



# Product Preference Analysis of Low Multi-Purpose Vehicle Using Topic Modeling

Trisna Guruminda<sup>1\*</sup> and Teguh Widodo<sup>2</sup>

<sup>1</sup> Master of Management Student, Telkom University, Bandung, Indonesia

<sup>2</sup> Master of Management Lecturer, Telkom University, Bandung, Indonesia  
trisnaminda@gmail.com

\*Corresponding author

**Abstract.** As the most prominent car market in Southeast Asia, the Indonesian car market has made the competition between car manufacturer companies even tighter. The in-depth analysis of the gap between the knowledge possessed by the manufacturer's brand and the consumer's actual preferences in the automotive market for cars in Indonesia, specifically in the best-selling car segment, Low-Multi Purpose Vehicles (LMPV), is a new development of previous research that focused only on the consumer's point of view with an analytical method that has limitations in the amount of data and focuses on a general car scale. This study aims to explain the characteristics of an ideal LMPV-type car according to Indonesian netizens and car manufacturer brands, also comparing those perspectives. There are 10,788 user comments and 129 brand text-marketing, including the official website and e-brochure used as data sources, thus analyzed by Latent Dirichlet Allocation (LDA) topic modeling. The study results show the gap between both points of view on how they represent their ideal characteristics, as the features, comfort, design, and engine factors are the priority characteristics of the car manufacturer's ideal LMPV. Therefore the design, features, price, comfort, reliability, and engine factors are the ideal LMPV characteristic of Indonesian Netizens. The results of this study can be utilized by car manufacturers in Indonesia, both on a national scale through Agen Tunggal Pemegang Merek (ATPM) and regional distributors in product development and marketing activities, such as inter-region car procurement, custom configuration, to content marketing transition from hard-selling to soft-selling.

**Keywords:** Automotive, Customer Preference, Marketing, Multi-Purpose Vehicle, Text Analytic, User-Generated Content.

## 1 INTRODUCTION

Indonesia's market potential for the automotive car industry has a high opportunity. The total domestic sales of car vehicles in Indonesia in 2021 reached 887,200 car units. This figure has a sales increase rate of up to 67% compared to 2020, which was only 523,407 car units (Arief,

202). This phenomenon places Indonesia as the first-ranked car industry market in the Southeast Asia (ASEAN) region, where out of a total of 2.97 million car units sold in the ASEAN region. Indonesia leads with a cumulative car sales rate of 887,202 units, followed by Thailand, with an incremental level of car sales is 754,254 units. In third place is Malaysia, with a cumulative level of car sales of 508,911 units (Dihni, 2022).

As the best-seller category in the Indonesian consumer car market, Multi-Purpose Vehicle (MPV) has a relevant characteristic between the vehicle characteristics and Indonesian demographics. MPV specification emphasizes maximizing passenger and goods capacity combined with the comfort-oriented engine, transmission, and suspension configuration (Jazar, 2017). MPV is relevant to the Indonesian market demographic with many family members and the "Gotong Royong" psychology behavior (Lahardi, 2019).

MPV can be classified into three categories based on engine capacity and overall vehicle dimension: low, medium, and premium MPV (Nindya, 2022). The Low Multi-Purpose Vehicle (LMPV) category has dominated the wholesale level of all consumer car categories in Indonesia (GAIKINDO, 2022). The combination of sufficient engine capacity of around 1500 cm<sup>3</sup> (cc) balances power output and fuel efficiency (Rasyidin, 2020). The small dimension makes it easy to travel around Indonesian streets with a characteristic dominated by the small size of the two-way road (Nindya, 2022). Also, a low-level government tax made an LMPV car ideal for operational or first-time cars and has a better resale value than other consumer vehicle categories (Saragih, 2018).

The rapid level of development has changed people's way of searching for information, including automotive-car related. Formerly people used to purchase magazines or newspapers with a limited one-way point of view of the journalist and visualization of the product. Thanks to the development of social media have changed how people consume information. Be able to receive a more flexible two-way point of view from the reviewer and another user, rich in content format from text, picture, and video to expand the visualization of the product.

