Analysis of Customer Rating Quality at Kangen Boutique Hotel Based On Customer Reviews by Using LDA

THESIS

Submitted to International Program Department of Industrial Engineering in Partial Fulfillment of the Requirement for the Degree of Sarjana Teknik Industri at Universitas Islam Indonesia

Arranged By :

Muhammad Faisal Zain

(15522129)

INTERNATIONAL PROGRAM

DEPARTMENT OF INDUSTRIAL ENGINEERING

FACULTY OF INDUSTRIAL TECHNOLOGY

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

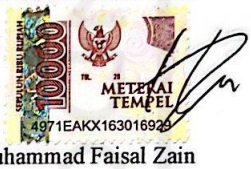
2022

AUTHENTICITY STATEMENT

AUTHENTICITY STATEMENT

In the name of Allah, I hereby confess that the research conducted is based on my works except for the cited references mentioned. If there is something wrong with my confession and dishonest resulting in academic misconduct or violation of papers regulation, then, i would likely to return my degree to Universitas Islam Indonesia.

Yogyakarta, December 2022



Muhammad Faisal Zain

ii

THESIS APPROVAL OF SUPERVISOR

THESIS APPROVAL OF SUPERVISOR

Analysis of Customer Rating Quality at Kangen Boutique Hotel Based On Customer Reviews by Using LDA

THESIS

Arranged by:

Name : Muhammad Faisal Zain
Student Number : 15522129

Yogyakarta, December 2022

Supervisor,

Muhammad Ridwan Andi Purnomo, S.T., M.Sc., P.hD.

iii

THESIS APPROVAL OF EXAMINATION COMMITTEE

THESIS APPROVAL OF EXAMINATION COMMITTEE

Analysis of Customer Rating Quality at Kangen Boutique Hotel Based On Customer Reviews by Using LDA

By

Name : Muhammad Faisal Zain

Student Number : 15522129

Was defended before Examination Committee in Partial Fulfillment of the requirements for the degree of Industrial Engineering Department

Universitas Islam Indonesia

Examination Committee

Muhammad Ridwan Andi Purnomo, S.T., M.Sc., P.hD.
Examination Committee Chair

Agus Mansur, S.T., M.Eng.Sc.
Member I

Dr. Drs. Imam Djati Widodo, M.Eng.Sc.
Member II

Acknowledged by,
Head of Department

(Muhammad Ridwan Andi Purnomo, S.T., M.Sc., P.hD)

iv

MOTTO

“Berbahagialah wahai para tersepelekan, karena dengan begitu kalian punya kesempatan besar untuk mengejutkan.”

- Farid Stevy

“If my mind can conceive it, and my heart can believe it, then I can achieve it”

- Muhammad Ali

“FORTIS FORTUNA ADIUVAT”

- Latin Proverb

“Victory Loves Preparation”

- Arthur Bishop

PREFACE

Assalamu'alaikum Wr. Wb.

Praise to Allah, The Most Grateful, and The Most Merciful, so that the author can finish the undergraduate thesis entitled Analysis of Customer Rating Quality at Kangen Boutique Hotel Based On Customer Reviews by Using Text Mining as the requirement to achieve the degree of Sarjana Teknik.

This undergraduate thesis would not have been completed properly without his input, assistance, and support from various parties. Therefore, the author would like to express gratitude and greetings to the following:

1. Mr. Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D. as the undergraduate thesis supervisor who gives guidance and assistance in the completion of the thesis.
2. My beloved parent and sisters who have been giving the support.
3. Mrs. Diana and Ms. Devi who have helped the author's administration.
4. Jeffry Suryadharma, Eky Fasich, and Rafif Elrienanto who have shared positive vibes and supported the author in finishing the study.
5. Ajeng Riva as my support system.
6. All Industrial Engineering IP batch 2015, thank you so much for the help, support, and friendship during the process.
7. All Industrial Engineering IP lecturers who have given useful knowledge and lessons.

The author believes that this undergraduate thesis contains imperfections that may be leaving wanders. Therefore, the author is looking forward to positive comments and constructive suggestions to cope with the thesis's shortcomings. Hopefully, the undergraduate thesis will be useful and beneficial for those parties who read this.

Wassalamua'alaikum Wr. Wb.

