

Article

# Analyzing Online Car Reviews Using Text Mining

En-Gir Kim and Se-Hak Chun \*

Department of Business Administration, Seoul National University of Science and Technology,  
232 Gongreung-Ro, Nowon-Gu, Seoul 01811, Korea; engir@mail.ru

\* Correspondence: shchun@seoultech.ac.kr; Tel.: +82-2-970-6487

Received: 27 February 2019; Accepted: 13 March 2019; Published: 17 March 2019



**Abstract:** Consumer reviews on the web have rapidly become an important information source through which consumers can share their experiences and opinions about products and services. It is a form of text-based communication that provides new possibilities and opens vast perspectives in terms of marketing. Reading consumer reviews gives marketers an opportunity to eavesdrop on their own consumers. This paper examines consumer reviews of three different competitive automobile brands and analyzes the advantages and disadvantages of each vehicle using text mining and association rule methods. The data were collected from an online resource for automotive information, Edmunds.com, with a scraping tool “ParseHub” and then processed in R software for statistical computing and graphics. The paper provides detailed insights into the superior and problematic sides of each brand and into consumers’ perceptions of automobiles and highlights differences between satisfied and unsatisfied groups regarding the best and worst features of the brands.

**Keywords:** big data analytics; text mining; association rule; car review

---

## 1. Introduction

The rapid spread of the Internet has provided humanity with a new way to obtain information. It has now become the biggest source of information, with people conducting ever more searches on the Web. Alongside this, social media, another part of the Internet domain, has also captured the attention of netizens. Social media can take many different forms, one of which is product-review websites. These, along with other types of social media, provide a platform for consumers to share their experiences and opinions about the products they purchase and use, thereby providing other consumers with information about the pros and cons of these products. Such communication is also known as electronic word-of-mouth and has opened up new horizons in marketing. By reading consumer reviews, companies and marketers get to know their customers and thus obtain a better understanding of marketing opportunities, the competitive landscape, the market structure, and the features of their own and competitors’ products that customers discuss.

Nowadays, with data spreading dramatically, many organizations as well as researchers strive to find patterns among data using datamining methods. Text mining is one of these methods and is used to analyze consumer reviews. For instance, we can take as an example a situation where consumers write comments or reviews about a mobile telephone they purchase and discuss their best or worst experiences. Consumer 1 purchases a mobile telephone from Company A while Consumer 2 purchases a mobile telephone from Company B. The autonomy of a mobile telephone from Company A is better than Company B, while the camera quality of the mobile phone from Company B is better than Company A. Hence, Consumer 2 is likely to write a review that will contain positive feedback about the camera. However, he is also likely to leave negative feedback about the autonomy of the camera. Thus, extracting meaningful information from consumer reviews such as the most frequent words and the relationship between them provides companies with an insight into the superior features of

given brands as well as problematic features companies need to address and improve in future. This approach to gathering data and necessary information is sometimes much faster than administering a questionnaire survey.

Many studies have analyzed online reviews for various product categories such as books, movies, fashion, cosmetics, hotels, airlines, and restaurant services. However, there have been very few studies on car purchasing behavior, although some studies have identified the features of a car that affect purchasing behavior [1,2]. To the best of our knowledge, there has been no research to date on online car reviews using a text mining approach. This paper therefore analyzes consumer reviews for automobile brands and compares three competitive brands: Hyundai, Honda, and Ford. Car review websites such as Edmunds, Kelley Blue Book, and Motor Trend usually ask consumers to fill in two different subfields: “best features” and “worst features.” Finding the most frequent words used for each subfield therefore gives us some information about the weak and strong aspects of given models. Using R programming, this paper will address three research questions using data from one of the best car review websites, Edmunds.com.

Firstly, this paper determines which terms occur most frequently in the “best features” and “worst features” reviews for each car and discusses their significance. Secondly, it analyzes eight essential car features by applying the association method for consumer reviews on each of three competitive automobile brands in 2012. The association method is then used to discuss the relationships between the most frequent terms and the eight different features of competing vehicles. This provides comprehensive information as to which features reviewers are most interested in and discuss. Finally, this paper compares the frequency and ratio of the terms for eight different features between satisfied and unsatisfied groups and presents relevant implications for the strategies used by car manufacturers.

The remainder of the paper is organized as follows. Section 2 reviews previous literature and presents the theoretical background to the study. Section 3 then presents research methodology and Section 4 presents the results. Section 5 concludes and discusses future research directions.

## 2. Research Background

### 2.1. Big Data Analytics and Business Value

With the rapid speed of the Internet and smart mobile devices, consumers can easily generate reviews and discussion on Web resources such as blogs, product review websites, chat rooms, and brand communities. These activities have exponentially grown in recent years and are increasing communication channels between consumers and firms. It has now become important to obtain meaningful information from Web resources such as overall product ratings and product reviews. Thus, organizations and companies are always looking for ways to use the power of big data analytics (BDA) to improve their decision making [3]. Big data analytics (BDA) improves data-driven decision making and provides ways to organize, learn, and innovate [3,4]. It is critical for organizational success that advantage is taken of all available information using big data analytics [4]. This will enable organizations to improve the management of operational risk, reinforce customer relationship management, enhance operational efficiency, and improve firm performance in general [5]. Liu [6], for instance, found that companies using analytics software could decrease customer acquisition costs by about 47% and increase their revenue by about 8%. Wamba et al. [7] also examined the effects of big data analytics capability on firm performance and found both direct and indirect impacts.

However, there is insufficient understanding of how organizations need to be structured and how they should utilize their big data initiatives to generate business value [8–10]. Grover et al. [11] described the value proposition of big data analytics by delineating its components and discussed constructs and relationships that focus on the creation and realization of such value. Dong and Yang [12] explained how and why social media analytics create super-additive value through synergies in functional complementarity between social media diversity and big data analytics. They found that social media diversity and big data analytics have a positive interaction effect on market performance,

which is more salient for SMEs than for large firms [13]. Müller et al. [13] analyzed how firm performance is related to big data analytics and found that big data analytics (BDA) assets are associated with an average 3–7% improvement in firm productivity. It is necessary for organizations to understand the impact that big data quality has on firm performance [14]. Organizations should therefore establish specific processes and practices to realize value from their big data investments where different factors are emphasized depending on the context of examination [15].

## 2.2. Text Mining and Association Rule

Text mining is a data mining technique that obtains structured information from unstructured text information in order to summarize and classify textual data generated through traditional data mining and statistical techniques [16]. Text mining tasks are usually classified into five types of task: information extraction, text categorization, text clustering, document summarization, and association analysis [17–19]. The information extraction technique involves finding important words from social media and brand names, such as those relating to product features. These are then sorted into defined categories or topics in the categorization task. Using computer programs, text categorization treats texts as a bag of words and counts word frequencies while text clustering combines similar documents in groups without predefined categories [16]. Document summarization summarizes the most important concepts from a large collection of texts, enabling data analysts to identify changes in consumer preferences over time and market trends in general. Association analysis is an association rule which aims to find associations for a certain term and is based on counting co-occurrence frequencies [16].

The association rule refers to discovering unpredicted and unique rules from large datasets, finding correlations between elements in transactional databases, and then linking the information by discovering common relationships between the different factors [20]. It is an important data mining technique that is used to identify attribute-value conditions that frequently appear together in datasets [21–25]. There are two types of association rule methods: classical and relational rule mining. Classical association rules only consider co-occurrences between the attribute values while relational association rules are able to depict various types of relations between attribute values. From this perspective, relational association rule mining is an effective unsupervised learning model that can discover hidden patterns in data [25].

As text mining has gained increasing momentum in recent years, comment mining has much attention being given to sentiment analysis and opinion mining and becomes an important technique to obtain information from user-generated contents [13]. User-generated content can be applied to very different types of media where reviews have been researched and have been found to influence consumers to buy products. Consumers are likely to use reviews if they perceive the credibility of the source to be high [26,27]. Online social networks now enable users to share their own lives, generate and interact with vast amounts of multimedia content (text, audio, video, and images), and supplement these with feedback, comments, or feelings. The role of big data technologies is therefore becoming more important in obtaining meaningful information from users' interactive activities [28]. Amato et al. [29] developed a more effective and efficient mechanism for a text pre-processing task where each linguistic term is assigned with a weight that is computed using the well-known tf-idf formula. Yahav et al. [30] proposed an adjustment to tf-idf that accounts for this bias introduced by between-participant discourse to the study of comments in social media and illustrated the effects of both the bias and correction through data from seven Facebook fan pages.

## 2.3. Consumer Car Purchasing Behavior

Online consumer reviews have a significant impact on consumers and businesses. They are more reliable than information provided by sellers because they offer personalized advice as well as ratings of products or services [31]. Many studies have analyzed the relationship between online reviews and product sales for well-known websites such as eBay [32], Amazon [33–35], and Airbnb [36,37]

and various product categories such as books [33–35,38], movies [39–41], fashion [42], cosmetics [43], hotels [44–47], washing machines [48], online lectures [49], restaurants [50,51], and airlines [52].

However, there have been very few studies on consumer car reviews. Kulkarni et al. [53] examined whether Internet use is associated with different choice patterns for cars and found that Internet users rely more on ratings while non-Internet users rely more on recommendations. Sagar et al. [1] considered whether factors affecting car choice behavior such as competition, consumer preferences, and government policies are salient features. Kaushal [2] identified car purchasing behavior through 39 items and validated the usefulness of five factors: safety & security, quality, performance, value, and technology. In this paper, we use mining methods to analyze eight different features: performance, comfort, value, interior, reliability, safety, technology, and exterior for three competitive brands in the automobile market in 2012. Such an approach gives marketers the opportunity to eavesdrop on their own consumers and on consumers in the automotive market in general. In particular, engineers and marketers from automotive companies, based on the results of the review analysis, can obtain structured information about both the superior and problematic aspects of their vehicles and those of their rivals, thereby gaining a competitive advantage in the market. Consequently, the implementation of such an approach can trigger sales growth and improve firm performance.

