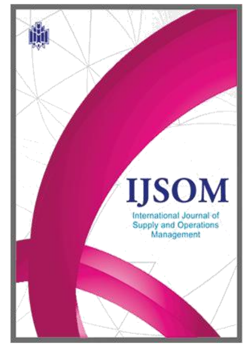
International Journal of Supply and Operations Management

IJSOM

2025, Volume 12, Issue 1, pp. 48-63
ISSN-Print: 2383-1359
ISSN-Online: 2383-2525
www.ijsom.com



Integrating Machine Learning and Text Mining to Enhance Customer Value Propositions in Hotel Supply Chain

Jie-Shin Lin [a] and Chih-Hao Tsai [b*]

[a] *Department of Public Policy and Management, I-Shou University, Kaohsiung, Taiwan*
[b] *Department of Hospitality, MICE Marketing Management, National Kaohsiung University of Hospitality and Tourism, Kaohsiung, Taiwan*

Abstract

The accelerated digital transformation in the contemporary business landscape, propelled by the Fourth Industrial Revolution, has fundamentally reshaped marketing research practices. This study leverages machine learning techniques and big data analytics to extract critical customer value propositions from extensive online reviews, aligning with predictive marketing strategies. Using a hybrid approach that combines qualitative and quantitative analyses, the research examines 8,290 customer reviews sourced from an online platform within the tourism industry. Two advanced analytical techniques were applied: clustering analysis to identify 20 distinct value components prioritized by tourists and associative rule mining to uncover seven essential patterns embedded in customer feedback. The results highlight the potential of big data and machine learning in accelerating marketing research processes, improving precision, and lowering operational costs. The findings emphasize the transformative role of digital tools in modern marketing practices, enabling businesses to enhance customer satisfaction, optimize services, and maintain competitive advantages in a data-driven economy.

Keywords: Big Data Analytics; Marketing Research; Industry 4.0; Machine Learning; Value Creation.

1. Introduction

The business environment is undergoing profound and fundamental transformations, with many business and economics experts recognizing these changes as indicative of a new era: the Fourth Industrial Revolution (Maar, 2016; Gupta et al., 2017).

Today, companies operating in a competitive business environment and grappling with big data face challenges such as making rapid decisions to enhance productivity. Many production systems are not equipped to manage big data due to a lack of intelligent analytical tools (Lee et al., 2014). The Fourth Industrial Revolution describes a world where communication technology bridges digital domains and offline reality, enabling people to live in and manage this interconnected environment (Shu et al., 2018).

The First Industrial Revolution fundamentally altered life and the economy, transitioning from an agrarian economy to one dominated by industry and machinery. The Second Industrial Revolution saw oil and electricity facilitating

*Corresponding author email address: billy.tsai88@gmail.com
DOI: 10.22034/ijsom.2025.110543.3212

mass production. The Third Industrial Revolution introduced information technology for production automation. Finally, the Fourth Industrial Revolution focuses on enhancing human cognitive production capabilities, a power that has been significantly bolstered (Shu et al., 2018). Given the various definitions and scientific discussions surrounding the first three industrial revolutions, it is evident that the Fourth Revolution, also known as the "Digital Revolution," commenced at the beginning of this century. This era is characterized by ubiquitous, mobile internet, smaller and more powerful sensors, artificial intelligence, and machine learning.

The rapid advancements driven by the Fourth Industrial Revolution are transforming industries, redistributing wealth, and reshaping knowledge systems. These changes are occurring at an unprecedented pace, requiring businesses and societies to remain vigilant and adaptive. Understanding and harnessing these advancements is critical for leveraging their full potential and ensuring sustainable growth and innovation (Shu et al., 2018).

The exponential growth of the Internet and social media platforms has revolutionized communication, collaboration, and interaction, leading to an unprecedented surge in user-generated content. This vast data pool is valuable for consumers and businesses, offering insights into customer preferences, behaviors, and expectations (Duan et al., 2013). Social media has transitioned communication from traditional one-way channels to dynamic, interactive dialogues, enabling real-time engagement and feedback (Halwani et al., 2019).

Given this data-driven landscape shaped by the Fourth Industrial Revolution, the current research focuses on identifying and analyzing customer value components through advanced big data techniques. The tourism industry is selected as the case study, leveraging online reviews to reflect travelers' experiences and satisfaction levels. These reviews provide a rich source of insights shaped by travelers' preferences, perceptions, and overall experiences at their destinations (Sanchez et al., 2019).

To conduct this research, data preprocessing operations are performed in the initial step to prepare the data. The second step analyses the data using the "data clustering" method to extract the value components from the reviews. Finally, in the third stage, the "associative rules" extraction method is employed to uncover hidden knowledge from the set of expressions and words used in tourists' opinions.

A review of internal and external studies reveals that, given the topic's novelty in terms of opportunities and threats in marketing research—impacted by the Fourth Industrial Revolution—researchers are at the nascent stages, and there is considerable scope for further investigation in this area. This issue has been particularly neglected domestically, with no independent research specifically focusing on the use of big data in Industry 4.0. The identified gaps underscore the importance and necessity of this study, which can pave the way for future research addressing these challenges.

Overall, this research centers on utilizing big data with machine learning and artificial intelligence (AI) and their impact on market research and marketing. Therefore, the main contributions of this research are as follows:

1. Development of a novel framework integrating machine learning and text mining for marketing research.

2. Demonstration of how big data analytics can reduce costs and improve the accuracy of market research.

3. Provision of actionable insights for enhancing customer satisfaction and loyalty in the tourism industry.

2. Theoretical Foundations and Literature Review

2.1. Theoretical Background

2.1.1. Value Creation

Each business creates and provides "value" to customers according to its business model. To survive and grow, a business must convince customers to pay for the perceived "value" it offers. In recent years, the study of the concept of "value" has become a focal point for managers and researchers in the field of management. The emergence of new tools for customer communication, including social networks, has introduced both challenges and opportunities for organizations to introduce and provide value to their stakeholders (Labrecque, 2014).

