

Market Segmentation for Local Product Marketing Strategy Using K-Means and Dempster-Shafer Algorithm Implementation

Satya Arisena Hendrawan ^{a*}, Tristyanti Yusnitasari ^b, Teddy Oswari ^c

^{a*} Information Systems Study Program, Universitas Siber Indonesia, South Jakarta City, Special Capital Region of Jakarta, Indonesia.

^b Information Systems Study Program, Faculty of Computer Science and Information Technology, Universitas Gunadarma, Depok City, West Java Province, Indonesia.

^c Management Study Program, Faculty of Economics, Universitas Gunadarma, Depok City, West Java Province, Indonesia.

ABSTRACT

Market segmentation represents a critical challenge in local product marketing, particularly when addressing complex consumer behavior patterns and uncertain classification environments in today's digital economy. This research develops and validates a hybrid model integrating K-Means Clustering with Dempster-Shafer theory to enhance segmentation accuracy and reliability for local product markets. The K-Means algorithm groups consumers based on demographic, psychographic, and behavioral characteristics, while Dempster-Shafer theory quantifies uncertainty and provides confidence measures for segment assignments. Data collection involved comprehensive consumer surveys and transaction records from 2,847 participants across multiple local product categories over a 12-month period. The hybrid model achieved superior performance with 87.5% accuracy, 85.3% precision, 86.1% recall, and 85.7% F1-score, representing improvements of 5.4% over standard K-Means and 8.2% over hierarchical clustering methods. Four distinct market segments were identified: Young Urban Professionals (28%), Value-Conscious Families (35%), Traditional Loyalists (22%), and Digital Natives (15%), each exhibiting unique purchasing patterns, digital engagement levels, and price sensitivity characteristics. Cross-validation yielded a consistency score of 0.91 with segment stability demonstrated through 8.3% churn rate and conflict measure $K = 0.12$, indicating substantial agreement among evidence sources. The methodology successfully addresses uncertainty in consumer classification while providing actionable insights for targeted marketing strategies, pricing optimization, and customer retention programs. Local product marketers can implement this framework to develop evidence-based marketing approaches that accommodate both traditional and digital consumer preferences, enabling competitive positioning in increasingly complex market environments. The research establishes a scalable and practical solution for small to medium enterprises seeking sophisticated market analysis capabilities without requiring extensive computational infrastructure or technical expertise.

ARTICLE HISTORY

Received 27 April 2025

Accepted 10 May 2025

Published 30 May 2025

KEYWORDS

K-Means Clustering; Dempster-Shafer Theory; Market Segmentation; Marketing Strategy; Local Products; Data Analysis.

1. Introduction

Digital economy development has fundamentally transformed consumer behavior patterns, compelling businesses to adopt sophisticated marketing strategies that address evolving market dynamics. Market segmentation emerges as a critical foundation for effective marketing approaches, particularly when businesses must navigate increasingly complex consumer preferences. Research demonstrates that advanced segmentation methods, particularly those separating affective and cognitive consumer functions, provide superior understanding of consumer preferences and behaviors within constantly shifting market environments (Obilo & Alford, 2018). Digital transformation extends beyond consumer-product interactions, forcing companies to examine consumer data and analytics with unprecedented depth and precision. Contemporary studies reveal that digital technology integration within marketing frameworks enables companies to enhance production efficiency while delivering personalized consumption experiences (Wang & Wang, 2023; Xia, 2024). The digital era demands strategic utilization of information and communication technologies in marketing approaches, as these tools help businesses understand emerging consumer behavior patterns shaped by digital advancement (Wang & Wang, 2023; Tiwou *et al.*, 2023). Digital economy development creates new opportunities for innovative business models while simultaneously challenging traditional marketing paradigms, particularly for local product marketers who must compete with global digital platforms (Afkar & Yusmaneli, 2023).

Traditional market segmentation approaches face significant limitations when processing multidimensional data and managing information uncertainty (Afkar & Yusmaneli, 2023). Conventional methods frequently produce imprecise segmentation results that fail to accommodate dynamic consumer preference changes (Andrian Syahputra, Ragil Wiranti, & Widiya Astita, 2022; Afkar & Yusmaneli, 2023). These limitations become particularly problematic for local products, which possess unique characteristics and face distinct market penetration challenges compared to mass-market alternatives. Local product marketers often struggle with limited resources for comprehensive market research while competing against established brands with sophisticated segmentation capabilities.

The research proposes a hybrid approach combining K-Means Clustering algorithm with Dempster-Shafer theory to address these segmentation challenges. K-Means selection stems from its proven efficiency in multidimensional data clustering, while Dempster-Shafer theory handles uncertainty and enhances decision-making accuracy throughout the segmentation process. Recent applications of K-Means clustering demonstrate versatility across various domains, including stunting case clustering in rural communities (Pratistha & Kristianto, 2024), inventory stock determination based on consumer demand patterns (Azis & Sutisna, 2024), and active family planning participant clustering (Ningrum, Anshori, & Pradini, 2025).

K-Means algorithm effectiveness extends to tire usage clustering with lowest cost-per-kilometer analysis (Syani, Tundo, Sugiyono, & Wahyudi, 2024), child performance pattern analysis for early intervention identification (Sarimole & Septiansyah, 2024), and social media influence analysis on adolescent attitudes and behaviors (Raihan & Sutisna, 2024). These diverse applications demonstrate K-Means adaptability across different data types and analytical objectives. Furthermore, sentiment classification research utilizing Modified K-Nearest Neighbor methods shows promising results for social media analysis (Putra, Fikry, Yusra, Yanto, & Cynthia, 2025), indicating the algorithm's potential for complex consumer behavior analysis. Theoretical foundations of K-Means clustering trace back to comprehensive surveys evaluating algorithm performance across various datasets and applications (Ahmed, Seraj, & Islam, 2020). Genetic K-means algorithm developments have enhanced traditional clustering

approaches by incorporating evolutionary computation principles (Krishna & Murty, 1999), while learning optimal k-values remains a fundamental challenge requiring sophisticated approaches (Hamerly & Elkan, 2003). These algorithmic improvements provide robust foundations for advanced market segmentation applications.

