Journal of Korea Trade Vol. 25, No. 2, April 2021, 22-45
https://doi.org/10.35611/jkt.2021.25.2.22
ISSN 1229-828X

Brand Personality of Global Automakers through Text Mining

JKT 25(2)

Received 11 September 2020
Revised 26 February 2021
Accepted 2 April 2021

Sungkuk Kim†
Department of International Trade, Sungkyunkwan University, South Korea

Abstract

Purpose – This study aims to identify new attributes by analyzing reviews conducted by global automaker customers and to examine the influence of these attributes on satisfaction ratings in the U.S. automobile sales market. The present study used J.D. Power for customer responses, which is the largest online review site in the USA.

Design/methodology – Automobile customer reviews are valid data available to analyze the brand personality of the automaker. This study collected 2,998 survey responses from automobile companies in the U.S. automobile sales market. Keyword analysis, topic modeling, and the multiple regression analysis were used to analyze the data.

Findings – Using topic modeling, the author analyzed 2,998 responses of the U.S. automobile brands. As a result, Topic 1 (Competence), Topic 5 (Sincerity), and Topic 6 (Prestige) attributes had positive effects, and Topic 2 (Sophistication) had a negative effect on overall customer responses. Topic 4 (Conspicuousness) did not have any statistical effect on this research. Topic 1, Topic 5, and Topic 6 factors also show the importance of buying factors. This present study has contributed to identifying a new attribute, personality. These findings will help global automakers better understand the impacts of Topic 1, Topic 5, and Topic 6 on purchasing a car.

Originality/value – Contrary to a traditional approach to brand analysis using questionnaire survey methods, this study analyzed customer reviews using text mining. This study is timely research since a big data analysis is employed in order to identify direct responses to customers in the future.

Keywords: Automaker, Brand Personality, Customer Review, Topic Modeling
JEL Classifications: C45, D12, M52

1. Introduction

The automotive industry is one of the world's largest economic sectors today. Even if people around the world have different languages and cultures, they have the widest range of demand in that automobile is a typical consumer product. Since the invention of automobiles in the 19th century, competition for automobile production and sales has been fierce as a representative consumer product, and companies has been running with various ideas. The automobile sales researches have been conducted in various ways to the extent that most of the management strategies and marketing strategies are mobilized.

In general, consumers use various methods in the process of processing information to judge products or make decisions (Mantel and Kardes, 1999). When purchasing a car, consumers' decision is made according to their preference, and many studies have paid attention to grasping the factors affecting their purchasing intention (Lave and Train, 1979; Erickson, Johansson and Chao, 1984; Baltas and Saridakis, 2013).

It has traditionally been known that an automobile's price, size, power, operating cost, transmission type, reliability, and body type were important factors to consider in

† **First and Corresponding author:** lloyds@skku.edu

determining preference. However, Erickson, Johansson, and Chao (1984) revealed that Multi-Attribute Product Evaluations take place as image variables of automobiles influence beliefs and attitudes. This is not limited to car purchases, it has to do with the brands used in most products. This is because a product or brand satisfies or fills a need of a consumer, and gives benefits such as convenience through its use. The brand's image plays an important role in a company's sales, profits, and market share, and has a positive effect on car purchases (Hocherman, Prashker, and Ben-Akiva, 1983; Berkovec and Rust, 1985; Mannering and Winston, 1985; Mannering, Winston and Starkey, 2002, Train and Winston, 2007).

For successful automobile sales, it is important to form a positive consumer attitude. Brands have a comprehensive and broad impact on product evaluation. Brands with high recognition have been associated with positive purchasing evaluations in the past and serve as an influential clue when consumers evaluate the quality of products (Maheswaran, Mackie, and Chaiken, 1992; Richardson, Dick, and Jain, 1994). Therefore, recent marketing strategists and brand planners of manufacturers are setting strategies to enhance brand image to succeed in the competition for preference against competing products or competitors (Aaker, 1997; Kotler, Keller, Brady, Goodman, and Hansen, 2009).

The measurement of intangible brand associations is often operationalized using measures of brand personality; such an approach has attracted controversy while it has proven helpful to both academics and practitioners in accounting for the results of brand associations (Eisend and Stokburger-Sauer, 2013). Content analysis is a well-established research methodology commonly used in social sciences to analyze communications (Holsti, 1969). Over the past three decades, content-analysis research has greatly benefited from the exponentially increasing volume of electronic data, including various types of media messages, interview transcripts, discussion boards in virtual communities, and texts from Web sites (Neuendorf, 2002; Rainer and Hall, 2003; Romano et al. 2003; Wickham and Woods, 2005).

On the other hand, post-purchase satisfaction or post-purchase evaluation has affected potential customer purchase intentions. Word-of-mouth (WOM) communication has been found to have a great impact on that intention as well. Online word-of-mouth is a non-commercial information transmission voluntarily delivered among consumers, such as advice and complaints about services that have been used or experienced, and has higher credibility than commercial information. Voss (1984) said that despite active marketing activities, more than 80% of consumers purchase products from the recommendations of a specific person rather than frommass media. When consumers evaluate a product, oral communication with reference groups such as family, relatives, and friends has a strong influence on purchasing attitude.

Consumers search for information on products or services through the Internet, and are proactive in providing information, clarifying their experiences through bulletin boards and comparison sites, and asking for help. Online word-of-mouth communication, which has anonymity, and transcends space and time, is free and convenient communication compared to traditional word-of-mouth (Chevalier and Mayzlin, 2003; Henning-Thurau, Gwinner, Walsh, and Gernler, 2004).

The present paper focuses on an issue related to brand personality (Aaker, 1997). Our main aim is to identify dimensions of brand personality that can be truly generic and applicable across all contexts by using the frequency of brand personality of customer reviews in big data and topic modeling. The author will examine whether brand personality can be used appropriately in measuring the automotive brand personality scale. Data was collected from Customer Reviews through J.D. Power in the U.S. using a web data mining crawler. Collected reviews were analyzed through text analysis.

The present study proposes a new methodological approach to conduct content analysis of electronic text data in a more efficient way. Text data are processed iteratively through software tools such as topic modeling SW, WordSmith Tools, and SPSS. This approach smooths the survey responses of the original text data, identification of the variables of interest, and the counting of the occurrences of these variables in the corresponding texts. It also permits the storage of word-frequency data derived from statistical packages.

2. Theoretical Background

2.1. Consumer Automotive Purchasing Behavior

Table 1 summarizes influential vehicle type choice models found in previous studies. In each study, a brief presentation of the explanatory variables entered the models, along with sample size, and the main findings have been provided.

Classically, in purchasing a car, the properties of furniture, vehicle characteristics, and gas prices have been considered as the influential variables. Lave and Train (1979) suggested purchase price, operating cost, number of seats, weight, horsepower to weight, and fuel efficiency as factors that American consumers consider when purchasing a vehicle. The relationship between consumer characteristics suggested that if the consumer's income is high, a large and expensive vehicle is purchased, and the second vehicle purchased has a tendency to be smaller than the existing vehicle.

Depending on the circumstances, major considerations for purchase have changed. In the second oil crisis of the 1970s, fuel economy became the most important consideration. For this reason, Japanese cars with better fuel economy than American cars sold explosively in the United States. In a study by Manski and Sherman (1980), above all, household income and income level were important considerations when purchasing a vehicle. It was confirmed that households with low incomes will hesitate to buy a car that requires extensive operating costs.

Vehicle fuel considerations were extended by Brownstone, Bunch, and Train (2000). By expanding the consideration of simple fuel costs, the relationship between the factors of preference for car purchase have been found by considering electric cars, natural gas cars, and methanol cars.

According to a consumer classification by Campbell, Ryley, and Thring (2012) early adopters preferentially paid attention to alternative fuels such as hybrids, biofuel, solar, and zero emission electric cars, and as countermeasures against global warming were implemented in the automotive industry, interest in alternative fuels also grew.

A study tried to demonstrate that factors considered in purchasing a vehicle varied depending on consumer preferences and styles, rather than on the mere selection of performance. Choo and Mokhtarian (2004) presented a study indicating that the designs of vehicles owned in a neighborhood may influence the purchase of vehicles. Since consumers usually select vehicles according to personal characteristics and lifestyles, those reluctant to travel were likely to buy luxury vehicles. On the contrary, consumers that often traveled long distances were likely to buy compact vehicles.

In the same vein as this opinion, Cao, Mokhtarian, and Handy (2006) showed that the designs of neighborhood vehicles may have an effect on purchase . Moreover, they also suggested that the types of vehicles were determined by considering commute distances, the size of yards, and off-street parking availability.