YouTube is one of the ideal social media for searching text-based data because a specific user comment feature for one video improves the relevance of comments to video content. YouTube is also the social media with the most significant number of users in Indonesia, reaching 139 million (Yunianto, 2022) and impacting the quantity of local-relevant data generated from it.

This research aims to explain the characteristics of an ideal LMPV based on Indonesian netizens' user comments on the review video of the top four best-sellers LMPV in Indonesia: Toyota Veloz, Mitsubishi Xpander, Daihatsu Xenia, and Suzuki Ertiga (Ihsan, 2022) in the Indonesian top three media automotive channels on YouTube based on subscribers' amount, specifically Motomobi, Oto Diver, and AutonetMagz channel (Social Blade, 2023). Another research aims to explain the characteristics of an ideal LMPV based on official brand manufacturers text-marketing, including website text content and e-brochures that are available on the brand's official website.

## **2 LITERATURE REVIEW**

### **2.1 Marketing**

Marketing can be defined as reaching out to a customer, having a good relationship with them, and creating a unique value that works between the company and the customer (Kotler &

Armstrong, 2018). Marketing also can be defined as the formulation of the marketing mix in determining the product's characteristics, such as service, price, and distribution facilities, to forms of communication that aim to create superior value in the eyes of consumers (Indrawati, 2019).

## **2.2 Customer Behavior**

Customer behavior analyzes how individuals, groups, and organizations select, purchase, use, and dispose of goods, experiences, and services to meet their needs and wants (Kotler & Keller, 2016). Customer behavior includes analyzing individuals and groups' selecting, using, and disposing of products and services, with the impact of each of these processes on the overall experience of consumers (Indrawati, 2019). Customer buying behavior can be classified into five main processes. The first stage is the need for recognition, triggered by internal and external stimuli. The second stage is information search, an information search activity by consumers using four information sources: personal, commercial, public, and experiential. The third stage is evaluating alternatives, an evaluation stage of attributes and personal beliefs that can meet needs and provide more consumer benefits. The fourth stage is the purchase decision, which consists of two situations: the attitudes of others and unexpected situational factors. And the fifth stage is the post-purchase decision which includes the relationship between the consumer's expectations and the product's perceived performance (Kotler & Armstrong, 2018).

## **2.3 Customer Preference**

Customer preferences are a factor that determines consumers in choosing a contender product they have nominated in the choice set the stage in the successive sets, which are found in stages of evaluation of alternatives in the consumer decision-buying process flow (Kotler & Keller, 2016). Several previous studies have concluded some of the attributes that are majorly considered by the customer in the automobile industry, such as the availability of the modern feature (O'Connor *et al.*, 2022), the automobile design concept (Jung *et al.*, 2021), and the engine construction, mainly the level of the performance and fuel-efficiency (Linzenich *et al.*, 2019)

## **2.4 Topic Modeling**

Modern text-based data types create various hidden patterns, one of which is the identification of several chat topics consisting of several related keywords with similar levels of correlation and weighting to one another. Topic modeling is a text analytics method capable of phenomenon-based rendering constructs, prominent features and data variables finding, including conceptual relationships originating from textual data (Hannigan *et al.*, 2019). One of the most popular topic modeling algorithms is Latent Dirichlet Allocation (LDA) based on the generative model, which is a probability distribution of the combined observed and hidden variables, relying on the Bag of Words (BOW) approach (Kherwa & Bansal, 2019).