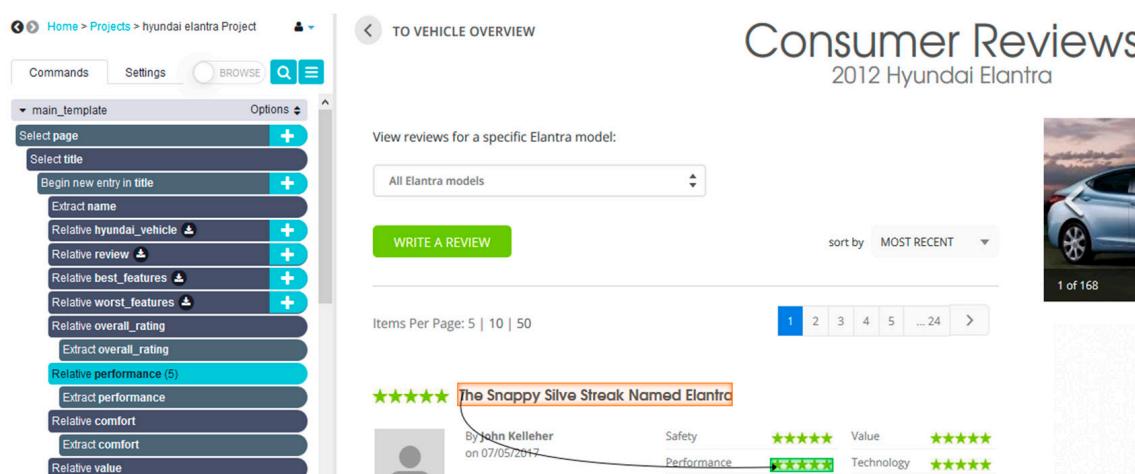
### 3. Research Methodology

#### 3.1. Text Mining Approach to Car Reviews

To obtain adequate and proper data, reviews were collected from one of the biggest online resources for automotive information, Edmunds.com. The process of mining is divided into several steps.

##### 3.1.1. Scraping Data from Websites

Data was collected using the scraping tool “ParseHub.” This is a useful instrument when dealing with information of any kind. It can be adapted to any website, and scholars can extract any piece of information, be it a text or an image (e.g., Figure 1). In this study, units such as title, model of vehicle, best features, worst features, ratings for 8 different features, and total ratings were collected. The process of scraping is illustrated in Figure 1.



**Figure 1.** Consumer reviews.

The results can be saved in JSON or CSV format. In this study, they were saved to CSV format, slightly corrected in an Excel program, and then changed to XSLX format for further usage. The output appears as shown in Figure 2.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
sample	name	vehicle	review	best_features	worst_features	overall_rating	performance	comfort	value	interior	exterior	reliability	safety	technology	
1	hyundai elantra 2012	2012 GLS Elantra w/preferred pkg	GLS 4dr Sedan (1.8L 4-cyl 6A)	So far, I love this car. I bought it 6/5/11 to replace a 2000	I like that adding the bluetooth with	4.75	5	4	5	5	5	5			
2				Hyundai??s safety recall	I only wish that the seats, at least the										
3	Did Hyundai Provide Me Great Customer Service?		GLS 4dr Sedan (1.8L 4-cyl 6M)	Hyundai??s safety recall Campaign 137 was on the		2	2	3	3	2	2	3			
4	Can NOT recommend this car		Limited 4dr Sedan (1.8L 4-cyl 6A)	Limited 4dr Sedan (1.8L 4-cyl 6A) I previously leased a 2010	It looks cool from the outside. But don't	1.75	1	1	1	4	4	1			
5	No buyers remorse!!		GLS 4dr Sedan (1.8L 4-cyl 6A)	GLS 4dr Sedan (1.8L 4-cyl 6A) We put our first 1000 miles on	The dashboard is a bit	4.75	5	4	5	5	5	5			
6	Very good, but not excellent		Limited 4dr Sedan (1.8L 4-cyl 6A)	I previously leased a 2010	easy to read and well										
7	Update		GLS 4dr Sedan (1.8L 4-cyl 6A)	Sorry about the last review. The editors at Edmunds didn't quite	+ The exterior styling. I love looking	5	4	5	4	5	5	5			
8	Pros and cons		GLS 4dr Sedan (1.8L 4-cyl 6A)	I purchased the 2012 Elantra about a month ago.	Price. Warranty. Appearance	4.625	4	4	5	5	5	5			
9	Nice Car.. Could use some improvements		Limited 4dr Sedan (1.8L 4-cyl 6A)	I bought this car in April new and now have had 17,000	The style of the car is	3.875	4	4	4	5	5	4			
10	My first Hyundai		Limited 4dr Sedan (1.8L 4-cyl 6A)	years ago. I handled was a joke. How times have changed.	Love the bluetooth. Leather interior. XM, headroom would be	4.125	4	4	4	5	5	4			
11	Awesome Car		Limited 4dr Sedan (1.8L 4-cyl 6A)	I got a new lease on the 2012 Elantra Limited a few days ago	I love the heated	4.5	4	4	5	5	5	5			
12	At 79 thousand no important complaints!		Limited 4dr Sedan (1.8L 4-cyl 6A)	I've seen that some people have gotten lemons, but most buyers											
13	Nice Car But!		Limited 4dr Sedan (1.8L 4-cyl 6A)	Suspension engineering and steering awful	Body design and interior layout very	3	3	4	4	5	5	2			
14	2012 Elantra Won't Start		GLS 4dr Sedan (1.8L 4-cyl 6A)	Brand new car. Intermittently won't start. Starts only maybe	Fix the suspension and steering issues.	4	4	3	4	5	5	1			
15	Car Failure with only 8 miles		GLS 4dr Sedan (1.8L 4-cyl 6A)	I was very excited to drive to my first Elantra 2012 with only	I love the looks and how the engine has	3.75	2	4	5	4	5	1			
16	Love my new Elantra - great looks and fuel efficiency		GLS 4dr Sedan (1.8L 4-cyl 6M)	I've read tons of other reviews and some people seem	Love the bluetooth. Audio system is a little	4.875	4	5	5	5	5	5			
17	Good, not great!		GLS 4dr Sedan (1.8L 4-cyl 6A)	So I purchased a new 2012 Elantra about 3 months ago. I	Bluetooth Steering wheel controls	3.625	4	3	4	4	5	3			

Figure 2. Extracted data.

All reviews were divided into two groups: satisfied and unsatisfied. The overall rating by consumers was used as the basis for this division. The average mean of overall ratings thus became a separating point for all three car samples and the next condition was set as follows:

```
If ((car = 1) and (overall_rating >= 4.30)) GD = 1
If ((car = 1) and (overall_rating < 4.30)) GD = 0
If ((car = 2) and (overall_rating >= 4.60)) GD = 1
If ((car = 2) and (overall_rating < 4.60)) GD = 0
If ((car = 3) and (overall_rating >= 3.70)) GD = 1
If ((car = 3) and (overall_rating < 3.70)) GD = 0
```

where Car #1 is a Hyundai Elantra, Car #2 is a Honda Civic, and Car #3 is a Ford Focus. GD is a variable for group diversity where GD = 1 relates to the satisfied group and GD = 0 relates to the unsatisfied group. The result of the data split is presented in Table 1.

Table 1. Frequency of the two groups.

Car Brands	Unsatisfied Group (GD = 0)	Satisfied Group (GD = 1)
Elantra (#1)	58	58
Civic (#2)	69	87
Focus (#3)	132	135

### 3.1.2. Input Data

For further analysis, data was input into the R program using the "xlsx" package. This enables the R user to read, write, and format Excel files. Best features, worst features, satisfied and unsatisfied variables, reviewers' id, name of automobile brand, and car id relative to automobile brand were included as input. Part of the input data is displayed in Figure 3.

In the output we acquired 539 reviews, each of which referred to either the satisfied group or unsatisfied group in accordance with the sat/unsat variable obtained from the average mean of the overall rating.

car_var	car	bestdf	worstdf	sat_unsat_var	id_var
1	hyundai	I like that adding the bluetooth with steering wh>	I only wish that the seats, at least the driver's>	1	1
1	hyundai	NA	NA	0	2
1	hyundai	It looks cool for a compact. But don't judge this>	The whole car and the company that makes it is no>	0	3
1	hyundai	The dashboard is easy to read and well designed. >	The bluetooth is a little sporadic in responding >	1	4
1	hyundai	+ The exterior styling. I love looking at my 2012>	Here are the things that I don't like: - The glove>	1	5
1	hyundai	Price. Warranty. Appearance compared to Civic and>	Fuel Economy: This is my least favorite thing abo>	1	6
1	hyundai	NA	Either raise the front of the vehicle or redesign>	0	7
1	hyundai	The style of the car is nice. I do get compliment>	Improvements would be vents in the back seat. Als>	0	8
1	hyundai	Love the bluetooth, leather interior, XM, sunroof>	Another inch of headroom would be nice. My hair b>	1	9
1	hyundai	Being from Michigan I love the heated front and r>	NA	1	10
1	hyundai	NA	NA	0	11
1	hyundai	Body design and interior layout very favorable.	Fix the suspension and steering issues.	0	12
1	hyundai	NA	NA	0	13
1	hyundai	I love the looks and power the engine has consider>	Make sure the car runs well before the dealer put>	0	14
1	hyundai	exterior and interior styling, fuel efficiency, r>	mpg display in the onboard computer is about 2 - >	1	15
1	hyundai	Bluetooth Steering wheel controls Styling Ride	Audio system is a little weak. Headlights are ver>	0	16
1	hyundai	Back up camera ( once you have one you cant imagi>	Maybe a bit more power in the low end would be ni>	1	17
1	hyundai	If you call space a feature, then it's on my list>	A few things -- the storage compartment where the>	0	18
1	hyundai	Seats were comfortable. Drove nice and smooth. Ex>	NA	0	19
1	hyundai	MPG!!!! The Bluetooth is a great feature, and wor>	This is my first car with the slanted windshield.>	1	20
1	hyundai	keyless entry, smart key, nav system, backup came>	Let's add climate controlled AC/heat and put a sp>	1	21
1	hyundai	Nice interior and exterior. The seats are comfort>	Improve sound control. Fix the shifter gear noise>	1	22
1	hyundai	Great looks inside and out Awesome plush interior>	Get rid of the base model with no options and hav>	1	23
1	hyundai	How it drives and the interior. I bought it for t>	Scratching my head on this one. I guess I'd sugge>	1	24
1	hyundai	exterior and interior design, interior room (feel>	AC vents in the center are positioned too low. Th>	1	25

Figure 3. Input data.

