UNITED KINGDOM · CHINA · MALAYSIA

Decision-making for high-involvement products: Topic modelling using online reviews

by

Chanyoung Lee

2020

A Dissertation presented in part consideration for the degree of MSc Business Analytics.

ACKNOWLEDGEMENT

First and foremost, heartily thankful to my supervisor, Professor Andrew Smith, whose guidance, supervision, constant encouragement and suggestions, enabled me to develop an understanding of my dissertation.

Secondly, I give thanks to other teaching staffs of the MSc Business Analytics that providing offline and online tutoring and thanks for the tremendous help I have received during 2019/2020 academic year.

Last but not least, I am sincerely grateful for the hidden support of my family.

Despite all the help I have received, the responsibility of any fault in the dissertation remains entirely mine.

ABSTRACT

High-involvement products require a complex decision-making process. Automobiles require time and effort and a financial budget to purchase, requiring the consideration of diverse characteristics of the vehicle before purchase. Research in this area has focused primarily on finding the most significant features when purchasing an automobile using the traditional statistical method with surveys. However, advanced linguistic technique analysis provides an opportunity to extract meaning from the diverse comments provided by owners. In this paper, the author identifies the key topics of the customer decision-making process from electric automobile owners using a topic modelling approach with a latent Dirichlet allocation (LDA) model combined with natural language processing techniques. The dataset includes 956 free-text customer online reviews for Tesla. In an exploratory analysis involving electric automobiles, LDA uncovered 10 comprehensive lists of topics discussed by customers. Topics are key for electric automobile companies to manage their interactions with customers by understanding the primary interests and features in terms of the decision-making process of current and future customers. The proposed approach and findings are beneficial to support understanding customer perceptions. Through this method, the marketing and business strategy can be improved to maintain current customers and attract future customers.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGUES

CHAPTER 1. INTRODCUTION

The automobile has become an indispensable part of life in the modern era, and the types and brands of an automobile are numerous. Consumers have a variety of options to consider, such as safety and comfort (Sakthivel et al., 2013), in purchasing automobiles; thus, they purchase an automobile after complicated consideration. Studies have been conducted, finding safety (Koppel et al., 2008); (Johansson-Stenman and Martinsson, 2006), corporate brand credibility (Li et al., 2011), environmentally friendly (Yusof et al., 2013), and accessibility (Narteh et al., 2012) to be significant features when purchasing an automobile. Furthermore, existing studies have focused on exploring the consumer car purchase decision-making process (Byun, 2001); (Sakthivel et al., 2013) using traditional statistical methods. However, studies face difficulty in reflecting all characteristics using the traditional models because these models cannot think like a human who considers many kinds of features at once.

The reason consumers require the complicated process of decision-making when purchasing an automobile is that the automobile is a high-involvement product. High-involvement products are risky, expensive, and involve a complex decision process (Jain, 2019). Thus, similar to complex problem-solving, consumers tend to have a high level of concern when purchasing a high-involvement product (Quester and Pascale Genevieve 2007). To make a wise decision, the consumer relies on various information and considers the diverse characteristics of a specific product, comparing alternatives. In modern society with a highly developed Internet, the easiest method of product research is to find information is from online reviews in which owners of a particular product have written about their experiences in detail. This paper focuses on electric vehicles (EVs) as a high-involvement product. The interest in EVs has increased rapidly over a few years due to technological development (Lassila et al., 2011). Several studies have investigated the reasons that the scale of the EV industry has become large and what opens the new opportunity for electric automobile purchases.

Because online reviews contain the most interesting characteristics of the product, analysing online reviews of EVs can facilitate determining the consumer decision-making process by assessing which features and details consumers consider when purchasing EVs. Moreover, potential customer decision-making processes can be deduced from the analysed results of the current owner reviews. Additionally, as the number of customer reviews continues to increase (Zhang and Zhu, 2013) and consumers consider the reviews to be reliable (Park et al., 2007), researchers have used online reviews as a dataset. However, the majority of the existing research has not focused on high-involvement products. To the best of the authors' knowledge, online consumer reviews have not been used to examine consumer behaviour in EVs despite their immense potential. In addition, EV managers continue to face challenges in understanding customer behaviour and experiences for effective management and marketing of EVs. The following essential questions have yet to be thoroughly answered: (1) What are the main topics that consumers discuss in online reviews about electric cars? (2) What are the critical technical aspects that influence the consumer experience? (3) Has COVID-19 affected consumer purchasing?

This study aims to fill in this research gap and presents an approach to analysing the experiences and behaviour of EV owners based on online reviews. The approach employs topic modelling, which is one of the techniques in artificial intelligence (AI) that focuses on language and aims to discover hidden semantic information in textual data, such as documents and reviews. Topic modelling has been applied in some recent studies using online reviews (Luo et al., 2020); (Özdağoğlu et al., 2018); (Goh et al., 2013). However, its performance in inspecting the consumer decision-making process regarding EVs has not been evaluated. The current study presents the analysis of the consumer experience at Tesla, which is a popular automobile brand specialising in EVs. The analysis is based on online reviews and reveals popularly discussed topics, which are grouped into categories. This study is beneficial for EV researchers and EV managers using online reviews to gain general insight into consumer decision-making regarding EVs.

The remainder of this paper is organised as follows. The Literature Review section discusses the decision-making process of automobile purchasing and the high-involvement product assessed in this study and presents a review of the existing literature on employing online reviews and topic modelling. The Methodology section describes the methodology for textual data processing and topic discovery. The Results section presents the results of the analysis using visualisation tools and topic grouping. Next, the Discussion section presents details about the result interpretation with practical implications and the limitations and directions for further study. Lastly, the Conclusion section provides a summary of this study and the conclusion.

CHAPTER 2. LITERATURE REVIEW

2.1 Automobile Purchase Decision Making Process

In the olden days, automobiles were only for wealthy people because of their high cost and limited supply. However, in the modern era, diverse automobile brands have been launched and many different types of car have been created, thus cars have become more affordable. As a result, automobiles have become an indispensable part of life. Purchasing a new vehicle is regarded as a decision that mainly reflects customer preferences. Before a customer visits a showroom in person to buy a new car in person, they should consider their expected expenditure and various options. Within their available budget, customers should consider important attributes including safety, fuel economy, comfort and convenience features, car options and insurance information.

As the automobile market becomes more competitive, a greater demand has been generated for innovation such that manufacturers are required to consider not just the price and fuel economy. Meanwhile, customers evaluate automobiles in terms of many extra features such as safety, comfort and convenience (Sakthivel et al., 2013). Constant changes in customer demands when purchasing privately owned vehicles lead manufacturers to produce innovative and improved vehicles. Several studies have examined what factors influence new-vehicle purchasing decisions. In the automobile context, the influence of numerous factors such as brand, country of origin, corporate brand credibility (Li et al., 2011) and price on consumer's purchase intention have been examined. Other factors include environmentally conscious consumer behaviour (McIntosh, 2009) and accessibility (Narteh et al., 2012).

(Koppel et al., 2008); (Boyle and Schulman, 1996) found that the importance of vehicle safety for privately owned vehicles had increased over the previous decade. (Johansson-Stenman and Martinsson, 2006) investigated Swedish drivers' perception of vehicle features, including safety, environmental friendliness, look, motor power, comfort, space, fuel consumption and reliability when purchasing a vehicle; the results of that study found that safety was rated as 'very important' (85%), followed by reliability and fuel consumption, while status was the least important characteristic (6%). (Vrkljan and Anaby, 2011) also argued that safety and reliability were considered the most important features when buying a car; the results of that study found that safety and reliability were rated as the most important features, whereas design and performance had the lowest rating. (Yusof et al., 2013) examined the purchase intention of environmentally friendly automobiles in terms of consumers' perceptions of environmental advertisements. The results indicated that consumer's environmental responsibility and values significantly influenced their purchase behaviour with regard to environmentally friendly vehicles.