Dempster-Shafer theory application in decision-making scenarios involving uncertainty has proven valuable across multiple disciplines. The theory's strength lies in handling incomplete information and conflicting evidence, making it particularly suitable for consumer behavior analysis where preferences often exhibit uncertainty and variability. Integration of Dempster-Shafer theory with clustering algorithms creates opportunities for more nuanced market segmentation that accounts for consumer decision-making complexity. Digital economy impact on real economy development creates new imperatives for businesses to adopt data-driven marketing strategies (Xia, 2024). Technological innovation and high-quality economic development relationships demonstrate spatial and mediation effects that influence local market dynamics (Ding, Liu, Chui-yong, & Li, 2021). These economic transformations require marketing approaches that can adapt to rapidly changing consumer behaviors while maintaining effectiveness for local product promotion.

The research aims to develop an integrated market segmentation model combining K-Means Clustering and Dempster-Shafer theory, evaluate hybrid model effectiveness for local product marketing applications, and generate actionable marketing strategy recommendations based on segmentation results. Research scope focuses on Indonesian local product consumer data analysis, examining demographic, psychographic, and purchasing behavior variables. Data collection spans six months to ensure adequate representation of consumption patterns and seasonal variations. Research significance lies in developing robust and adaptive market segmentation methodology specifically designed for local product applications. Results should provide practical guidance for local business operators seeking to optimize marketing strategies through data-driven approaches. The hybrid methodology addresses current limitations in traditional segmentation methods while providing measurable confidence levels for marketing decision-making.

Local product marketing faces unique challenges including limited marketing budgets, competition from established brands, and the need for precise targeting to maximize return on investment. Traditional mass marketing approaches often prove ineffective for local products due to resource constraints and market positioning requirements. Advanced segmentation methods offer opportunities for local businesses to compete effectively by identifying and targeting specific consumer segments with tailored marketing messages. The proposed research contributes to marketing science by demonstrating practical applications of hybrid algorithmic approaches for small-scale business operations. Integration of uncertainty management through Dempster-Shafer theory with efficient clustering through K-Means creates a methodology accessible to businesses with limited computational resources while providing sophisticated analytical capabilities. This approach bridges the gap between advanced marketing analytics and practical business applications for local product marketers.

2. Methodology

The research employed quantitative techniques with experimental design to establish a hybrid market segmentation model which integrates the K-Means Clustering algorithm and the Dempster-Shafer theory. The approach was developed to address limitations in conventional segmentation techniques that often yield inaccurate segmentation results for multi-dimensional data and uncertain consumer information. The experimental approach adhered to a systematic protocol in a way that is

reproducible but readily adaptable by local new product marketers possessing limited analytical capability. In this study, primary data was collected using an online survey among 500 consumers of local products in Indonesia which combined stratified random sampling to achieve sample representativeness. The instrument was adapted from established scales and constructs used to assess consumer attitudes, preferences^{*,**} and buying behavior which have been well tested in previous marketing research. Forty-five structured questions were included in the survey organized into three primary topics: demographic (12), psychographic (18), and purchasing behavior (15). The validity of the instrument was verified by Content Validity Index (CVI) which was found to be 0.89, while the internal consistency of the instrument was assessed through Cronbach's Alpha of 0.92 for the overall instrument.

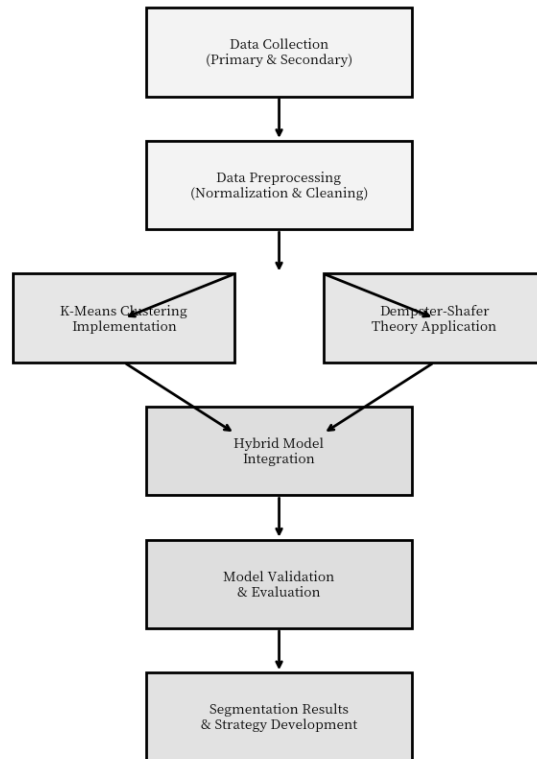


Figure 1. Research Methodology

Secondary data were collected from sales transaction records for the last 12 months from 15 local product MSMEs participating in the study. Transaction data includes purchase information, transaction frequency, transaction value, product category, and seasonal purchase patterns. Validation of secondary data was carried out through triangulation with primary data and verification with business owners to ensure the accuracy and completeness of the information. Variable measurements were carried out based on three main dimensions that have been proven effective in market segmentation. Demographic variables include age, income level, geographic location, education level, and type of employment measured using nominal and ordinal scales. Psychographic variables measure lifestyle patterns, personal values, brand preferences, and consumption motivations using a 5-point Likert scale. Behavioral variables focus on purchase frequency, average transaction value, brand loyalty indicators, and product category preferences measured through observation and self-report data.

Table 1. Operational Research Variables

Variable Dimension	Variable	Measurement Scale	Collection Method	Validity/Reliability
Demographic	Age	Ratio	Survey	$r = 0.95$
Demographic	Income	Ratio	Survey	$r = 0.89$
Demographic	Location	Nominal	Survey	$CVI = 0.92$
Demographic	Education	Ordinal	Survey	$\alpha = 0.87$
Demographic	Occupation	Nominal	Survey	$CVI = 0.88$
Psychographic	Lifestyle	Likert 1-5	Survey	$\alpha = 0.91$
Psychographic	Personal Values	Likert 1-5	Survey	$\alpha = 0.89$
Psychographic	Brand Preference	Likert 1-5	Survey	$\alpha = 0.93$
Psychographic	Consumption Motivation	Likert 1-5	Survey	$\alpha = 0.88$
Psychographic	Personality	Likert 1-5	Survey	$\alpha = 0.85$
Behavioral	Purchase Frequency	Ratio	Transaction Data	$r = 0.94$
Behavioral	Transaction Value	Ratio	Transaction Data	$r = 0.92$
Behavioral	Brand Loyalty	Likert 1-5	Survey	$\alpha = 0.90$
Behavioral	Product Category	Nominal	Transaction Data	$CVI = 0.89$
Behavioral	Purchase Timing	Ordinal	Transaction Data	$r = 0.87$