The concept of value has been defined in various contexts with differing definitions. From the customer's perspective, value is created when the benefits obtained from consuming a product or service exceed the costs incurred (Zhang et al., 2019). In today's dynamic market, customers expect organizations to offer the most value at the most appropriate price, and businesses are continually seeking new ways to create and deliver this "value" (La and Condompoli, 2004). To gain a competitive advantage, businesses must shift their focus from merely selling goods or services to creating value for customers (Kahns and Ingwald, 2016).

A review of existing research indicates that the Customer Value Proposition (CVP) concept is crucial in determining how value is delivered to customers. However, this concept remains poorly understood (Payne et al., 2017).

2.1.2. Fourth Industrial Revolution

The Fourth Industrial Revolution represents a transformative era characterized by the convergence of digitalization, information and communication technologies, robotics, machine learning, and artificial intelligence. These technologies collectively enhance decision-making capabilities, enabling a shift from human-driven processes to machine-based automation (Siam and Sharma, 2018). The impact of these advancements extends across industries, particularly marketing research and sales management, where data-driven strategies have become central to competitiveness.

Key elements shaping business concepts under Industry 4.0 include:

Enhanced Customer Expectations: Consumers now demand personalized, efficient, and enhanced services tailored to their specific needs.

Digital Integration: Embedding digital technologies into products and services elevates customer value and satisfaction.

Collaborative Innovation: Leveraging data-driven insights to improve customer experiences, streamline operations, and foster innovation.

Platform Economy: Emerging global platforms necessitate changes in talent management, organizational structures, and cultural paradigms to accommodate new business models (Schwab, 2015).

Figure 1 illustrates the framework of Industry 4.0, highlighting its key components and their interdependencies in transforming modern businesses.

Figure 1. Framework of Industry 4.0 (Qomariyah & Priandoyo, 2020)

2.1.3. Big Data

Organizations today gather and store vast amounts of data, anticipating that it will yield valuable insights in the future. This abundance of data introduces challenges, particularly in selecting the appropriate data and extracting actionable knowledge to support managerial decision-making (Amado et al., 2018). The concept of "big data" refers to datasets with high velocity, volume, and variety, which exceed the processing capabilities of traditional methods (Elgendy & Elragal, 2014). Big data encompasses structured data, like traditional databases, and unstructured data, such as user-generated content on social media and mobile applications, significantly influencing customer decisions and brand development (Lansley & Longley, 2016; Moro et al., 2016).

Big data analysis enhances the decision-making process for marketers by providing precise market insights. It helps address key questions, including identifying the most suitable product for a market, determining optimal advertising strategies, selecting communication channels, and devising effective sales promotions (Bendle & Wang, 2016). Given these advantages, marketing has been one of the first fields to adopt and leverage big data, highlighting its pivotal role in modern business strategies.

2.1.4. Machine Learning

Machine learning is utilized in data mining to build predictive models and analyze big data. The term "machine learning," coined by Arthur Lee Samuel in 1959, refers to the use of computers to respond to situations without being explicitly programmed. Samuel, a pioneer in artificial intelligence, envisioned machine learning as a promising approach for artificial intelligence to reach human-level intelligence (Siam and Sharma, 2018). In essence, machine learning involves developing and discovering methods and algorithms that enable computers to learn.

Machine learning is generally divided into two main categories: supervised learning and unsupervised learning. Supervised learning, or "supervised statistical learning," involves creating a statistical model to predict or estimate an output based on one or more inputs. In contrast, unsupervised learning involves inputs without predetermined or limited outputs (James et al., 2013). In supervised learning, the data comes with labels, indicating which class each data point belongs to, thus providing supervision. In unsupervised learning, the data lacks labels, meaning there is no supervision of the data (Rasheka and Mirjalali, 2019).

Compared to traditional statistical methods, machine learning methods often perform better in forecasting because they can leverage non-linear and complex relationships to extract the values of output variables from input variables. However, a drawback of machine learning methods is the difficulty in interpreting the results compared to traditional models (Hastie et al., 2017).

One sub-branch of machine learning is text mining, an extension of data mining aimed at extracting meaningful patterns from textual documents. Text mining of social media pages helps businesses better understand customers. Businesses use text mining of customer opinions to predict sales trends, manage future customer relationships, gather competitor intelligence, assess brand performance, analyze customer sentiments and opinions, and make informed, knowledge-based decisions (Yi Liao and Pei Tan, 2014).

2.2. Literature Review

Currently, many countries emphasize the development of the tourism industry as a key driver of economic growth. In Iran, the tourism industry enhances economic prospects, leading to an increase in GDP per capita. Additionally, the development of tourism-related industries amplifies the positive impact of tourism on the country's economic growth.

Big data has various applications, such as modifying products, selecting distribution channels, implementing electronic marketing strategies, and creating content. Research findings indicate that the use of big data optimizes and personalizes marketing strategies, thereby improving the marketing and sales process. This leads to increased customer satisfaction and loyalty, ultimately balancing organizational products with customer needs and yielding greater profits and value for both businesses and customers.