Several studies have investigated car purchase decision processes. The analytic hierarchy process (AHP) is an intuitive approach for formulating, analysing and then solving decision-making problems (Saaty, 1980); (Saaty and Kearns, 2014). This is achieved by evaluating a group of main and sub-criteria elements through complex pairwise comparisons. AHP has been applied in numerous types of decision problems (Byun and Suh, 1996); (Lai et al., 1999); (Liberatore and Stylianou, 1994); (Zahedi, 1986). The AHP method has also used to make car purchase decisions. (Byun, 2001) proposed an AHP model for the car selection problem. This paper explores the use of an extension of AHP on car purchases by focusing on two issues. First, pairwise comparison is implemented using a five-point Likert rating scale. Second, the group weight is applied to a reciprocal consistency ratio. Nevertheless, AHP is widely used in decision-making problems and conventional AHP cannot reflect the style of human thought processes (Mikhailov, 2003); (Chan, 2003). Another method is the multi-criteria decision-making (MCDM) approach that considers diverse criteria such as safety, comfort, economic aspect, exterior and convenience (Sakthivel et al., 2013). Nonetheless, these findings still have flaws with the current state of knowledge and methods regarding vehicle purchase decisions. For instance, vehicle purchase decision-making generally involves trade-offs among price, reliability, safety and intended use; however, few studies and little information have attempted to reflect all characteristics (Koppel et al., 2008). In this paper, instead of implementing AHP or MCDM to examine decision-making process, natural language processing (NLP), which is AI that mainly do with language is used as a tool.

2.2 High-Involvement Products and Electric Vehicles

When purchasing a high-involvement product, the process of decision-making is more complicated and intensive compared to buying a low-involvement product. (Jain, 2019) claimed that the three most important characteristics of high-involvement products are that they are risky, expensive and involve a complex decision process. The decision-making process is complex because the buyer has to consider the cost, the risk, the possibility of malfunction and the available alternatives to minimise potential regret after the purchase. Many studies have investigated consumers' approaches to decision-making with respect to high-involvement purchases (Nayeem and Casidy, 2013) and found that the buyer's degree of involvement is a significant factor in the process (Jain, 2019). Consumers' decision-making for high-involvement purchases is similar to complex problem-solving and requires a sequential process comprising problem recognition, information search, brand evaluation and selection, purchase and post-purchase actions (Quester and Pascale Genevieve 2007). In this sense, while purchasing a high-involvement product, the consumer tends to be more rational (Kotler and Philip 2007).

While purchasing a high-involvement product, the consumer may rely on varied information sources and consider multiple criteria and carefully evaluate numerous alternatives. For example, a consumer may obtain significant information related to their future purchase from sources such as friends, family members, product dealers, the internet and magazines, among others (Nayeem and Casidy, 2013). In the past when the internet wasn't ubiquitous, consumers experienced difficulty in identifying all needs, factors and information before purchasing a product. In the present era, technological development has replaced the traditional internal and external methods of information collection by the internet. The internet is an extremely efficient tool for searching and locating valuable information that is difficult to find in the real world. The impact of online websites has been substantial for both consumers and sellers (Gu et al., 2012). The internet and associated technologies enable young-generation buyers to incorporate real-time information in their decision-making process for purchasing high-involvement products (Santandreu and Shurden, 2017). This study found that durability, reliability and price were the main factors in young-generation buyers' decision-making process, rather than brand name, discount and service. In contrast, (Raj and Roy, 2015) argued that positive brand image is the most influential factor in the purchasing decision of high-involvement and technology products. Positive brand image can be regarded as psychological assurance that develops trust within customers for reputed brands.

There are several benefits to consumers when they purchase high-involvement products online. First, the imbalance of information power between the consumer and the dealer/salesperson is reduced (Molesworth and Suortti, 2002). Second, the internet provides more opportunities and channels to buy a product and allows increased interaction with a number of dealers (Wootten, 2003). Lastly, online consumer reviews provide information and recommendations. The quality and quantity of reviews influence consumers during their decision-making process (Park et al., 2007).

There are diverse types of products in the high-involvement product category. This paper focuses on EVs. Rapid technological development in recent years has opened a new opportunity for electric cars (Lassila et al., 2011). The economic and environmental benefits of EVs, which substitute fossil fuels with electricity to run the engine, have accelerated their sales (Larson et al., 2014). EVs reduce a significant amount of greenhouse gases and other emissions, and this feature promotes the use of renewable energy (Egbue and Long, 2012). The primary reasons for purchasing electric cars are their financial and technical attributes, including utility; purchase, operation and maintenance cost; and vehicle performance (Liao et al., 2017). The battery recharging time and after-sale warranties are also important factors to consider when purchasing an EV (Junquera et al., 2016). Many governments around the world have introduced policies to encourage EVs production and adoption (Sierzchula et al., 2014).

As EVs are a state-of-the-art technology in the vehicles industry, their development and adoption are sensitive to shifts in policies, brands and various technologies. When a customer plans to purchase an EV, the decision-making process is that of a high-involvement product. The customer has to actively engage in information search and assess the risks. As mentioned earlier, the purchase of high-technological and high-involvement products is associated with uncertainties, complications and risks (Raj and Roy, 2015).

2.3 Natural Language Processing and Topic Modelling on Customer Reviews

Topic modelling, which a technique in NLP is employed in this paper. With NLP techniques, one can explore hidden semantic patterns in unstructured text data, such as books, reports, academic dissertations and virtual objects including web pages, e-mails and online reviews. The main concept of topic modelling is grouping relatively similar documents or words within the same cluster. In general, NLP and text mining methods help examine a given text's meaning by extracting, uncovering and synthesizing information. NLP can handle a large volume of textual documents (Uys et al., 2008). There is an overlap between NLP and text mining, and their definitions can be vague. Different methods are implemented for different goals. Classic text mining techniques include basic text analyses, such as text clustering, categorisation and sentiment analysis, that mainly deal with the text itself (Han et al., 2012); (Pang and Lee, 2008). The NLP techniques of opinion mining and topic modelling enable the discovery of the underlying meaning of a text (Conrad and Schilder, 2007); (Meng et al., 2012).

Topic modelling can be structured in different mathematical frameworks and approaches. The following are some of the most popular topic modelling methods.

- Latent semantic indexing, which uses singular value decomposition on the document-term matrix (Deerwester et al., 1990); (Papadimitriou et al., 2000); (Dumais, 2004)
- Probabilistic latent semantic indexing (pLSI) (Hofmann, 1999)
- Mixture of unigrams model (Nigam et al., 2000)
- Non-negative matrix factorisation (NMF) (Arora et al., 2012)
- Latent Dirichlet allocation (LDA) (Blei et al., 2003)

The LDA model is a generative probabilistic model that was proposed by David Blei, Andrew Ng and Michael Jordan in 2003 (Blei et al., 2003). LDA enables discovering underlying semantic information from a massive amount of unstructured text data. An unsupervised method, LDA has become the most common method for topic modelling (Guo et al., 2017) because it can handle a large volume of data, disaggregated time periods and sparse data (Blei et al., 2003). As LDA is a probabilistic mixture over all the words of the vocabulary for the application of statistical techniques, it overcomes the disadvantages of earlier models such as pLSI and NMP by assigning numerous topics to each document.