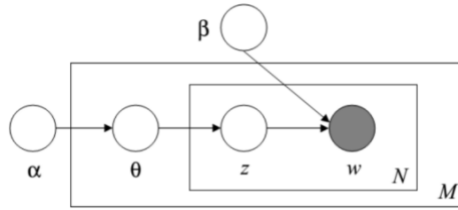


Fig. 1. LDA Graphical Representation

The  $\alpha$  and  $\beta$  parameters act as global parameters used in the entire corpus in the text-based dataset. The  $\theta$  component acts as an interest which evaluates the results of the  $\alpha$  parameter output in the form of a sequence of documents that appears in the  $M$  level and a sequence of words that appears in the  $N$  level, while the  $z$  component is a collection of topics that are formed from the joint probability process of the  $w$  component which is a collection of sequential words (Mifrah & Benlahmar, 2020)

### 3 RESEARCH METHODOLOGY

This research can be classified into qualitative research, which uses data containing words, uses open-ended questions, and aims to generate valid inferences from the entire dataset with many conclusions (Sekaran & Bougie, 2016). This research contains three major steps, as included in the figure below:

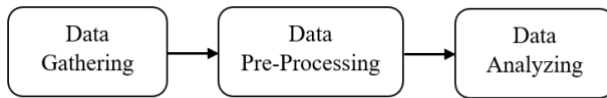


Fig. 2. Research Steps

#### 3.1 Data Gathering

This study uses the data scraping method to carry out the data-gathering process. The data scraping process, carried out to collect data sourced from social media Youtube, can be done using a third-party web service called Coberry, accessed at (coberry.com/youtube). Coberry's third-party web service can gather 10,788 User-Generated Content (UGC) and data from YouTube user comments from Toyota Veloz, Mitsubishi Xpander, Daihatsu Xenia, and Suzuki Ertiga review videos from Oto Driver, Motomobi, and AutonetMagz YouTube channels. To use the Coberry service, users could start a new export and include the link from the YouTube video that will be the data source. Apart from using third-party web Coberry, another data scraping software is Parsehub version 54.0.1. Parsehub software can collect brand text-marketing content, including official website marketing content and e-brochures published on brand official websites, totaling approximately 192 brand text-marketing data. The users can configure website data scraping settings on the Parsehub interface consisting of select, which collects data for category titles, and relative select, which contains data related to predefined category titles. Both output data from Coberry and Parsehub are saved in Comma-Separated Value (CSV) formats.

### 3.2 Data Pre-Processing

After the data-collecting process, the subsequent analysis technique is data pre-processing. Agarwal (2015: 30) suggests that data pre-processing transform raw data into an understandable format because raw data has incomplete, inconsistent, redundant, and noisy characteristics. The data pre-processing phase is able to reduce the total quantity of the research data from 10,788 data to only 6,346 data by using the Google Colab tools. Several data pre-processing phases have been done, including removing duplicates and special characters, that aim to filter duplicate data and special characters with identical content characteristics. The next phase is lowercase, which aims to consistently convert capital letters in the raw data into lowercase letters to prevent discordance in the meaning of the raw data (Rocha & Paredes-Calderón, 2020). The third phase is tokenizing, a breakdown of the format of the user review from paragraph form to sentence form. The sentence form will be continued to break down to per-word form, while sentences that contain less than three words will be automatically deleted.

### 3.3 Data Analyzing

The following research step is data analysis, done with topic modeling text-analytic method with Latent Dirichlet Allocation (LDA) algorithm using the Google Colab tool. The quantity of topics used will depend on which topics quantity have the highest topic coherence score for maintaining the good interpretability and capability of those topics for representing the whole dataset (Syed & Spruit, 2017). The second configuration was max\_df and min\_df configuration, which helps delete terms that have too many occurrences, while the min df setting allows deleting words that have too few existences (Markham, 2016). The max\_df setting is set at 0.75, which has the interpretation that terms that have an occurrence rate of more than 75% in one dataset will be automatically deleted. Meanwhile, the min\_df setting is set at 10, which has the interpretation that terms that have an occurrence rate of fewer than ten times in one dataset will be automatically deleted, continued by determining the alpha value ( $\alpha$ ), which functions as a control value for the combined topics formed with the setting level of the  $\alpha$  value of 0.5. The third configuration is to determine the value of the relevance metric ( $\lambda$ ), which functions as a value hyperparameter in controlling the distribution of words on each topic that is formed. The value of  $\lambda$  is set to 1, sort the terms in each topic that appears based on the most relevant terms. Lastly, each topic is ranked based on the percentage proportion of their token's occurrence by the entire dataset.