Utilizing the potential capabilities of information systems is a critical success factor for competition in the hospitality industry (Spark et al., 2016; Ragiso et al., 2017). Given the significant investment required in the tourism industry, it is essential to investigate and analyze the characteristics that travellers describe or recreate in their minds about the services they receive and share in their online comments (Sanchez et al., 2019). The advent of social media, Web 2.0, and digital channels has led to the proliferation of online recommendations, reviews, and customer feedback, significantly influencing marketing through word-of-mouth (WoM). This phenomenon has given rise to electronic word-of-mouth marketing (eWoM), a new form of WoM that occurs via the Internet, electronic devices, and social networks (Leung et al., 2019).

Research by Chong et al. (2018) underscores the importance of opinions and comments discussed online, showing that most travellers consider previous travelers' opinions as crucial sources of information for decision-making. Additionally, this research highlights the importance and utility of analyzing this data for predictive marketing. Gil Seto et al. (2019) analyzed customer perceptions regarding environmental protection methods using online reviews. Their results show that customers recognize hotel actions in six areas: 1) energy, 2) water, 3) shopping, 4) waste, 5) site and training, and 6) innovation. However, customers do not recognize hotels' overall level of environmental commitment.

Sanchez et al. (2019) used supervised machine learning to analyze customer opinions about hotels and introduced solutions to enhance service quality and guest satisfaction. They extracted opinions from 33 hotel customers on Yelp (a popular social network in the United States) and presented a model to assist hotel managers in improving customer service quality.

Ren (2024) explored strategies for enhancing collaboration between leading online travel agencies and hotels, focusing on various partnership models, including price-based agreements, cost-sharing arrangements, and revenue-sharing frameworks. The study revealed that both cost-sharing and revenue-sharing approaches outperform price-only agreements in motivating hotels to adopt more environmentally sustainable practices. Meanwhile, Pertusa-Ortega et al. (2025) investigated how organizational agility—encompassing customer, partner, and operational agility—mediates the relationship between quality management and performance within the hospitality sector. Using survey data collected from multiple hotels in Spain, their research identified partial mediation, indicating that quality management indirectly boosts hotel performance by strengthening partner and operational agility.

2.3. Contributions of this research

This study makes significant contributions to the application of machine learning and text mining in tourism marketing within the framework of the Fourth Industrial Revolution. By employing advanced analytical techniques to analyze large datasets of customer reviews, it enhances the understanding of customer value propositions in the hospitality industry. The research introduces an innovative methodological framework that improves the accuracy and depth of market research, offering valuable insights into drivers of customer satisfaction and feedback patterns.

From a practical perspective, the findings provide hotel managers and marketers with actionable strategies to enhance service quality and strengthen customer engagement. The study demonstrates the strategic role of big data in uncovering key factors that shape customer perceptions and loyalty, enabling the development of targeted marketing initiatives aligned with consumer expectations. Moreover, this research contributes to academic and policy discussions on leveraging digital transformation to promote sustainable tourism practices. It highlights the importance of integrating digital technologies to boost operational efficiency and elevate customer service standards in the hospitality sector.

3. Methodology

The current research is a descriptive-survey in terms of practical purpose and employs a mixed-methods (qualitative-quantitative) approach with an inductive methodology. Methods and algorithms for analyzing large textual data sets have been utilized to examine the content of customer comments. The data set consists of 8,290 comments recorded by customers on the internet, specifically in the hotel supply chain, and is in the Persian language. RapidMiner software and the Python programming language were used to conduct data mining, text mining, and big data analysis. The research was implemented in three main stages, as detailed below.

3.1. Data Preprocessing

In text mining, the input data is predominantly unstructured text documents; therefore, preprocessing is required. Initially, the text must be converted into discrete elements such as words, symbols, or other meaningful units, referred to as "tokens" (Wise et al., 2015). This process involves removing punctuation marks and other non-textual characters to transform the data into a sequence of words. Figure 2 illustrates the "Word Cloud Distribution" of tourists' comments about the hotel following the data preprocessing phase.

Figure 2. Word cloud representation of comments in textual data

3.2. Clustering Data

In the next phase, the clustering method extracts "value" components from large customer comments. Clustering involves dividing the data into groups, or clusters, where the items within each group share similarities while being distinct from items in other groups. Each cluster contains similar objects that are dissimilar to those in other clusters (Raschka and Mir Jalali, 2019). Figure 3 depicts the process of data clustering.

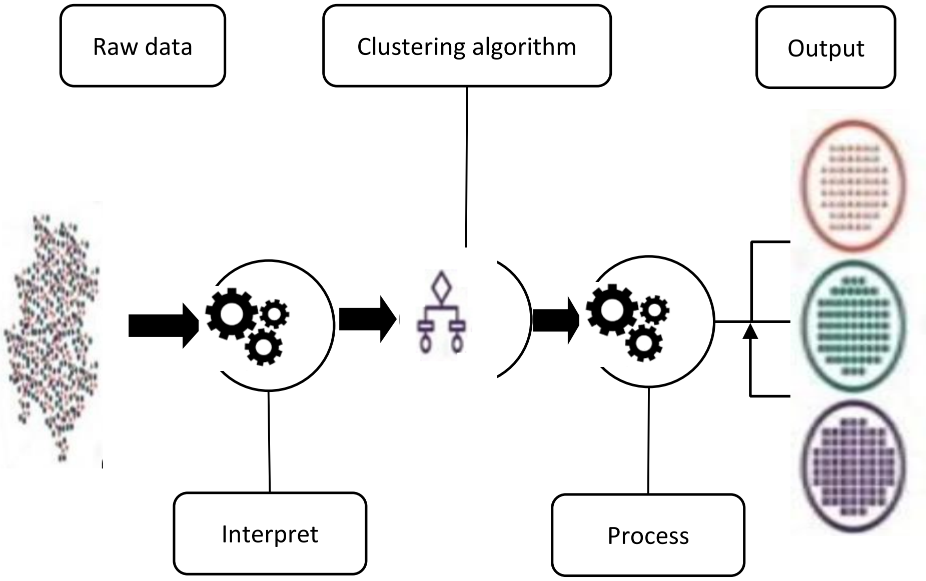


Figure 3. Schematic display of clustering (separation) of topics of text documents

In text mining, clustering is a technique used to group similar documents by topic. This method becomes essential when dealing with a large volume of text documents, where manually categorizing them based on their topics would be impractical or excessively time-consuming. Among various clustering methods, the "k-means" algorithm is widely accepted. This method focuses on the similarity of text documents and the relationships between keywords (Weiz et al., 2015). Figure 4 demonstrates how RapidMiner performs data clustering in this research.