This paper concentrates on a methodology for LDA topic modelling on customer review data. As the information about specific products from other consumers is considered reliable (Park et al., 2007) and the number of customer reviews continues to increase (Zhang and Zhu, 2013), researchers in diverse domains have applied topic modelling and text mining on online reviews. Examples include using opinion extraction in the banking domain to gain insight into public opinions and sentiments of conversations in Twitter (Kolyshkina et al., 2013), analysing the behaviours and experiences of theme park visitors based on online reviews (Luo et al., 2020) and identifying patterns in the relation between news reports about corporate social responsibility and operating performance forecasting (Lin and Hsu, 2018). Reviews can be studied to build representative profiles of customers and to analyse words in order to interpret customer behaviours, such as needs, satisfaction levels and feelings (Özdağoğlu et al., 2018). In this paper, it analysed the 'digital voice' of customers and extracted customer needs by applying LDA-based topic modelling. The data in the text was unstructured; therefore, topic modelling was an appropriate and useful tool for identifying and characterising the embedded information about consumer opinions on a particular product (Uys et al., 2008). Traditional statistical methods cannot analyse this kind of data (Guo et al., 2017); (Lu and Stepchenkova, 2015).

Several studies exist on text mining that examine online customer reviews on websites and transform them into specific information (Zhao et al., 2005); (Park and Lee, 2011). Some papers have applied text mining and topic modelling to online customer reviews to extract product features and customer satisfaction, sentiments and behaviours (Goh et al., 2013); Luo, Vu et al. 2020; (Zhang and Zhu, 2013); (Zhan et al., 2009); (Zhao et al., 2005); (Chen et al., 2009). Studies have used supervised machine learning models to discover the most critical information or the most mentioned topics (Titov and McDonald, 2008); (Zhan et al., 2009); (Mukherjee and Liu, 2012) as well as used topic modelling in unsupervised and probabilistic models to capture topics and sentiments in text (Mei et al., 2007); (Cheung et al., 2009); (Yu et al., 2010). LDA topic modelling with NLP techniques was used to discover the main topics from 21,580 restaurant customer reviews (Yu et al., 2010). (Moghaddam and Ester, 2012) proposed an opinion mining method based on grammatical relationships and reported its performance compared to different models on 500,000 reviews.

CHAPTER 3. METHODOLOGY

This section presents the methodology for the electric vehicle review analysis using topic modelling. Topic modelling refers to the process of finding topics in a set of documents. It is used to identify topics in documents, such as search engines, customer complaint systems, and so on. Data are extracted from online review websites and text pre-processing is implemented using modules of the Natural Language Toolkit (NLTK) (www.nltk.org) in the Python programming environment. Thereafter, the LDA topic modelling technique was applied to examine the hidden topics within the reviews.

Analysing text in the reviews requires several steps before applying topic modelling techniques. First, reviews are a free-text form that has no structure. Most customers write reviews using casual language. Second, some content in the reviews does not contain information about the product and must be removed. Lastly, the words in all reviews must be transformed into numerical form to be analysed (Tirunillai and Tellis, 2014).

3.1 Reviews Collection

EV reviews are available on various online review websites. This study web scraped from Car.com, ComsumerAffairs.com, and Edmunds.com as the data sources because these webpages provide owner reviews, not reviews by experts. The majority of the prior studies have used online reviews with topic modelling techniques to discover underlying semantic information, but previous studies have not examined customer behaviour in high involvement products, especially in EVs. Manual data extraction using Python programming was performed to extract reviews on EVs from three websites. Web scraping extracts textual review comments and other associated data, such as scores. The reviews were in a free-text structure, which requires processing before further analysis.

3.2 Text Pre-processing

Several steps are involved in pre-processing the text to transform the reviews into a clean, easy-to-analyse format. These steps are quite similar to those adopted in prior studies (Guo et al., 2017); (Tirunillai and Tellis, 2014); (Lee and Bradlow, 2007)

- Replacement of URLs: Web scraping using URLs with the corresponding pages that contain reviews and scores. The HTML code was also extracted. The URLs and HTML were replaced using BeautifulSoup and urlib.
- Replacement of pronouns: All the pronouns in the text were replaced with their respective object names to improve clarity in the sentences.
- Lowercasing of alphabets: All letters were converted to lowercase (e.g. 'Car' to 'car' and 'Tesla' to 'tesla') that contain the same spelling but different cases.
- Tokenisation: The text was broken into individual words using regular expressions tokenisation.

- Part-of-speech (POS) tag: Part-of-speech tagging was applied to retain only words that are nouns, verbs, adverbs, and adjectives—that is, words that have meaningful information about the product.
- Removal of stop words: All stop words (e.g. 'the', 'and', and 'is') that frequently occur in all contexts and are used for connection and grammar but carry little significance were removed. The built-in list in NLTK of 179 stop words and a few meaningless words were added that show in the text, and discarded negative words (e.g. 'no', 'nor', 'not', 's', and 'x') was used by removing commonly used stop words.
- Combine two words into a single word: To include the names of the Tesla model, the term 'model' with an 's' or 'x' was combined into a single word.
- Replacement of common negatives of words (e.g. 'no' and 'nor') by prefixing a 'not' to the token words that follow: To gain more precise information about the product, negative words must be included during analysis.
- Lemmatisation: The words were lemmatised based on POS tagging, which only allows nouns, verbs, adverbs, and adjectives using the spaCy lemmatiser. Through lemmatisation, the word is converted to its meaningful base form by considering the context.
- Bigram and trigram: Two-word and three-word sequences of words are combined using Genism's phrases. Based on the conditional probability, separate words are linked into one, which improves the interpretability of the results.

- Removal of stop words: All stop words were removed again before applying LDA topic modelling. The NLTK built-in stop words was used to remove words in text.

For example, an original review appears as:
"From the head turning falcon wing doors to the exhilarating torque and acceleration that the all electric dual motors provide, this suv is a dream. With frequent updates, this car gets BETTER all the time. The Autopilot keeps improving and will some day fully drive itself."

After pre-processing of the text, the review became.
"head, turn, falcon_wing_door, exhilarate, torque, acceleration, electric, dual_motor, provide, suv, dream, frequent, update, car, get_well, time, autopilot, keep, improve, day, fully, drive, autopilot"

- Pruning: Words were removed that less than 2 occurrence or more than 10 per cent total number of text using Gensim's dictionary filter extremes.

3.3 Topic Modelling

A large sample was collected by web scraping online reviews. In this section, the dimensions influencing customer behaviour towards EVs were extracted. The typical dimensionality reduction method is principal component analysis, which aims to detect the correlation between variables and uses this to reduce the dimensionality. Using the principal components, which comprise a linear combination of original features/axes that spread the variance in the data, the method squashes the original data and makes a new feature/axis. This helps combat the curse of dimensionality (Verleysen and François, 2005).

However, this traditional method is unfeasible in the context analysis for several reasons. First, online reviews comprise numerous different words by diverse customers. As shown in Figure 1., the corpus of words is very large and highly skewed. The text shows a long tail (Anderson, 2006), which leads to the curse of dimensionality. Next, customer express opinions based on their personal experiences. Thus, each review does not contain all the dimensions that are representative of all consumers. As a result, extracting meaningful dimensions is challenging using the traditional dimension reduction method.