Yogyakarta, Desember 2022

Muhammad Faisal Zain

ABSTRACT

Customer satisfaction is always the mission of an organization that wants to stay on track to be competitive for a long period of time. The organization innovates their products and keeps developing them to satisfy the customer, moreover, in the era where online digital platforms have been growing rapidly. Google review is one online review platforms that shows many reviews from customers about a hotel. The reviews found online are very useful for a potential customer to make a consideration before the purchase. Based on this problem, it needs to conduct topic modeling that can process the number of reviews to obtain useful information for the organization's improvements. The object used in this study is Kangen Boutique Hotel. The process of topic modeling will be done using Latent Dirichlet Allocation (LDA), and the process of providing the improvement for the organization is using Kansei Engineering. The test results showed that there are 4 key factors that influence the rating qualities, which are, "Staff Service, "Room Condition", "Food", and "Room Facilities". Several causes of the factors above are the lack of help from the staff, limited choices of food, and some facilities which are not working properly. The organization should overcome these factors by training the staff with correct manners and procedures and also the maintenance of the hotel building to boost comfort during the customer's visits.

Keywords: online review, topic modeling, LDA, Kangen boutique hotel, Kansei engineering.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

1.1 Background

In this industry 4.0 era, industry revolutions have started playing vital roles. In accordance with the industrial revolution, technological development has been growing. It is forcibly pleaded the established company to invest in the latest or recent technology to keep on the competitive track. Many companies compete in the industrial competition in order to stay winning or be bold amongst the other company. The rise of competition level is a threat for the company that is not following the recent developments. In consequence, established company has anticipated the competition, and each of them got its own strategy to be capable of surviving the tough competition. The industrial revolution also shift society's mindset toward a product. It results in a high consumption level of the latest technology electronics. This reason increases the competitiveness in the market. It leads to points where the company also gives the best possible efforts to the honored customer, not only in the product but also services wise in order to maintain the relationship.

Customer satisfaction is always the mission for an organization that wants to stay on track to be competitive for a long period of time. The organization innovates their products and keeps developing them to be able to satisfy the customer. However, the products received by the customer are not always expected by the customer. Most of the time, the customer finds that the products are not fulfilling their expectations. Customer reviews of the organizationare really a concern to them. Customer reviews can be found everywhere on the internet, moreover, in the era where online digital platforms have been growing rapidly. The reviews found online are very useful for the potential customer to make a consideration before the purchase (Madhumita, 2022).

The present study in this research is going to talk about customer reviews on online digital platforms. Many customers use the digital platform to purchase and give reviews toward the organization, which in this case is a hotel. Using text mining approaches, the study

examines and analyses customer reviews and their overall satisfaction. The researcher uses customers' reviews to measure their satisfaction (Xun, 2020).

Text Mining is the process of discovering unknown knowledge through the automated extraction of information from a large amount of unstructured text. Text mining is applied in industry, the web, applications, and others. Application such as search engines, customer reviews, and product suggestions uses text mining for sentiment analysis (Aggarwal, 2012).

Derived from the background that has been stated above, further analysis needs to be taken toward customer reviews on Kangen Boutique Hotel on the online digital platform. This research is aimed to examine the factors that influence customer reviews on the organization in order to be the evaluation references for the organization to improve and develop their services or products. Therefore, the organization is capable of fulfilling the customer's expectations and earning customer overall satisfaction that results in beneficial feedback for the organization.

1.2 Problem Formulation

As the background has been delivered, the problem formulation of the research is as follows,

1. What factors influence the customers' rating on organization?
2. What improvements must be madeby the organization to enhance the customer experience?

1.3 Research Objectives

The objectives of the research are stated as follows,

1. Determining what kind of factors possibly cause the rating quality of customers towards the organization's product or services.
2. Determining the right actions for improvement to enhance the customer experience.

1.4 Benefits of Research

The expected benefit of this research to some involved parties are:

1. For the author

   The researcher earns the experience and implements the theories that have been studied during the time at the university. Also, the study fulfills the obligation of the author as an undergraduate student to accomplish the industrial engineering degree.

2. For organization

   The output of the research can be useful input for the organization as a reference for decision-making for improvement on their products or services.

1.5 Research Limitation

During the research, the author finds limitations of the research, as follows,

1. The data for the analysis were taken from the google review platform in the reviews section.

CHAPTER II

LITERATURE REVIEWS

2.1 Inductive Study

Alzate et al. (2022) conducted a study entitled “Mining the text online of online consumer reviews to analyze the brand image and positioning”. This research proposed an unified, unstructured, and easy-to-implement procedure for the text analysis review with the goal of studying brand image and positioning. The research method used in the study is text mining based on a lexicon-based approach in order to provide the researcher insight to brand associations. There are proposed procedures to obtain insight into business derived from the lexicon-based approach, such as Detecting consumers' positive or negative emotions, understanding how they perceive the brand, and product development.