Table 1. Summary of Automotive Purchase Models

Reference	Sample Size	Vehicle Characteristics Examined	Main Findings
Lave and Train (1979)	541 new car buyers	Purchase price, operating cost, number of seats, weight, horsepower to weight, fuel efficiency	(a) Larger households are more likely to choose subcompact vehicles. (b) Households with longer driving distances are more likely to select larger vehicles. (c) Older people tend to choose larger vehicles. (d) Households with higher incomes are likely to select larger and more expensive vehicles. (e) Vehicle price negatively affects the selection of each of vehicle types. (f) Households possessing more than two vehicles have a tendency to select smaller vehicles when they buy another.
Manski and Sherman (1980)	1,200 Households from a consumer panel survey	Purchase price, operating cost, number of seats, weight, luggage space, acceleration time, vehicle age, turning radius, braking distance, noise level, scrappage rate, search cost, country of origin	(a) Both seating and luggage space had positive effects on the selection of vehicle type, especially in households with larger single-vehicles. (b) Scrappage rates (a proxy for the probability of mechanical failure in vehicles) had a negative effect on the selection of vehicles. (c) Heads of households older than 45 tended to consider weight in selecting the types of vehicles. (d) Households with lower incomes were less likely to select vehicles with higher operating costs. (e) Acceleration time has a significantly positive effect on the selection of vehicle type.
Hocherman et al. (1983)	A sample of 500 households that did not buy a car, and 800 households that bought a car in 1979	Purchase price, operating cost, vehicle size, engine size, luggage space, horsepower to weight, transaction cost, vehicle age	(a) Purchase price, operating cost, and vehicle age negatively influenced the selection of vehicle type. (b) The size of vehicle negatively affected the selection of vehicle types in urban areas as opposed to rural areas. (c) The values of horsepower were higher for the age group of 45 or younger. (d) Ford and foreign manufacturers were significantly positively valued; other domestic vehicle brands were significantly negatively valued.
Berkovec and Rust (1985)	237 single-vehicle households	Purchase price, operating cost, number of seats, vehicle age, turning radius, horsepower to weight, manufacturer, transaction	(a) Operating cost, purchase price, and vehicle age had negative effects on the selection of vehicle type. (b) The size of vehicle negatively affected the selection of vehicle type in urban areas as opposed to rural areas. (c) The value of horsepower was higher for those aged 45 or younger. (d) Ford and foreign manufacturers were significantly positively valued; other domestic vehicle brands were significantly negatively valued.
Mannering and Winston (1985)	Sample of 3,842 single-vehicle and two-vehicle households	Purchase price, operating cost, vehicle age, shoulder room, luggage space, horsepower to engine, displacement	(a) The variables of household brand loyalty had positive effects on the selection of particular vehicle makers. (b) Capital and operating costs negatively affected the selection of vehicle type.

Table 1. (Continued)

Reference	Sample Size	Vehicle Characteristics Examined	Main Findings
Brownstone et al. (2000)	Sample of 4,747 Households that successfully completed a vehicle-choice experiment	Vehicle range, purchase price, home and service station refueling time, home and service station refueling cost, service station availability, acceleration time, top speed, tailpipe emissions, vehicle size, luggage space	(a) The resulting preference data appeared to be critical for understanding selected body type and scaling information, but there were problems in both multicollinearity and difficulties with measuring vehicle attributes. (b) The preference data were critical for obtaining information regarding attributes which cannot be used in the market place. (c) The use of the preference models alone may result in implausible forecasts
Mannering et al. (2002)	654 households that bought new vehicles between 1993 and 1995	Purchase price, operating cost, passenger side airbag, horsepower, turning radius, vehicle reliability, vehicle residual value, vehicle size	(a) Regardless of acquisition type, house-holds were more likely to select vehicles with higher brand loyalty and residual values. (b) When households decide to obtain vehicles via leases, they tended to put greater value on vehicle attributes such as passenger side airbags and horsepower, but were also are more likely to select larger vehicles or SUVs
Choo and Mokhtarian (2004)	Sample of 1,904 respondents	Choice among 9 alternatives based on size and body type	(a) Travel attitudes, personality, and lifestyle were important in the selection of vehicle type. (b) People that resided in very dense urban areas were more likely to drive luxury vehicles or SUVs. (c) Unwillingness for travel in general was associated with driving a luxury vehicle (a luxury vehicle would probably be selected to change an undesirable activity into more pleasant one). (d) People who found they often traveled long distances were likely to drive compact vehicles

Source: Baltas and Saridakis (2013).

Brand loyalty is an important factor in purchasing vehicles. Hocherman, Prashker, and Ben-Akiva (1983), Berkovec and Rust (1985), Mannering and Winston (1985), and Mannering, Winston, and Starkey (2002) have continuously supported brand loyalty as a factor considered important in purchasing vehicles after the 1980s. The brand of a vehicle has been increasingly important as vehicle-related technology advances and competition among automobile companies intensifies. Automobile brands have diversified enough to represent consumer personalities and social statuses. In order to differentiate from competitors, brands try to win consumer trust through a variety of identity factors, and by sharing culture. Automobile makers spend a huge amount of money and make great efforts to establish strong brands today.

Brand loyalty is a positive for automakers. Not only does it mean a person is more likely to return and spend more money with the original equipment manufacturer (OEM), it means that a repeated buyer is also likely to be a cheerleader that introduces others to the brand. For that reason, J.D. Power[1] (2019) has started tracking automotive brand loyalty, and the results of its first study have just been published. J.D. Power unveiled the results of its first automotive brand loyalty study (see Table 2). The group calculated the percentages in its study based on transactions from June 2018 to May 2019, including all model years of trade-in vehicles. The resulting values represent the percentage of buyers that bought or leased a new vehicle from an automaker after trading in an existing car.

Table 2. Automotive Brand Loyalty Study

Luxury Automotive Brand		Mass Market Automotive Brand	
Brand	**Loyalty Percentage**	**Brand**	**Loyalty Percentage**
Lexus	47.6%	Subaru	61.5%
Mercedes-Benz	44.2%	Toyota	59.5%
BMW	43.6%	Honda	57.7%
Porsche	43.5%	RAM	56.2%
Audi	43.3%	Ford	54.0%
Land Rover	40.3%	Kia	49.4%
Maserati	38.0%	Chevrolet	49.0%
Acura	36.1%	Nissan	45.8%
Lincoln	35.5%	Hyundai	44.8%
Cadillac	34.1%	Jeep	40.9%
Volvo	33.3%	Volkswagen	38.1%
Infiniti	32.1%	Mazda	38.0%
Jaguar	20.6%	GMC	37.5%

Source: J.D. Power (2019).

2.2. Brand Personality

Brands, designed by companies to identify products, and regarded as images in the mind of consumers and other target groups, change based on consumer demand (Grönroos, 1996, 1997; Kotler, Keller, Brady, Goodman, and Hansen, 2009). In this line of thought, Kotler et al. (2009) strengthened the importance of preserving existing customers through relationships, saying that a "relation is a focus on building long-term relationships with consumers rather than a focus on new customers as the growth potential." Understanding the meaning of customer brand relationships and how to manage these relationships is a success-factor. Swaminathan, Page, and Gürhan-Canli (2007) argued that "consumer-brand relationships can be formed based on individual- or group-level connections." To illustrate the meaning of customer brand relationship, Swaminathan et al. (2007) took Mercedes as an example, arguing that customer relationships with this brand might be based on a desire to express an individual-level of uniqueness and an exclusive identity. Palmatier (2008) gives another example of the luxury market segment, stating that firms that offer poor interactions with contact employees should recognize that some efforts may be wasteful in building customer relationships (expensive advertising, loyalty points, and rebate programs) from a relationship viewpoint.

[1] J.D.Power is an American-based data analytics and consumer intelligence company founded in 1968 by James David Power III. The company is now a global leader in consumer insights, data, analytics, and advisory services to help clients drive growth and profitability. The company's industry benchmarks and reputation have established this company as one of the world's most well-known and trusted brands.

It is up to consumers to vote with their wallets. They can influence not only what they buy but also what others buy. With the help of social networks and digital devices, consumers can increasingly dictate when, where, and how they engage with brands. Consumers show how both critics and creators are requesting more personalized services. When a voice has been given to consumers, they make that voice heard. They are willing to share their opinions and experiences with others.

In personality studies, Goldberg (1993) created the Big Five model through empirical research, and suggested more specific personality traits. This model derives 35 variables of personality structure, as suggested by Cattell (1943), and the consequent personality work of many researchers (Fiske, 1949; Tupes and Christal, 1961; Norman, 1967). Later, the 35 personality variables were consolidated into five categories. Goldberg (1993) named the model "the Big Five". The application of the Big Five to brands may be found in advertising and marketing literature (Aaker, 1997; Aaker, 1999; Caprara, Barbaranelli, and Guido, 2001). Although earlier literature suggested that brand personality operates differently from human personality (Aaker, 1997), the application of human personality traits to brands still appear valid.

Table 3 shows brand personality dimensions and typical items the author identified in 21 studies conducted from 1997 to 2016. They were derived in various contexts of respondent types (customers, consumers, students, commercial buyers and sellers, and employees and potential employees); branded entity types (consumer brands, company/corporate brands, cities, and countries); and country/culture (USA, UK, Spain, Japan, Holland, Canada, Germany, Croatia, Turkey, India, Brazil, Belgium, and China).