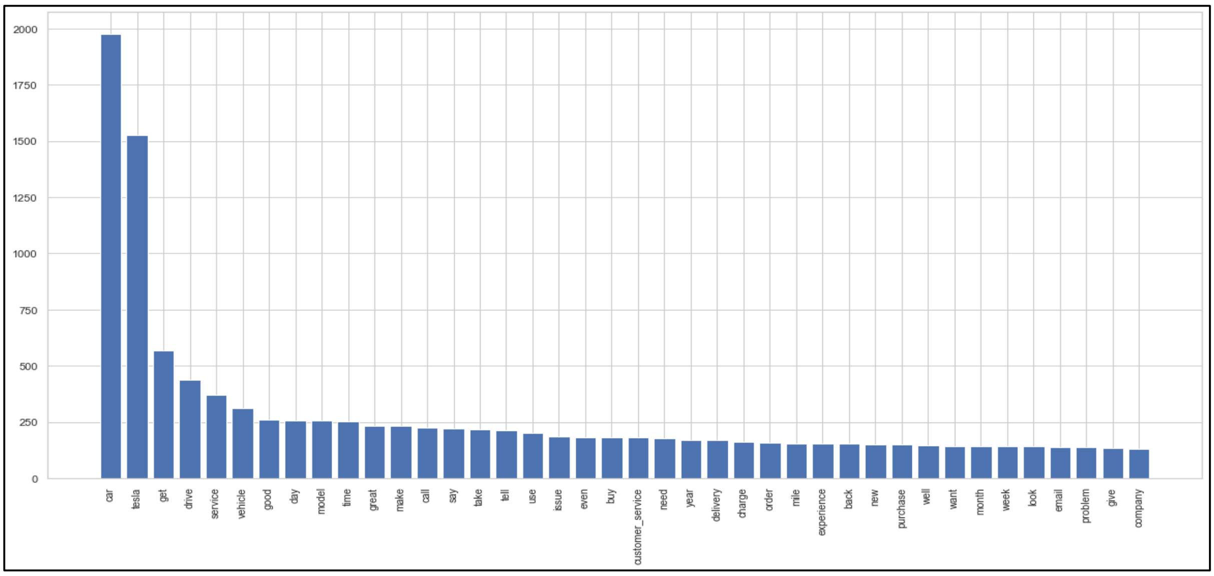


Figure 1. Word Counts (5,253 words) for Tesla reviews (Total 46,486 words).

This study uses LDA (Blei et al., 2003) which, which is a generalisation of pLSI (Hofmann, 1999). In addition, LDA is the most common method for topic modelling in the fields of machine learning and NLP to effectively extract dimensions from a large amount of textual data. Using the LDA approach with the optimum number of dimensions can discover topics (i.e., customer behaviour towards EVs) from a massive number of documents (i.e. online reviews). A 'dimension' is defined as a 'topic' in the LDA literature and is defined as a latent distribution over a vocabulary of words that consumers use to describe their electric vehicle experience.

Moreover, LDA is a generative probabilistic model of a corpus. Documents are represented as random mixtures over latent topics (Blei et al., 2003). Comprehensively, a set of documents is input into the LDA, and a variable *k* is set to determine the number of topics. The LDA uses the bag-of-words or term frequency-inverse document frequency matrix, a frequency-based expression method that disregards the sequence of words. In other words, only the existence of words is used under the assumption of exchangeability. Finally, a topic is extracted by estimating the probability that a word exists in a particular topic and the probability that a particular topic exists in a document.

In other words, the association between the words and documents becomes a word-topic and topic-document association. In addition, LDA assumes the word-topic and topic-document association independently. Word-topic association demonstrates how words can create a topic based on the relative frequencies of the words for a certain topic. Topic-document association indicates how many topics are represented in a specific document, such as customer feelings, experiences, and thoughts about products. Afterwards, topic-document association is used to determine the main documents for each topic.

Moreover, LDA assumes that *N* is the number of words in a review, which is referred to as a 'document' in the literature, $W = \{w_1, w_2, \ldots, w_N\}$, where *M* is the total number of reviews in the corpus *D*, $D = \{W_1, W_2, \ldots, W_M\}$. Moreover, *k* is the total number of topics, which is a hyperparameter in the LDA model. In this section, *k* is referred to as the dimension. The only variable that can be observed is the j^{th} word in w_{ij} appearing in the i^{th} reviews. All potential variables should be estimated except hyperparameters α and β using the w_{ij} information alone.

The LDA model consists of three hierarchies, as illustrated in Figure 2. The square is the number of repeats, and the circle is the variables. The shadow circle *W* represents the observed variables, which include the word, and assigns words that appear in documents. Moreover, θ and *Z* refer to latent variables, where θ is the topic-document association representing the proportion of topics in the i^{th} document. It is as long as the total number of topics *k* and is affected by *α*. In addition, *Z* is the word-topic association that assigns the i^{th} document's j^{th} word into a specific topic. Moreover, φ is the vector corresponding to the k^{th} topic with the length at as many words as in the whole corpus, which is affected by β. Finally, the parameters α and β are hyperparameters, α is a Dirichlet parameter, and β is the topic hyperparameter. The hyperparameter ɑ and β require to be inferred by learning using method such as the variation expectation maximization algorithm (Blei et al., 2003), the maximum likelihood estimation method (Asuncion et al., 2012), or Gibbs sampling methods (Griffiths and Steyvers, 2004).

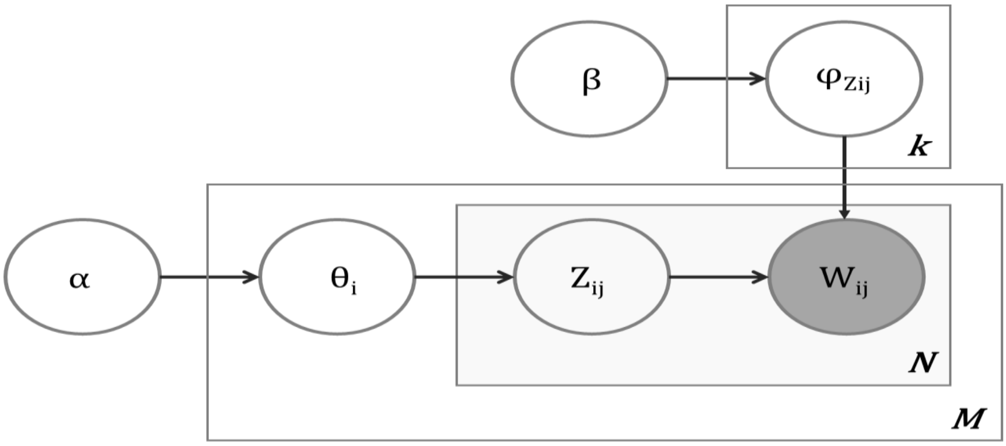


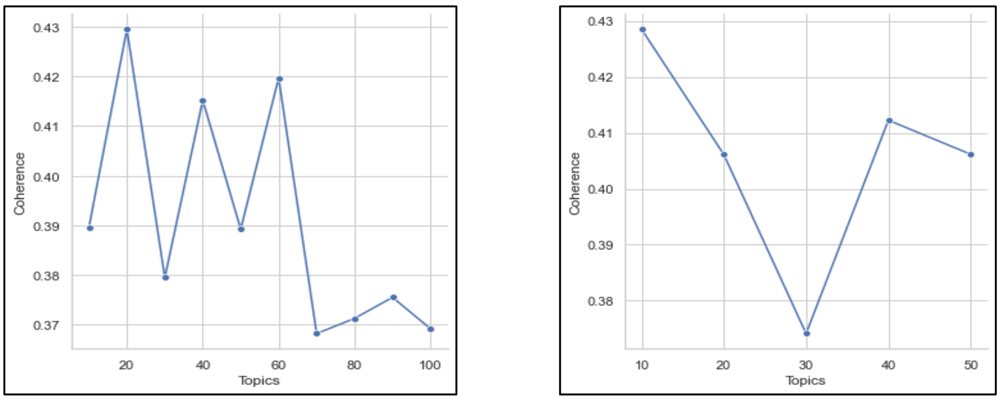
Figure 2. LDA model with plate notation.