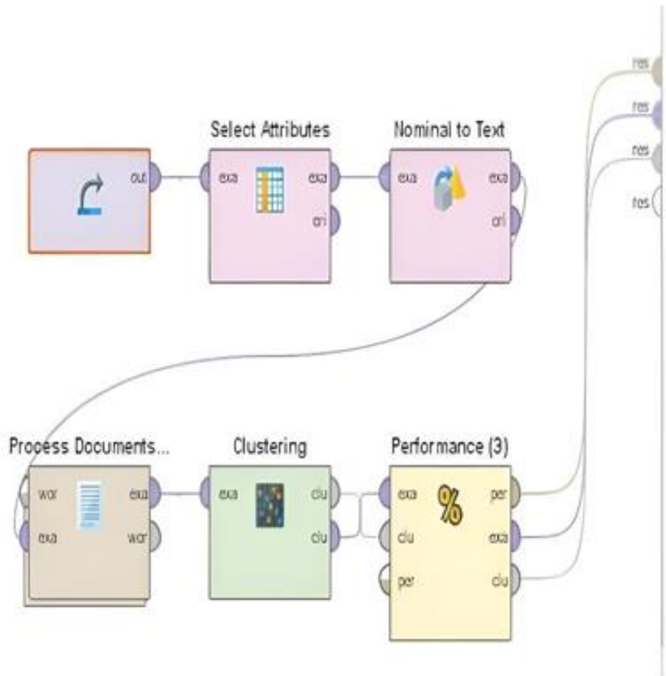


Figure 4. Clustering of comments by RapidMiner

In this research, the number of clusters and the validation of clustering are determined using the "Dan index" method, which has demonstrated superior performance compared to other methods in big data analysis (Kasambara, 2017). The Dan index evaluates clustering based on two criteria: compression, which measures the distance between members within a cluster, and separability, which assesses the distance between clusters. The index uses two metrics, "distance" and "Qatar," to calculate the levels of compression and resolution. DI index is defined by Equation (1), where $d(C_i, C_j)$ represents the distance between clusters i and j, and $silent(C_j)$ signifies the representative of the cluster C_j.

$$DI = \left\{ \left\{ \frac{d(C_i, C_j)}{silent\ (C_i)} \right\} \right\} \tag{1}$$

3.3. Association Rules

Association rules are a data mining technique used to uncover frequent patterns, correlations, and causal relationships within large datasets, aiding in analyzing and predicting user behavior (Shalini & Lal, 2016). By identifying connections between different data elements, these rules reveal recurring conditions within a dataset and describe the likelihood of certain items (entities) being present based on the presence of others. In the context of text mining, the application of association rules focuses on discovering relationships between words, enabling the identification of patterns that suggest the co-occurrence or conditional presence of specific terms within a text.

In a database if $I = \{I_1, I_2, \ldots, I_m\}$ is a set of items and D is a set of transaction databases where each transaction has an identifier as well as a set of items such as $T \subseteq I$ Is.

If A is a set of items, then a transaction such as T contains A if we have $A \subseteq T$ An association law in the form $A \to B$ It is stated that in it $A \subset I$، $B \subset I$ ، $A \neq \varphi$، $B \neq \delta$، $A \cap B = d$ (Ibid et al., 2011) An associative law states in its general form that if event A occurs, then event B will also occur. From the symbol $A \to B$ It is used to show an associative rule; A is called the premise of the law and B is the conclusion of the law. The source of the generation of associative rules are those times that have been repeated enough that they are called "set of repeated items" (Nasser et al., 2017). $A \to B$ It has an index called "support" that indicates how often this has been repeated. The amount of legal support $A \to B$ It is calculated using equation 2. The fraction's numerator is equal to the number of simultaneities of A and B, and the fraction's denominator refers to the total number of occurrences.

$$Support\ (A \to B) = N(A\&B)/N \tag{2}$$

After the rules with a minimal amount of support are generated, the probability of the rules being correct can be calculated. This index, which is called "confidence value", indicates how many per cent of the times that event A has occurred, event B has also occurred. The certainty value of the law $A \to B$ is calculated with the help of equation 3 (Li et al., 2017).

$$onfidence\ (A \to B) = N(A\&B)/N(A) \tag{3}$$

The Apriori and FP-Growth algorithms are two of the most prominent and practical methods for extracting association rules. In this study, the FP-Growth algorithm was chosen for extracting association rules after a thorough evaluation and comparison of the efficiency and speed of both methods. The process of extracting these association rules using RapidMiner is depicted in Figure 5.