## 4 RESULT / FINDING

### 4.1 Indonesian Netizen Dataset

Calculating the level of coherence score is needed to determine the ideal number of topics to be used in discussing the characteristics of the ideal LMPV-type car according to Indonesian netizens. It is identified that the highest coherence score is obtained using six topics with a value of 0.5904.

**Table 1.** Indonesian Netizen Topic Modeling

Topics	Top Words					Dataset Portion's
1	Mobil	Beli	Keren	Liat	Facelift	19.8%
2	CVT	Fitur	Safety	Nyaman	TSS	18.5%
3	FWD	RWD	Lampu	Nanjak	Jalan	17.8%
4	Fitur	Model	Design	Bagus	Harga	16.4%
5	Harga	Mobil	Tipe	Rem	Ganti	13.8%
6	Hybrid	Mesin	Ganteng	Body	Diesel	13.6%

Sources: Processed Data, 2022.

Topic one is the topic with the highest level of popularity, with the most relevant term reaching 19.8% of the Indonesian netizens' user comment dataset. This topic mainly discusses Indonesian netizens' main concern of primary design LMPV type car-related factors, including the design similarity and proportion between the LMPV and other car vehicles. The LMPV's level of design similarity with other car vehicle products with a premium class positioning will be appreciated by Indonesian netizens. On the other hand, if the design of an LMPV car reminds them of a product with an entry-level class positioning, the Indonesian netizens will scold it.

Topic two is the topic with the second-highest level of popularity, with the most relevant term reaching 18.5% of the Indonesian netizens' user comment dataset. This topic mainly discusses Indonesian netizens' concern about the comfort level provided by LMPV products, primarily related to suspension travel, handling quality, transmission technology, and overall noise insulation. The other factors discussed by Indonesian netizens are the advanced semi-autonomous safety features commonly applied to LMPV products, especially the usage, sensitivity, and reliability level.

Topic three is the topic with the third-highest level of popularity, with the most relevant term reaching 17.8% of the Indonesian netizens' user comment dataset. This topic mainly discusses Indonesian netizens' concern about the reliability and ruggedness level provided by LMPV products, primarily related to drivetrain technology. The more common Front-Wheel Drive (FWD) drivetrain is mentioned to be more comfortable, fuel-efficient, and provide more cabin space than the more conventional Rear-Wheel Drive (RWD) drivetrain. The downside of FWD drivetrains is that they are unreliable and rugged enough for travelling around challenging route schemes such as extreme mountainous or road contours made of non-asphalt such as dirt tracks and laterite, commonly found in various regions in Indonesia.

Topic four have the most relevant term portion reaching 16.4% of the entire dataset. This topic mainly discusses Indonesian netizens' evaluation of LMPV's overall value between the On-The-Road (OTR) prices and features provided by LMPV products, including multimedia, safety, and essential features. This topic also mentions other factors related to the design factors of LMPV products, such as popular sporty and cross-over LMPV design genres.

Topic five have the most relevant term portion reaching 13.8% of the Indonesian netizens' user comment dataset. This topic mainly discusses Indonesian netizens' concerns about the maintenance factor of LMPV products, including the levels of availability and prices of spare parts. This topic also discourses the braking technology that is applied to the LMPV, including electronic-related safety Anti-lock Braking System (ABS), Equal Brake Distribution (EBD), and Brake Assist (BA). The braking construction system also mentioned that more advanced disc brakes are better at heat release than more conventional drum brakes.

Topic six have the most relevant term portion reaching 13.6% of the Indonesian netizens' user comment dataset. This topic mainly converses Indonesian netizens' regarding the engine type factors implemented on LMPV products, including engine technology, such as typical naturally-aspirated gasoline engine, more responsive but noisy diesel engine, and more advanced, fuel-efficient but highly cost-maintenance hybrid engine. Those engine types' power characters were also examined by Indonesian netizens based on paper-sheet Horsepower (Hp), Torque (Nm), Revolution per Minute (RPM), and their experience on daily or holiday trips.