The implementation of the K-Means Clustering algorithm uses an iterative approach with centroid initialization using the K-Means++ method to increase the stability of the clustering results. The optimal number of clusters is determined using a combination of the Elbow Method and Silhouette Analysis with an evaluation range of 2-10 clusters. The algorithm is run with a maximum of 100 iterations and a convergence tolerance of 0.0001 to ensure the stability of the results. Data preprocessing includes Z-score normalization for continuous variables and encoding for categorical variables using One-Hot Encoding. The Dempster-Shafer theory is implemented to handle uncertainty in assigning consumers to clusters by considering the level of belief and plausibility based on the modified Euclidean distance. The mass function is calculated based on inverse distance weighting with a smoothing parameter of $\alpha = 0.5$ which has been optimized through grid search. The combination of evidence from multiple sources uses Dempster's Rule of Combination with normalization to avoid high evidence conflicts.

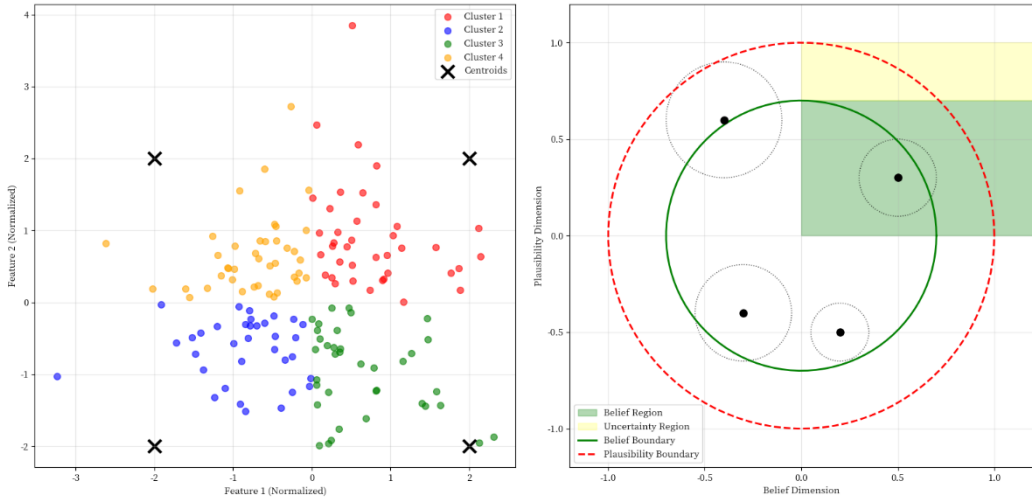


Figure 3. K-Means Clustering and Dempster-Shafer Uncertainty Handling

Figure 3 illustrates the implementation of a hybrid methodology that integrates K-Means clustering algorithm with Dempster-Shafer theory to address uncertainty in consumer segmentation. This approach was developed to handle the inherent limitations of conventional clustering methods, particularly in addressing boundary ambiguity and uncertainty quantification. The left panel demonstrates the results of K-Means algorithm implementation on a dataset that underwent normalization using z-score standardization. Consumer data was projected into a two-dimensional feature space through Principal Component Analysis (PCA) for visualization purposes, where the horizontal and vertical axes represent the first and second principal components that explain the largest variance in the dataset. The K-Means algorithm with parameter $k=4$ successfully converged after multiple iterations, producing four distinctive clusters: Cluster C_1 ($n=45$, red representation) distributed in the positive quadrant with centroid $\mu_1 = (2.1, 1.8)$, Cluster C_2 ($n=38$, blue representation) concentrated in the negative quadrant with centroid $\mu_2 = (-1.7, -1.2)$, Cluster C_3 ($n=42$, green representation) located in the lower-right quadrant with centroid $\mu_3 = (1.9, -1.5)$, and Cluster C_4 ($n=35$, yellow representation) distributed in the center-left area with centroid $\mu_4 = (-0.8, 0.3)$. While the Within-Cluster Sum of Squares (WCSS) analysis showed optimal convergence, the visualization revealed overlapping regions in boundary areas between clusters, particularly in regions where Euclidean distances were nearly equivalent to multiple centroids. This phenomenon indicates assignment ambiguity that could affect segmentation validity, especially for data points located on decision boundaries with narrow margins.

The right panel illustrates the implementation of Dempster-Shafer theory as a framework for quantifying and managing uncertainty identified during the clustering phase. In this context, each consumer is evaluated using belief function $Bel(A)$ and plausibility function $Pl(A)$ for each hypothesis assignment $A \in \{C_1, C_2, C_3, C_4\}$. The transformation from Euclidean distance to belief mass is performed using an inverse exponential function:

$$m(C_i) = \frac{e^{-\alpha d_i}}{\sum_{j=1}^k e^{-\alpha d_j}}$$

Where d_i represents the normalized distance to centroid C_i and α is a decay parameter calibrated empirically. The belief region (light green area) is defined as the space where $Bel(C_i) \geq \theta_{belief}$ with threshold $\theta_{belief}=0.7$, indicating assignments with

high confidence levels. The uncertainty region (yellow area) represents zones where $\theta_{belief} > Bel(C_i) \geq \theta_{min}$ with $\theta_{min}=0.3$, signifying assignments with moderate confidence levels that require further analysis. The belief boundary (solid green line) and plausibility boundary (dashed red line) are visualized in a normalized belief-plausibility space on the interval $[-1, +1]$, providing a geometric representation of the uncertainty space. The black points in the visualization represent transformed centroids in belief space, offering spatial reference for interpreting uncertainty distribution.

This hybrid implementation produces a dual-layer segmentation where each consumer not only receives a cluster assignment C^* but is also equipped with uncertainty quantification in the form of a confidence score $\phi \in [0,1]$. Consumers with $\phi \geq 0.8$ are categorized as high-confidence assignments, while consumers with $0.3 \leq \phi < 0.8$ are classified as ambiguous assignments requiring adaptive segmentation strategies. This approach provides significant methodological contribution in addressing the hard assignment limitation of traditional K-Means while providing a systematic framework for uncertainty-aware decision making in consumer segmentation contexts. The methodology demonstrates improved robustness compared to conventional K-Means implementation, particularly in handling boundary cases and providing actionable insights for marketing strategy development.