Kumar et al. (2021) conducted a study about the applications of text mining in service management. This study mainly discusses the implementation of text mining. A systematic literature methodology about text mining is used for the study. The study that used text mining methods like sentiment analysis, topic modeling ,and language processing is selected. Also, the visualization methods and tools for text mining is applied, and the association is to understand the dominant themes and relationship. The result of the analysis in this study showed that social media analysis, market analysis, and competitive intelligence are the most dominant themes.

Zhang et al. (2021) conducted a study about the impact of customer-perceived value on repeat purchase intention based on online reviews. The study aims to clarify the relationship between customer perceived value and repeat purchase intention. This paper used big data technology to crawl online reviews of peer-to-peer (P2P) accommodation platforms. The method applied for this study is LDA (Latent Dirichlet Allocation) topic model and sentiment analysis to construct the measurement indicator of perceived values on online reviews. The study identified that perceived value has a positive impact on ”consumers’ repurchase intention in P2P accommodation. Also, social relationship value was considered the most important influencing factor.

Villeneuve et al. (2020) conducted a study about the online digital platform for Airbnb to explore indoor environmental quality. Tex mining is applied to public guest reviews toward temporary accommodations. It is applied on 1.35 million Canada Airbnb reviews to obtain the top causes. The results indicate 5% of reviews complained of IEQ issues, while 0.25% complained about certain forms of IEQ.

Chung et al. (2022) conducted a study on online customer reviews on music streaming services. Text mining is applied to find overall customer satisfaction. This study investigates the factors of customer satisfaction by using social media data. Online review apps on five music streaming are analyzed by using text mining. This study states that customers deliver comments on several aspects related to the environment, price plans, and content. The determinant factors mentioned above have a significant impact on customer satisfaction on each review platform.

Lucini et al. (2020) conducted a study focusing on online customer reviews of airline customer satisfaction. The study aims to measure customer satisfaction in the airline industry. Text mining methodology is applied to Online Customer Reviews (OCR) to provide guidelines to improve competitiveness. The study analyzes 55.775 OCR, covering 419 airlines and passengers from 171 countries. The study showed the variation in distributions according to passenger characteristics and preferences. There are several suggestions for the airline industry to maximize customer satisfaction. They should focus on customer service to the first class passenger, comfort, checking luggage, and waiting time. Also, the analysis showed that cabin staff, onboard service, and value for money are the top three satisfaction determinants.

Guerreiro and Rita (2020) conducted a study focusing on online reviews. This study aimed to predict the recommendation of online reviews. The study focused on the factor that may drive customers or reviewers to make direct comments in the form of texts. The methodology uses text mining to obtain the factors that drive of explicit customer recommendations. The result from the analysis found that the competences shortages from the provider and negative attitudes are triggers of negatives point of view. On the other hand, a positive point of view predicts positive recommendation in the body review.

Li et al. (2020) conducted a study analyzing of restaurant reviews. The purpose of the study is to explore the effect of emotional intensity on how other consumers perceived review usefulness on review length and reviewer expertise. The data was taken

from 300 popular restaurants in the US obtained from Yelp. There are 600.686 reviews taken from the restaurant reviews. The methodologies used in the study are the text mining approach and econometric analysis. The analysis results showed that positive emotional intensity has a negative impact on perceived review usefulness. Otherwise, the negative emotional intensity has a positive impact on perceived review usefulness. Also, reviews that express anger are more useful than those expressing anxiety. Last, the content of the review moderates the effect of emotional intensity on perceived usefulness.

Verkijika and Neneh (2021) studied the recommendation of mobile payment apps. The study aims to explain why users provide positive and negative recommendations about mobile payment systems. The study analyzed the user reviews from the Google Play store to unveil the reasons why users provide positive or negative recommendations towards mobile payment apps. There are total 27 mobile payment apps that contain reviews taken from the Google Play store. The limitation of the study for the data is taken only from January 2015 to December 2020. The data contain a total of 805.707 reviews. The result from the text mining showed a total of 13 themes,: ease of use, usefulness, convenience, security, reliability, satisfaction, transaction speed, time saving, customer support, output quality, perceived cost, usability, and trust.

Jung and Suh (2019) conducted a study using the voice of employees to identify and analyze job satisfaction factors from online employee reviews. The data in this study are taken from 35.063 online employee reviews provided on jobplanet.co.kr. The methodology used is Latent Dirichlet Allocation (LDA). Then followed by analyses based on the factors. The sentiment, and importance of each job satisfaction factor are measured. The results of the study provide the business manager into finding the best possible decisions for managing the job satisfaction of their employees.