## 4.2 Manufacturer Brands Dataset

Calculating the level of coherence score is needed to determine the ideal number of topics to discuss the characteristics of the ideal LMPV-type car, according to manufacturer brands. It is identified that the highest coherence score is obtained using six topics with a value of 0.7236

**Table 2.** Manufacturer Brands Topic Modeling

Topics	Top Words					Dataset Portion's
1	New	Design	Interior	Features	Family	31.0%
2	Lamp	Engine	New	Auto	Vehicle	18.8%
3	Better	Comfort	Design	Interior	Steering	15.5%
4	Brake	Vehicle	Features	CVT	Electronic	14.0%
5	Rear	Door	Parking	Safety	Seat	12.8%
6	Fuel	Engine	Power	Dual	Seat	8.0%

Sources: Processed Data, 2022.

Topic one is the topic with the highest level of popularity, with the most relevant term reaching 31% of the manufacturer brands dataset. This topic mainly discusses manufacturer brands highlighting the components and features contained in the interior of the LMPV car, such as the combination of soft-touch materials on the dashboard and door trim components, to the use of leather materials for the passenger seat components combined with beige or black colours giving rise to an elegant, luxurious impression. The presence of exclusive multimedia features such as a head unit with Android Auto or Apple CarPlay operating system, integrated with a rear-seat entertainment system, can accommodate family and colleagues' entertainment needs better along the journey.

Topic two is the topic with the second-highest level of popularity, with the most relevant term reaching 18.8% of the manufacturer brands dataset. This topic emphasizes the importance of using the latest technology in a vehicle lighting system, such as head-lamp, rear-lamp, and sign-lamp, by using Light-Emitting Diode (LED) technology that provides better illumination range and clarity. This advanced lighting technology has better energy efficiency and improves component lifetime expectancy, such as bulbs and batteries.

Topic three is the topic with the third-highest level of popularity, with the most relevant term reaching 15.5% of the manufacturer brands dataset. This topic accentuates the importance of comfort not only limited to the passenger but mostly the driver. The availability of custom driving configurations, such as tilt and telescopic steering wheel and seat height, position, and reclining, provide better driver control and less fatigue for long-hour driving sessions. The advancement of

steering-wheel controls makes less distraction for the driver while operating multimedia and Multi-Information Display (MID) features.

Topic four have the most relevant term portion reaching 14% of the manufacturer brands’ entire dataset. This topic points out several electronic components that provide a better braking performance as one of the most critical car control mechanisms and significantly increase the convenience experience for the driver and passenger. The advancement of ABS, EBD, and BA technology dramatically improves the driver control of the car by preventing the brake locking phenomenon and maintaining traction, preventing serious accidents. Those technologies, combined with electronic parking brakes and auto hold features, increase convenience by making it easier to operate and make less noise than manual parking brakes.

Topic five have the most relevant term portion reaching 12.8% of the manufacturer brands dataset. This topic explains the importance of safety feature that prevents collision, especially in the back section of a car, as the most significant blind spot and a hard-to-see area for the driver. A parking sensor that sounds variably depends on the closeness distance of the car’s back component and nearest obstacle, to the presence of a 360<sup>0</sup> degree camera placed in the grille, mirrors, and back door components provides the driver with an eye-bird like eyesight for better surrounding vision. Those technologies, combined with semi-autonomous features such as brake assist and rear-cross traffic alert, provide additional protection from an unseen collision.

Topic six have the most relevant term portion reaching 8% of the manufacturer brands dataset. This topic describes exclusive technology implemented in the LMPV engine that can fuse two factors that commonly contradict each other regarding engine power output and fuel efficiency. The engine-exclusive technology usually has a similar mechanism. Still, it is different in naming, such as standard valve timing technology Dual VVT-I, MIVEC, and VVT. Another exclusive technology applied to the engine is hybrid configuration technology, which provides gasoline and electric motors to increase fuel efficiency and instant acceleration significantly.

