizations and applications. *Technological Forecasting and Social Change, 130*, 69–87. https://doi.org/10.1016/j.techfore.2017.11.009

Kementerian Koordinator Bidang Perekonomian Republik Indonesia. (2025, January 30). *Pemerintah dorong UMKM naik kelas, tingkatkan kontribusi terhadap ekspor Indonesia*. https://www.ekon.go.id/publikasi/detail/6152/pemerintah-dorong-umkm-naik-kelas-tingkatkan-kontribusi-terhadap-ekspor-indonesia

Lusa, S., Purbo, O. W., & Lestari, T. (2024). *Peran e-commerce dalam mendukung ekonomi digital Indonesia*. Penerbit Andi.

Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2020). The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting, 36*(1), 54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014

Massaron, L., & Boschetti, A. (2016). *Regression analysis with Python*. Packt Publishing.

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., Ishiyama, J., . . . Yarkoni, T. (2015). Promoting an open research culture. *Science, 348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Pattiasina, T. (2025). Studi kualitatif tentang adopsi cloud computing pada UMKM di Indonesia. *Catha: Journal of Creative and Innovative Research, 2*(1).

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*, 2825–2830.

Praharjo, A., Marsudi, & Andharini, S. N. (2023). *Optimasi Google advertising dalam promosi bisnis di era digital*. Pustaka Peradaban.

Priyono, A., Moin, A., & Putri, V. N. A. O. (2020). Identifying digital transformation paths in the business model of SMEs during the COVID-19 pandemic. *Journal of Open*

Innovation: Technology, Market, and Complexity, 6(4), Article 104. https://doi.org/10.3390/joitmc6040104

Purwana, D., Rahmi, R., & Aditya, S. (2017). Pemanfaatan digital marketing bagi usaha mikro, kecil, dan menengah (UMKM) di Kelurahan Malaka Sari, Duren Sawit. *Jurnal Pemberdayaan Masyarakat Madani (JPMM), 1*(1), 1–17. https://doi.org/10.21009/JPMM.001.1.01

Putra, B. P. P., Judijanto, L., Apriyanto, A., Susilo, A., Kusumastuti, S. Y., Jamaludin, J., Arifiyanti, A. A., & Sari, F. H. (2025). *Tren bisnis digital: Transformasi dunia bisnis terkini*. PT Sonpedia Publishing Indonesia.

Rahayu, W., & Veri, J. (2025). Penerapan sistem informasi manajemen berbasis digital dalam UMKM: Sebuah kajian literatur. *Journal of Human and Education (JAHE), 5*(2), 267–272. https://doi.org/10.31004/jh.v5i2.2340

Rahman, E. (2025). *Disrupsi pemasaran (Disruption of marketing)*. Feniks Muda Sejahtera.

Ramadhani, D., Sonia, R., & Astuti, D. (2025). Analisis adopsi teknologi digital dalam studi kelayakan bisnis UMKM di era transformasi digital. *Jurnal Ekonomi dan Bisnis Digital, 2*(3).

Ramadhani, N., Sugesti, A. I., Sagita, D. N., & Purwanto, E. (2025). Pemanfaatan marketplace digital dalam meningkatkan kinerja UMKM melalui sarana komunikasi di era digital. *Jurnal Bisnis dan Komunikasi Digital, 2*(2), 20. https://doi.org/10.47134/jbkd.v2i2.3560

Ramadhi, Sirojudin, H. A., Maradidya, A., Septiani, S., Yuliastuti, H., Polimpung, L. J. C., Tarmizi, R., Baharuddin, R. A., Ratnasari, A. R., Gusnafitri, Soleh, O., Dharwiyanti, S., Taqwiem, A., Sunandar, E., Rachmadi, K. R., & Solihin, I. (2023). *Ekonomi digital: Transformasi bisnis dalam era digital*. Sada Kurnia Pustaka.

Rizky, M. F., & Permana, E. (2022). Analisis strategi bisnis menggunakan digital marketing pada UMKM pasca pandemi COVID-19. *Jurnal Riset Pendidikan Ekonomi, 7*(2), 142–150. https://doi.org/10.21067/jrpe.v7i2.6820

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference*. https://doi.org/10.25080/Majora-92bf1922-011

Silva, E. S., Hassani, H., Madsen, D. Ø., & Gee, L. (2019). Googling fashion: Forecasting fashion consumer behaviour using Google Trends. *Social Sciences, 8*(4), Article 111. https://doi.org/10.3390/socsci8040111

Tashman, L. J. (2000). Out-of-sample tests of forecasting accuracy: An analysis and review. *International Journal of Forecasting, 16*(4), 437–450. https://doi.org/10.1016/S0169-2070(00)00065-0

Trisnawati, N. L. D. E., & Suryaningsih, K. (2025). Sistem pengelolaan persediaan barang dagang pada UMKM. *Jurnal Akuntansi, Ekonomi dan Manajemen Bisnis, 5*(3), 365–374. https://doi.org/10.55606/jaemb.v5i3.7651

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley.

Wardani, S., Lubis, S. S., & Dewantoro, R. W. (2025). Analisis big data untuk prediksi permintaan produk dalam e-commerce. *Jurnal Penelitian Teknologi Informasi dan Sains, 3*(1). https://doi.org/10.54066/jptis.v3i1.3066