Table 3. Brand Personality Dimensions

Dimensions (and typical items)	**Study Reference Number**																				
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**
Sincerity (honest, genuine, and cheerful)	X	X	X		X	X	X	X	X	X	X	X	X		X	X	?	X	X	?	X
Competence (reliable, dependable, and efficient)	X		X	X	X	X	X		X	X	X		X	X	X	?	X		X	X	
Excitement (daring, imaginative, and up-to-date)	X	X	X	X	X	X	X		X	X	X	?	X	X	X	?	X	X	X	X	X
Sophistication (glamorous, charming, and romantic)	X	X	X		X	?	X	X		X		?				X	X			X	
Ruggedness (tough, strong, and rugged)	X			X	?	?		X						?							
Ruthlessness (controlling and aggressive)					X				X	?		?		X				?			
Thrift (poor, sloppy, and low-class)		X				X															
Peacefulness (gentle, mild, and peaceful)		X	X	X										X							X
Unpleasant (annoying, irritating, and outmoded)				X			X						X								
Simplicity (ordinary and simple)					?							X									

Table 3. (Continued)

Dimensions (and typical items)	**Study Reference Number** 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Sensitivity (delicate, sensitive, and romantic)								?									X				
Conformity (religious, spiritual, and traditionalist)										X											
Prestige (reputable and successful)																			X	X	X
Cosmopolitan (international and cosmopolitan)																			X		
Materialism (selfish, materialistic, and pretentious)																				X	
Conspicuousness (special and extravagant)																					X

Notes: X indicates that the dimension is apparent in the study; ? means that it may be present; and a blank indicates that it was not present. Studies: 1 = Aaker (1997); 2 = Aaker et al. (2010); 3 = Aaker et al. (2001); 4 = Smit et al., 2002; 5 = Davies et al. (2004); 6 = Slaughter et al. (2004); 7 = d'Astous and Levesque (2003); 8 = Venable et al. (2005); 9 = Bosnjak et al. (2007); 10 = d'Astous and Boujbel (2007); 11 = Milas and Mlačić,Mlarcic (2007); 12 = Geuens et al. (2009); 13 = Chen and Rogers (2006); 14 = Kaplan et al. (2010); 15 = Herbst and Merz (2011); 16 = Das et al. (2012); 17 = Muniz and Marchetti (2012); 18 = Rojas-Méndez et al. (2013a, 2013b); 19 = Rauschnabel et al. (2016); 20 = Sung et al. (2015); and 21 = Willems et al. (2011).
Source: Davies, Rojas-Méndez, Whelan, Mete, and Loo (2018).

2.3. Word of Mouth

On the basis of customer reviews on online sites, customers engage in more reliable purchase decision-makings of goods by securing more credible information through external searches for word-of-mouth information, and therefore aim for more reasonable consumption. Word-of-mouth includes positive and negative information, and the latter have a negative effect on purchase (Brister, 1991).

Word-of-mouth has a direct effect on customer behavior. Most behavior-related models, including the Theory of Reasonable Action, demonstrate that word-of-mouth behavior corresponds with purchase intention. Online word-of-mouth is therefore a measure used to easily and rapidly spread much more information than traditional word-of-mouth (Chatterjee, 2001; Schindler and Bickart, 2005). Online word-of-mouth not only contributes to building trust in internet commerce but also has a huge effect on products and their images, so companies must pay attention to the efficient management and activation of word-of-mouth. The word-of-mouth effect is a factor helping consumers reliably purchase products, and is an important measure in forming trust.

Most studies on the effect of online word-of-mouth depend on results from questionnaire surveys of customers with purchase experience on online sites. Although studies examining multi-factors with plural questions have high external validity, they precisely analyze only the effects of variables that researchers attempts to analyze. Qualitative studies using questionnaire surveys were previously valid, but they had limitations in finding variables that researchers could not discover.

Text data mining, a new study method for understanding all themes contained in customer reviews, is therefore required. As Chatterjee (2001) indicated, big data analysis is adequate, as a large amount of positive and negative information is necessary given the characteristics of online word-of-mouth.

Customer reviews are very important as they influence customer trust in companies, regardless of whether they are positive or negative. The reliability of any information can be determined by agreement (Schindler and Bickart, 2005). In general, it can be evaluated by the number of 'Likes' and reviews with similar opinions.

Since customers think that online sites with customer reviews are more reliable, they are very important marketing means, regardless of whether of positivity. These days, most buyers check related customer reviews before purchasing vehicles, so these are important sources of information for companies. Companies may maximize the effect of word-of-mouth by strategically managing customer reviews. Different from traditional offline word-of-mouth, online word-of-mouth is characterized by exchanging information via online sites. Online word-of-mouth is the act or process of communication in which positive or negative information derived from direct and indirect consumer experiences with certain products or services via e-mail or hypertext (Nguyen, Calantone, and Krishnan, 2020).

Some studies for enhancing customer satisfaction have been conducted recently using user reviews expressing satisfaction, dissatisfaction, and needs. Text mining to extract relationship data refined with natural language processing and a morphological analysis of atypical data in the large scale form of text has emerged as an analysis method for customer reviews.

Coughlan (2013) indicated a limit of existing questionnaire methods. Questionnaire surveys via e-mail, a widely used data gathering method, could not easily collect a variety of samples because of low response rates and offline customers.

The use of user reviews increased to complement the efficiency of previous questionnaire surveys, which were limited to responses to various questions, and review crawling has been established as an area of research (Mudambi and Schuff, 2010; Archak, Ghose, and Ipeirotis, 2011; Kostyra, Reiner, Natter, and Klapper, 2016).

Studies on reviews search keywords by conducting a content analysis, or discriminate positive opinions from those negative using a sentiment analysis (Mudambi and Schuff, 2010).

Kim Yong-Hwan, Kim Ja-Hee, Park Ji-Hoon, and Lee Seung-Jun (2016) conducted a partial least squares (PLS) regression analysis of important main factors discovered with content analysis. In this regard, others have also tried to conduct a Latent Dirichlet Allocation (LDA) topic analysis by introducing text mining (Chae Seung-Hoon, Lim Jay-Ick, and Kang Ju-Yong, 2015; Kim Kwang-Kook, Kim Yong-Hwan, and Kim Ja-Hee, 2018).

3. Methodology and Hypothesis

3.1. Text mining

A central idea of quantitative content analysis is that "many words of text can be classified into much fewer content categories" (Weber, 1990). The methodology of extracting content categories from the text, counting occurrences in sampled text blocks, and analyzing associations between categories using a frequency matrix was developed in the mid-20th century, primarily by a group of Harvard researchers, and is often referred to as contingency analysis (Pool, 1959; Roberts, 2000). George (1959), one of the pioneers of content analysis, criticized the use of contingency analysis, saying that the contingency method was not sensitive enough for the intended meaning. Indeed, contingency analysis assumes that "what an author says is what he means" (Pool, 1959), and it cannot take into account such text features as figures of speech or irony. George's opinion was supported by Shoemaker and Reese (1996), who argued that the process of reducing large volumes of text to quantitative data "does not provide a complete picture of meaning and contextual codes, since texts may contain many other forms of emphasis besides sheer repetition." Newbold, Boyd-Barrett, and Van den Bulck (2002) agreed that "there is no simple relationship between media texts and their impact, and it would be too simplistic to base decisions in this regard on mere figures obtained from a statistical content analysis." Moreover, quantitative content analysis does not always account for source credibility, the political or social context of the messages being examined, and audience characteristics such as age, sex, or education (Macnamara, 2003). However, despite its limitations, quantitative content analysis has long been employed in social studies due to its clear methodological reasoning based on the assumption that the most frequent theme in the text is the most important, as well as to the ability to incorporate such scientific methods as "a priori design, reliability, validity, generalizability, replicability, and hypothesis testing" (Neuendorf, 2002).

Text mining has become an exciting research field as it aims to discover valuable information from unstructured texts. Computers cannot simply use unstructured texts in further processing. Thus, exact processing methods, algorithms, and techniques are vital in order to extract this valuable information, which is completed by text mining. Text mining has become an important research focus. A large amount of information is stored in different places in unstructured compilations. Approximately 80% of the world's data is in unstructured text (Ramanathan and Meyyappan, 2013). This unstructured text cannot be easily used by computers in deeper processing. Therefore, there is a need for a technique that is useful in extracting valuable information from unstructured text. These pieces of information are then stored in a text database format that contains structured and a few unstructured fields. The raw text data can be sited in mails, chats, short message service (SMS) records, newspaper articles, journals, product reviews, and organizational records (Vidya and Aghila, 2010). Key information is stored in electronic form by almost every institution, government sector, organization, and industry. There are a variety of names for text mining, such as text data mining, knowledge discovery (Gupta and Lehal, 2009), and retrieved from textual databases. Analysis of intelligent text refers to extracting or retrieving valuable information from unstructured text. Text mining discovers new pieces of information from text data that was previously unidentified or unknown information by extracting it via different techniques. Text mining is a multidisciplinary field concerning the retrieval of information, analysis of text, extraction of information, categorization, clustering, visualization, mining of data, and machine learning.