## 5 DISCUSSION

### 5.1 Indonesian Netizen – Manufacturer Brands Comparison

While compared, those Indonesian netizens’ user comments dataset can be divided into several main factors. Firstly, design factors as the most prevalent factors discussed the LMPV cars, by appearing in topics one, two, and four. The second most discussed factors are features implemented on LMPV cars, as occurred in topics two and five. The third factor is LMPV tag price, known as OTR prices, which emerged in topics four and five. The other three critical factors present in the LMPV Indonesian netizen user comment dataset are comfort (topic two), reliability (topic five), and lastly engine factor (topic six).

**Table 3.** Priority Factors Ranking.

Dataset	Priority Factors	Topics		
Indonesian Netizens	(1.) Design	1	2	4
	(2.) Feature	2		5
	(3.) Price	4		5
	(4.) Comfort	2		



	(5.) Reliability	5	
	(6.) Engine	6	
Manufacturers Brands	(1.) Feature	2	4 5
	(2.) Comfort	3	5 6
	(3.) Design	1	3
	(4.) Engine	2	6

Sources: Processed Data, 2022

Slightly different from the factor that appears on the LMPV manufacturer brands' text-marketing datasets. The most frequently discussed topic is related to the exclusive feature of brands, both electronically and mechanically, presented in topics two, four, and five. This feature factor, followed by the comfort factor as the second most recurrent topic, emerged in topics three, five, and six. The other two critical factors in the LMPV manufacturer brands are design (topics one and three) followed by engine factors (topics two and six).

Several main differences exist between the Indonesian netizens and manufacturer brands dataset in representing an ideal LMPV car vehicle. First, it's about the order of priority factors that make LMPV vehicles ideal according to each other's points of view. While some common factors are identified in both datasets, the order of priority factors is slightly different between those datasets according to each of the top words and most relevant terms portion that appeared on Indonesian netizen and manufacturer brands datasets.

Secondly, the main differences occurred in how both points of view represent their understanding of the ideal LMPV vehicle should be created. The manufacturer brands assume that highlighting all their unique and exclusive possessed technology makes their vehicle more attractive to the consumer and implicitly state that other LMPVs without exclusive features are not ideal. In contrast, Indonesian netizens tend to judge the unique and exclusive features of the LMPV product by their user knowledge and experience before determining whether those LMPV products are ideal or just impractical gimmicks.

For example, manufacturer brands proclaim that their unique, exclusive, and advanced semi-autonomous features are crucial for the consumer and can significantly scale down the probability of a fatal accident. Also, the availability of advanced semi-autonomous features in LMPV class car products is the best deal for the consumer because these features are usually only available for the premium class product. In comparison, the Indonesian netizen argues that sometimes semi-autonomous features are important factors. Indonesian netizens consider that some of these semi-autonomous features are too protective, thus eliminating the driving experience, and sometimes provide feedback with a sensitivity level that is too high, thus providing more distraction for drivers. The potential for damage to some of the fragile spare parts of the semi-autonomous system results in a higher maintenance burden for long-term use of the LMPV car vehicle.

Another difference can be identified in converting the more conventional RWD to the FWD drivetrain systems phenomenon. The manufacturer's brand explains that the conversion process can provide a broader level of space in the car's cabin, better suspension configurations, and more economical fuel consumption. In contrast, Indonesian netizens proclaim that the conversion process causes LMPV vehicles to become less reliable, especially in areas with inadequate road infrastructure conditions or incline terrain with extreme slopes. This phenomenon causes unnecessary anxiety and higher maintenance costs due to more significant

potential damage to the drive components on FWD compared to the RWD drivetrain on LMPV cars.

## 6 CONCLUSION AND RECOMMENDATION

Based on the results of the topic modeling analysis that has been carried out, four main factors support the ideal characteristics of the LMPV car according to the claims of the manufacturer brands. The first factor is complete security, convenience, and multimedia features. The second factor is the comfort level of the LMPV car, which is related to seat configuration, driving control, and electronic control. The third factor is the design of the LMPV car, including interior and exterior design. The fourth factor is the engine of the LMPV car, which is related to exclusive technology and the level of fuel efficiency of the engine components of each manufacturer's brand.