In this paper, coherence score is used to explore ɑ, β and k. Evaluate each of LDA models using topic coherence, it found the optimal number of ɑ, β and k. Coherence is a measure of how interpretable the topics are to human's point of view. It essentially measures how similar these words are to each other. There are numerous methods for using coherence, however, in this paper, the method named c_v is used. This measurement found to be the most highly correlated when performed a comparative analysis of various methods, correlating them to human judgements (Röder et al., 2015).

CHAPTER 4. RESULTS

For topic modelling using LDA, users need to specify the number of topics, k, and hyperparameters, α and β, to be learned from the review collection. Topic coherence (c_v) (Röder et al., 2015) was employed to evaluate model performance, where superior models produce higher coherence. Two tuning steps were adopted for hyperparameters α and β with k. First, k was set to 10, 20, …, 100 to identify optimal α and β based on coherence. Optimal coherence occurred for k = 20, asymmetric α, and β = 0.91. Second, the range for k was narrowed from 10 to 50, with step size 10, and final optimal settings were k = 10, α = 0.91, and β = 0.91, producing coherence = 0.49.

Figure 3(a). shows coherence outcomes with respect to several selected topics for the first step, identifying the initial optimal asymmetric α and β = 0.91, and Figure 3(b). shows the case for a narrowed range of topics with optimal α, and β. Thus, k = 10 was sufficient to accurately model the topics in the text corpus while reducing computation time.



(a) (b)

Figure 3. Coherence score for several topics

Table 1. Result of LDA analysis.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Delivery request	General feature of tesla vehicle	Technology	Electric car	Delivery process	Mobile service	Security	Exterior look	High tech car	Comparing with other brands
delivery	tire	change	world	june	appt	bag	tyre	oscar	smog_producer
email	auto pilot	tech nology	electric_car	price	part	police	fabulous	auto	nothing
customer	wheel	battery	elon_musk	march	uber	laptop	hour	auburn_way	ford
ask	cost	door	much	text	guy	return	change	high_tech_car	gas
problem	seat	replace	fuel	customer	item	leave	minute	thank	ever
sale	feel	owner	minor	promise	arrive	ask	supply	patient	honda
phone	power	open	tech	trade	book	staff	exterior_look	help	friendly
people	gas	door_handle	excellent	delivery	tell_mobile_service	thing	entire	professional	comfortable
someone	free	sensor	motor_manufacturer	offer	work	unhelpful	else	card	easy
bad	feature	break	interior	reserve	right	safe	let	steep	driving_experience

Table 1 shows several topics and popular words within various topics. Reviewers often used delivery, email, customer, and ask when discussing delivery requests; whereas change, technology, battery, and door were more commonly employed for technology topics. Figure 4. shows 10 example topic labels and corresponding probabilities. The most frequently mentioned topics were delivery request, general features of tesla vehicles, and technology. All topics were grouped in three categories: General discussion, Technology, and Service to assist interpretation, as shown in Table 2.

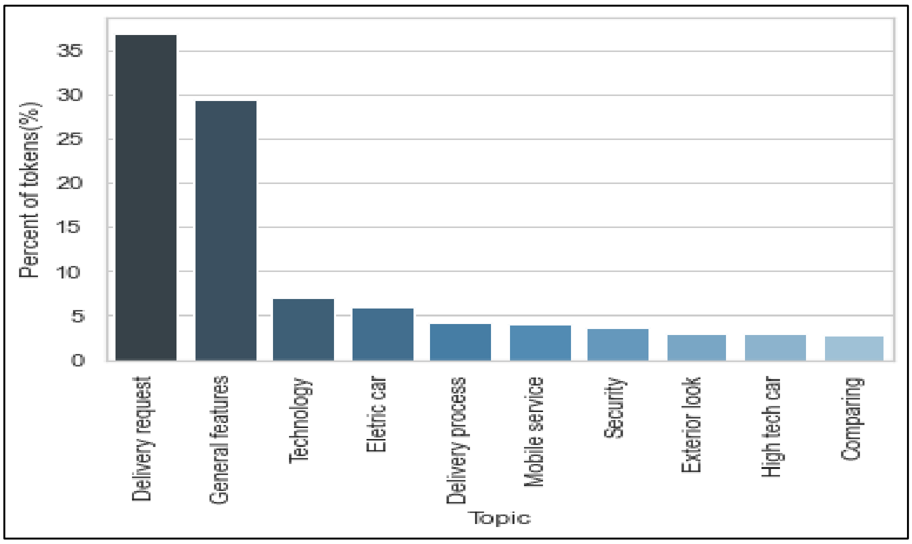


Figure 4. Probability distribution of topics.

Table 2. Topic grouping.

General discussion	Technology	Service	Example words
General feature of tesla vehicle	Technology	Delivery request	autopilot, battery, interior
Security	Electric car	Delivery process	bag, laptop, secure
Exterior appearance	High technology car	Mobile service	tyre, fabulous, supply
Comparing with other brands			Ford, Honda, Mitsubishi, gas, driving experience, environmental

Reviewers mentioned many general EV features, including security and exterior appearance, and commonly compared Tesla EVs with other brands in terms of comfortability, driving experience and environmental issues.

The underlying technology was a frequent review topic, covering

- technology: battery, door handle, senor;
- electric car: tech, electric, sustainable, Elon Musk;
- high tech car: learning, patience, knowledgeable, auto;

and other unique Tesla features. Some reviewers also commented about the patience required to learn a new technology.

Tesla service was frequently mentioned in the reviews with delivery, customer service, and mobile service being the most common terms, e.g.

- delivery request: email, problem, phone, customer;
- delivery process: months, trade, offer, delivery; and
- mobile service: appointment, book.

Several keywords were commonly linked with negative words, implying a potentially widespread problem with customer service.

CHAPTER 5. DISCUSSION

Using LDA analysis based on customer reviews, this study demonstrated a topic modelling approach to discovering the behaviour of EV owners. A comprehensive list of topics discussed by owners describing their EV experience was constructed using the LDA analysis of customer reviews that were not found via traditional means. Unlike the majority of previous studies employing topic modelling (Blei et al., 2003); (Guo et al., 2017); (Kim et al., 2019); (Mazanec, 2017) where the researchers predetermined the topic number, in this study the models were evaluated to determine the optimal number of topics for modelling. The relative importance of the extracted meaningful topics was identified according to the topic of coherence value. Ideally, when more data are available, the generalisations are more accurate. Our LDA analysis was calculated based on 956 reviews created by 2,765 unique tokens and words. To compensate for the limitation of the number of datasets, customised preprocessing was applied to make the results more reliable and accurate. The findings in the present study are intuitive and straightforward, considering the review addresses EVs. For example, 956 reviews primarily discussed three topics: the general features of Tesla vehicles, technology, and service.

Several studies have argued that purchasing a high-involvement product requires the complicated and intensive process of decision-making (Kotler and Philip 2007) and a sequential process for problem recognition for purchase and post-purchase actions (Quester and Pascale Genevieve

2007). One of the post-purchase actions is writing a review about the purchase on the owner's website or a particular review website. From the reviews, the customers' general interests, prior concerns before purchase, poster problems about the product, and diverse sentiments regarding the product are available. By analysing the reviews, the consumers' decision-making process and what features and details the consumers considered can be intuited. Moreover, from their opinions about the product after purchase, the decision-making process of other potential customers who are willing to buy EVs can be deduced. From the LDA analysis, the 10 most-discussed topics were categorised into three groups: *general discussion, technology*, and *service*. Thus, general features of vehicles, technology and service are the main topics that consumers discuss in online reviews about electric cars. Based on the results, the decision-making process can be linked. The general customers interested in EVs tend to consider the general features of Tesla vehicles, technology features, and customer service.