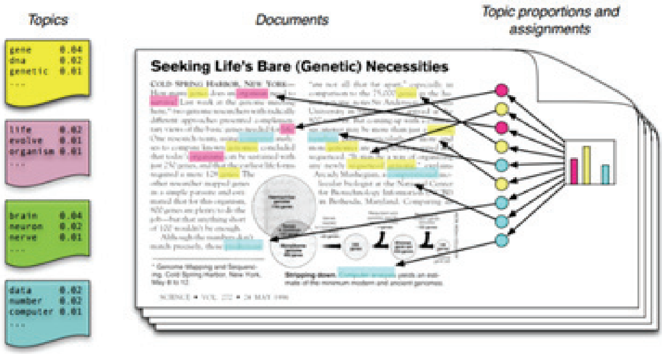
Text mining is the process of analyzing a large collection of unstructured texts for the purpose of exploring interesting and significant patterns and behaviors. There are many domain specific applications of text mining. For example, companies use text mining to locate occurrences and instances of key terms in large blocks of text such as articles, Web pages, customer reviews, or complaint forums (Godbole and Roy, 2008). Unstructured data formats are converted into topic structures and semantic networks by data drilling tools. By studying a semantic network, one can learn the general tone of complaints, as well as and the reasons for these complaints. It also finds common words used in complaints and their relationships to other words in the text via semantic weight (Chen, 2009).

A certain set of words or terms that are commonly used by respondents can been analyzed to describe the pros and cons of product or service. As per the responses of customers, industries take advantage of this for marketing (Grimes, 2005).

3.2. Topic Modeling and Hypothesis

Topic modeling is a machine learning technique to automatically analyze text data in order to cluster words for a set of documents. Topic modeling is a text-mining tool to dig out hidden semantic structures in a text body. Intuitions are the big idea behind latent Dirichlet allocation. As seen in Fig. 1 (far left), a number of "topics" are distributions over words. Each document is assumed to be grouped as follows. First, a distribution is chosen from the topics, as seen in Fig. 1 (the histogram at right). For each word, a topic assignment is chosen as seen in Fig. 1 (the colored coins). Finally, a word from the corresponding topic is selected.

Fig. 1. Explanation of the LDA Process



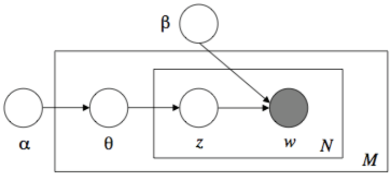
Source: Blei (2011).

Latent Dirichlet allocation (LDA), the most common topic model currently in use, is a generalization of probabilistic latent semantic analysis (Blei 2011). In natural language processing, LDA is frequently used to classify text in a document in accordance with a particular topic. LDA is an instance of a topic model. It belongs to the machine learning toolbox as well as to the artificial intelligence toolbox in a wider sense.

Plate notations are often used to represent probabilistic graphical models. They concisely capture dependencies among many variables. The boxes are "plates" representing replicates, which are repeated entities. The outer plate represents documents, while the inner plate represents repeated word positions in a given document; each position is associated with a choice of topic and word. Variable names are defined as follows (see Fig. 2, Blei, Ng, and Jordan, 2003):

M denotes the number of documents
N is the number of words in a given document (document i has N_i words)
α is the parameter of the Dirichlet prior on per-document topic distributions
β is the parameter of the Dirichlet prior on the per-topic word distribution
θ is the topic distribution for document i
φ is the word distribution for topic k
z_{ij} is the topic for the j-th word in document i
w_{ij} is the specific word.

Fig. 2. Plate Notations Representing the LDA Model



Source: Blei, Ng, and Jordan (2003).

Seo Min-Kyo, Yang Oh-Suk, and Yang Yoon-Ho (2020) examined the relationship between consumer ratings and user reviews, both of which are posted on the Google Play store. The study demonstrated that user reviews were related with positive ratings based on a big data analysis using a Neural Network Model. In addition, it was suggested that customer reviews play the most important role in spreading word-of-mouth. Customer reviews can serve the most significant role in vehicle sales.

A large amount of text-based customer reviews provide opportunities for marketers to grasp customer thoughts. An analysis on customer reviews used to predict consumer purchasing behaviors can be used as a main approach from the perspective of text-mining. Kim, En-Gir, and Se-Hak Chun (2019) analyzed reviews of vehicles with text-mining based on the occurrence frequency of determinants of purchasing vehicles by brand. This study it compared satisfaction with dissatisfaction using frequently occurring words by brand. Such words are, however, were often associated with the performance factors of vehicles, so their use alone is limited in understanding brand image as stressed by the automobile industry.

Topics such as brand personality extracted from customer experience after buying a car seem to have an effect on the overall evaluation of customer reviews. It is meaningful to identify topics like brand personality because automakers have traditionally managed brands competitively. In this research, brand reputation is used as an important clue when consumers evaluate a product or service (Nguyen and Leblanc, 2001). The following is the postulation of the null hypothesis.

Hypothesis: Customer reviews on brand personality with attributes by topic will have a significant impact on overall satisfaction.

4. Topic Modeling Results

4.1. Data Collection and Description

To secure data, materials provided by J.D. Power (http://www.jdpower.com) were used as shown in Fig. 3. The data derived from the reviews include purchase time, driving distance, time of review, and the ratings of four factors: reliability, interior, exterior and driving by vehicle model. This study attempts to consider the relationship between reviews and total ratings, where detailed reviews are qualitatively presented. Moreover, an analysis of the reviews with data mining was also attempted, as it is supposed that they contain brand personality.

The data used in this study were 2,998 evaluations by vehicle brand, which were secured from J.D. Power as the source of customer reviews. Table 4 evaluates 1,440 reviews on luxury automotive brands, including 288 reviews of BMW, which accounts for 48% of the total. It also analyzes 1,558 reviews on mass market automotive brands, including 281 reviews of

Hyundai, which accounts for 52% of the total. The data were used in analysis as they were not biased in favor of a certain brand, and the quantity was adequate to be used.

Fig. 3. Research Data via J.D. Power

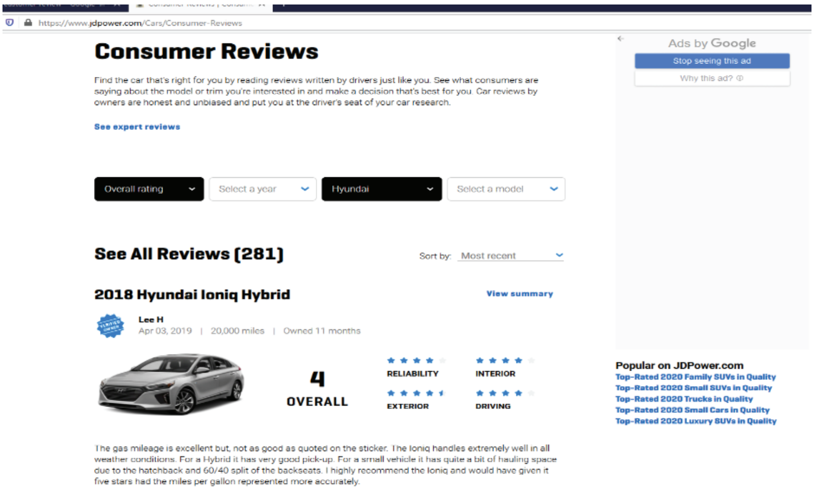


Table 4. Descriptive Statistics

Luxury Automotive Brand			Mass Market Automotive Brand		
Brand	**N**	**%**	**Brand**	**N**	**%**
Acura	131	4.4	Chrysler	93	3.1
Audi	203	6.9	Honda	279	9.3
BMW	288	9.6	Hyundai	281	9.4
Genesis	79	2.6	Kia	331	11.0
Infiniti	169	5.6	Mazda	192	6.4
Lexus	251	8.4	Mitsubishi	156	5.2
Lincoln	146	4.9	Subaru	226	7.5
Volvo	168	5.6			
Sub total	1,440	48.0	Sub total	1,558	52.0

Note: Author accessed http://www.jdpower.com on May 1 to 25, 2020.

The recent interest in corpus linguistics has created a need for software packages that allow researchers to conduct corpus-based investigations. These corpus-based investigations can be used to provide evidence for the quality of products so that customers are exposed to real language rather than artificial text (Biber, Conrad and Reppen, 1998; McEnery, Xiao, and Tono, 2006).

WordSmith Tools is, along with several other software products similar in nature, an internationally popular program for work based on the corpus-linguistic methodology. WordSmith Tools is a software package primarily for linguists, in particular for work in the field of corpus linguistics (Reppen, 2001).

This present study uses 2,998 automotive customer reviews conducted by J.D. Power that have uniquely complex customer experiences. The 2,998 automotive customer reviews consist of 171,589 tokens and 7,598 different types (see Table 5). Here, token is used to refer to running words, and type is used to refer to different words.

Table 5. Automotive Customer Review General Statistics

Number of Reviews	Tokens	Types	TTR	STTR	Sentences	Mean in Words
2,998	171,589	7,598	4.43%	41.42%	11,764	14.59

Note: TTR as type/token ratio; STTR as standardized type/token ratio.

Keywords are typical traits of any text or group of texts. They are extracted by statistically calculating which words are more or less frequent than expected according to some norm. That is, they are usually calculated using two word lists, one from the study corpus that the author investigates, and the other from the normally larger reference corpus that acts as a standard of comparison with the study corpus, or provides background data for keyword calculation. A keyword normally indicates a significant word from a title or document used as an index for the content. In corpus-based linguistic studies, however, the notion is defined as a word "whose frequency is unusually high in comparison with some norm" (Scott, 1997; 2016)[2].