According to claims from Indonesian netizens, based on the results of the topic modeling analysis, six factors support the ideal characteristics of the LMPV car. The first factor lies in the level of proportionality and similarity of the overall design of the LMPV car with other car vehicles, plus the type of design genre adopted by the LMPV car. The second factor is prioritizing and sorting safety, convenience, and multimedia features. The third factor is the official price for the LMPV car product, including OTR and consumer reference prices. The fourth factor, the comfort factor, consists of the quality of the suspension's rebound, stability, and various car noises disruption, including rain, engine, ambience, and road noises. The fifth factor is the reliability factor associated with the LMPV car's capability through demanding travel terrain. The sixth factor and closing of all the factors Indonesian netizens consider regarding the ideal LMPV vehicle include engine factors, which are related to the level of power, torque, maintenance, and the price and availability of fuel needed by the engine type.

Future research can expand the range of research objects to include other types of consumer car segments officially available in Indonesia's automotive car market, such as hatchbacks, SUVs, and MPVs, both by doing a profound explanation of one segment and by making head-to-head comparisons between segments of the car vehicle. Future research is expected to use research subjects with a more significant number of user-generated content covering some of the most popular social media. It can use other big data analytics that focuses on prediction, classification, and association so that it is expected to conduct a comprehensive research analysis.

## REFERENCES

1. Agarwal, V. (2015). Research on Data Preprocessing and Categorization Technique for Smartphone Review Analysis. *International Journal of Computer Applications*. 131(4). <https://www.researchgate.net/publication/291019609>. (Accessed 05 July 2022)
2. Arief, A.M. (2022). *Penjualan Mobil 2021 Hampir 900 Ribu Unit, Ini 10 Merek Paling Laris*. [online]. KataData. <https://katadata.co.id/maesaroh/berita/61dd632f4c99d/penjualan-mobil-2021-hampir-900-ribu-unit-ini-10-merek-paling-laris> (Accessed 08 April 2022).
3. Dihni, V.A. (2022). *Penjualan Mobil di ASEAN Tembus 2,97 Juta Unit Selama 2021, Indonesia Terbanyak*. [online]. KataData. <https://databoks.katadata.co.id/datapublish/2022/02/09/penjualan-mobil-di-asean-tembus-297-juta-unit-selama-2021-indonesia-terbanyak> (Accessed 08 April 2022).