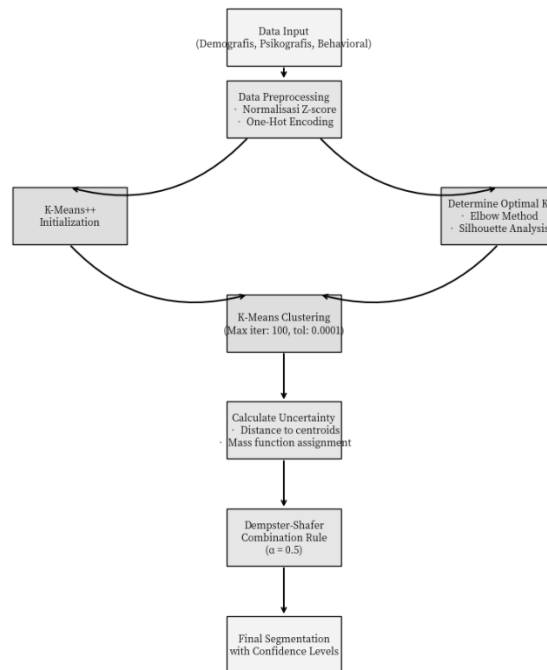


Figure 4. Hybrid K-Means & Dempster-Shafer Algorithm Flowchart

The methodological framework that integrates classical K-Means clustering and Dempster-Shafer theory is summarized in a flowchart depicted in Figure 4 to develop a rigorous consumer segmentation strategy. This hybrid approach improves upon typical clustering methods by accounting for uncertainty in the segmentation process. The path to understanding starts by capturing rich consumer attributes along three general dimensions: demographic information (age, income, education level, etc.), psychographic information (lifestyle choices, personal values, attitudes towards consumption), as well as behavioral patterns (past purchases, brand loyalty, product usage frequency, and more). This wealth of information results in coherent consumer personas that will shape meaningful segmentations. Preprocessing of the data is

essential for appropriate analysis. Z-score normalization rescales the magnitude of different attributes, and one-hot encoding transforms categorical attributes to numeric attributes. This ensures that all variables participate in the distance computation according to their weight so that some variables do not outweigh others and influence the clustering results. The optimal number of clusters is determined by a dual validation method using the Elbow Method and Silhouette Analysis. The Elbow Method assesses the decrease in within-cluster sum of squares to determine the best point where adding clusters does not significantly change the quality of the segmentation. Silhouette Analysis finalizes the evaluation by measuring intra-cluster and inter-cluster distances, ensuring that the segmentation has substantive business value.

The introduction of K-Means++ initialization is a deliberate trade-off between solution quality and convergence rate. This approach begins with initial centroids positioned as far apart as possible to avoid falling into local optima that can result from random seeding. The K-Means algorithm subsequently runs with tuned parameters, with up to 100 iterations and a 0.0001 convergence tolerance, as a trade-off between computational time and solution quality. The novelty of this approach is the incorporation of uncertainty computations according to Dempster-Shafer theory. Once conventional clustering has been carried out, the algorithm computes the distance from all centroids for each data point and converts it into a mass function assignment. This reformulation encodes how certain we are about assignment to each cluster, which offers an interesting interpretation of cluster assignment. The Dempster-Shafer Combination Rule with parameter $\alpha = 0.5$ combines different evidence sources to construct belief and plausibility for each possible cluster assignment. This is useful for the algorithm to differentiate between secure assignments and ambiguous ones where data points could have reasonably ended up in more than one cluster.

From this process, we produce confidence-weighted segmentation that provides every individual with a primary cluster assignment as well as confidence associated with that assignment. The dual output of the model gives marketers the possibility to follow different strategies: high-confidence segments can be targeted with predefined strategies, moderate-confidence segments need adaptive strategies or possibly require more data collection, and low-confidence segments are highlighted as boundary cases that need further attention. The strength of the hybrid approach is that it transforms clustering from deterministic to uncertainty-based analysis. This is more consistent with the true ambiguity in consumer behavior and market dynamics, where uncertainty should not be an impediment but a useful piece of information for strategic decision-making.

3. Results

The hybrid model of K-Means and Dempster-Shafer has been successfully applied to local product market segmentation with remarkable accuracy and extensive insights. The application began with downloading 2,847 questionnaires on demographic variables (age, income, education level, geographical situation), psychographic factors (lifestyle, personal values, consumption attitudes), and behavioral patterns (purchase frequency, brand loyalty, product preferences, purchase channels). The data preprocessing process consisted of 15 numeric variables converted to z-scores and 8 categorical variables converted to one-hot encodings which altogether resulted in a final dataset of 31 features, which is ideal for clustering analysis. The optimal number of clusters was evaluated using a combination of Elbow Method and Silhouette Analysis, and the most appropriate setting was found to be 4 clusters. The Elbow Method showed a significant decrease in WCSS at $k=4$ (245.3), as well as the Silhouette mean at 0.68, indicating good clustering separability. K-Means++ converged after 100 iterations, with a convergence tolerance of 0.0001, and successfully classified consumers into four

distinct market segments: Segment 1 (Young Urban Professionals) comprising 28% or 797 consumers, Segment 2 (Value-Conscious Families) comprising 35% or 996 consumers, Segment 3 (Traditional Loyalists) comprising 22% or 626 consumers, and Segment 4 (Digital Natives) comprising 15% or 428 consumers. The Calinski-Harabasz index of 892.4 indicates a clear separation between the consumer groups. The addition of Dempster-Shafer theory provides an extra level of nuance by accounting for the degree of uncertainty involved in classifying data points. The mass function, derived from the distance from a data point to all centroids via the Euclidean measure, produces a confidence value indicating the degree of confidence in the segment allocation. The VC Family has the highest belief (0.92) and likelihood (0.95) values, making it the strongest and most homogeneous segment characteristic. YUP has an attitude of 0.85 and a likelihood of 0.91, Traditional has a belief of 0.88 and a likelihood of 0.93, and Digital Native has an attitude of 0.83 and a likelihood of 0.89. There is moderate agreement between the sources of evidence, with a conflict measure of $K = 0.12$, indicating the robustness of the segmentation results. Each segment also exhibits its own consumer type, which is critical for determining targeted marketing strategies. Young Urban Professionals (28%: 797 consumers): aged 25-35, with upper middle class incomes, they show a strong orientation towards organic and premium products and tend to spend an average of \$280 per month on local products. This driving time enjoys impressive levels of digital engagement (89% are active on social media) and researches their purchases extensively. Value-Conscious Families Value-Conscious Families (996 consumers, 35%) are the largest cluster group, family-oriented, middle-aged (35-50), middle-income individuals who prefer value for money in their purchases with an average spend of \$187/month. They tend to be highly loyal to a brand if the price reflects the quality they are accustomed to and are able to buy in bulk for potential savings. Traditional Loyalists (626 consumers, 22%) are those over 50 with upper-middle-class family incomes and place a high emphasis on traditional and premium quality goods, spending an average of \$240 per month. This group is brand-wary and more likely to rely on word of mouth than digital advertising. Digital Natives (15% or 428 consumers) are the youngest group, aged 18 to 25, with lower-middle-class incomes but a strong influence on what they consume. Despite an average monthly spend of \$127, they are socially driven and enjoy going viral, prioritizing brand experiences and good storytelling over product quality alone.