This paper used The Open American National Corpus (OANC) as a reference corpus. The OANC is a large electronic collection of American English of spoken and written data collected from 1990 onward. OANC contains roughly 15 million words of contemporary American English with automatically-produced annotations for a variety of linguistic phenomena.

In the keyword list, the BIC Score is effectively an alternative to P scores. It uses the log-likelihood score and the size of the two corpora in its formula. BIC scores will help, especially where the comparison corpus is fairly small, as it tends to note more negative keywords reflecting the nature of the comparison corpus. Gabrielatos (2018) suggested that BIC scores can be interpreted as follows: below 0 = not trustworthy, 0-2 = only worth a minimal mention, 2-6 = positive evidence, 6-10 = strong, and more than 10 = very strong.

Table 6 lists the top 20 keywords sorted by the higher BIC and groups of per 100 keywords up to 500 keywords. The top 20 keywords covered 12.92% of cumulative frequency occurring 22,175 times in the customer reviews as the study corpus, while the same words covered 1.36% occurring 211,037 times in OANC as the reference corpus. The top 500 keywords occurred 121,207 times and covered 70.64% in the study corpus, while the same words occurred 1,055,841 times and covered 6.78% in the reference corpus. Content words out of the top 20 keywords are listed as country names of global automakers, such as JAPAN, KOREA, and GERMANY, brands such as KIA and BMW, automotive performance evaluation such as VEHICLE, DRIVE, DRIVING, MILEAGE, FEATURES, INTERIOR, GAS, and SEATS, and complimentary adjectives, verbs, and adverbs such as POPULAR, GREAT, COMFORTABLE, LOVE, and VERY. These content words allow interferences about of automotive customer reviews.

[2] This keyword analysis is called a traditional keyword analysis. Most recently, Egbert and Biber (2019) proposed that text dispersion keywords can be computed by comparing the number of texts where each word is found in both the study corpus and the reference corpus. In this study, we followed the traditional keyword analysis as proposed by Scott (2016).

Table 6. Keywords Extracted from Automotive Customer Reviews

Keywords	Freq.	%	RC. Freq.	RC. %	BIC	Probability
1 POPULAR	1,727	1.01	1,882	0.01	10,636.66	6.68053E-22
2 JAPAN	1,405	0.82	875	0.01	9,664.44	8.91069E-22
3 VEHICLE	1,123	0.65	428	0.00	8,315.02	1.40034E-21
4 GREAT	1,375	0.80	5,962	0.04	5,461.90	4.95704E-21
5 MY	2,243	1.31	24,598	0.16	5,366.75	5.22628E-21
6 KOREA	690	0.40	250	0.00	5,136.17	5.96486E-21
7 DRIVE	957	0.56	1,853	0.01	5,068.58	6.20754E-21
8 LOVE	1,025	0.60	2,785	0.02	4,871.27	6.99574E-21
9 COMFORTABLE	707	0.41	486	0.00	4,771.20	7.44687E-21
10 MILEAGE	534	0.31	104	0.00	4,244.60	1.05911E-20
11 FEATURES	748	0.44	1,403	0.01	3,995.27	1.27101E-20
12 VERY	1,378	0.80	11,244	0.07	3,980.35	1.28541E-20
13 I	4,949	2.88	155,911	1.00	3,927.11	1.33865E-20
14 INTERIOR	550	0.32	483	0.00	3,537.17	1.83467E-20
15 GAS	621	0.36	988	0.01	3,471.52	1.94127E-20
16 SEATS	482	0.28	256	0.00	3,392.55	2.08075E-20
17 DRIVING	572	0.33	778	0.00	3,330.22	2.20043E-20
18 GERMANY	497	0.29	716	0.00	2,849.14	3.52324E-20
19 KIA	321	0.19	5	0.00	2,833.01	3.58416E-20
20 BMW	271	0.16	30	0.00	2,238.10	7.30532E-20
1-20th Keywords	22,175	12.92	211,037	1.36		
1-100th Keywords	41,025	23.91	334,472	2.15		
1-200th Keywords	47,542	27.71	396,861	2.55		
1-300th Keywords	52,279	30.47	440,173	2.83		
1-400th Keywords	55,676	32.45	479,358	3.08		
1-500th Keywords	121,207	70.64	1,055,841	6.78		

Note: BIC scores are 0: not trustworthy; 0-2: only worth a bare mention; 2-6: positive evidence; 6-10: strong; and more than 10: very strong.

4.2. LDA Analysis Results

Fig. 4 shows the convergence of perplexity versus iteration for the equilibrium distribution using the study corpus as the data set and six topics. For the data set, the author set alpha(α)=0.1, sigma(σ)=1, beta(β)=0.001, and the number of topics K to 6. The data set was run for 99,999 iterations. As seen in Fig. 3, as a result of machine learning for topic modeling, the increasing iteration continually reduces perplexity, and the perplexity values of the data set generally converge to 2,100 by 10 iterations.

When analyzing a topic model using the LDA algorithm, the author set the number of topics *K* to 6, as shown in Table 7 below.

If topics are identified as brand personality after the author has assembled high-ranking words extracted from each topic, Topic 1 was named “Competence” (reliable, dependable, and efficient) in that several words, such as seat, back, space and room, related to automobile structure occurred in the top 30 ranked words of Topic 1. Topic 2 was named “Sophistication” (glamorous, charming, and romantic) as several words, such as love, vehicle, and drive, related to automobile performance occurred in the top 30 ranked words of Topic 2. Topic 3 was named "Ruggedness" (tough, strong, and rugged) as several words. MPG, power, and turbo, related to automobile performance technology occurred in the top 30 ranked words of Topic 3. Topic 4 was called "Conspicuousness: (special and extravagant) as words like camera, navigation aids, and phone related to automobile options occurred in the top 30 ranked words of Topic 4. Topic 5 was named "Sincerity" (honest, genuine, and cheerful) in that several words, such as reliable, safety, and safe, related to automobile reliability occurred in the top 30 ranked words of Topic 5. Topic 6 was called "Prestige" (reputable and successful) as several words, BMW, LEXUS, AUDI, and VOLVO, related to luxury cars occurred in the top 30 ranked words of Topic 6.