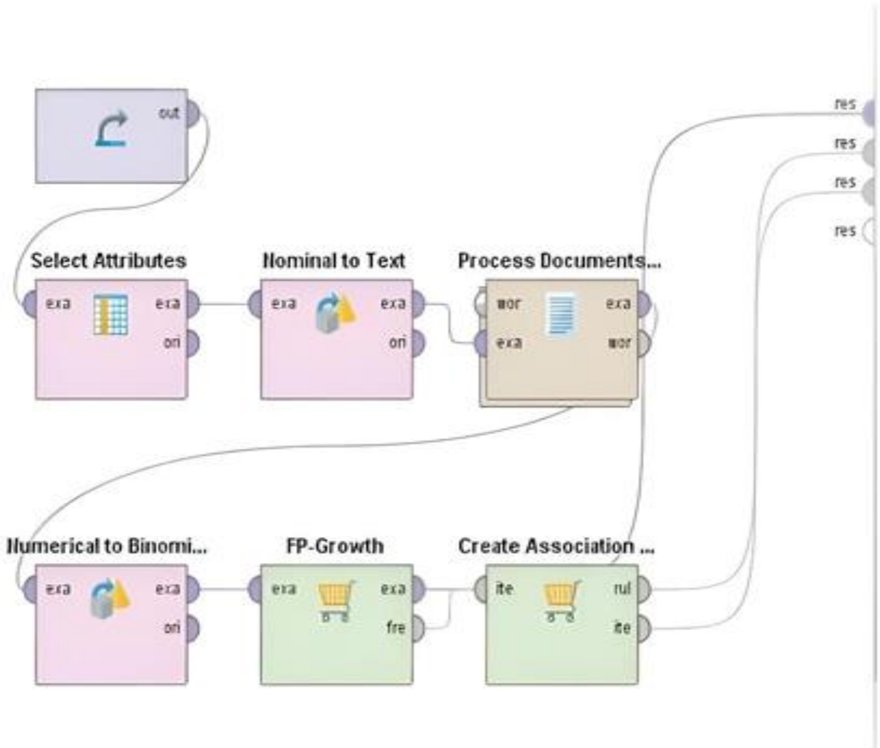


Figure 5. Mining community rules by RapidMiner

4. Research Findings

The findings derived from text-mining customer opinions regarding the tourism industry are presented in "Opinion Clustering" and "Association Rules Extraction."

4.1. Findings of the Clustering Process

In this section, the components of customer expected value are extracted by clustering the collected opinions using the k-means algorithm. The most frequent words within that cluster are analyzed to identify the topics or features mentioned in each cluster. Each cluster is named based on the logical relationships among the most recurrent words within each category, reflecting the prevalent topics (Guo et al., 2017).

The results of clustering tourists' comments, along with the eight most frequent words in each cluster, are presented in Table 1 and Figure 6. These frequent words help elucidate the topics or features that customers consider significant within each cluster. Applying the clustering algorithm to tourists' comments revealed 20 distinct clusters. Examining each cluster and its most frequently occurring words, the predominant topic for each cluster is identified. Of the 20 topics or features extracted, the top 11 pertain to positive value components, while the remaining nine refer to negative value components from the tourists' perspectives.

Table 1. Topics extracted from comment clustering and eight words with higher probability

Clusters	Words with the highest frequency (probability) in each cluster								
Cluster 1	Other people's comments about the hotel	Comments	Friends	site	Hotel	when	Selection	I read	Star
Cluster 2	Overall cleanliness and hygiene of the hotel	room	Hotel	service	but	Breakfast	clean	sanitary	Restaurant
Cluster 3	Personnel encounter	Hotel	Thank	Mr	Lady	Personnel	Collision	Management	Good
Cluster 4	Proximity to shopping and entertainment centers	Hotel	on foot	Market	Beach	Center	minutes	Close	Buy
Cluster 5	Adequate bathroom and toilet	Bathroom	service	Turn it around	clean	sound	sanitary	WC	Good
Cluster 6	Room cleanliness (bedding, table cloth, etc.)	room	Good	clean	To	the bed	white	bedding	Table
Cluster 7	The beauty of space and interior layout	Hotel	Excellent	Beautiful	room	Restaurant	space	Very	stylish
Cluster 8	Regional facilities for recreation	Park	we went	USD	Restaurant	Ship	attack	has	cult
Cluster 9	Quality and variety of breakfast	Cheese	Good	the egg	jam	Butter	Breakfast	lens	Sausage
Cluster 10	Transfer service and room delivery	room	the watch	Hotel	To	Delivery	transfer	we became	Airport
Cluster 11	Additional facilities	room	Phone	Refrigerator	Tea	enough	Internet	Speed	Coffee
Cluster 12	Expensive entertainment outside the hotel	USD	Park	Price	Ship	Restaurant	Ticket	cult	Expensive
Cluster 13	power outage	room	Electricity	the rooms	the darkness	to cut	Hello	Area	Electricity

Clusters	Words with the highest frequency (probability) in each cluster								
Cluster 14	The star of the hotel does not fit with its quality	Hotel	room	Star	Collision	Quality	service	dirty	stay
Cluster 15	flight delay	the watch	late	USD	Delay	we became	Airport	Delivery	Opportunity
Cluster 16	The presence of annoying insects in the room	mother in law	peace of mind	Morning	Insect	annoying	more important	Of all	boil
Cluster 17	Low quality sanitary ware	Turn it around	Quality	soap	Toothbrush	shampoo	Bad	the bell	person
Cluster 18	Entering annoying noise from outside	room	sound	Floor	the outside	Tall	window	Voice	annoying
Cluster 19	Not cleaning the room and...	room	dirty	service	Bathroom	WC	damaged	Toilet	Turn it around
Cluster 20	expensive	USD	Cost	agency	room	Flight	Net	Money	Expensive