Several studies have discovered that diverse reasons accelerate EV purchases and increase the scale of the industry (Larson et al., 2014) (Liao et al., 2017). Previous studies have argued that safety is one of the most important features when purchasing an automobile (Koppel et al., 2008); (Boyle and Schulman, 1996); (Vrkljan and Anaby, 2011). In this study, safety was not one of the primary topics that reviewers discussed; however, safety was primarily discussed in the *general discussion category* as a *security topic* and was discussed as a feature in the *general features of*

Tesla vehicle topic. Because Tesla is considered a pioneer in EVs, customers' priority interest is more about advanced technology, including the battery, distinct characteristics from other brands, and exterior appearance.

(Johansson-Stenman and Martinsson, 2006) examined fuel consumption, which is another important characteristic, and this feature was mentioned in the *technology category*. Because this study was focused on EVs, the technology related to electricity is substantial. The terms *fuel, gas, electric, tech,* and *sustainable* were included in the *electric car topic*, which is related to the electric battery that is the substitute for the fuel consumption engine. Moreover, in the topic *comparing with other brands*, the reviewers compared Tesla with other brands using the terms *gas* and *friendly*, which can be interpreted as relating to the fuel consumption and electric battery. (Yusof et al., 2013) argued that consumers who demonstrate environmental responsibility and values and tends to purchase environmentally friendly vehicles, which is also presented in this study. The reviews were discussed regarding the battery, environment, and sustainability in many topics, such as the *general features of Tesla vehicles, comparing with other brands,* and the *technology category*. Based on Tesla EV owners who bought EVs with responsibility regarding the environment in mind, future EV customer decision-making processes can be predicted for those who also value and are interested in environmentally friendly vehicles.

The electric battery and environment have an inseparable relationship in terms of EVs. In the results of the LDA analysis, environmental benefits, such as electric battery engines, which have been mentioned in previous studies, were discovered in several topics, including the *general features of the Tesla vehicle, comparing with other brands, technology,* and *electric car*. Especially in *comparing with other brands,* Ford, Honda, and Mitsubishi are compared with Tesla, which uses an electric battery instead of gas or fuel, regarding driving experience and environmental issues. Because Tesla's major characteristic is the electric-based car, the technology and environmental discussions are highly related and are generally paired with each other in a specific topic. For instance, in the electric car topic, the words 'tech' and 'electric' are used regarding technology and the environment along with the term 'sustainable'. As a results, the critical technical aspects that influence the consumer experience are general aspects of technology, battery, electricity and driving experiences related to autopilot. Furthermore, this study had not find any significant results to whether Covid-19 affected customers' purchase, as one of the research questions. There were reviews that included the word 'covid-19', but according to the analysis results, it was not a big part of each topic.

The analysis indicates that 10 topics can be categorised into three groups. The general discussion and technology categories include multiple advantages of Tesla EVs; thus, targeted marketing strategies can be developed to attract consumers to a specific characteristic. For example, marketing that targets future consumers can include information on state-of-the-art unique techniques, such as *autopilot, driving experience, battery,* and *door handle,* to attract customers to Tesla EVs. Security characteristics can be incorporated into marketing materials to target consumers, as indicated by one of the topics in the general discussion category. Additionally, Tesla should be compared with other automobile brands, such as Ford, Honda, and Mitsubishi, in terms of driving experience and environmental features because this has been demonstrated to be a major topic in the results.

Another benefit of this study is that the consumers' major problems with Tesla were also included in the reviews and were analysed with the topics, providing valuable information and the opinions of Tesla owners, allowing the staff and managers to improve EV performance. For instance, the staff and managers should considerably focus on improving customer service. Additional effort should be exerted to improve the consumer perception of the delivery process. Because the terms *problem, bad,* and *promise* can be found in delivery-related topics, improving customer service could attract more consumers by providing an improved delivery experience. Furthermore, customer service must also be a focus. Because Tesla sells vehicles online and offline, mobile services are one of the major methods to communicate with the company staff and managers. By improving the quality of mobile service, both future potential customers and current owners can experience better customer service.

5.1 Limitations and Further Research

This study has some important limitations. First, the number of reviews was 956. Thus, the total amount of data is small, decreasing the reliability of the study. However, Tesla EVs launched in 2016, and only three models are available at the moment. Therefore, to compensate for the problem, precise preprocessing, which added a few more steps, was applied to make the data more interpretable and more suitable for analysis. Second, an analysis performed without considering types of EVs and years of cars. The findings may not be generalised to each model in Tesla. Since each model has different features that had built to target specific customers, analysing based on types by years may give different results in each model. Third, an analysis was performed only for Tesla, the results of which may be inapplicable to other brands of EVs (e.g. Hyundai, BMW, and Audi). The findings and recommendation may not be suitable to other EVs brands because they may have different characteristics and customer's experiences compared with those included in the current study. Hence, other EVs brands' manager should conduct further study on their brands for appropriate practical applications. Fourth, this study only used reviews written in English due to the language barrier. The identified topics may not cover all possible EVs' owners decision-making process and experiences, especially those from countries where English is not the first language. Non-English reviews require other text preprocessing techniques for a comprehensive understanding. Fifth, the authors determined the labels of the topics to describe the topic meanings based on the results that only provide words separately. Thus, each label may not be the best labels for the topics. However, these labels should have captured the general meanings of the topics because the wording of the labels is determined based on the most popular words that appear for each topic. The number of topics to be specified for the LDA model is subject to the nature of the collected data set. A larger data set may capture more topics, whereas a smaller data set may contain fewer topics. Nevertheless, determining the appropriate topic numbers should be evaluated using topic coherence. Sixth, we do not analyse rare or infrequent words in the long tail of the distribution. We have deleted that less than two occurrences or more than 10 per cent total number of text. However, these words could reflect consumer preferences and decision-making process that could be very helpful in understanding customers who purchase EVs. Seventh, the LDA analysis is computationally intensive and expensive. Especially in this study, when it tuning hyperparameters, it demands overnight with less than 1,000 reviews. However, with the current advances in computing and the increasing adoption of largescale computing techniques, this limitation will dissolve over time. Eighth, the LDA model is sensitive to the values of hyperparameters α and β of the Bayesian priors, which could influence the results in terms of the number of topics extracted. Moreover, the number of topics k, need to be considered with hyperparameters α and β. These limitations could be rich avenues for further research.

Future studies can investigate other brands of EVs to generalize the findings of the decision-making process of EVs. Reviews from other review sites and platforms could also be considered in future studies to increase the reliability of dataset and results. Moreover, text processing techniques for languages other than English can be implied in further studies. This may provide more reviews from more customers and generate further results that are representative and comprehensive.

CHAPTER 6. CONCLUSION

Analysing an online reviews corpus to gain an overview of the addressed concepts can be a tedious process when performed through traditional approaches, such as manual evaluation and statistical models. This paper illustrated how a topic modelling approach, specifically the LDA analysis, assesses underlying semantic information, how topic modelling can provide an overview of the topics underlying a document collection, and how each topic is labelled based on the most popular words that appear for each topic. From the topic modelling results, the consumer decision-making process can be deduced, such as the consideration of features and details of EVs, which is the most common topic that Tesla EVs owner discussed. From the current customer opinions, the decision-making process of potential customers who will purchase EVs can be assumed.