Fig. 4. On the Study Corpus Using K = 6 Topics through Machine Learning

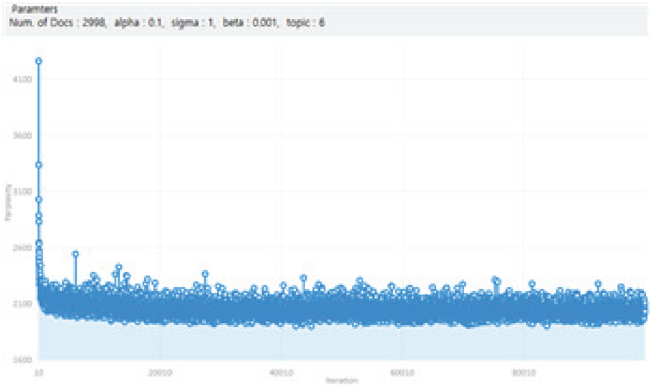


Table 7. Topic Results of top 30 Ranked Words

Topic 1		Topic 2		Topic 3		Topic 4		Topic 5		Topic 6	
Word	**Probability**	**Word**	**Probability**	**Word**	**Probability**	**Word**	**Probability**	**Word**	**Probability**	**Word**	**Probability**
seat	0.0443	one	0.0176	drive	0.0230	feature	0.0310	great	0.0501	drive	0.0289
comfort	0.0299	purchase	0.0147	gas	0.0198	system	0.0224	very	0.0438	vehicle	0.0235
back	0.0174	love	0.0145	mileage	0.0150	seat	0.0180	drive	0.0358	very	0.0181
space	0.0172	vehicle	0.0142	get	0.0149	like	0.0160	good	0.0278	look	0.0167
great	0.0168	year	0.0127	mpg	0.0128	camera	0.0123	gas	0.0248	feature	0.0166
drive	0.0167	time	0.0111	power	0.0124	control	0.0122	love	0.0245	comfort	0.0166
very	0.0154	new	0.0097	very	0.0111	safety	0.0102	vehicle	0.0241	bmw	0.0163
love	0.0152	will	0.0096	handle	0.0104	love	0.0100	comfort	0.0241	great	0.0147
room	0.0145	buy	0.0092	mile	0.0099	interior	0.0098	mileage	0.0233	handle	0.0137
ride	0.0113	subaru	0.0087	engine	0.0098	use	0.0093	look	0.0207	interior	0.0134
easy	0.0108	kia	0.0086	road	0.0096	great	0.0093	feature	0.0158	lexus	0.0132
vehicle	0.0103	because	0.0084	use	0.0096	sound	0.0086	reliable	0.0158	best	0.0119
cargo	0.0099	another	0.0081	good	0.0093	navig	0.0083	interior	0.0142	quality	0.0118
need	0.0092	first	0.0081	accelerate	0.0090	rear	0.0083	handle	0.0138	perform	0.0117
trip	0.0090	just	0.0079	highway	0.0085	light	0.0079	like	0.0135	audi	0.0112
like	0.0085	get	0.0077	great	0.0084	vehicle	0.0078	get	0.0121	luxury	0.0112
get	0.0080	problem	0.0076	can	0.0076	also	0.0076	well	0.0121	love	0.0111
row	0.0079	want	0.0075	sport	0.0076	heat	0.0075	kia	0.0120	volvo	0.0097
road	0.0079	only	0.0075	well	0.0074	wheel	0.0075	easy	0.0102	style	0.0097
use	0.0078	drive	0.0074	like	0.0074	cruise	0.0074	fun	0.0102	reliable	0.0096
fit	0.0078	issue	0.0073	little	0.0071	spot	0.0072	ride	0.0101	safety	0.0094
lot	0.0078	honda	0.0070	turbo	0.0071	side	0.0066	nice	0.0098	ride	0.0088
small	0.0076	never	0.0070	mode	0.0065	lane	0.0066	price	0.0091	technology	0.0081
can	0.0073	reliable	0.0069	speed	0.0065	blind	0.0066	safety	0.0079	fun	0.0079
well	0.0070	bought	0.0069	overall	0.0060	phone	0.0065	safe	0.0077	feel	0.0074
passenger	0.0069	look	0.0066	fun	0.0060	only	0.0065	recommend	0.0076	well	0.0071
suv	0.0067	any	0.0059	need	0.0057	steer	0.0064	suv	0.0075	excel	0.0066
family	0.0066	replace	0.0056	take	0.0057	option	0.0061	best	0.0073	many	0.0061
handle	0.0066	like	0.0055	vehicle	0.0056	can	0.0061	exterior	0.0073	sport	0.0061
trunk	0.0066	service	0.0053	bit	0.0055	back	0.0061	feel	0.0073	like	0.0061

From the 2,998 reviews collected, the average of total ratings were estimated to be 4.628, as shown in Table 8. An independent sample t-test showed that there were no differences in the averages between both brands at p=0.001. It is thus possible to examine the portion of each topic extracted from the topic analysis, with the overall ranking as a dependent variable.

Table 8. T-test

Variable		N	Mean	Std. Deviation	*t*	*p*
Overall	Mass Market	1,558	4.599	0.741	-2.246	.003**
	Luxury	1,440	4.659	0.708		
	Total	2,998	4.628	0.725		

*p<0.1, **p<0.05, ***p<0.01

The author set models, as shown in Table 9, in order to verify the hypothesis through regression analysis. Table 10 shows the results of the regression equation as follows: Overall = 4.558 + 0.006 (Topic 1) - 0.007 (Topic 2) - 0.004 (Topic 3) - 0.006 (Topic 4) + 0.010 (Topic 5) + 0.0149 (Topic 6).

According to this analysis, Topic 1 (Competence), Topic 5 (Sincerity), and Topic 6 (Prestige) attributes have a positive effect overall in customer reviews, whereas Topic 2 (Sophistication) and Topic 4 (Conspicuousness) attributes have a negative effect on overall customer reviews. Interestingly, Topic 3 (Ruggedness) did not have any statistical effect in this research.

Table 9. Model Summary and ANOVA*

Model		R	R Square	Adjusted R Square	Std. Error of the Estimate
1		.183	.034	.032	.71414
Model		**Sum of Squares**	**df**	**Mean Square**	**F.**	**Sig.**
1	Regression	52.923	6	8.821	17.295	.000**
	Residual	1525.392	2991	.510		
	Total	1578.315	2997			

* Dependent Variable: Overall
**Predictors: (Constant), Topic 1, Topic 2, Topic 3, Topic 4, Topic 5, Topic 6

Table 10. Regression Coefficient

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	4.558	.026		175.121	.000***
	Topic 1	.006	.002	.058	3.179	.001***
	Topic 2	-.007	.002	-.075	-4.115	.000***
	Topic 3	-.004	.002	-.046	-2.507	.012
	Topic 4	-.006	.002	-.057	-3.082	.002***
	Topic 5	.010	.002	.087	4.539	.000***
	Topic 6	.014	.002	.144	7.568	.000***

Dependent Variable: Overall
*p<0.1, **p<0.05, ***p<0.01

The hypothesis was tested to show that automobile reviews state brand personality across 6 topics. Among these, five topics, competence, sophistication, conspicuousness, sincerity, and prestige, influenced the total ratings of vehicles. As demonstrated by a previous study on vehicle purchasing behavior, several factors, including engines, may have no significant effect on purchasing behavior as vehicle performance has improved. The analysis has an academic implication in that it can categorize sentences contained in customer reviews into topics using data mining. Previous studies could respond to customer review variables , set by the researcher, but others that the researcher did not define were excluded from study models. Text mining is useful in discovering variables that researchers might otherwise ignore, as it can extract meaningful types from customer reviews.

As suggested by Schindler and Bickart (2005), the increase of customer reviews contributes to better information, winning customer trust, and creating a positive effect on vehicle purchasing behavior. It was thus found that online word-of-mouth played a valid role in inducing positive purchasing behavior.

5. Results and Conclusion

This paper tried to identify the brand personalities of global automakers using topic modeling analysis through text mining. Lately, global automakers have focused on brand management in order to increase the value of their brands. Automobile brands have attentively been managed through customer reviews for customer satisfaction. Customer reviews seem to have greatly reflected brand personality to provide reliable information for customers that want to buy cars.

The reviews provided by customers are the word-of-mouth method most efficiently used in marketing today. Since the effect of online word-of-mouth has been grown in importance, customer reviews reflecting brand personality have become strong marketing tools.

In the present study, after identifying attributes of brand personality using customer reviews, the author investigated whether or not such attributes had an effect on the overall evaluation of customer satisfaction. The author postulated a hypothesis and verified it using large and sophisticated datasets consisting of customer reviews from J.D. Power in the USA. Contrary to a traditional approach to brand analysis using questionnaire survey methods, this present study analyzed customer reviews using text mining.

Existing questionnaire surveys were designed based on variables researchers grasped in advance with items adjusted to study models. Other variables that researches could not define, despite level of importance, were not reflected. Text mining as a big data analysis method overcomes this limit. Word-of-mouth data, such as reviews written by customers, can be understood with topic modeling analysis.

This study is timely research a big data analysis is employed in order to identify direct responses to customers in the future. This study, however, has research limitations that should be supplemented since it did not distinguish brands and extracted attributes of brand personality from the dataset.

This study conducted text mining of all customer reviews without distinguishing mass market automotive brands from luxury automotive brands.

In future studies, there is a need to subdivide brands of automakers into luxury and mass market automotive brands, and then compare and analyze these brands. According to Choo and Mokhtarian (2004), it is necessary to reflect the differences between the two groups, as there are separate characteristics influencing the purchase of luxury vehicles.

References

Aaker, J. L. (1997), “Dimensions of Brand Personality”, *Journal of Marketing Research*, 34(3), 347-356.

Aaker, J. L. (1999), “The Malleable Self: The Role of Self-Expression in Persuasion”, *Journal of Marketing Research*, 36(1), 45-57.

Aaker, J. L., K. D. Vohs and C. Mogilner (2010), “Nonprofits are seen as warm and for-profits as competent: firm stereotypes matter”, *Journal of Consumer Research*, 37(2), 224-237.

Aaker, J. L., V. Benet-Martinez and J. Garolera (2001), “Consumption symbols as carriers of culture: a study of Japanese and Spanish brand personality constructs”, *Journal of Personality and Social Psychology*, 81(3), 492-508.

Archak, N., A. Ghose and P. G. Ipeirotis (2011), “Deriving the pricing power of product features by mining consumer reviews,” *Management Science*, 57(8), 1485-1509. https://doi.org/10.1287/mnsc.1110.1370

Arnould, E. J. and L. L. Price (1993), “River Magic: Extraordinary Experience and the Extended Service Encounter”, *Journal of Consumer Research*, 20(1), 24-45.

Baltas, G. and C. Saridakis (2013), “An empirical investigation of the impact of behavioral and psychographic consumer characteristics on car preferences: An integrated model of car type choice”, *Transportation Research Part A: Policy and Practice*, 54, 92-110. https://doi.org/10.1016/j.tra.2013.07.007.

Belk, R. W. (1988), “Profession and Extended Self”, *Journal of Consumer Research*, 15(7), 139-169.

Berkovec, J. and J. Rust (1985), “A nested logit model of automobile holdings for one vehicle households”, *Transportation Research Part B: Methodological*, 19, 275-285.

Biber, D., S. Conrad and R. Reppen (1998), *Corpus Linguistics: Investigating language structure and use*, Cambridge, UK: Cambridge University Press.

Blei, D. M. (2011), “Introduction to Probabilistic Topic Models”, *Communications of the ACM*, 55, 1-16.

Blei, D. M., A. Y. Ng and M. I. Jordan (2003), “Latent Dirichlet Allocation”, *Journal of Machine Learning Research*, 3(January), 993-1022.