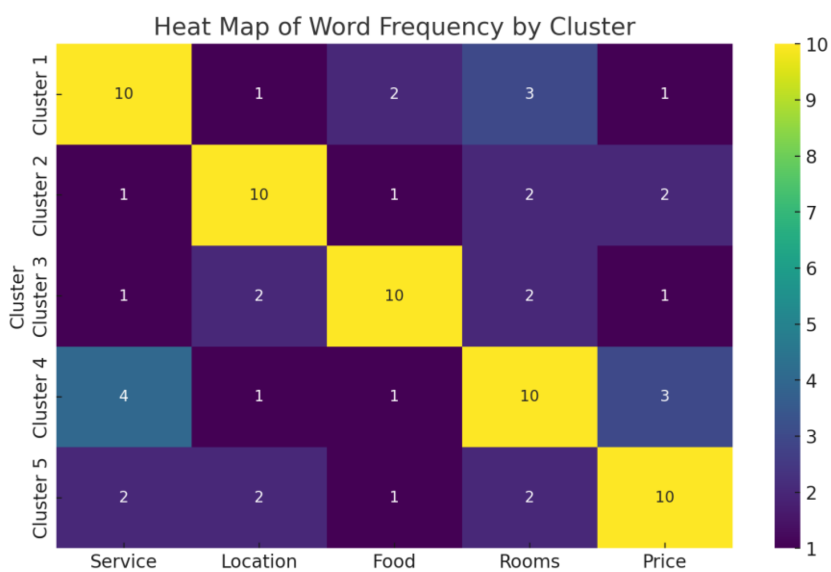


Figure 6. Heat Map of Word Frequency by Cluster

4.2. Findings of the Associative Rules Process

Identifying association rules aims to reveal relationships among words within a vast dataset of customer comments. This technique involves uncovering hidden insights by analyzing words or phrases that frequently appear together in customer feedback. Table 2 illustrates the extracted association rules, accompanied by two critical metrics: support level and confidence level.

Table 2. The results of extracting the associative rules

Introduction (law)	Result (law)	Confidence level	Backup amount	Law
Collision	Accommodation, hotel	0.9650	0.3901	[hotel accommodation] → [Collision]
service	Accommodation, hotel	0.9679	0.2221	[hotel accommodation] → [service]
room	Accommodation, hotel	0.9686	0.5174	[hotel accommodation] → [room]
Restaurant	Accommodation, hotel	0.9689	0.3008	[hotel accommodation] → [Restaurant]
Breakfast	Accommodation, hotel	0.9734	0.4023	[hotel accommodation] → [Breakfast]
clean	Accommodation, hotel	0.9746	0.2317	[hotel accommodation] → [clean]
Star	Accommodation, hotel	0.9847	0.2726	[hotel accommodation] → [Star]

Associative rules are formulated in the "if → then" format. This method reveals hidden relationships and dependencies within large data sets, uncovering significant and useful patterns. By providing valuable insights, associative rules facilitate the decision-making process for managers, enabling them to make informed decisions based on the discovered relationships in the data.

4.3. Sentiment Analysis Results

To effectively present the results of sentiment analysis, Table 3 is provided.

Table 3. Distribution of Sentiments in Customer Reviews

Sentiment Category	Percentage	Number of Reviews
Positive	63%	5,223
Neutral	21%	1,741
Negative	16%	1,326
Total	100%	8,290

Table 3 provides a clear breakdown of sentiments across the reviews collected. It quantifies how many reviews fall into each sentiment category, offering a straightforward view of the overall sentiment landscape.

Figure 7 illustrates the trend of sentiments over the review period. It provides a visual representation of how customer sentiments—positive, neutral, and negative—fluctuated over time. This analysis helps identify potential shifts in customer satisfaction and service quality, offering insights into periods of improvement or decline. Such patterns are invaluable for businesses to assess the impact of changes in service strategies, marketing campaigns, or operational adjustments. Moreover, the trend analysis highlights recurring seasonal patterns or external influences affecting customer perceptions, enabling businesses to proactively address issues and refine their approaches to enhance customer satisfaction.

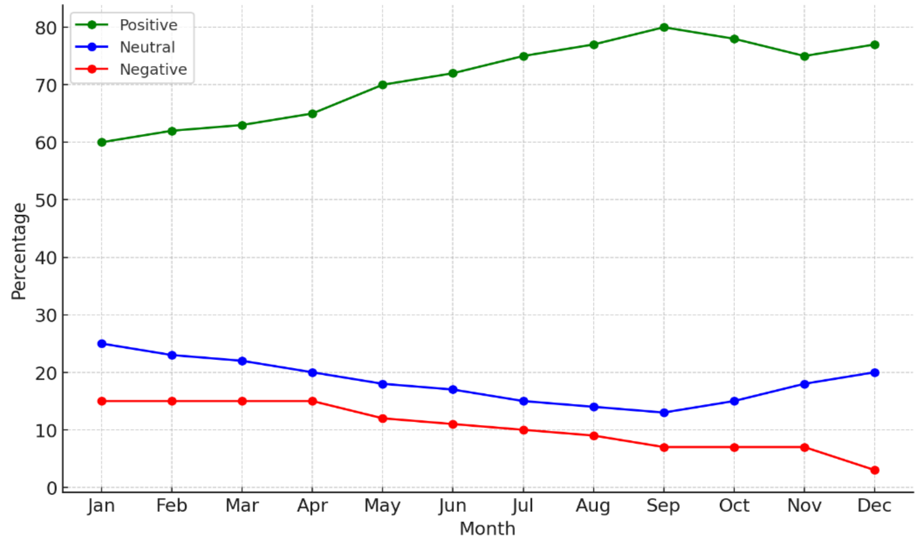


Figure 7. Sentiment Trend Across Review Period

Table 4. Detailed Sentiment Analysis by Review Category

Review Category	Positive Reviews	Neutral Reviews	Negative Reviews	Total Reviews
Room Amenities	70%	20%	10%	2,000
Staff Service	65%	25%	10%	1,800
Food & Dining	60%	30%	10%	1,500
Location	75%	15%	10%	2,990

Table 4 segments the sentiments by specific categories or aspects of the hotel experience. This breakdown helps identify strengths and areas for improvement within different facets of the service offered.