### 3.1.3. Data Manipulation

After the data were superficially arranged, R program tools were used to process the data and divide the reviews into four different groups: satisfied best features, satisfied worst features, unsatisfied best features, and unsatisfied worst features. Such segregation can be deciphered as meaning that, even though consumers may be satisfied or unsatisfied, they still identify some best and worst features of the car they are reviewing. An example of the output is presented in Figure 4, where rows are organized randomly to increase clarity.

car_var	car	best_and_worst_features	sat_unsat	sat_unsat_var	id_var	group	group_var
2	honda	6-speed Torquey at low rpm Handling Efficient thr>	unsatisfied	0	186	UNsatisfied-BEST	3
2	honda	NA	satisfied	1	220	satisfied-WORST	2
2	honda	The only complaint I have is the rattles on the i>	satisfied	1	149	satisfied-WORST	2
3	ford	Sporty exterior design.	unsatisfied	0	505	UNsatisfied-BEST	3
1	hyundai	NA	unsatisfied	0	27	UNsatisfied-BEST	3
3	ford	Honesty from FMC; customer service that means som>	unsatisfied	0	490	UNsatisfied-WORST	4
2	honda	NA	satisfied	1	119	satisfied-WORST	2
3	ford	NA	unsatisfied	0	423	UNsatisfied-WORST	4
2	honda	I had my windows tinted, which helps a lot due to>	satisfied	1	121	satisfied-WORST	2
3	ford	NA	unsatisfied	0	359	UNsatisfied-BEST	3
3	ford	We had a little trouble figuring out some of the the >	satisfied	1	374	satisfied-WORST	2
2	honda	Econ mode-great idea! I was also able to add an e>	unsatisfied	0	173	UNsatisfied-BEST	3
3	ford	Ford Manufacturing will need to either recall thi>	unsatisfied	0	348	UNsatisfied-WORST	4
2	honda	NA	satisfied	1	188	satisfied-WORST	2
2	honda	The EV only mode; its the first Honda IMA hybrid >	unsatisfied	0	226	UNsatisfied-BEST	3
1	hyundai	- You can feel the road bumps more than in the Cr>	satisfied	1	43	satisfied-WORST	2
3	ford	NA	satisfied	1	338	satisfied-WORST	2
2	honda	Better mileage, more room than previous generatio>	satisfied	1	159	satisfied-BEST	1

Figure 4. Output data.

### 3.1.4. Data Cleansing

To make it possible to count words and identify their co-occurrence in reviews, text must first undergo the cleaning process. Some features of text are thus removed, such as numbers, white spaces, punctuation, and common English words that have no semantic meaning. As well as converting text to lower case, fixing contractions and text stemming is also essential to obtain accurate and valuable data. We therefore received all the words occurred in all 1078 reviews, which totaled 2299 words. A sample of the output is presented in Figure 5.

row.names	1	2	3	4	5
1 ability	0	0	0	0	0
2 able	0	0	0	0	0
3 abs	0	0	0	0	0
4 absolutely	0	0	0	0	0
5 accelerate	0	0	0	0	0
6 accent	0	0	0	0	0
7 acceptable	0	0	0	0	0
8 access	0	0	0	0	0
9 accessories	0	0	0	0	0
10 accident	0	0	0	0	0
11 accord	0	0	0	0	0
12 accumulated	0	0	0	0	0
13 accurate	0	0	0	0	0
14 aheat	0	0	0	0	0
15 achieve	0	0	0	0	0
16 across	0	0	0	0	0
17 action	0	0	0	0	0
18 active	0	0	0	0	0
19 acts	0	0	0	0	0
20 actual	0	0	0	0	0
21 acura	0	0	0	0	0
22 add	1	0	0	0	0
23 addition	0	0	0	0	0
24 addons	0	0	0	0	0
25 address	0	0	0	0	0
26 adequate	1	0	0	0	0

**Figure 5.** Terms that occurred after review cleaning.

### 3.1.5. Data Mastering

For further data analysis, the final step was to combine our primary database and database, which consisted of occurring words. Thus, with the help of R tools, a master table was created, as shown in Figure 6.

car_var	car	best_and_worst_features	sat_unsat	sat_unsat_var	id_var	group	group_var	ability	able	abs	absolutely	accelerate	accent
1 hyundai	I like that adding the bluetooth with steering wheel con...	satisfied		1	1	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	The dashboard is easy to read and well designed. Gas m...	satisfied		1	4	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	+ The exterior styling. I love looking at my 2012 Elantra ...	satisfied		1	5	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Price, Warranty, Appearance compared to Civic and othe...	satisfied		1	6	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Love the bluetooth, leather interior, XM, sunroof and I...	satisfied		1	9	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Being from Michigan I love the heated front and rear se...	satisfied		1	10	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	exterior and interior styling, fuel efficiency, ride comfort	satisfied		1	15	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Back up camera ( once you have one you cant imagine lif...	satisfied		1	17	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	MPG!!!! The Bluetooth is a great feature, and works reall...	satisfied		1	20	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	keyless entry, smart key, nav system, backup camera, gre...	satisfied		1	21	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Nice interior and exterior. The seats are comfortable and...	satisfied		1	22	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Great looks inside and out Awesome plush interior w/pr...	satisfied		1	23	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	How it drives and the interior. I bought it for the mix of f...	satisfied		1	24	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	exterior and interior design, interior room (feels like a la...	satisfied		1	25	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Everything for the money.	satisfied		1	28	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Style, layout, many standard features, larger trunk, goo...	satisfied		1	31	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	My favorite features are the comfortable seats, amazing ...	satisfied		1	34	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Mileage, warranty, inside room, looks, and price	satisfied		1	37	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	Love the blue tooth, interior design, good on gas \$35 b...	satisfied		1	39	satisfied-BEST	1	0	0	0	0	0	0
1 hyundai	-The exterior is very stylish. -The seats are comfortable. -...	satisfied		1	40	satisfied-BEST	1	0	0	0	0	0	0

**Figure 6.** Master table.

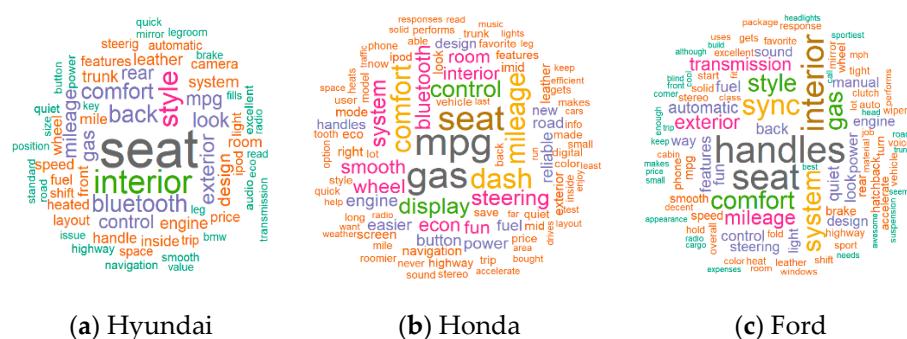
## 4. Results of the Study

### 4.1. Frequency Analysis

Once the master table was created, the actual analysis could be conducted. For this we utilized “WordCloud,” one of R program’s utilities. This is a visualization method that displays how frequently words appear in a given sample of text, and the way it works is quite simple. The more frequently a specific word appears in a database, the bigger and bolder it appears in the word cloud. The results of our cases will now be discussed.

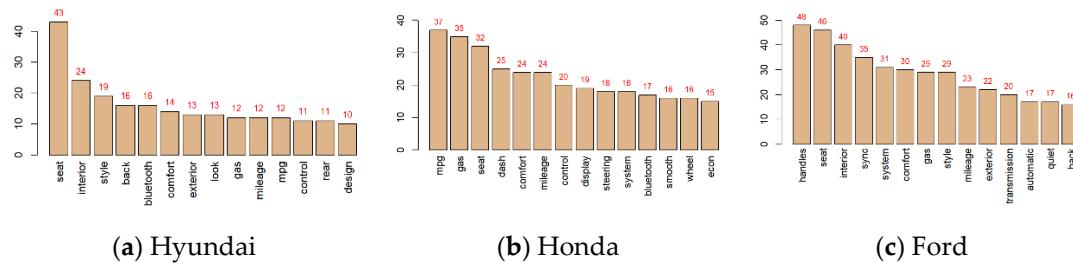
#### 4.1.1. Best Features

Figure 7 shows the best features of three cars. Words with the highest co-occurrence are represented in this word cloud, with the most frequent and important words located in the center and the least frequent words located on the edges. Hence, the closer the words are to edges, the less frequent they are. In the case of Hyundai, “seat” is the most frequent word, followed by “interior,” then “style,” and then the rest. In the case of Honda, the most frequent words are, in order, “mpg,” “gas,” “seat,” “comfort,” “mileage,” “dash,” “control,” “display,” “smooth,” “steering,” “wheel,” “econ,” “fun,” and so on. In contrast with Hyundai, where the main advantage was design and style, consumers mostly emphasize characteristics related to value, technology, and movement on a road. In the case of Ford, the most frequent words are “handles,” “seat,” “interior,” “system,” “sync,” “style,” “comfort,” “gas,” “transmission,” “exterior,” “mileage,” and so on.