4. Gaikindo.or.id (2022). *Automobile Industry Data*. [online]. Gaikindo. <https://www.gaikindo.or.id/indonesian-automobile-industry-data/> (Accessed 17 May 2022).
5. Hannigan, T.R., Haans, R.F.J., Vakili, K., Tchalian, H., Glaser, V.L., Wang, M.S., Kaplan, S., and Jennings, P.D. (2019). Topic Modeling in Management Research: Rendering New Theory from Textual Data. *Academy of Management Annals Journal*, 13. [online]. <https://journals.aom.org/doi/10.5465/annals.2017.0099>. (Accessed 10 June, 2022).
6. Indrawati., Hutami, R. R. F., and Octavia, D. (2019). *Marketing for Non-Marketing Managers*. Indonesia: PT. Elex Media Komputindo.
7. Ihsan, M. A. F. (2022). Ini 10 Mobil Terlaris di Indonesia Sepanjang 2021 [online]. <https://otomotif.kompas.com/read/2022/01/14/163163100615/ini-10-mobil-terlaris-di-indonesia-sepanjang-2021> (Accessed 10 June 2022).
8. Jazar, R.N. (2017). *Vehicle Dynamics Theory and Application*, 3rd ed. Switzerland: Springer International Publishing AG.
9. Jung, J., Yeo, S., Lee, Y., Moon, S., and Lee, D.J. (2021). Factors Affecting Consumers' Preferences for Electric Vehicle: a Korean Case. *Research in Transportation Business & Management*, 41. [online]. <https://www.sciencedirect.com/science/article/abs/pii/S2210539521000493> (Accessed 27 July 2022).
10. Kherwa, P., and Bansal, P. (2019). Topic Modeling; A Comprehensive Review. *EAI Journal*, 1. [online]. <https://journals.aom.org/doi/10.5465/annals.2017.0099> (Accessed 10 June 2022).
11. Kotler, P., and Armstrong, G. (2018). *Principles of Marketing*, 17th ed. United States: Pearson.
12. Kotler, P., and Keller, K.L. (2016). *Marketing Management*, 15th ed. United States: Pearson.
13. Lahardi, Y. (2019). Kenapa MPV Laris di Indonesia? Ini Kata Pengamat. *Otodriver Article*. <https://otodriver.com/berita/2019/kenapa-mpv-laris-di-indonesia-ini-kata-pengamat-kenbhjccmat> (Accessed 10 May 2022).
14. Linzenich, A., Arning, K., Bongartz, D., Mitsos, A., and Ziefle, M. (2019). What Fuels the Adoption of Alternative Fuels? Examining Preferences of German Car Drivers for Fuel Innovations. *Journal of Applied Energy*, 249. [online]. <https://www.sciencedirect.com/science/article/abs/pii/S030626191930666X>. (Accessed 17 July 2022).
15. Mifrah, S., and Benlahmar, E.H. (2020). Topic Modeling Coherence: A Comparative Study between LDA and NMF Models using Covid 19 Corpus. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(4). [online]. <https://www.researchgate.net/publication/344238129>. (Accessed 25 July 2022).
16. Markham, K. (2016). Understanding min\_df and max\_df in Scikit CountVectorizer. Stackoverflow Discussion. <https://stackoverflow.com/questions/27697766/understanding-min-df-and-max-df-in-scikit-countvectorizer> (Accessed on 20 July 2022).
17. Nindya, H. (2022). 15 Mobil MPV Terbaik Lengkap dengan Harganya. Otoklix Article. <https://otoklix.com/blog/mobil-mpv/> (Accessed on 10 July 2022).
18. O'Connor, P.J., Moss, J., Adams, J., Matemberere, C., and Kaya, M. (2022). What Drives Consumer Automobile Choice? Investigating Personality Trait of Predictors of Vehicle Preference Factors. *Journal Personality and Individual Differences*, 184. [online]. <https://www.sciencedirect.com/science/article/abs/pii/S0191886921005997> (Accessed on 25 July 2022).
19. Rasyidin, B (2020). Tipe Motor Mesin CC Terbaik Mana Yang Cocok Untuk Kamu Gunakan di Jalanan. *Priceprice Article*. <https://id.priceprice.com/motor/news/Motor-Mesin-CC-Terbaik-11177/> (Accessed 20 May 2021).

20. Rocha, A., and Paredes-Calderón, M. (2020). Developments and Advances in Defense and Security. *Proceedings of MICRADS 2019* [online]. <https://link.springer.com/book/10.1007/978-981-19-7689-6>. (Accessed 11 July 2022)
21. Saragih, F. A. (2018). Dunia Minta Sedan, Kenapa Indonesia Cuma Jago MPV? *Kompas Otomotif*. <https://otomotif.kompas.com/read/2018/01/17/094400515/dunia-minta-sedan-kenapa-indonesia-cuma-jago-mpv-> (Accessed 10 July 2022).
22. Sekaran, U., dan Bougie, R. (2016). *Research Methods for Business A Skill-Building Approach* (7th ed.). United Kingdom: John Wiley & Sons Ltd.
23. Social Blade. (2023). Motomobi, Oto Driver, and Autonetmagz, YouTube Stats (Summary Profile). [online]. <https://socialblade.com/search/search?query=motomobi> (Accessed 18 May 2022).
24. Syed, S., and Spruit, M. (2017). Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation. *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*. [online]. <https://ieeexplore.ieee.org/document/8259775> (Accessed 20 June 2022).
25. Yunianto, T.K. (2022). Penetrasi Internet Masif, Pengguna YouTube di RI Tembus 139 Juta [online]. Available: <https://www.marketeers.com/penetrasi-internet-massif-pengguna-youtube-di-ri-tembus-139-juta/> (Accessed 10 July 2022).

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