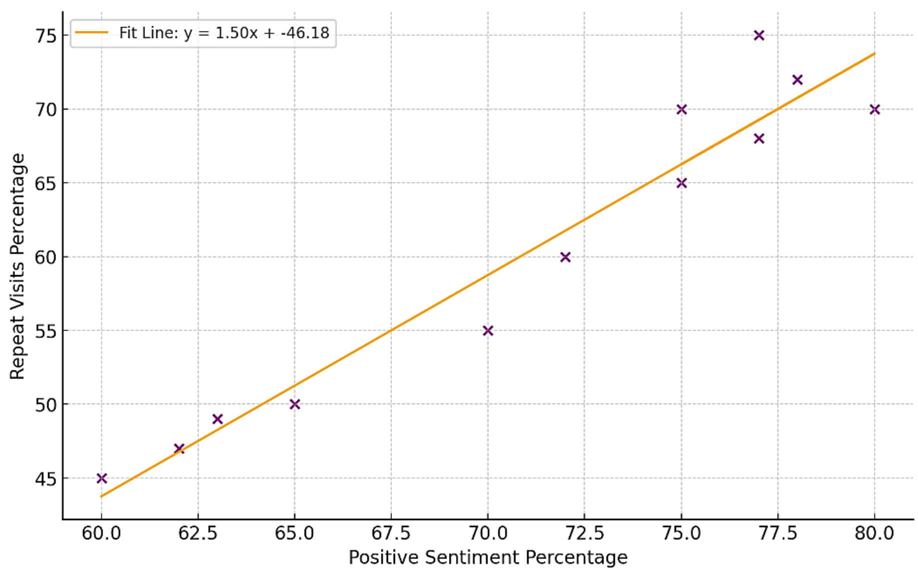


Figure 8. Correlation Between Sentiment and Repeat Visits

Figure 8 explores the relationship between overall sentiment and the likelihood of repeat visits. A scatter plot or bar graph can effectively show how positive sentiments correlate with returning customers, which is crucial for understanding the impact of customer emotions on loyalty. By illustrating the distribution, trends, and detailed categories of sentiments, these visual aids help convey complex data in an accessible and interpretable format, providing clear evidence to support the research findings.

5. Discussion and Conclusion

Businesses operate based on defined models that outline how they create and deliver value to customers. At the core of these models is the value proposition, representing the goods or services offered to meet customer needs (Sheehan and Bruni Bossio, 2015). Effectively designing and delivering this value is critical to achieving competitiveness and profitability.

In response to the opportunities introduced by the Fourth Industrial Revolution, this study develops a framework leveraging big data analytics to refine value propositions. The framework employs machine learning techniques, including clustering and associative rule extraction, to analyze customer reviews and identify key value components perceived by tourists.

The findings reveal 20 distinct value components categorized into 11 positive aspects (e.g., cleanliness, staff behavior) and 9 negative aspects (e.g., high costs, poor room quality). Moreover, seven critical factors influencing tourist satisfaction were identified, including staff attitude, service quality, and cleanliness. These insights empower businesses to enhance service quality, meet customer expectations, and strengthen their competitive positioning.

The results align with prior studies (Sanchez et al., 2019; Chong et al., 2018) emphasizing the importance of customer reviews in influencing decision-making. Unlike Gil Seto et al. (2019), who highlighted gaps in environmental awareness among customers, this study found no such issue in the Iranian context.

From a practical perspective, this research underscores the importance of leveraging big data to validate business strategies and improve service quality. Analyzing social media data helps businesses identify weaknesses and focus on areas with the greatest impact on customer satisfaction.

Future studies can extend this research by examining customer value in other sectors and integrating traditional methods with machine learning techniques to enhance accuracy. Limitations such as unsupervised clustering methods and the scarcity of Persian-language reviews may be addressed through hybrid approaches and broader sampling methods to improve data reliability and generalizability.

References

Amado, A., Cortez, P., Rita, P., & Moro, S. (2018). Research trends on Big Data in marketing: A text mining and topic modeling based literature analysis. *European Research on Management and Business Economics, 24*(1), 1-7.

Bendle, N. T., & Wang, X. S. (2016). Uncovering the message from the mess of big data. *Business Horizons, 59*(1), 115-124.

Chong, A. Y. L., Khong, K. W., Ma, T., McCabe, S., & Wang, Y. (2018). Analyzing key influences of tourists' acceptance of online reviews in travel decisions. *Internet Research, 28*(3), 564-585.

Duan, W., Cao, Q., Yu, Y., & Levy, S. (2013). Mining online user-generated content: Using sentiment analysis technique to study hotel service quality. In *2013 46th Hawaii International Conference on System Sciences* (pp. 3119–3128). IEEE.

Elgendy, N., & Elragal, A. (2014). Big data analytics: A literature review paper. In P. Perner (Ed.), *Advances in data mining: Applications and theoretical aspects* (Vol. 8557, pp. 214–227). Springer.

Gil-Soto, E., Armas-Cruz, Y., Morini-Marrero, S., & Ramos-Henríquez, J. M. (2019). Hotel guests' perceptions of environmentally friendly practices in social media. *International Journal of Hospitality Management, 78*, 59-67.

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.