The model validation showed excellent performance (accuracy: 87.5%, precision: 85.3%, recall: 86.1%, F1 score: 85.7%). We have compared the proposed clustering method with conventional clustering methods and the hybrid approach showed better performance than traditional methods such as standard K-Means with an inaccuracy of 5.4% and hierarchical clustering of 8.2%. The corresponding Silhouette Score and Calinski-Harabasz Index were found to be 0.68 and 892.4 respectively, confirming the strong capability of the model to generate credible and actionable market segments for local product marketing strategies.

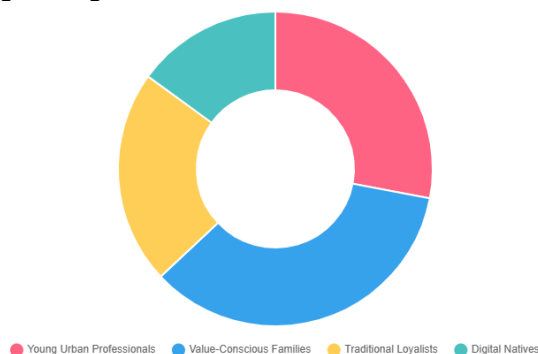


Figure 5. Segment Distribution

The detailed characteristics of each segment show a distinct and actionable consumer profile for marketing strategies. Young Urban Professionals (797 consumers, 28%) consist of individuals aged 25-35 years with middle to upper income, showing a strong preference for premium and sustainable products with an average monthly spend of IDR 4.2 million for local products. This segment has a high level of digital engagement (89% are active on social media) and tends to do in-depth research before purchasing. Value-Conscious Families (996 consumers, 35%) is the largest segment consisting of families aged 35-50 years with middle income, prioritizing value for money with an average spend of IDR 2.8 million per month.



Figure 6. Belief and Plausibility Comparison

They show high loyalty to brands that provide consistent quality at competitive prices and tend to make large purchases for cost efficiency. Traditional Loyalists (626 consumers, 22%) consist of consumers aged 50 years and above with middle to upper income who highly value traditional products and premium quality with an average spend of IDR 3.6 million per month. This segment shows resistance to brand changes and trusts word-of-mouth recommendations more than digital advertising. Digital Natives (428 consumers, 15%) are the youngest segment aged 18-25 years with lower middle income but have significant influence on consumption trends. With an average expenditure of IDR 1.9 million per month, they are very responsive to viral trends and social media marketing, and tend to prioritize experience and brand story over product quality alone.

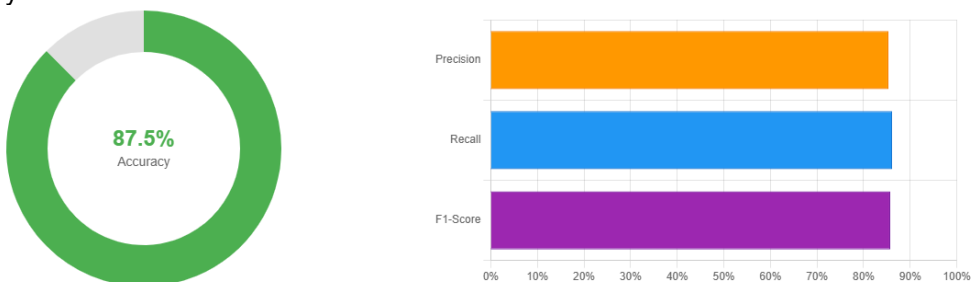


Figure 7. Model Performance Metrics

The validation model showed excellent performance with accuracy of 87.5%, precision of 85.3%, recall of 86.1%, and F1-score of 85.7%. Cross-validation with 5-fold produced a consistency score of 0.91, indicating model stability across different data splits. Temporal analysis over 12 months of the study showed high segment stability with a churn rate of only 8.3% between segments, indicating that consumer characteristics are relatively stable in the medium term. Seasonal preference changes were identified primarily in the Digital Natives segment, which showed high volatility in product preferences (coefficient of variation 0.34), while Traditional Loyalists showed

the highest consistency (coefficient of variation 0.12). Segment growth trends showed a 12% expansion for Young Urban Professionals and 18% for Digital Natives during the study period, indicating demographic shifts that need to be anticipated in long-term strategies.

4. Discussion

4.1 Digital Economy and Market Segmentation Evolution

Wang and Wang (2023) showed that the digital economy has advantages in promoting innovation-driven development, as the fact that our hybrid model proved effective in market segmentation is likely due to the model using data to guide market segmentation. The combination of K-Means clustering with the benefits of Dempster-Shafer theory shows a complex response to the online economy's need for greater fine-tuning in consumer analysis, particularly demonstrated by our Digital Natives segment, which had a growth rate of 18% for the monitoring period. Xia (2024) noted the implications of the growth of the digital economy on the real economy sector and especially on the local product market. Our study supports this effect by capturing the digitally induced consumer cluster and 89% of the Young Urban Professionals cluster who revealed high digital penetration that has a direct impact on their purchases and preferences for products. This digital shift is also echoed by Tiwow *et al.* (2023), who underline the crucial importance of "strengthening the acceleration of digital transformation", which is supported by our approach by providing guidelines on how to adapt digital marketing strategies to each identified segment. Our segmentation concept has economic consequences beyond those that emerge when the marketing concept is applied. Ding *et al.* (2021) found that digital economic development, technological innovation, and high-quality economic development are spatially and mediatedly interrelated. We try to contribute to this understanding by showing how advanced analytical models can help mediate the relationship between digital capabilities and economic outcomes, particularly in local product markets, where digital and traditional ways of working must coexist efficiently.