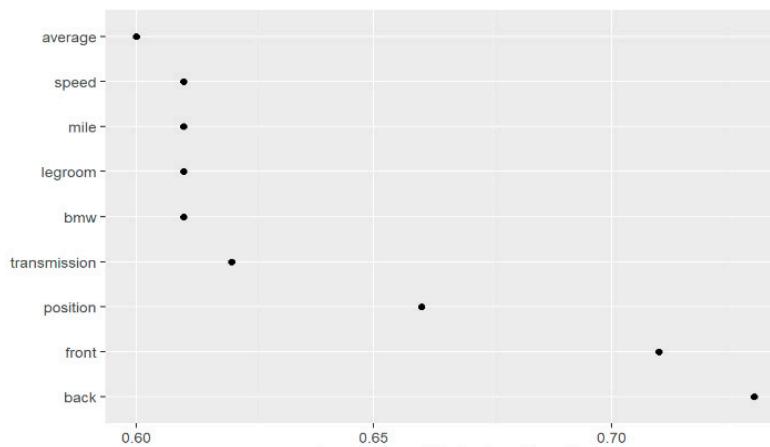


**Figure 7.** Best features (Hyundai, Honda, and Ford) using word cloud.

In the case of Hyundai, after filtration only 14 words remain. As shown in the barplot in Figure 8, there are many words relating to appearance. These are “interior” (24), “style” (19), “exterior” (13), “look” (13), and “design” (10). The interpretation of this result is that consumers mostly liked the car design. The most frequent word is “seat,” which occurs 43 times. Because this word has such high frequency, association analysis was conducted to determine its significance. We performed correlation analysis on the most frequent word, as shown in Figure 9. For example, in the case of Hyundai, the term “seat” has a high correlation with words such as “position,” “front,” and “back.” Hence, we can assume that this word refers to the convenience and comfort consumers felt when they sat in a Hyundai Elantra. Consistent with this interpretation, words such as “back,” “comfort,” and “rear” might also refer to comfort, which was one of the best features for consumers. Similarly, the occurrence of words such as “mpg” and “gas” means that consumers were satisfied with Hyundai’s fuel consumption. For the word “control,” the most closely associated word was “steering” with a correlation of 0.78. This means consumers were likely to be satisfied with their control over the movement of a vehicle.



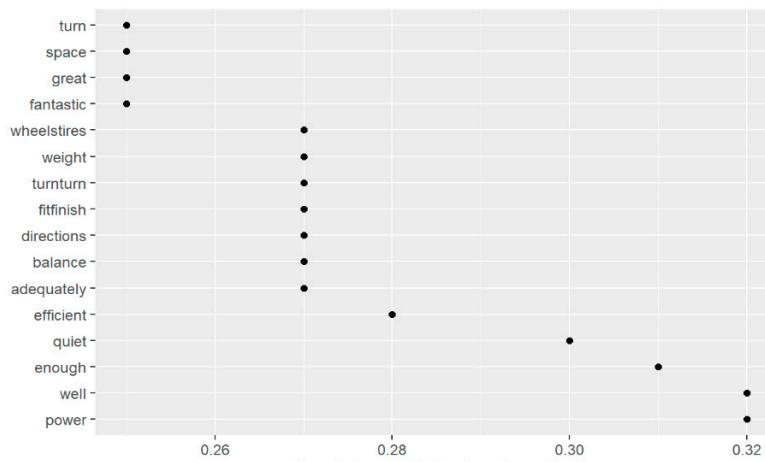
**Figure 8.** Best features (Hyundai, Honda, and Ford) (barplot).



**Figure 9.** Correlation graph for the term “seat” (Hyundai).

In the case of Honda, the first two words are “mpg” and “gas” which means this car has very low fuel consumption. Additionally, for some consumers, seats seem to be very comfortable. The rest of the words are related to the dashboard and technological features such as “dash,” “control,” “display,” “steering,” “system,” “bluetooth,” and “econ.” The word “steering” refers more to technology than holding the road because the correlated words were “wheel,” “control,” “electronic assist,” and “dash.” The word “econ” was correlated with words such as “mode” and “feature.” This is explained by the fact that the Honda Civic has an econ button as a special function, which has become one of its most favored features.

In the case of Ford, the most frequent word is “handles,” which occurred 48 times. Because it is quite difficult to interpret this word, an association analysis was conducted as shown in Figure 10. We can assume that the word “handles” does not refer to the means by which a thing is held, carried, or controlled, but how easily a car is to handle on a road. Such words as “turn,” “directions,” “balance,” “quiet,” “turn,” and others can help to precisely interpret the meaning of this word. Another frequent word was “sync,” which are correlated with some words such as voice, system, phone, ipod, control, navigation, and so on.

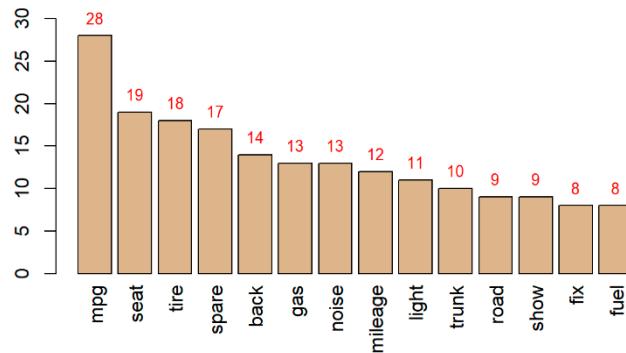


**Figure 10.** Correlation graph for the term “handles” (Ford).

#### 4.1.2. Worst Features

The same analysis was then conducted using reviews that contained the worst features of three brands. For a Hyundai Elantra, the most frequent word is “mpg,” which occurred 28 times, as shown in Figure 11. Although “mpg” also occurred in the results for best features, it is not impossible for the

same term to appear in worst features. Here, we can assume that many consumers were not satisfied with fuel consumption and that these consumers outnumber those who were satisfied.



**Figure 11.** Hyundai's worst features—barplot.

Moreover, if we consider the correlation analysis for “mpg” in Figure 12, we can see that there are highly correlated words such as “show,” “computer,” “onboard,” and “display.” Therefore, we assume there might be some problem related to displaying the mpg on the onboard computer. This hypothesis was checked manually, and it was found that many consumers were complaining about an incorrect mpg display.

\$mpg	show	computer	efficiency	fil	fillup	onboard	got
	0.90	0.68	0.68	0.68	0.68	0.68	0.64
average	display						
	0.61	0.61					

**Figure 12.** Correlation for the term “mpg” (Hyundai).

Consistent with this finding, the correlation for the word “gas” yielded a similar result as shown in Figure 13, so we can assume that the words “estimates” and “misleading” are referring to the same problem.

\$gas	mileage	eco	estimates	misleading
	0.73	0.52	0.52	0.52

**Figure 13.** Correlation for the term “gas” (Hyundai).

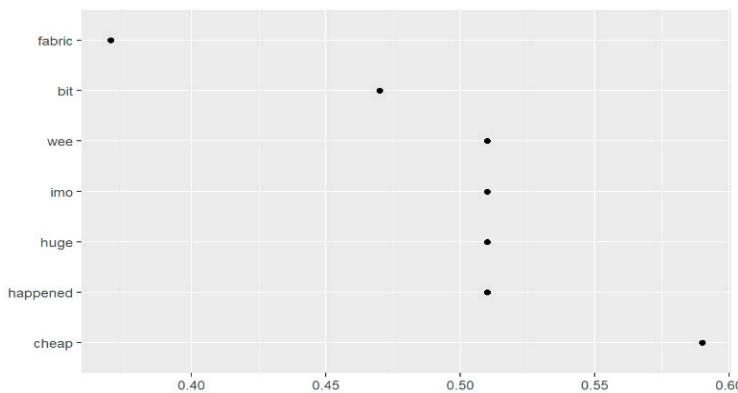
The barplot for worst features shows that consumers were unsatisfied with spare tire (“spare” and “tire” were the most highly correlated with = 0.89), noise on the road, fog lights, trunk, mpg efficiency, as well as the mpg display and seats. As mentioned previously, “mpg,” “gas,” “fuel,” as well as “seat,” occurred inconsistently in the results for best features. Such phenomena could be accounted for by the differing preferences of every individual. Furthermore, based on the proportion of words for both groups, mpg and mileage are more likely to be considered poor rather than superior features because the sum of occurrences in best features is 36; in worst features it is 61.

In the case of Honda, we can see that one of the most frequent words in terms of worst features is “interior,” as shown in Figure 14.

The correlation graphic shows that this is highly correlated with “cheap,” as shown in Figure 15. Therefore, we can assume that some customers did not like the quality and appearance of their interior, viewing it as a drawback rather than an advantage. Additionally, “fabric” is correlated with “interior.” Although the correlation is quite low, we can still assume that reviewers were unsatisfied with the material of their interior. Furthermore, some consumers were not satisfied with back or front seats. Moreover, fog lights, mirrors, and noise on roads became some of the worst features in the Honda Civic as shown in Figure 14.