A total of 946 Tesla EV reviews were collected from websites. Preprocessing was implemented to transform them into an easy-to-analyse format. After tuning the LDA model hyperparameters and the number of topics, the optimal settings were $k = 10$, $\alpha = 0.91$, and $\beta = 0.91$ with coherence = 0.49. This paper used the topic coherence score as an evaluation method to measure how similar words are to each other in the results. Ten topics with the most popular words were representative of Tesla reviews. After labelling each topic based on the most popular words that appeared for each topic, they were categorised into three groups: general discussion, technology, and service. The general features of the topics of Tesla vehicles, security, exterior appearance, and brand comparison with other topics were in the general discussion category. This category indicated terms such as *comfortability*, *driving experience*, and *environmental*. Next, the technology, electric, and high technology car topics were categorised in the technology category with words like *battery*, *electric*, and *patience*. Finally, the delivery request, delivery process, and mobile service topics were in the service category. The service category includes such words as *delivery*, *problem*, and *appointment*. From the results, in the decision-making of EVs, features of the vehicle, technology, and service are included in significant characteristics.

The results of the analysis reveal several benefits to Tesla. First, marketing strategies can be a focus of one of the categories in the results: general features of tesla vehicles, technology, and service. For instance, for those who are interested in the technology of EVs, Tesla can target these individuals to provide unique technology from Tesla in terms of marketing. Moreover, enhancing the security characteristics in the marketing strategy can attract more customers. Another benefit of this study is that topics related to service contain negative words, such as problem and bad, which indicates the potential for increasing sales and improving customer experience by managing the problems.

This study has several limitations. The number of reviews was 956; thus, the total amount of data is small, decreasing the reliability of the study. Therefore, to compensate for the problem, with basic preprocessing it added a few more steps to make more precise. Moreover, this study collected data written only in English from three websites. Overcoming these limitations could be rich avenues for further research.

REFERENCES

ANDERSON, C. 2006. *The long tail: Why the future of business is selling less of more*, Hachette Books.

ARORA, S., GE, R. & MOITRA, A. Learning topic models--going beyond SVD. 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science, 2012. IEEE, 1-10.

ASUNCION, A., WELLING, M., SMYTH, P. & TEH, Y. W. 2012. On smoothing and inference for topic models. *arXiv preprint arXiv:1205.2662*.

BLEI, D. M., NG, A. Y. & JORDAN, M. I. 2003. Latent dirichlet allocation. *Journal of machine Learning research,* 3**,** 993-1022.

BOYLE, J. M. & SCHULMAN, R. 1996. National Highway Traffic Safety Administration 1995 customer satisfaction survey. United States. National Highway Traffic Safety Administration.

BYUN, D.-H. 2001. The AHP approach for selecting an automobile purchase model. *Information & Management,* 38**,** 289-297.

BYUN, D.-H. & SUH, E.-H. 1996. A methodology for evaluating EIS software packages. *Journal of Organizational and End User Computing (JOEUC),* 8**,** 21-31.

CHAN, F. S. 2003. Interactive selection model for supplier selection process: an analytical hierarchy process approach. *International Journal of Production Research,* 41**,** 3549-3579.

CHEN, J., HUANG, H., TIAN, S. & QU, Y. 2009. Feature selection for text classification with Naïve Bayes. *Expert Systems with Applications,* 36**,** 5432-5435.

CHEUNG, D., SONG, I., CHU, W., HU, X. & LIN, J. 2009. *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, Association of Computing Machinery.

CONRAD, J. G. & SCHILDER, F. Opinion mining in legal blogs. Proceedings of the 11th international conference on Artificial intelligence and law, 2007. 231-236.

DEERWESTER, S., DUMAIS, S. T., FURNAS, G. W., LANDAUER, T. K. & HARSHMAN, R. 1990. Indexing by latent semantic analysis. *Journal of the American society for information science,* 41**,** 391-407.

DUMAIS, S. T. 2004. Latent semantic analysis. *Annual review of information science and technology,* 38**,** 188-230.

EGBUE, O. & LONG, S. 2012. Barriers to widespread adoption of electric vehicles: An analysis of consumer attitudes and perceptions. *Energy policy,* 48**,** 717-729.

GOH, K.-Y., HENG, C.-S. & LIN, Z. 2013. Social media brand community and consumer behavior: Quantifying the relative impact of user-and marketer-generated content. *Information Systems Research,* 24**,** 88-107.

GRIFFITHS, T. L. & STEYVERS, M. 2004. Finding scientific topics. *Proceedings of the National academy of Sciences,* 101**,** 5228-5235.

GU, B., PARK, J. & KONANA, P. 2012. Research note—the impact of external word-of-mouth sources on retailer sales of high-involvement products. *Information Systems Research,* 23**,** 182-196.

GUO, Y., BARNES, S. J. & JIA, Q. 2017. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet allocation. *Tourism Management,* 59**,** 467-483.

HAN, J., KAMBER, M. & PEI, J. 2012. Data mining: concepts and techniques, Waltham, MA. *Morgan Kaufman Publishers,* 10**,** 978-1.
HOFMANN, T. Probabilistic latent semantic indexing. Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval, 1999. 50-57.
JAIN, M. 2019. A study on consumer behavior-decision making under high and low involvement situations. *IJRAR-International Journal of Research and Analytical Reviews,* 6.
JOHANSSON-STENMAN, O. & MARTINSSON, P. 2006. Honestly, why are you driving a BMW? *Journal of Economic Behavior & Organization,* 60**,** 129-146.
JUNQUERA, B., MORENO, B. & ÁLVAREZ, R. 2016. Analyzing consumer attitudes towards electric vehicle purchasing intentions in Spain: Technological limitations and vehicle confidence. *Technological Forecasting and Social Change,* 109**,** 6-14.
KIM, K., PARK, O., BARR, J. & YUN, H. 2019. Tourists' shifting perceptions of UNESCO heritage sites: lessons from Jeju Island-South Korea. *Tourism Review*.
KOLYSHKINA, I., LEVIN, B. & GOLDSWORTHY, G. 2013. Using Social Media Data for Comparing Brand Awareness, Levels of Consumer Engagement, Public Opinion and Sentiment for Big Four Australian Banks. *Data Mining and Analytics 2013 (AusDM'13)***,** 59.
KOPPEL, S., CHARLTON, J., FILDES, B. & FITZHARRIS, M. 2008. How important is vehicle safety in the new vehicle purchase process? *Accident analysis & prevention,* 40**,** 994-1004.
LAI, V. S., TRUEBLOOD, R. P. & WONG, B. K. 1999. Software selection: a case study of the application of the analytical hierarchical process to the selection of a multimedia authoring system. *Information & Management,* 36**,** 221-232.
LARSON, P. D., VIãFARA, J., PARSONS, R. V. & ELIAS, A. 2014. Consumer attitudes about electric cars: Pricing analysis and policy implications. *Transportation Research Part A: Policy and Practice,* 69**,** 299-314.
LASSILA, J., HAAKANA, J., TIKKA, V. & PARTANEN, J. 2011. Methodology to analyze the economic effects of electric cars as energy storages. *IEEE Transactions on smart grid,* 3**,** 506-516.
LEE, T. & BRADLOW, E. T. 2007. Automatic construction of conjoint attributes and levels from online customer reviews. *University Of Pennsylvania, The Wharton School Working Paper*.
LI, Y., WANG, X. & YANG, Z. 2011. The effects of corporate-brand credibility, perceived corporate-brand origin, and self-image congruence on purchase intention: Evidence from China's auto industry. *Journal of Global Marketing,* 24**,** 58-68.
LIAO, F., MOLIN, E. & VAN WEE, B. 2017. Consumer preferences for electric vehicles: a literature review. *Transport Reviews,* 37**,** 252-275.
LIBERATORE, M. J. & STYLIANOU, A. C. 1994. Using knowledge-based systems for strategic market assessment. *Information & Management,* 27**,** 221-232.
LIN, S.-J. & HSU, M.-F. 2018. Decision making by extracting soft information from CSR news report. *Technological and Economic Development of Economy,* 24**,** 1344-1361.
LU, W. & STEPCHENKOVA, S. 2015. User-generated content as a research mode in tourism and hospitality applications: Topics, methods, and software. *Journal of Hospitality Marketing & Management,* 24**,** 119-154.