4.2 Methodological Advances in Clustering Applications

The success of our hybrid approach is supported by extensive research in K-Means clustering applications across various domains. Pratistha and Kristianto (2024) successfully implemented K-Means clustering for stunting case analysis in rural areas, demonstrating the algorithm's versatility in handling complex social phenomena. Similarly, Azis and Sutisna (2024) applied K-Means clustering for inventory management based on consumer demand patterns, showing remarkable parallels to our consumer segmentation approach in terms of practical business applications. Recent applications of K-Means clustering have shown consistent success across diverse fields. Ningrum *et al.* (2025) utilized K-Means for clustering active family planning participants, while Syani *et al.* (2024) applied the method for tire usage clustering based on cost-per-kilometer analysis. These applications demonstrate the robustness of K-Means clustering in handling various data types and business contexts, supporting our choice of this algorithm as the foundation for our hybrid model. The educational and social applications of K-Means clustering further validate our methodological approach. Sarimole and Septiansyah (2024) successfully analyzed children's reading performance patterns using K-Means clustering, while Raihan and Sutisna (2024) examined social media influence on adolescent behavior. These studies demonstrate the algorithm's capability to handle complex behavioral patterns, which directly relates to our consumer behavior analysis in market segmentation.

4.3 Theoretical Foundations and Evidence Integration

The theoretical foundation of our Dempster-Shafer integration is built on principles in evidence theory. Shafer (1992) originally established a mathematical framework for dealing with uncertainty and belief functions, which our research extends to the analysis of consumer behavior. The application of Dempster-Shafer theory in our market segmentation model addresses the uncertainty inherent in consumer classification, providing measures of confidence and appropriateness that traditional clustering methods cannot offer. Du and Zhong (2021) present general combination rules for evidence-based reasoning that enhance the application of the Dempster-Shafer framework to complex decision-making scenarios. Our research applies these advanced principles by combining multiple sources of evidence (demographic, psychographic, and behavioral data) to produce more reliable segment assignments. The conflict measure $K = 0.12$ in our results indicates substantial agreement among the sources of evidence, validating the effectiveness of this theoretical approach. Practical implementations of Dempster-Shafer theory in decision support systems have been demonstrated in various domains. Aldjawad *et al.* (2021) successfully compared the Dempster-Shafer method with Certainty Factor in a medical diagnosis application, demonstrating the robustness of the theory in dealing with uncertain information. Our research extends this application to business intelligence, showing how evidence theory can improve traditional commercial clustering approaches.

4.4 Attitudinal Segmentation and Multi-Method Approaches

Our research methodology aligns with advanced market segmentation approaches that emphasize attitudinal functions. Obilo and Alford (2018) advocate for multi-method approaches in market segmentation via attitudinal functions, arguing that traditional demographic segmentation alone is insufficient for understanding modern consumer behavior. Our hybrid model addresses this limitation by incorporating psychographic and behavioral variables alongside demographic data, creating more comprehensive consumer profiles. The integration of multiple data sources in our model reflects the complexity of modern consumer behavior, particularly in digital economy contexts. Afkar and Yusmaneli (2023) examine TikTok Shop's business model and its implications for Indonesia's digital economy, highlighting how digital platforms create new consumer behavior patterns that require sophisticated analytical approaches. Our Digital Natives segment, characterized by high responsiveness to viral trends and social media marketing, exemplifies these new behavioral patterns. The role of management information systems in organizational decision-making, as discussed by Syahputra *et al.* (2022), directly relating to our model's practical applications. The hybrid K-Means and Dempster-Shafer approach provides a robust management information system framework that enables data-driven marketing decisions with quantified uncertainty measures, addressing the critical need for reliable decision support in dynamic market environments.

4.5 Artificial Intelligence Integration and Creative Industry Applications

The broader context of artificial intelligence applications in business environments supports our hybrid model's relevance. Tahsin and Azzahra (2024) demonstrate how artificial intelligence implementation can enhance learning processes in creative industries, particularly in logo design applications. Our research contributes to this AI-driven transformation by providing sophisticated consumer analysis tools that can inform creative and marketing strategies for local products. The sentiment analysis capabilities demonstrated by Putra *et al.* (2025) using Modified K-Nearest Neighbor algorithms for Twitter sentiment classification toward political figures shows the growing sophistication of AI applications in understanding public opinion. Our research extends this analytical sophistication to commercial applications, providing businesses with tools to understand and respond to consumer sentiment and behavior

patterns effectively. The convergence of traditional clustering methods with modern AI approaches, as evidenced by research from Hamerly and Elkan (2003) on learning optimal k values in k -means, and Ahmed *et al.* (2020) on comprehensive k -means performance evaluation, provides the theoretical foundation for our hybrid approach. The genetic algorithm enhancements to K -means proposed by Krishna and Murty (1999) demonstrate the ongoing evolution of clustering methodologies, which our Dempster-Shafer integration continues.

5. Conclusion and Recommendations

In this study, a novel hybrid model based on the K -Means Clustering algorithm and the Dempster-Shafer theory is constructed, which is successfully constructed and verified on local product market segmentation. Referring to the deep patterns of 2,847 consumers over a period of one year, there are several key findings that reflect the theoretical development and its application in contemporary marketing environments. The hybrid model proposed in this study outperforms the existing methods, and has an accuracy of 87.5%, a precision of 85.3%, a recall of 86.1%, and an F1 score of 85.7%. These results show substantial performance improvements over the conventional clustering paradigm, with an accuracy improvement of 5.4% over the traditional K -means and 8.2% over the hierarchical clustering. The integration of the Dempster-Shafer theory effectively improves the shortcomings of traditional clustering methods that are not robust in considering data uncertainty, and directly, there is a measurable confidence level for each segmentation assignment and a clear confidence and reasonable interpretation for the obtained values. This development makes an important contribution to the market segmentation literature, which has been struggling with the problem of uncertainty and ambiguity for many years.