Bosnjak, M., V. Bochmann and T. Hufschmidt (2007), “Dimensions of brand personality attributions: a person- centric approach in the German cultural context”, *Social Behavior and Personality*: An International Journal, 35(3), 303-316.

Brakus, J., B. Schmitt and L. Zarantonello (2009), “Brand Experience: What Is It? How Is It Measured? Does It Affect Loyalty?”, *Journal of Marketing*, 73(5), 52-68.

Brister, J. M. (1991), “Word of mouth communication and their effects in consumer network”, *Advances in Consumer Research*, 18, 155-169.

Brownstone, D., D. Bunch and K. Train (2000), “Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles”, *Transportation Research Part B: Methodological*, 34, 315-338.

Campbell, A. R., T. Ryley, and R. Thring (2012), “Identifying the early adopters of alternative fuel vehicles: a case study of Birmingham, United Kingdom”, *Transportation Research Part A: Policy and Practice*, 46, 1318-1327.

Cao, X., P. L. Mokhtarian and S. L. Handy (2006), “Neighbourhood design and vehicle type choice: evidence from Northern California”, *Transportation Research Part D: Transport and Environment*, 11, 133-145.

Caprara, G. V., C. Barbaranelli and G. Guideo (2001), “Brand Personality: How to Make the Metaphor Fit”, *Journal of Economic Psychology*, 22(3), 377-395.

Cattell, R. B. (1943), “The Description Of Personality: Basic Traits Resolved Into Clusters”, *Journal of Abnormal And Social Psychology*, 38(4), 476-506.

Chae, Seung-Hoon, Jay-Ick Lim and Kang Ju-Yong (2015), “A Comparative Analysis of Social Commerce and Open Market Using User Reviews in Korean Mobile Commerce,” *Journal Intelligence Information System*, 21(4), 53-77. http://dx.doi.org/10.13088/jiis.2015.21.4.053

Chatterjee, P. (2001), "Online review: Do consumers use them?", *Advances in Consumer Research*, 28, 129-133.

Chen, K. C. (2009), “Text Mining e-Complaints Data From e-Auction Store with Implications for Internet Marketing Research”, *Journal of Business & Economics Research*, 7(5), 42-49.

Chen, Q. and S. Rogers (2006), “Development of an instrument to measure web site personality”, *Journal of Interactive Advertising*, 7(1), 4-46.

Chevalier, J. A. and D. Mayzlin (2003), *The Effect of Word of Mouth on Sales: Online Book Reviews* (Working Paper w10148), Cambridge, MA: National Bureau of Economic Research, Available from https://www.nber.org/system/files/working_papers/w10148/w10148.pdf

Choo, S. and P. L. Mokhtarian (2004), “ What type of vehicle do people drive? The role of attitude and lifestyle in influencing vehicle type choice”, *Transportation Research Part A: Policy and Practice*, 38, 201-222.

Coughlan, M., P. Cronin and F. Ryan (2013), “Survey research: Process and limitations,” *International Journal of Therapy and Rehabilitation*, 16(1), 9-15. https://doi.org/10.12968/ijtr.2009.16.1.37935

d’Astous, A. and L. Boujbel (2007), “Positioning countries on personality dimensions: scale development and implications for country marketing”, *Journal of Business Research*, 60(3), 231-239.

d’Astous, A. and M. Levesque (2003), “A scale for measuring store personality”, *Psychology and Marketing*, 20(5), 455-469.

Das, G., B. Datta and K. K. Guin (2012), “Impact of retailer personality on consumer-based retailer equity: an empirical study of retail brands”, *Asia Pacific Journal of Marketing and Logistics*, 24(4), 619-639

Davies, G., J. I. Rojas-Méndez, S. Whelan, M. Mete and T. Loo (2018), “Brand personality: theory and dimensionality”, *Journal of Product & Brand Management*, 27(2), 115-127.

Davies, G., R. Chun, R. V. da Silva and S. Roper (2004), “A corporate character scale to assess employee and customer views of organisation reputation”, *Corporate Reputation Review*, 7(2), 125-146.

Egbert, J. and D. Biber (2019), “Incorporating text dispersion into keyword analyses”, *Corpora*, 14(1), 77-104.

Eisend, M. and N. E. Stokburger-Sauer (2013), “Brand personality: a meta-analytic review of antecedents and consequences”, *Marketing Letters*, 24(3), 205-216.

Erickson, G., J. Johansson and P. Chao (1984), “The Image Variables in Multi-Attribute Product Evaluations: Country-of-Origin Effects”, *Journal of Consumer Research*, 11, 694-699, https://doi.org/10.1086/209005

Fiske, D. W. (1949), “Consistency of the Factorial Structures of Personality Ratings from Different Sources”, *Journal of Abnormal and Social Psychology*, 44(3), 329-344.

Gabrielatos, C. (2018), "Keyness Analysis: nature, metrics and techniques". In C. Taylor and A. Marchi (Eds.), Corpus Approaches to Discourse: A critical review. Chapter 12, Oxford, UK: Routledge.

George, A. (1959), “Quantitative and Qualitative Approaches to Content Analysis”. In A. Raftery (Ed.), *Sociological Methodology*, Oxford: Basil Blackwell, 35-144.

Geuens, M., B. Weijters and K. de Wulf (2009), “A new measure of brand personality”, *International Journal of Research in Marketing*, 26(2), 97-107.

Godbole, S. and S. Roy (2008), “Text to Intelligence: Building and Deploying a Text Mining Solution in the Services Industry for Customer Satisfaction Analysis”, *2008 IEEE International Conference on Services Computing*, Honolulu, HI., 441-448,

Goldberg, L. R. (1993), “An Alternative Description of Personality: The Big Five Factor Structure”, *Journal of Personality and Social Psychology*, 59(6), 1216-1229.

Goyette, I., L. B. Ricard and F. Marticotte (2010), “E-WOM Scale: Word of Mouth Measurement Scale for E-Services Context”, *Canadian Journal of Administrative Science*, 27, 5-23.

Grimes, S. (2005), “The developing text mining market”, white paper, Text Mining Summit Alta Plana Corporation, Boston, MA., 1-12.

Grönroos, C. (1996), “The value concept and relationship marketing”, *European Journal of Marketing*, 30(2), 19-30.

Grönroos, C. (1997), “Value-driven relational marketing: From products to resources and competencies”, *Journal of Marketing Management*, 13(5), 407-419.

Gupta, V. and G. Lehal (2009), “A Survey of Text Mining Techniques and Applications”, *Journal of Emerging Technologies in Web Intelligence*, 1(1), August 2009, 60-76.

Henning-Thurau, T., K. P. Gwinner, G. Walsh and D. D. Gernler (2004), "Electronic word of mouth viva consumer opinion platforms: what motivates consumers to articulate themselves on the internet", *Journal of Interaction Marketing*, 18(1), 38-52.

Herbst, U. and M. A. Merz (2011), “The industrial brand personality scale: building strong business-to-business brands”, *Industrial Marketing Management*, 40(7), 1072-1081.

Hocherman, I., J. Prashker and M. Ben-Akiva, M. (1983), “Estimation and use of dynamic transaction models of automobile ownership”, *Transportation Research Record*, 944, 134-141.

Holsti, O. R. (1969), *Content Analysis for the Social Sciences and Humanities*, Reading, MA: Addison-Wesley.

J.D.Power (2019), U. S. Automotive Brand Loyalty Study (2019), http://www.jdpower.com/pr-id/2019173 (accessed June 25, 2020)

Kaplan, M. D., O. Yurt, B. Guneri and K. Kurtulus (2010), “Branding places: applying brand personality concept to cities”, *European Journal of Marketing*, 44(Nos 9/10), 1286-1304.

Keller, K. L. (1993), “Conceptualizing, Measuring, and Managing Customer-Based Equity”, *Journal of Applied Psychology*, 85(5), 751-765.

Kim, En-Gir and Se-Hak Chun (2019), "Analyzing Online Car Reviews Using Text Mining", *Sustainability*, 11(6), 1611. https://doi.org/10.3390/su11061611

Kim, Kwang-Kook, Yong-Hwan Kim and Ja-Hee Kim (2018), “A Study on Customer Satisfaction of Mobile Shopping Apps Using Topic Analysis of User Reviews”, *The Journal of Society for e-Business Studies*, 23(4), 41-62. https://doi.org/10.7838/jsebs.2018.23.4.041

Kim, Yong-Hwan, Ja-Hee Kim, Ji-Hoon Park and Seung-Jun Lee (2016), “Analysis of Success Factors for Mobile Commerce using Text Mining and PLS Regression,” *Journal of the Korea Society of Computer and Information*, 21(11), 127-134.

Kostyra, D. S., J. Reiner, M. Natter and D. Klapper (2016), “Decomposing the effects of online customer reviews on brand, price, and product attributes,” *International Journal of Research in Marketing*, 33(1), 11-26. https://doi.org/10.1016/j.ijresmar.2014.12.004

Kotler, P., K. L. Keller, M. Brady, M. Goodman and T. Hansen (2009), Marketing *Management*, Harlow, UK: Pearson Education.

Kozinets, R. V. (2002), “Can Consumers Escape the Market? Emancipatory Illuminations from Burning Man”, *Journal of Consumer Research*, 29(1), 20-38.