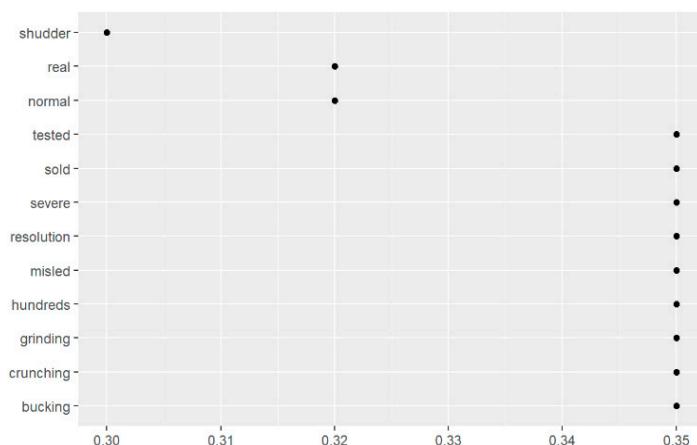


**Figure 14.** Honda's worst features using word cloud.



**Figure 15.** Correlation for the term "interior" (Honda).

In the case of Ford, the worst features were "transmission," "seat," "back," "control," "fix," "issue," "shift," "rear," "wheel," "system," and so on. The most frequent word "transmission" occurred 57 times, comparatively larger than other terms. To understand the transmission flaw in the Ford Focus, an association graph was built, as shown in Figure 16.



**Figure 16.** Correlation for the term "transmission" (Ford).

Among the terms highly correlated with "transmission" were "severe," "grinding," "crunching," and "bucking." Therefore, we can assume there is a problem with the transmission, as it is perceived as making strange sounds and being inconvenient to use. In addition, the terms "issue," "fix," "manual," "problem," and "shift" also correlated with "transmission," which means that it is probably the most significant problem with the Ford Focus. Ford also seems to have problems in terms of technology. For instance, the terms "control" and "wheel" were highly correlated with "steering," "device," "equipment," "aux," "cruise," "dashboard," and other words, which can be interpreted as

Ford exhibiting a deficiency in equipment. There are also consumers who are certainly not satisfied with the seats and space in a cab, both front and back (rear).

#### 4.1.3. Implications and Discussion

Based on the results, we can assume that the biggest strength of the Hyundai Elantra car is its design. This is supported by the fact that the most frequent words are related to car appearance. These include "interior," "style," exterior," "look," and "design." The worst features for Hyundai appear to be gas consumption and some problems with technology, such as the mpg display on the onboard computer and problems with a spare tire. In contrast to the Hyundai car, the biggest strengths of the Honda car are low gas consumption, dashboard, controls on the steering wheel, and the "econ" mode, which improves fuel efficiency. The worst feature for Honda appears to be its interior, which reviewers emphasized as cheap. Furthermore, they were unsatisfied with the material it was made of. In the case of Ford, the biggest strength appeared to be manipulation of the car. This is consistent with the high frequency of the word "handles." The other best features for Ford were the interior and exterior, noiselessness during the ride, and the "Ford Sync" system which allowed users to control automotive functions using their voice. The biggest disadvantage for Ford was found to be transmission. This was supported by the high frequency of the word "transmission" and other frequent yet negative words, such as "issue," "fix," "manual," "problem," and "shift." Another important disadvantage relates to technology, specifically a problem with the controls on the steering wheel. In addition, in all three cases one of the most frequent words was "seat," which appeared in both "best features" and "worst features" categories, suggesting that reviewers are divided in their opinions. Hence, it can be concluded that it is difficult for all three car brands to find favor in the eyes of all consumers.

#### 4.2. Analysis of Car Features Using the Association Rule

According to Edmunds.com, eight different consumers rate features whereby each one of eight features refers to certain terms and involves some form standard conception. Otherwise, every individual might have a different conception about each of the features relative to other individuals.

- Performance involves terms such as acceleration, braking, road holding, and shifting.
- Comfort relates to front seats, rear seats, getting in/out, and noise/vibration.
- Value involves fuel economy, maintenance cost, purchase cost, and resale value.
- Interior implies cargo/storage, instrumentation, interior design, and the logic of controls.
- Reliability relates to repair frequency, dealership support, engine, and transmission.
- Safety consists of headlights, outward visibility, parking aids, and rain/snow traction.
- Technology stands for entertainment, navigation, Bluetooth, and USB ports.
- Exterior stands for exterior design.

The question that arises is: how are the most frequent terms for each car related to the eight different features and what is their frequency? To answer this question, both groups of reviews, which contain best features and worst features, were combined and analyzed using text mining tools. The aim was to determine the frequency of every word that occurs in reviews. The process was conducted for all three car brands: Hyundai Elantra 2012, Honda Civic 2012, and Ford Focus 2012. The 24 most frequent terms were chosen as a sample and, using the same association approach as in previous research, the relationship between the terms and eight different features were found. The results are presented in Table 2.

**Table 2.** Relationship between terms and eight features.

Hyundai Elantra			Honda Civic			Ford Focus		
Terms	Freq	Features	Terms	Freq	Features	Terms	Freq	Features
seat	62	Comfort	seat	61	comfort	seat	82	comfort
mpg	40	Value	mpg	46	value	transmission	77	reliability
back	30	Comfort	dash	40	tech	sync	56	Tech
interior	26	Interior	gas	40	value	interior	52	interior
gas	25	Value	interior	35	interior	handle	51	Perform/comfort
mileage	24	Value	mileage	30	value	back	48	comfort
bluetooth	20	Tech	comfort	29	comfort	system	48	perform/tech
style	20	interior/exterior	control	26	perform/comfort/tech	control	44	perform/comfort/tech
tire	19	Interior	display	23	interior/tech	gas	40	Value
trunk	19	Interior	wheel	23	perform/tech	comfort	36	comfort
rear	18	comfort/interior	road	22	perform/comfort	manual	34	reliability
light	17	safety/tech	steering	22	perform/tech	mileage	33	Value
spare	17	Interior	back	21	comfort	rear	32	comfort/interior
control	16	Perform/comfort/tech	bluetooth	21	tech	automatic	31	reliability
engine	16	Reliability	system	21	perform/tech	style	31	interior/exterior
look	16	interior/exterior	power	20	perform	wheel	28	perform/tech
system	16	perform/tech	room	19	comfort/interior	issue	27	reliability
comfort	15	Comfort	design	18	interior/exterior	look	27	interior/exterior
design	15	interior/exterior	engine	18	reliability	exterior	26	exterior
front	15	Comfort	econ	17	value/tech	front	26	comfort
fuel	15	Value	front	17	comfort	shift	26	reliability
noise	14	comfort	light	17	safety/tech	speed	26	perform
speed	14	perform	look	17	interior/exterior	light	25	safety/tech
exterior	13	exterior	noise	17	comfort	steering	25	perform/tech

Although the frequency for the most frequent terms was found, the total number of all reviews for each vehicle brand was different; specifically, 116 for the Hyundai Elantra, 156 for the Honda Civic, and 267 for the Ford Focus, respectively. Thus, we have to adjust the numbers to the common denominator to interpret the comparison between three brands more clearly. To do this, the following formula was applied:

$$F = \frac{\text{frequency of a given term}}{\text{total quantity of reviews for a given car}}$$

where  $F$  is approximate occurrence of a given term in one review.

Thus, all words according to a specific feature were summed, and their frequency before and after adjustment was determined.

#### 4.2.1. Analysis Results for Eight Features

As shown in Table 3, the highest frequency was for terms related to comfort and interior features in Hyundai, comfort and technological features in Honda, and comfort in Ford (Criteria:  $F \geq 1$ ). Therefore, reviewers were mostly interested in these features and discussed these most heatedly. These can now be scrutinized more closely for each case.

##### (1) Hyundai

If we compare the frequency of terms for interior and the average rating score of consumers for interior we see that terms related to interior appear more often in the satisfied group and in best features because the score for interior is quite high. Therefore, the Hyundai company has won the favor of consumers in respect to its interior. The comfort score is 3.97, which is neither high nor low. This suggests there might be some factors reviewers were not satisfied with. Thus, a more precise analysis is needed. Another feature worth considering is the exterior feature. Compared to other automobile brands, the frequency of terms related to the exterior is significantly higher than for Honda

and Ford. It can therefore be assumed that the exterior is also the strongest feature for Hyundai. Its score is highest among the scores for all features and the likelihood that terms related to an exterior would mostly occur in the satisfied group and best features is very high.

**Table 3.** Frequency before and after adjustment.

Features	Hyundai			Honda			Ford		
	Freq (before)	Freq (after)	Score	Freq (before)	Freq (after)	Score	Freq (before)	Freq (after)	Score
performance	46	0.40	3.72	134	0.86	4.29	222	0.83	3.33
comfort	170	1.47	3.97	212	1.36	4.26	319	1.19	3.70
value	104	0.90	4.13	133	0.85	4.52	73	0.27	3.63
interior	150	1.29	4.39	112	0.72	4.14	142	0.53	3.80
exterior	64	0.55	4.70	35	0.22	4.40	84	0.31	4.47
reliability	16	0.14	3.90	18	0.12	4.66	195	0.73	3.19
safety	17	0.15	4.05	17	0.11	4.10	25	0.09	2.95
technology	69	0.59	3.70	210	1.35	4.10	226	0.85	2.84

## (2) Honda

Comparing the frequencies and scores for comfort and technology, it is very likely that terms related to these will occur mostly in the satisfied group and best features because the score for both groups is pretty high. Furthermore, the frequency and score for technology is the highest among all three automobile brands, which can be interpreted as Honda being a technological leader.

## (3) Ford case

Although the frequency of terms related to comfort is high, the score is quite low. Although it is not the lowest score compared to other features, the result suggests that such terms would appear in both satisfied and unsatisfied groups and in both the best and worst features groups. Furthermore, reliability is also worth mentioning, because the frequency of terms related to this feature is much higher than for Hyundai and Honda. With a very low score for reliability, we can assume that terms will mostly appear in worst features for both satisfied and unsatisfied groups.

### 4.2.2. Comparison of Two Groups' Reviews

In this section, we compare the reviews of both groups and find terms whose influence is greater than others. We also compare the differences between satisfied and unsatisfied groups of reviewers. What, therefore, are the frequency and ratio of terms for eight different features between satisfied and unsatisfied groups and what are the implications of this?