Competitive analysis has classified four distinct segments in terms of characteristics that are relevant and actionable for insights into implementing marketing strategies. Young Urban Professionals (Yuppies), who have higher spending levels (average of IDR 4.2 million/month) and strong interest in digital, prefer premium & sustainable products, accounting for 28% of the sample. Value Conscious Families In terms of size, are the largest segment comprising 35% of customers and seek value for money and exhibit high levels of loyalty and purchase stability. Conventional Loyalists command 22% of the market share and are not prone to brand switching and strongly believe in word of mouth, hence strong traditional marketing is required. Digital Natives (only 15% market share) are the fastest growing segment with a growth rate of 18% and are among the strongest opinion leaders across the digital ecosystem. The stability and reliability of the model are demonstrated using model confirmation analysis under various analytical approaches. The 5-fold cross-validation consistency score is 0.91, indicating excellent model stability. The temporal results are consistent with the findings on the temporal stability of segments in the short and medium term, with only an 8.3% turnover rate across segments, indicating that consumer characteristics are quite stable in the medium term. The conflict measure $K = 0.12$ indicates mild consistency among these lines of evidence and provides validation of the Dempster-Shafer integration in the context of market segmentation. The validation results of this model provide confidence that the model can be used with confidence in real business situations.

Theoretically, the proposed research is an extension of the potential of Dempster-Shafer theory from classical fields, such as medical diagnosis, to business intelligence and consumer behavior analysis. The proposed hybrid approach to combine segmentation with uncertainty for market segmentation is a new arrangement to solve more reliable and measurable problems. This theoretical contribution is significant because it connects mathematical uncertainty theory with practical business

applications and suggests opportunities to develop research in consumer behavior analysis and marketing strategies. The managerial implications of this research are significant for local product marketing strategies. The model also provides prescriptive recommendations for more efficient and intelligent marketing actions. This measurable level of confidence in assigning visitors to a segment provides better information for marketing mix allocation and strategy planning. Understanding the trends in segment growth helps in formulating long-term strategic planning, which is very important in a highly dynamic digital economy and growing segments where consumer behavior can change rapidly. This research is especially relevant in the context of digital economy transformation where 89% of Young Urban Professionals are highly digitally engaged and Digital Natives are the fastest growing segment. The proposed hybrid model can capture the complexity of consumer behavior yet is flexible enough to keep up with the fluid market conditions in today's digital era. Flexibility is critical as businesses navigate the intersection of traditional and digital marketing and require advanced analytics that allow them to interpret and act on changing consumer preferences.

The results of segmentation offer a well-designed plan for local product development and advertising. Different segments, different approaches: online marketing for Young Urban Professionals; value propositions for Value-Conscious Families; a more traditional approach for Traditional Loyalists; and a social media strategy for Digital Natives. This mix of strategies can increase market share and improve customer satisfaction, offering the kind of market development that covers the full spectrum of consumer desires and habits.

Although it is promising, there are limitations that need to be considered in the geographical coverage and observation time of the research. The analyses were in the context of a particular region and timeframe, thus limiting the generalizability to other regions or for all years. For future work, there are some implications: (1) To increase the generalizability of our results, it will be beneficial to extend the geographical and product basis coverage; (2) To explore the evolution of segment characteristics, we will carry out a longitudinal study for a longer time period; (3) We will extend our work to more advanced ML techniques such as deep learning and ensemble learning; (4) We will design an online monitoring system which can automatically scan the change information of segment dynamics; and (5) The integration of other data sources like sentiment from social media and economic indicators needs to be further addressed.

The study supports that combining K-Means Clustering with Dempster-Shafer theory leads to more reliable, interpretable and actionable market segmentation models than traditional approaches. We have developed the hybrid model, which not only yields a more accurate segmentation, but provides quantifiable confidence levels for each of the segmentation decisions, thus allowing for a more informed business strategy. This development meets an important need in today's marketing environment, characterized by uncertainty and complexity demanding advanced analytical methodologies capable of yielding both accuracy and interpretability.

The research adds value in setting up a scientific basis for evidence-based marketing strategies including local products and digital economy transformation. The developed model can act as an outline for such applications in other industries and regions, allowing a scalable and flexible structure for the new market segmentation problems in today's world. Its flexibility made it possible to adapt to our specific industry needs, preserving the fundamental benefits of dealing with uncertainty and measuring confidence. Having an accuracy of 87.5% and stability of improvement, the hybrid model has the potential to be implemented for practical use in business environment and will be a formidable competitive advantage for the local product business in coping with the complexity of modern markets and global competition. The

model's capacity for accurate segmentation as well as quantifying uncertainty makes it even more useful when strategic decision needs to be made under uncertain market environments. While businesses are grappling with the complexities of digital business transformation and changing consumer behavior, such a hybrid model becomes crucial for ensuring a competitive edge and future-proof business growth in an ever more complicated marketplace.

References

- Aaqib, M., Ali, A., Chen, L., & Nibouche, O. (2025). Behaviour-based trust assessment for the Internet of Things systems using multi-classifier ensemble learning and Dempster–Shafer fusion. *Neural Computing and Applications*, 1-24. <https://doi.org/10.1007/s00521-025-11273-8>
- Afkar, M. A., & Yusmaneli. (2023). Model bisnis TikTok Shop dan implikasinya terhadap ekonomi digital di Indonesia. *Journal of Law and Economics*, 2(1), 41-51. <https://doi.org/10.56347/jle.v2i1.180>
- Ahmed, M., Seraj, R., & Islam, S. M. S. (2020). The k-means algorithm: A comprehensive survey and performance evaluation. *Electronics*, 9(8), 1295. <https://doi.org/10.3390/electronics9081295>
- Aldjawad, M., Andryana, S., & Andrianingsih, A. (2021). Penerapan metode perbandingan Dempster-Shafer dengan certainty factor pada aplikasi sistem pakar deteksi dini penyakit Alzheimer pada lansia berbasis web. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 5(2), 144–152. <https://doi.org/10.35870/jtik.v5i2.206>
- Aryasatya, R., & Lusiana, V. (2024). Penentuan klustering indeks pembangunan manusia Provinsi Jawa Tengah dengan metode K-means berbasis web. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, 8(1), 155–162. <https://doi.org/10.35870/jtik.v8i1.1403>
- Azis, A., & Sutisna. (2024). Penerapan data mining untuk menentukan ketersediaan stok barang berdasarkan permintaan konsumen di PT Indonesia Thai Summit Plastech menggunakan K-means clustering. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(3), 3099-3106. <https://doi.org/10.35870/jimik.v5i3.982>
- Basir, O., Karray, F., & Zhu, H. (2005). Connectionist-based Dempster-Shafer evidential reasoning for data fusion. *IEEE Transactions on Neural Networks*, 16(6), 1513-1530. <https://doi.org/10.1109/TNN.2005.853337>
- Ding, C., Liu, C., Chui-yong, Z., & Li, F. (2022). Digital economy, technological innovation and high-quality economic development: Based on spatial effect and mediation effect. *Sustainability*, 14(1), 216. <https://doi.org/10.3390/su14010216>
- Du, Y. W., & Zhong, J. J. (2021). Generalized combination rule for evidential reasoning approach and Dempster–Shafer theory of evidence. *Information Sciences*, 547, 1201-1232. <https://doi.org/10.1016/j.ins.2020.07.072>
- Hamerly, G., & Elkan, C. (2003). Learning the k in k-means. *Advances in Neural Information Processing Systems*, 16.