Lave, C. A., and K. Train (1979), “A disaggregate model of auto-type choice”, *Transportation Research Part A: Policy and Practice*, 13, 1-9.

Macnamara, J. (2003), “Media Content Analysis–Its Uses, Benefits and Best Practice Methodology”, *Asia Pacific Public Relations Journal*, 6(1), 1-34.

Maheswaran, D., D. M. Mackie and S. Chaiken (1992), “Brand name as a heuristic cue: The effects of task importance and expectancy confirmation on consumer judgments”, *Journal of Consumer Psychology*, 1(4), 317-336. https://doi.org/10.1016/S1057-7408(08)80058-7

Mannering, F. and C. Winston (1985), “A dynamic empirical analysis of household vehicle ownership and utilization”, *Rand Journal of Economics*, 16, 215-236.

Mannering, F., C. Winston, C. and W. Starkey (2002), “An exploratory analysis of automobile leasing by US households”, *Journal of Urban Economics*, 52, 154-176.

Manski, C. F. and L. Sherman (1980), “An empirical analysis of household choice among motor vehicles”, *Transportation Research Part A: Policy and Practice*, 14, 349-366.

Mantel, S. P. and F. R. Kardes (1999), “The role of direction of comparison, attribute-based processing, and attitude-based processing in consumer preference”, *Journal of Consumer Research*, 25(4), 335-352.

McEnery, T., A. McEnery, R. Xiao and Y. Tono (2006), *Corpus-based Language Studies: An*

Advanced Resource Book, Oxford, UK: Routledge.

Meyer, C. and A. Schwager (2007), "Understanding Customer Experience", Harvard Business Review, 85(2), 116-126.

Milas, G. and B. Mlačić (2007), "Brand personality and human personality: findings from ratings of familiar Croatian brands", *Journal of Business Research*, 60(6), 620-626.

Mudambi, S. M. and D. Schuff (2010), "What makes a helpful review? A study of customer reviews on Amazon. com.," *MIS quarterly*, 34(1), 185-200.

Muniz, K. M. and R. Z. Marchetti (2012), "Brand personality dimensions in the Brazilian context", *Brazilian Administration Review*, 9(2), 168-188.

Neuendorf, K. A. (2002), *The Content Analysis Guidebook*, Thousand Oak, CA: Sage.

Newbold, C., O. Boyd-Barrett and H. Van den Bulck (2002), *The Media Book*, London: Arnold.

Nguyen, H., R. Calantone and R. Krishnan (2019), "Influence of Social Media Emotional Word of Mouth on Institutional Investors' Decisions and Firm Value", *Management Science*, 66(2), 887-910. https://doi.org/10.1287/mnsc.2018.3226

Nguyen, N. and G. Leblanc (2001), "Corporate Image and Corporate Reputation in Customers' Retention Decisions in Services", *Journal of Retailing & Consumer Services*, 8(4), 227-236.

Norman, W. T. (1967), *2,800 Personality Trait Descriptors: Normative Operating Characteristics for a University Population*, Ann Arbor, MI: Department of Psychology, University of Michigan.

Palmatier, R. W. (2008), *Relationship Marketing*, Cambridge, UK: Marketing Science Institute.

Pine, B. J. and Gilmore, J. H. (1999), *The Experience Economy: Work Is Theatre and Every Business A Stage*, Boston, MA: Harvard Business School Press.

Pool, I. de S. (Ed.) (1959), *Trends in Content Analysis*, Champaign: University of Illinois Press.

Rainer, A. and T. Hall (2003), "A Quantitative and Qualitative Analysis of Factors Affecting Software Processing", *Journal of Systems and Software*, 66(1), 7-13.

Ramanathan, V. and T. Meyyappan (2013), "Survey of Text Mining", *International Conference on Technology and Business and Management*, March 2013, 508-514.

Rauschnabel, P. A., N. Krey, B. J. Babin and B. S. Ivens (2016), "Brand management in higher education: the university brand personality scale", *Journal of Business Research*, 69(8), 3077-3086.

Reisenbichler, M. and T. Reutterer (2018), "Topic modeling in marketing: recent advances and research opportunities", *Journal of Business Economics*, 1-30.

Reppen, R. (2001), "Review of MonoConc Pro and WordSmith Tools", *Language Learning & Technology*, 5(3), 32-36.

Richardson, P. S., A. S. Dick and A. K. Jain (1994), "Extrinsic and intrinsic cue effects on perceptions of store brand quality", *Journal of Marketing*, 58(4), 28–36. https://doi.org/10.2307/1251914

Roberts, C. W. (2000), "A Conceptual Framework for Quantitative Text Analysis", *Quality & Quantity*, 34, 259-274.

Rojas-Méndez, J. I., S. A. Murphy and N. Papadopoulos (2013a), "The US brand personality: a Sino perspective", *Journal of Business Research*, 66(8), 1028-1034.

Rojas-Méndez, J. I., S. A. Murphy and N. Papadopoulos (2013b), "Measuring and positioning nation brands: a comparative brand personality approach", *Corporate Reputation Review*, 16(1), 48-65.

Romano, N. C., Jr., C. Donovan, H. Chen and J. F. Nunamaker Jr. (2003), "A Methodology for Analyzing Web-Based Qualitative Data", *Journal of Management Information Systems*, 19 (4), 213-246.

Schindler, R. M. and B. Bickart (2005), "Published Word of Mouth: Referable, Consumer-Generated Information on the Internet". In C. Hauvgedt, M. Karen and R. Yalch (Eds.), Online consumer psychology: Understanding and influencing consumer behavior in the virtual world, Mahwah, NJ: Lawrence Erlbaum Associates, 35-61.

Scott, Mike (1997), "PC Analysis of Key Words", *System*, 25(2), 233-245.

Scott, Mike (2016), *WordSmith Tools Help*, Stroud, UK: Lexical Analysis Software.

Seo, Min-Kyo, Yang Oh-Suk and Yang Yoon-Ho (2020), "Global Big Data Analysis Exploring the Determinants of Application Ratings: Evidence from the Google Play Store", *Journal of Korea Trade*, 24(7), 1-28. https://doi.org/10.35611/jkt.2020.24.7.1

Simmons, C. J. and J. G. Lynch Jr. (1991), "Inference Effects Without Inference Making? Effects of Missing Information on Discounting and Use of Presented Information", *Journal of Consumer Research*, 17(4), 477-491.

Slaughter, J. E., M. J. Zickar, S. Highhouse and D. C. Mohr (2004), "Personality trait inferences about organizations: development of a measure and assessment of construct validity", *Journal of Applied Psychology*, 89(1), 85-102.

Smit, E. G., E. Van den Berge and G. Franzen (2002), "Brands are just like real people! The development of SWOCC's brand personality scale", In F. Hansen and L. B. Christensen (Eds.), *Branding and Advertising*, Copenhagen, Denmark: Copenhagen Business School Press, 22-43.

Sung, Y., S. M. Choi, H. Ahn and Y. A. Song (2015), "Dimensions of luxury brand personality: scale development and validation", *Psychology & Marketing*, 32(1), 121-132.

Swaminathan, V., K. L. Page and Z. Gürhan-Canli (2007), ""My" Brand or "Our" Brand: The Effects of Brand Relationship Dimensions and Self-Construal on Brand Evaluations", *Journal of consumer research*, 34(2), 248-259.

Train, K. and C. Winston (2007), "Vehicle choice behavior and the declining market share of U.S. automakers", *International Economic Review*, 48, 1469-1496.

Tupes, E. C. and R. E. Christal (1961), "Recurrent Personality Factors Based on Trait Ratings", *Journal of Personality*, 60(2), 225-251.

Venable, B. T., G. M. Rose, V. D. Bush and F. W. Gilbert (2005), "The role of Brand personality in charitable giving: an assessment and validation", *Journal of the Academy of Marketing Science*, 33(3), 295-312.

Vergne, J. P. and T. Wry (2014), "Categorizing Categorization Research: Review, Integration, and Future Directions", *Journal of Management Studies*, 51(1), 56-94.

Verhoef, P. C., K. N. Lemon, A. Parasuraman, A. Roggeveen, M. Tsiros and L. A. Schlesinger (2009), "Customer Experience Creation, Determinants, Dynamics, and Management Strategies", *Journal of Retailing*, 85(1), 31-41.

Vidya, K. A. and G. Aghila (2010), "Text Mining Process, Techniques and Tools: an Overview", *International Journal of Information Technology and Knowledge Management*, 2(2), 613-622.

Voss, P. Jr. (1984), "Status shifts to Peer Influence", *Advertising Age*, 17, 1-10.

Weber, R. P. (1990), *Basic Content Analysis* (2nd ed.), Newbury Park, CA: Sage.

Wickham, M. and M. Woods (2005), "Reflecting on the Strategic Use of CAQDAS to Manage and Report on the Qualitative Research Process", *The Qualitative Report*, 10(4), 687-702.

Willems, K., G. Swinnen, W. Janssens and M. Brengman (2011), "Fashion store personality: scale development and relation to self-congruity theory", *Journal of Global Fashion Marketing*, 2(2), 55-65.