#### Performance

As shown in Table 4, in the case of Hyundai consumers rarely mentioned words related to performance in comparison to Honda and Ford consumers.

The satisfied group mentioned the words “control,” “system,” and “speed” more often than the unsatisfied group, although it is difficult to say they were definitely satisfied with these factors because these terms appeared in both best features and worst features. The frequency of these terms in the unsatisfied group is very low. Along with a fairly low score for performance, we can assume that reviewers were unsatisfied due to factors other than “control,” “system,” and “speed.” In the case of Honda, consumers were generally satisfied with the performance because the most frequent words for this feature mostly appeared in best features, and, looking at the performance score of 4.29 in Table 3,

we can assume they were significant to a certain degree. The result for Ford is ambiguous, but we can say with confidence that reviewers like how the car handles as the term “handle” appeared much more frequently than other terms and, in 95% of cases, appeared in best features for both satisfied and unsatisfied groups.

**Table 4.** Performance (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
control	0.09/0.04	0.01/0.00	0.14	0.08/0.01	0.04/0.03	0.17	0.04/0.07	0.01/0.04	0.16
system	0.07/0.03	0.01/0.03	0.14	0.06/0.01	0.05/0.01	0.13	0.07/0.04	0.04/0.02	0.18
speed	0.06/0.05	0.01/0.00	0.12				0.04/0.04	0.00/0.02	0.10
wheel				0.06/0.01	0.04/0.01	0.15	0.03/0.05	0.00/0.01	0.10
road				0.06/0.03	0.01/0.04	0.14			
steering				0.06/0.01	0.06/0.02	0.14	0.04/0.03	0.01/0.01	0.09
power				0.05/0.02	0.03/0.03	0.13			
handle							0.15/0.01	0.03/0.00	0.19
Sum	0.22/0.12	0.03/0.03	0.40	0.37/0.08	0.24/0.17	0.86	0.39/0.25	0.10/0.10	0.83

## Comfort

As shown in Table 5, Hyundai drivers felt comfortable in the car, and most were very satisfied with the seats. However, it seems that it has some problems with noise as the average score for comfort is low. Looking at the results for Honda, consumers who were satisfied with the Honda Civic purchase felt very comfortable in a cab and were pleased with the space provided, but it is likely that both satisfied and unsatisfied groups were unsatisfied with the back seats. In the case of Ford, there were people who found it comfortable and people who did not. The unsatisfied group did not discuss the comfort feature as much as the satisfied group, and it seems that individuals from the satisfied group liked neither the front nor the back seats. The most positive aspect of Ford mentioned by reviewers is the manipulation of the car.

**Table 5.** Comfort (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
seat	0.27/0.13	0.10/0.03	0.53	0.12/0.09	0.08/0.10	0.39	0.13/0.10	0.04/0.03	0.31
back	0.09/0.08	0.04/0.04	0.26	0.04/0.05	0.01/0.04	0.13	0.04/0.09	0.01/0.03	0.18
control	0.09/0.04	0.01/0.00	0.14	0.08/0.01	0.04/0.03	0.17	0.04/0.07	0.01/0.04	0.16
comfort	0.09/0.01	0.03/0.00	0.13	0.12/0.01	0.03/0.02	0.19	0.09/0.00	0.02/0.02	0.13
rear	0.08/0.03	0.02/0.00	0.16				0.03/0.05	0.01/0.03	0.12
front	0.06/0.03	0.01/0.03	0.13	0.02/0.05	0.01/0.03	0.11	0.01/0.07	0.01/0.01	0.10
noise	0.01/0.07	0.00/0.04	0.12	0.00/0.03	0.03/0.05	0.11			
road				0.06/0.03	0.01/0.04	0.14			
handle							0.15/0.01	0.03/0.00	0.19
room				0.08/0.02	0.01/0.01	0.12			
Sum	0.69/0.39	0.21/0.18	1.47	0.52/0.29	0.22/0.33	1.36	0.52/0.39	0.13/0.15	1.19

## Value

As shown in Table 6, most Hyundai holders were not satisfied with the fuel consumption of this car. Nevertheless, some of the satisfied group felt that Hyundai’s mpg was not bad. The reason

for this might depend on individual satisfaction levels in relation to mpg assessment. In the case of Honda, all terms related to fuel consumption constantly appeared in best features for both satisfied and unsatisfied groups. Occurring several times in worst features was the word minimal, which means Honda's mpg is very high and probably best among the three brands. Ford holders did not mention fuel consumption as much as owners of the other cars, but it is likely that Ford does not have any problems with fuel consumption and is credibly even better than Hyundai's mpg. Hence, there are other factors that resulted in the low score for value.

**Table 6.** Value (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
gas	0.09/0.03	0.01/0.09	0.22	0.17/0.02	0.06/0.01	0.26	0.08/0.04	0.03/0.00	0.15
mpg	0.09/0.22	0.02/0.03	0.34	0.12/0.04	0.12/0.02	0.29			
mileage	0.09/0.04	0.13/0.02	0.21	0.08/0.02	0.08/0.02	0.19	0.06/0.02	0.03/0.01	0.12
fuel	0.06/0.07	0.00/0.00	0.13		0.07/0.01	0.03/0.01	0.11		
econ									
Sum	0.33/0.35	0.04/0.17	0.90	0.44/0.08	0.28/0.06	0.85	0.14/0.06	0.06/0.02	0.27

## Interior

As shown in Table 7, the interior was most often discussed in the Hyundai case, and it is clear that reviewers from both satisfied and unsatisfied groups were greatly satisfied with this attribute. However, there is also a problem with a spare tire, which often appeared in worst features for both satisfied and unsatisfied groups. Opinions about the interior for Honda were divided among the satisfactory group, but the common element for both groups is that they liked the dashboard display. Additionally, the satisfied group often mentioned "room," which means they were satisfied with this feature. The interior results for Ford holders were good rather than bad, but it seems this was not the most important feature for reviewers.

**Table 7.** Interior (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
interior	0.15/0.00	0.06/0.02	0.22	0.08/0.04	0.01/0.10	0.22	0.11/0.03	0.04/0.02	0.19
style	0.09/0.01	0.08/0.00	0.17				0.09/0.00	0.02/0.00	0.12
tire	0.01/0.09	0.00/0.07	0.16						
trunk	0.07/0.07	0.01/0.02	0.16						
rear	0.08/0.03	0.02/0.03	0.16				0.03/0.05	0.01/0.03	0.12
spare	0.00/0.10	0.00/0.04	0.15						
look	0.09/0.02	0.02/0.01	0.14	0.04/0.02	0.01/0.03	0.11	0.03/0.03	0.03/0.01	0.10
design	0.08/0.03	0.01/0.02	0.13	0.05/0.02	0.01/0.03	0.12			
display				0.08/0.01	0.04/0.01	0.15			
room				0.08/0.02	0.01/0.01	0.12			
Sum	0.56/0.34	0.19/0.20	1.29	0.33/0.11	0.10/0.19	0.72	0.26/0.11	0.10/0.06	0.53

## Exterio

As shown in Table 8, in all three cases, customers were plenty satisfied with the exterior, especially Hyundai and Ford users. For instance, the occurrence of terms related to the exterior was very high for the Hyundai Elantra and the term "exterior" never appeared in worst features in either of the two

groups. In the case of Ford, relative to other features, the exterior was the only factor with which consumers were satisfied. It is also the only factor which has a high score in the Ford sample. However, looking at the frequency of the term, we can assume that it was not the hottest topic for discussion compared to Hyundai. For the Honda Civic, the term “exterior” did not appear at all, which means it was not the main factor in determining whether customers purchased this car.

**Table 8.** Exterior (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
style	0.09/0.01	0.08/0.00	0.17				0.09/0.00	0.02/0.00	0.12
look	0.09/0.02	0.02/0.01	0.14	0.04/0.02	0.01/0.03	0.11	0.03/0.03	0.03/0.01	0.10
design	0.08/0.03	0.01/0.02	0.13	0.05/0.02	0.01/0.03	0.12			
exterior	0.09/0.00	0.02/0.00	0.11				0.03/0.01	0.05/0.01	0.10
Sum	0.35/0.05	0.12/0.03	0.55	0.10/0.04	0.03/0.06	0.23	0.15/0.04	0.10/0.02	0.32

### Reliability and Safety

As shown in Table 9, among the 24 most frequent terms for Hyundai and Honda, only one word, “engine,” was related to reliability. There might be other words such as “engine” that could influence a decision to rate reliability, but these did not appear among the most frequent words. Therefore, it is hard to determine the extent to which the term “engine” affected the reliability score, but many people from the satisfied group for Honda were satisfied with its engine and mentioned it a few times. In the Hyundai group, the term “engine” occurred almost equally in best features and worst features. In the satisfied group it occurred more often in best features while in the unsatisfied group it appeared more frequently in worst features, which makes sense. Thus, we can assume there was an approximately equal number of people who were satisfied and unsatisfied with the engine.

**Table 9.** Reliability and safety (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
Reliability feature									
engine	0.06/0.02	0.02/0.04	0.14	0.05/0.01	0.03/0.03	0.12			
transmission							0.06/0.09	0.01/0.12	0.29
manual							0.05/0.06	0.00/0.01	0.13
automatic							0.06/0.03	0.01/0.02	0.12
issue							0.00/0.04	0.01/0.04	0.10
shift							0.03/0.04	0.00/0.03	0.10
Sum	0.06/0.02	0.02/0.04	0.14	0.05/0.01	0.03/0.03	0.12	0.19/0.27	0.04/0.23	0.73
Safety feature									
light	0.04/0.05	0.01/0.04	0.15	0.03/0.04	0.01/0.03	0.11	0.04/0.04	0.01/0.01	0.09

Reliability was actively discussed in the Ford group. It is clear that Ford has serious problems in this field, mostly to do with transmission. The frequency of the term “transmission” was 0.29, the highest among all terms in the reliability group and more than three times higher than the frequency of the term “engine” in the Hyundai and Honda groups. Furthermore, the frequency of terms “manual,” “automatic,” “issue,” and “shift” was also high. Based on frequency analysis, where a link was found between these terms and “transmission,” we can say that both satisfied and unsatisfied groups criticized transmission, and the occurrence of these terms in total was 0.73, which is very high for just one specific part of a car. The low score of 3.19 in Table 3 is consistent with the results for

frequency, so Ford must solve this problem in order to secure clients' trust. In the case of safety, it is difficult to interpret the results as there is only one word, "light," that, after association analysis, was correlated with several features such as "safety" and "technology." Hence, it would be a mistake to judge the significance of the relationship between safety scores for all three car brands and the term "light" as well as its frequency in the satisfied and unsatisfied groups.