Gupta, S., Keen, M., Shah, A., Verdier, G., & Walutowy, M. F. (Eds.). (2017). *Digital revolutions in public finance*. International Monetary Fund.

Halawani, F. M., Soh, P. C., & Muthaiyah, S. (2019). The effect of social media on hotels' business performance in the Lebanese hotel sector. *Journal of Electronic Commerce in Organizations, 17*(3), 54-70.

Han, J., Pei, J., & Kamber, M. (2011). *Data mining: Concepts and Techniques*. Elsevier.

Hastie, T., Tibshirani, R., & Friedman, J. (2017). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2017). An introduction to statistical learning. Springer.

Kans, M., & Ingwald, A. (2016). Business model development towards service management 4.0. *Procedia CIRP, 47*, 489-494.

Kassambara, A. (2017). *Practical guide to cluster analysis in R: Unsupervised machine learning* (Vol. 1). STHDA.

La, K. V., & Kandampully, J. (2004). Market oriented learning and customer value enhancement through service recovery management. *Managing Service Quality: An International Journal, 14*(5), 390-401.

Labrecque, L. I. (2014). Fostering consumer-brand relationships in social media environments: The role of parasocial interaction. *Journal of Interactive Marketing, 28*(2), 134-148.

Lansley, G., & Longley, P. (2016). Deriving age and gender from forenames for consumer analytics. *Journal of Retailing and Consumer Services, 30*, 271-278.

Lee, E. B., Kim, J., & Lee, S. G. (2017). Predicting customer churn in the mobile industry using data mining technology. *Industrial Management & Data Systems, 117*(1), 90-109.

Lee, J., Kao, H. A., & Yang, S. (2014). Service innovation and smart analytics for Industry 4.0 and big data environment. *Procedia CIRP, 16*, 3-8.

Leong, L. Y., Hew, T. S., Ooi, K. B., & Lin, B. (2019). Do electronic word-of-mouth and elaboration likelihood model influence hotel booking? *Journal of Computer Information Systems, 59*(2), 146-160.

Marr, B. (2016, April 5). Why everyone must get ready for the 4th industrial revolution. *Forbes*. Retrieved from https://www.forbes.com/sites/bernardmarr/2016/04/05/why-everyone-must-get-ready-for-4th-industrial revolution/#26be9e2f3f90

Moro, S., Rita, P., & Vala, B. (2016). Predicting social media performance metrics and evaluation of the impact on brand building: A data mining approach. *Journal of Business Research, 69*(9), 3341-3351.

Payne, A., Frow, P., & Eggert, A. (2017). The customer value proposition: Evolution, development, and application in marketing. *Journal of the Academy of Marketing Science, 45*(4), 467-489.

Qomariyah, N. N., & Priandoyo, A. (2020, February). Industry 4.0 strategic alignment framework: Multilevel perspective of digital transition in Indonesia. In 2020 International Conference on Smart Technology and Applications (ICoSTA) (pp. 1-6). IEEE.

Raguseo, E., Neirotti, P., & Paolucci, E. (2017). How small hotels can drive value their way in infomediation: The case of 'Italian hotels VS. OTAs and Trip Advisor'. *Information & Management, 54*(6), 745-756.

Raschka, S., & Mirjalili, V. (2019). *Python machine learning: Machine learning and deep learning with Python, scikit-learn, and TensorFlow 2*. Packt Publishing Ltd.

Ren, Y. (2024). Optimal format design in green tourism supply chain in the presence of a dominant online travel agency. *Heliyon, 10*(9).

Pertusa-Ortega, E. M., Tarí, J. J., Molina-Azorín, J. F., & Pereira- Pertusa-Ortega, E. M., Tarí, J. J., Molina-Azorín, J. F., & Pereira-Moliner, J. (2025). Agility as a mediator in the relationship between quality management and hotel performance. *Service Business, 19*(1), 1-30.

Sánchez-Franco, M. J., Navarro-García, A., & Rondán-Cataluña, F. J. (2019). A naive Bayes strategy for classifying customer satisfaction: A study based on online reviews of hospitality services. *Journal of Business Research*, *101*, 499-506.

Schwab, K. (2015). *The Fourth Industrial Revolution: What it means and how to respond*. SNAPSHOT, December 12.

Shalini, S., & Lal, K. (2016). Improved pseudo-association rules technique. In 2016 International Conference on Computing, Communication and Automation (ICCCA) (pp. 890-895). IEEE.

Sheehan, N. T., & Bruni-Bossio, V. (2015). Strategic value curve analysis: Diagnosing and improving customer value propositions. *Business Horizons, 58*(3), 317-324.

Sparks, B. A., So, K. K. F., & Bradley, G. L. (2016). Responding to negative online reviews: The effects of hotel responses on customer inferences of trust and concern. *Tourism Management*, *53*, 74-85.

Syam, N., & Sharma, A. (2018). Waiting for a sales renaissance in the fourth industrial revolution: Machine learning and artificial intelligence in sales research and practice. *Industrial Marketing Management, 69*, 135-146.

Weiss, S. M., Indurkhya, N., & Zhang, T. (2015). *Fundamentals of predictive text mining*. Springer.

Xu, M., David, J. M., & Kim, S. H. (2018). The fourth industrial revolution: Opportunities and challenges. *International Journal of Financial Research*, *9*(2), 90-95.

Yee Liau, B., & Pei Tan, P. (2014). Gaining customer knowledge in low-cost airlines through text mining. *Industrial Management & Data Systems, 114*(9), 1344-1359.

Zhang, T. C., Gu, H., & Jahromi, M. F. (2019). What makes the sharing economy successful? An empirical examination of competitive customer value propositions. *Computers in Human Behavior*, *95*, 275-283.