LUO, J. M., VU, H. Q., LI, G. & LAW, R. 2020. Topic modelling for theme park online reviews: analysis of Disneyland. *Journal of Travel & Tourism Marketing,* 37**,** 272-285.
MAZANEC, J. A. 2017. Determining long-term change in tourism research language with text-mining methods. *Tourism Analysis,* 22**,** 75-83.
MCINTOSH, M. J. 2009. *Determinants of environmentally conscious consumer behaviors: Measuring the value consumer environmentalism and predicting behavioral intention to purchase environmentally friendly products*, University of Massachusetts Amherst.
MEI, Q., LING, X., WONDRA, M., SU, H. & ZHAI, C. Topic sentiment mixture: modeling facets and opinions in weblogs. Proceedings of the 16th international conference on World Wide Web, 2007. 171-180.
MENG, X., WEI, F., LIU, X., ZHOU, M., LI, S. & WANG, H. Entity-centric topic-oriented opinion summarization in twitter. Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining, 2012. 379-387.
MIKHAILOV, L. 2003. Deriving priorities from fuzzy pairwise comparison judgements. *Fuzzy sets and systems,* 134**,** 365-385.
MOGHADDAM, S. & ESTER, M. On the design of LDA models for aspect-based opinion mining. Proceedings of the 21st ACM international conference on Information and knowledge management, 2012. 803-812.
MOLESWORTH, M. & SUORTTI, J. P. 2002. Buying cars online: the adoption of the web for high-involvement, high-cost purchases. *Journal of Consumer Behaviour: An International Research Review,* 2**,** 155-168.
MUKHERJEE, A. & LIU, B. Aspect extraction through semi-supervised modeling. Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), 2012. 339-348.
NARTEH, B., ODOOM, R., BRAIMAH, M. & BUAME, S. 2012. Key drivers of automobile brand choice in sub-Saharan Africa: the case of Ghana. *Journal of Product & Brand Management*.
NAYEEM, T. & CASIDY, R. 2013. The role of external influences in high involvement purchase behaviour. *Marketing Intelligence & Planning*.
NIGAM, K., MCCALLUM, A. K., THRUN, S. & MITCHELL, T. 2000. Text classification from labeled and unlabeled documents using EM. *Machine learning,* 39**,** 103-134.
ÖZDAĞOĞLU, G., KAPUCUGIL-İKIZ, A. & ÇELIK, A. F. 2018. Topic modelling-based decision framework for analysing digital voice of the customer. *Total Quality Management & Business Excellence,* 29**,** 1545-1562.
PANG, B. & LEE, L. 2008. Opinion mining and sentiment analysis. *Foundations and trends in information retrieval,* 2**,** 1-135.
PAPADIMITRIOU, C. H., RAGHAVAN, P., TAMAKI, H. & VEMPALA, S. 2000. Latent semantic indexing: A probabilistic analysis. *Journal of Computer and System Sciences,* 61**,** 217-235.
PARK, D.-H., LEE, J. & HAN, I. 2007. The effect of on-line consumer reviews on consumer purchasing intention: The moderating role of involvement. *International journal of electronic commerce,* 11**,** 125-148.
PARK, Y. & LEE, S. 2011. How to design and utilize online customer center to support new product concept generation. *Expert Systems with Applications,* 38**,** 10638-10647.
RAJ, M. P. M. & ROY, S. 2015. Impact of brand image on consumer decision-making: A study on high-technology products. *Global Business Review,* 16**,** 463-477.

RøDER, M., BOTH, A. & HINNEBURG, A. Exploring the space of topic coherence measures. Proceedings of the eighth ACM international conference on Web search and data mining, 2015. 399-408.

SAATY, T. 1980. The Analytic Hierarchy Process Mcgraw Hill, New York. *Agricultural Economics Review,* 70.

SAATY, T. L. & KEARNS, K. P. 2014. *Analytical planning: The organization of system*, Elsevier.

SAKTHIVEL, G., ILANGKUMARAN, M., NAGARAJAN, G., RAJA, A., RAGUNADHAN, P. & PRAKASH, J. 2013. A hybrid MCDM approach for evaluating an automobile purchase model. *International journal of information and decision sciences,* 5**,** 50-85.

SANTANDREU, J. & SHURDEN, M. C. 2017. Purchase decisions for high involvement products: the new generation of buyers. *Journal of Marketing Development and Competitiveness,* 11**,** 88-92.

SIERZCHULA, W., BAKKER, S., MAAT, K. & VAN WEE, B. 2014. The influence of financial incentives and other socio-economic factors on electric vehicle adoption. *Energy Policy,* 68**,** 183-194.

TIRUNILLAI, S. & TELLIS, G. J. 2014. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research,* 51**,** 463-479.

TITOV, I. & MCDONALD, R. Modeling online reviews with multi-grain topic models. Proceedings of the 17th international conference on World Wide Web, 2008. 111-120.

UYS, J., DU PREEZ, N. & UYS, E. Leveraging unstructured information using topic modelling. PICMET'08-2008 Portland International Conference on Management of Engineering & Technology, 2008. IEEE, 955-961.

VERLEYSEN, M. & FRANéOIS, D. The curse of dimensionality in data mining and time series prediction. International work-conference on artificial neural networks, 2005. Springer, 758-770.

VRKLJAN, B. H. & ANABY, D. 2011. What vehicle features are considered important when buying an automobile? An examination of driver preferences by age and gender. *Journal of safety research,* 42**,** 61-65.

WOOTTEN, G. 2003. Channel conflict and high involvement internet purchases–a qualitative cross cultural perspective of policing parallel importing. *Qualitative Market Research: An International Journal*.

YU, C., ZHANG, X. & LUO, H. Mining Hot topics from free-text customer reviews An LDA-based approach. 2010 Seventh Web Information Systems and Applications Conference, 2010. IEEE, 85-89.

YUSOF, J. M., SINGH, G. K. B. & RAZAK, R. A. 2013. Purchase intention of environment-friendly automobile. *Procedia-Social and Behavioral Sciences,* 85**,** 400-410.

ZAHEDI, F. 1986. The analytic hierarchy process—a survey of the method and its applications. *interfaces,* 16**,** 96-108.

ZHAN, J., LOH, H. T. & LIU, Y. 2009. Gather customer concerns from online product reviews–A text summarization approach. *Expert Systems with Applications,* 36**,** 2107-2115.

ZHANG, Y. & ZHU, W. Extracting implicit features in online customer reviews for opinion mining. Proceedings of the 22nd International Conference on World Wide Web, 2013. 103-104.

ZHAO, K., LIU, B., TIRPAK, T. M. & XIAO, W. Opportunity map: a visualization framework for fast identification of actionable knowledge. Proceedings of the 14th ACM international conference on Information and knowledge management, 2005. 60-67.