### Technology

Among the three brands, technology was the most frequently discussed by Honda owners, quite frequently by Ford owners, and least often in the Hyundai group as shown in Table 10. Along with comfort, technology was the hottest topic for discussion in the Honda group. We can say with confidence that Honda holders from both satisfactory groups greatly enjoyed using the steering wheel, Bluetooth, econ function, and inward system. Opinions about the dashboard varied, as there were reviewers in both satisfied and satisfied groups who liked or did not like this feature. Hence, Honda consumers were very satisfied with the technological side, and the technological level is probably the highest among the three brands. In the case of Ford, satisfied and unsatisfied groups mentioned terms related to technology in both best features and worst features, and the ratio was quite similar. Indeed, one of the most frequent terms was "sync," which refers to Ford's special feature. Looking at the results, it seems that, regardless of the satisfactory group, some reviewers enjoyed using this system and some did not. Therefore, the technological side of the Ford Focus was worth paying attention to, but it is unclear whether this is beneficial or disadvantageous for an automotive company. Given the very low score in Table 3 for the technology, we can presume that reviewers who mentioned these words in worst features evaluated it very negatively, while reviewers who mentioned these words in best features did not evaluate it highly. If we look at the results for Hyundai, we can say that consumers liked the Bluetooth system, but we cannot say the same about other terms. Therefore, we suppose there are other factors that resulted in the low score for technology.

**Table 10.** Technology (best/worst features of each group).

Words	Hyundai			Honda			Ford		
	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum	Satisfied Best/Worst	Unsatisfied Best/Worst	Sum
system	0.07/0.03	0.01/0.03	0.14	0.06/0.01	0.05/0.01	0.13	0.07/0.04	0.04/0.02	0.18
control	0.09/0.04	0.01/0.00	0.14	0.08/0.01	0.04/0.03	0.17	0.04/0.07	0.01/0.04	0.16
bluetooth	0.09/0.03	0.04/0.01	0.17	0.07/0.02	0.04/0.01	0.13			
light	0.04/0.05	0.01/0.04	0.15	0.03/0.04	0.01/0.03	0.11	0.04/0.04	0.01/0.01	0.09
dash			0.09/0.04	0.07/0.05	0.26				
display			0.08/0.01	0.04/0.01	0.15				
wheel			0.06/0.01	0.04/0.03	0.15	0.03/0.05	0.00/0.01	0.10	
steering			0.06/0.01	0.06/0.02	0.14	0.04/0.03	0.01/0.01	0.09	
econ			0.07/0.01	0.03/0.01	0.11				
sync						0.10/0.06	0.03/0.02	0.21	
um	0.29/0.16	0.07/0.08	0.59	0.60/0.16	0.38/0.21	1.35	0.34/0.29	0.10/0.11	0.85

#### 4.2.3. Implications and Discussion

Based on the results of this research, the following propositions can be stated. Firstly, among the three car brands, the Hyundai Elantra car has the best marks in relation to "interior" and "exterior," but, in terms of the interior, there is a problem with a spare tire that needs to be solved. A few people were also unsatisfied with gas consumption, so it would be better for engineers from Hyundai to improve the mpg index. Furthermore, Hyundai has problems in terms of technology, one of which is incorrect mpg displays. Moreover, despite consumers' satisfaction with comfort, there was a problem

with noise on roads. This seems to be the main reason for a relatively low score for comfort. Hence, Hyundai should reconsider the value and technological particularities of the Elantra car to make it more competitive on the market. Secondly, among the three car brands, the Honda Civic received positive feedback for all features and was found to be best in terms of value and technology. It has the best mpg index and the best technological equipment compared to Hyundai and Ford. Despite satisfaction with all features, Honda engineers should pay attention to the interior, because many consumers criticized it for its cheapness. In addition, Honda should consider the issue of comfort, because back seats were also found to be a weak point. Thirdly, the Ford car was evaluated very poorly regarding all features with the exception of the exterior, where it received a high score. However, according to the results, this was not the most discussed topic among reviewers. Among the 24 most frequent terms, negative terms were in the majority. These were related to the topic of reliability, where reviewers severely criticized transmission and found this to be the biggest problem in the Ford automobile. Apart from problems with transmission, most consumers were quite unsatisfied with both back and front seats. The only feature reviewers were truly pleased with, according to the results, was manipulation of the car. In previous research, reliability was found to be one of the most significant factors for buyers. Therefore, looking at the poor evaluation of this car, reviewers were greatly disappointed with its reliability and, for this reason, the scores for other features were slightly biased. Hence, Ford marketers and engineers must completely reconsider and reassess their car from all sides, starting with the reliability feature.

In addition, the results for several features, such as safety, were unclear and ambiguous, thereby making interpretation difficult. This can be explained by the lack of terms chosen for the analysis. To fill such blind spots, a more extended analysis is needed.

## 5. Conclusions

In this paper, consumer reviews of three different competitive automobiles—the Hyundai Elantra, the Honda Civic, and the Ford Focus—were examined. The results can be summarized as follows:

Firstly, each car model was analyzed in terms of its best and worst features, thereby underlining the superior features of a given car as well as its problematic features. It was reached by virtue of finding the words appearing most frequently in corresponding reviews. In terms of best features, the Hyundai Elantra has car design, seats, interior, bluetooth, and steering control; the Honda Civic has low gas consumption, seats, dashboard, and technological equipment such as Bluetooth and “econ” mode; and the Ford Focus has car manipulation, exterior, interior, a quiet ride, and the “Ford Sync” function. In terms of the worst features, the Hyundai Elantra has an incorrect mpg display, problems with a spare tire, noise on the road, fog lights, trunk, and mpg inefficiency; the Honda Civic has a cheap interior, seats, fog lights, mirrors, and noise on the road; and the Ford Focus has problems with transmission, seats, space in the cab, and controls on the steering wheel. For the Ford Focus, some features such as seats, gas consumption, dashboard, and “Ford Sync” systems were found in both best and worst features, which can be interpreted as a difference of opinion among reviewers.

Secondly, eight specific yet different features were analyzed using consumers’ reviews of best and worst features. The results showed that consumers actively discussed the comfort feature for all three brands. In particular, Hyundai reviewers emphasized the interior and exterior, Honda reviewers were interested in technology, and Ford reviewers paid attention to reliability.

Thirdly, the ways in which the views of both satisfied and unsatisfied groups differed were analyzed. The results showed that Hyundai received the best marks in terms of design and interior but needs to reconsider the value and technology features to make the Elantra more competitive on market. The Honda Civic does not have any critical issues relating to any factors. It has the best mpg index and technological equipment compared to Hyundai and Ford. However, it should consider its cheap interior and comfort feature. In contrast, Ford should completely reconsider and reassess its car, starting with its reliability.

This paper has the following limitations. Firstly, even though the Edmunds website is one of the biggest online resources for automotive information, there was still a general lack of reviews. To obtain more significant results, a larger amount of reviews will be needed. Furthermore, it would be useful to obtain data on sales figures for all three car models to make a comparison between sales and the results of this study. In addition, because only 24 of the most frequent terms were chosen for the analysis, the results for some features, such as safety, were unclear and ambiguous, thereby making interpretation difficult. Hence, it is necessary to increase the number of terms to find more related to safety features.

**Author Contributions:** Conceptualization, E.-G.K. and S.-H.C.; formal analysis, E.-G.K. and S.-H.C.; methodology, S.-H.C.; software, E.-G.K. and S.-H.C.; visualization, E.-G.K.