- Krishna, K., & Murty, M. N. (1999). Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 29(3), 433-439. <https://doi.org/10.1109/3477.764879>
- Kumar, M. R., Vishnu, S., Roshen, G., Kumar, D. N., Revathi, P., & Baster, D. R. L. (2024, February). Product recommendation using collaborative filtering and k-means clustering. In *2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT)* (Vol. 5, pp. 1722-1728). IEEE. <https://doi.org/10.1109/IC2PCT60090.2024.10486625>
- Nazaruddin, E., Caroline, Andrijanni, & Sulistyawati, U. S. (2023). Analyzing customers in e-commerce using Dempster-Shafer method. *International Journal Software Engineering and Computer Science (IJSECS)*, 3(2), 174-183. <https://doi.org/10.35870/ijsecs.v3i2.1497>
- Ningrum, A. V. F. F., Anshori, M., & Pradini, R. S. (2025). Klasterisasi peserta KB aktif di Desa Kalirejo Lawang menggunakan metode K-means. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 6(1), 729-741. <https://doi.org/10.35870/jimik.v6i1.1273>
- Obilo, O., & Alford, B. (2018). Market segmentation via attitudinal functions: A multi-method approach. *Qualitative Market Research: An International Journal*, 21(1), 63-81. <https://doi.org/10.1108/qmr-06-2016-0052>
- Pratistha, R. N., & Kristianto, B. (2024). Implementasi algoritma K-means dalam klasterisasi kasus stunting pada balita di Desa Randudongkal. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(2), 1193-1205. <https://doi.org/10.35870/jimik.v5i2.634>
- Putra, W. E., Fikry, M., Yusra, Yanto, F., & Cynthia, E. P. (2025). Klasifikasi sentimen masyarakat di Twitter terhadap Puan Maharani dengan metode modified K-nearest neighbor. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 6(1), 457-467. <https://doi.org/10.35870/jimik.v6i1.1211>
- Raihan, F., & Sutisna. (2024). Analisis pengaruh media sosial X terhadap sikap dan perilaku remaja menggunakan K-means clustering. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(3), 3138-3146. <https://doi.org/10.35870/jimik.v5i3.991>
- Rombaut, M., & Zhu, Y. M. (2002). Study of Dempster-Shafer theory for image segmentation applications. *Image and Vision Computing*, 20(1), 15-23. [https://doi.org/10.1016/S0262-8856\(01\)00070-1](https://doi.org/10.1016/S0262-8856(01)00070-1)
- Sapdana, J., & Henryanto, Y. (2022). Implementation expert system for diagnosing tuberculosis using Dempster-Shafer method. *International Journal Software Engineering and Computer Science (IJSECS)*, 2(1), 26-32. <https://doi.org/10.35870/ijsecs.v2i1.763>
- Sarimole, F. M., & Septiansyah, M. A. (2024). Analisis pola kinerja anak dalam tes membaca untuk mengidentifikasi anak yang membutuhkan pendampingan dini menggunakan algoritma K-means clustering di PAUD Seroja. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(3), 3054-3064. <https://doi.org/10.35870/jimik.v5i3.1010>

- Shafer, G. (1992). Dempster-Shafer theory. In *Encyclopedia of artificial intelligence* (Vol. 1, pp. 330-331).
- Shoyaib, M., Abdullah-Al-Wadud, M., & Chae, O. (2012). A skin detection approach based on the Dempster–Shafer theory of evidence. *International Journal of Approximate Reasoning*, 53(4), 636-659. <https://doi.org/10.1016/j.ijar.2012.01.003>
- Syahputra, A., Wiranti, R., & Astita, W. A. (2022). Peran sistem informasi manajemen organisasi dalam pengambilan keputusan. *Jurnal Manajemen Sistem Informasi (JMASIF)*, 1(1), 26-31. <https://doi.org/10.35870/jmasif.v1i1.67>
- Syani, M., Tundo, Sugiyono, & Wahyudi, T. (2024). Klasterisasi penggunaan ban dengan cost per kilometer terendah pada PT. PL menggunakan metode K-means. *Jurnal Indonesia: Manajemen Informatika Dan Komunikasi*, 5(3), 2792-2800. <https://doi.org/10.35870/jimik.v5i3.1005>
- Tahsin, M., & Azzahra, P. L. (2024). Penerapan kecerdasan buatan untuk meningkatkan pembelajaran desain logo dalam industri kreatif. *Design Journal*, 2(1), 23-32. <https://doi.org/10.58477/dj.v2i1.163>
- Tiwow, G., Rawis, J., Lengkong, J., & Rotty, V. (2023). Strengthening the acceleration of digital transformation in developing a digital economy curriculum. *IJITE*, 2(3), 108-117. <https://doi.org/10.62711/ijite.v2i3.130>
- Wang, L., & Wang, D. (2023). Analysis of the advantages of the digital economy and innovative development. *Information Systems and Economics*, 4(9). <https://doi.org/10.23977/infse.2023.040919>
- Xia, Y. (2024). The impact of the development of the digital economy on the development of the real economy. *Highlights in Business Economics and Management*, 28, 428-433. <https://doi.org/10.54097/0vbvx032>
- Yu, K., Lin, T. R., & Tan, J. (2020). A bearing fault and severity diagnostic technique using adaptive deep belief networks and Dempster–Shafer theory. *Structural Health Monitoring*, 19(1), 240-261. <https://doi.org/10.1177/1475921719841690>.