**Funding:** This research received no external funding.

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Sagar, A.D.; Chandra, P. *Technological Change in the Indian Passenger Car Industry, Energy Technology Innovation Policy Discussion Paper: BCSIA Discussion Paper 2004–05*; Energy Technology Innovation Project; Kennedy School of Government, Harvard University: Cambridge, MA, USA, 2004.
2. Kaushal, S.K. Confirmatory factor analysis: An empirical study of the fourwheeler car buyer’s purchasing behavior. *Int. J. Glob. Bus. Manag. Res.* **2014**, *2*, 90–104.
3. Janssen, M.; van der Voort, H.; Wahyudi, A. Factors influencing big data decision-making quality. *J. Bus. Res.* **2017**, *70*, 338–345. [[CrossRef](#)]
4. Olszak, C.M. Toward better understanding and use of business intelligence in organizations. *Inf. Syst. Manag.* **2016**, *33*, 105–123. [[CrossRef](#)]
5. Kiron, D. Organizational alignment is key to big data success. *MIT Sloan Manag. Rev.* **2013**, *54*, 54307.
6. Liu, Y. Big data and predictive business analytics. *J. Bus. Forecast.* **2014**, *33*, 40–42.
7. Wamba, S.F.; Gunasekaran, A.; Akter, S.; Ren, S.J.; Dubey, R.; Childe, S.J. Big data analytics and firm performance: Effects of dynamic capabilities. *J. Bus. Res.* **2017**, *70*, 356–365. [[CrossRef](#)]
8. Vidgen, R.; Shaw, S.; Grant, D.B. Management challenges in creating value from business analytics. *Eur. J. Oper. Res.* **2017**, *261*, 626–639. [[CrossRef](#)]
9. Günther, W.A.; Mehrizi, M.H.R.; Huysman, M.; Feldberg, F. Debating big data: A literature review on realizing value from big data. *J. Strateg. Inf. Syst.* **2017**, *26*, 191–209. [[CrossRef](#)]
10. Mikalef, P.; Pappas, I.O.; Krogstie, J.; Giannakos, M. Big data analytics capabilities: A systematic literature review and research agenda. *Inf. Syst. e-Bus. Manag.* **2018**, *16*, 547–578. [[CrossRef](#)]
11. Grover, V.; Chiang, R.H.L.; Liang, T.; Zhang, D. Creating Strategic Business Value from Big Data Analytics: A Research Framework. *J. Manag. Inf. Syst.* **2018**, *35*, 388–423. [[CrossRef](#)]
12. Dong, J.Q.; Yang, C.-H. Business value of big data analytics: A systems-theoretic approach and empirical test. *Inf. Manag.* **2018**, in press. [[CrossRef](#)]
13. Müller, O.; Fay, M.; Brocke, J.V. The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *J. Manag. Inf. Syst.* **2018**, *35*, 488–509. [[CrossRef](#)]
14. Corte-Real, N.; Ruivo, P.; Oliveira, T. Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? *Inf. Manag.* **2019**, in press.
15. Mikalef, P.; Boura, M.; Lekakos, G.; Krogstie, J. Big data analytics and firm performance: Findings from a mixed-method approach. *J. Bus. Res.* **2019**, *98*, 261–276. [[CrossRef](#)]
16. Tang, C.; Guo, L. Digging for Gold with a Simple Tool: Validating Text Mining in Studying Electronic Word-of-Mouth (eWOM) Communication. *Mark. Lett.* **2015**, *26*, 67–80. [[CrossRef](#)]
17. Pennebaker, J.; Mehl, M.; Niederhoffer, K. Psychological aspects of natural language: Our words, our selves. *Annu. Rev. Psychol.* **2003**, *54*, 547–577. [[CrossRef](#)]
18. Gupta, V.; Lehal, G.S. A survey of text mining techniques and applications. *J. Emerg. Technol. Web Intell.* **2009**, *1*, 60–76. [[CrossRef](#)]
19. Ramanathan, V.; Meyyappan, T. Survey of text mining. *Proc. Int. Conf. Technol. Business Manag.* **2013**, 508–514.

20. Khader, N.; Lashier, A.; Yoon, S.W. Pharmacy robotic dispensing and planogram analysis using association rule mining with prescription data. *Expert Syst. Appl.* **2016**, *57*, 296–310. [[CrossRef](#)]
21. Ren, J.; Li, W.; Wang, Y.; Zhou, L. Graph-mine: A key behavior path mining algorithm in complex software executing network. *Int. J. Innov. Comput. Inf. Control* **2015**, *11*, 541–553.
22. Calders, T.; Dexters, N.; Gillis, J.J.; Goethals, B. Mining frequent itemsets in a stream. *Inf. Syst.* **2014**, *39*, 233–255. [[CrossRef](#)]
23. Han, J. *Data Mining: Concepts and Techniques*; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 2005.
24. Tan, P.N.; Steinbach, M.; Kumar, V. *Introduction to Data Mining*, 1st ed.; Addison-Wesley Longman Publishing Co.; Inc.: Boston, MA, USA, 2005.
25. Czibula, G.; Czibula, I.G.; Miholca, D.; Crivei, L.M. A novel concurrent relational association rule mining approach. *Expert Syst. Appl.* **2019**, *125*, 142–156. [[CrossRef](#)]
26. Müller, J.; Christandl, F. Content is king—But who is the king of kings? The effect of content marketing, sponsored content & user-generated content on brand responses. *Comput. Hum. Behav.* **2019**, *96*, 46–55.
27. Ayeh, J.; Au, N.; Law, R. Do we believe in TripAdvisor? Examining credibility perceptions and online travelers' attitude toward using user-generated content. *J. Travel Res.* **2013**, *52*, 437–452. [[CrossRef](#)]
28. Amato, F.; Moscato, V.; Picariello, A.; Sperlì, G. Multimedia social network modeling: A proposal. In Proceedings of the 2016 IEEE Tenth International Conference on, IEEE Semantic Computing, ICSC, Laguna Hills, CA, USA, 3–5 February 2016; pp. 448–453.
29. Amato, F.; Moscato, V.; Picariello, A.; Sperlì, G. Diffusion Algorithms in Multimedia Social Networks: A preliminary model. In Proceedings of the IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Sydney, Australia, 31 July–3 August 2017; pp. 844–851.
30. Yahav, I.; Shehory, O.; Schwartz, D. Comments Mining With TF-IDF: The Inherent Bias and Its Removal. *IEEE Trans. Knowl. Data Eng.* **2019**, *31*, 437–450. [[CrossRef](#)]
31. Bickart, B.; Schindler, R.M. Internet forums as influential sources of consumer information. *J. Interact. Mark.* **2001**, *15*, 31–40. [[CrossRef](#)]
32. Resnick, P.; Zeckhauser, R. *Trust among Strangers in Internet Transactions: Empirical Analysis of eBay's Reputation System*; Baye, M.R., Ed.; The Economics of the Internet and Ecommerce; Emerald Group Publishing Limited: Bingley, UK, 2002; pp. 127–157.
33. Chevalier, J.A.; Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *J. Mark. Res.* **2006**, *43*, 345–354. [[CrossRef](#)]
34. Schneider, M.J.; Gupta, S. Forecasting sales of new and existing products using consumer reviews: A random projections approach. *Int. J. Forecast.* **2016**, *32*, 243–256. [[CrossRef](#)]
35. Mudambi, S.M.; Schuff, D. What makes a helpful review? A study of customer reviews on amazon.com. *MIS Q.* **2010**, *34*, 185–200. [[CrossRef](#)]
36. Lawani, A.; Reed, M.R.; Mark, T.; Zheng, Y. Reviews and price on online platforms: Evidence from sentiment analysis of Airbnb reviews in Boston. *Reg. Sci. Urban Econ.* **2019**, *75*, 22–34. [[CrossRef](#)]
37. Cheng, M.; Jin, X. What do Airbnb users care about? An analysis of online review comments. *Int. J. Hosp. Manag.* **2019**, *76*, 58–70. [[CrossRef](#)]
38. Sohail, S.S.; Siddiqui, J.; Ali, R. Feature extraction and analysis of online reviews for the recommendation of books using opinion mining technique. *Perspect. Sci.* **2016**, *8*, 754–756. [[CrossRef](#)]
39. Zhang, X.; Dellarocas, C. The lord of the ratings: Is a movie's fate influenced by reviews? In Proceedings of the 2006 ICIS, Milwaukee, WI, USA, 10–13 December 2006; pp. 1959–1978.
40. Reinstein, D.A.; Snyder, C.M. The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *J. Ind. Econ.* **2005**, *53*, 27–51. [[CrossRef](#)]
41. Lee, J.H.; Jung, S.H.; Park, J.H. The role of entropy of review text sentiments on online WOM and movie box office sales. *Electron. Commer. Res. Appl.* **2017**, *22*, 42–52. [[CrossRef](#)]
42. Kawaf, F.; Istanbulluoglu, D. Online fashion shopping paradox: The role of customer reviews and facebook marketing. *J. Retail. Consum. Serv.* **2019**, *48*, 144–153. [[CrossRef](#)]
43. Kim, S.G.; Kang, J. Analyzing the discriminative attributes of products using text mining focused on cosmetic reviews. *Inf. Process. Manag.* **2018**, *54*, 938–957. [[CrossRef](#)]
44. Xu, X.; Li, Y. The antecedents of customer satisfaction and dissatisfaction toward various types of hotels: A text mining approach. *Int. J. Hosp. Manag.* **2016**, *55*, 57–69. [[CrossRef](#)]

45. Hu, Y.; Chen, Y.; Chou, H. Opinion mining from online hotel reviews—A text summarization approach. *Inf. Process. Manag.* **2017**, *53*, 436–449. [[CrossRef](#)]
46. Geetha, M.; Singha, P.; Sinha, S. Relationship between customer sentiment and online customer ratings for hotels—An empirical analysis. *Tour. Manag.* **2017**, *61*, 43–54. [[CrossRef](#)]
47. Lee, P.; Hu, Y.; Lu, K. Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telemat. Inf.* **2018**, *35*, 436–445. [[CrossRef](#)]
48. Wang, Y.; Lu, X.; Tan, Y. Impact of product attributes on customer satisfaction: An analysis of online reviews for washing machines. *Electron. Commer. Res. Appl.* **2018**, *29*, 1–11. [[CrossRef](#)]
49. Oza, K.S.; Naik, P.G. Prediction of Online Lectures Popularity: A Text Mining Approach. *Procedia Comput. Sci.* **2016**, *92*, 468–474. [[CrossRef](#)]
50. Nakayama, M.; Wan, Y. The cultural impact on social commerce: A sentiment analysis on Yelp ethnic restaurant reviews. *Inf. Manag.* **2019**, *56*, 271–279. [[CrossRef](#)]
51. Gao, S.; Tang, O.; Wang, H.; Yin, P. Identifying competitors through comparative relation mining of online reviews in the restaurant industry. *Int. J. Hosp. Manag.* **2018**, *71*, 19–32. [[CrossRef](#)]
52. Korfiatis, N.; Stamolampros, P.; Kourouthanassis, P.; Sagiadinos, V. Measuring service quality from unstructured data: A topic modeling application on airline passengers’ online reviews. *Expert Syst. Appl.* **2019**, *116*, 472–486. [[CrossRef](#)]
53. Kulkarni, G.; Ratchford, B.T.; Kannan, P.K. The Impact of Online and Offline Information Sources on Automobile Choice Behavior. *J. Interact. Mark.* **2012**, *26*, 167–175. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).