2.2 Deductive Study

2.2.1 Customer Satisfaction

Customer Satisfaction is the priority of every business management. It can be the reflection of the organization related to its products or services. Because it is the measure of how well such organizations manage and provide their business to meet customer expectations. A load of factors can be the reason for different customer

satisfaction. Factors like service quality, facilities, food, cleanliness, and hotel rating can influence customer satisfaction stated (Bi et al., 2020). This corresponds with the statement of Wu and Liang (2019) that different service contexts and encounters will result in different customer satisfaction. Therefore, each customer experiences their very own problem, whether positive or negative and leads to how the customer perceives the business. Customer satisfaction is important because it illustrates and reflects the running business whether meeting the customer's expectation or is far from there. Positive customer satisfaction will result in customer loyalty, brand image reputation, and then, most important thing, it can drive another possible customer to be the consideration to purchase the business the organization offered.

2.2.2 Online Review Platform

The application of online review platforms in tourism and hospitality has been valid proof for the customer. Customers always seek information through highly available online reviewss before purchasing products and services to meet their high expectations and avoid the risk of failure. Fu et al. (2022) stated online review is considered electronic word of mouth because customers can easily describe or give rate for business they may experience. Customer review is vital as they are provided online and accessible to others. It can be trustworthy from a cutomer point of view rather than the description given by the business management (Kwon, 2021). with the growth of technology and the digitalization era, an online review can be found easily on the online platform. There is an online platforms that provide reviews from customers toward products or services they experienced. It results in easily accessed reviews in the form of text or pictures from around the world. Thus, the role of the online review platform changes the perspective of the customer on making a decision to purchase product or service. Therefore, this study aims to understand the reviews section on online review platforms to identify what factors influence customer satisfaction by applying the text mining method.

2.2.3 Web Scrapping

Pereira et al. (2015) stated that web scrapping is a process of extracting unstructured documents from the internet with any method or technique. The objective of web scraping is to search for certain information and compile it into a file. It offers focused information or selected documents so that the finding process is easy and quick (Ayani et al. 2019). Web Scraping can change unstructured data into structured data that can be stored and validated in a database. The points of web scraping are collecting data, storing data, and validating data. The next important step is data analysis, thus, validated data can be interpreted into better information (Chandra et al., 2019).

2.2.4 Data Mining

Data mining is an activity of finding interesting patterns from large amounts of data. The data can be stored in the database, data warehouse, or other information storage. Data mining deals with other fields of science, such as database systems, data warehousing, statistics, machine learning, information retrieval, and high-level computing. In addition, data mining is supported by other sciences such as neural networks, pattern recognition, spatial data analysis, image database, and signal processing (Prasetya and Ratih, 2019).

2.2.5 Text Mining

Text mining is the process of transforming unstructured text data into machine-processable structured form to discover hidden patterns, also known as a knowledge discovery database from the text (KDT). It deals with the machine learning-supported analysis of textual data. Textual data is extracted from semistructured and unstructured datasets such as emails, full-text documents, and HTML files. Several text mining techniques are applied to analyze the pattern in a large text corpus. Some of the applications are document classification (text classification, document standardization), information retrieval (keyword search or querying and indexing), document clustering (phrase clustering), natural language processing (spelling correction, lemmatization, grammatical parsing, and word sense variant),

information extraction (relationship extraction/link analysis), and web link analysis (Thakur and Khumar, 2021).

2.2.6 Topic Modelling

Glaser et al. (2017) stated that topic models are statistical models designed to extract from a set of documents the relevant "topics". In turn, topic models represent each document within that "topic" or latent space. The most common application of topic models to the characterization of science and innovation uses probabilistic topic models. The method uses the likelihood of repeated words to assign unique subjects to each page. Latent Dirichlet is utilized in this work to define word repetition and topic probabilities. This algorithm is called Latent Dirichlet Allocation (LDA). LDA assumes that each document in a corpus is a distribution over topics, and a topic is a distribution over words. LDA is often used as a dimension reduction tool that extracts the latent structure hidden in the massive dataset (Zhao et al., 2020).

2.2.7 Latent Dirichlet Allocation

Hu et al. (2017) stated that Latent Dirichlet Allocation (LDA) is a top-notch probabilistic model that excels at topic modeling and has seen extensive use in academic studies. LDA assumes that a document is generated based on a set of subjects, and each word in the text is chosen at random from the vocabulary associated with that topic (Blei, Ng, & Jordan, 2004). LDA has been applied to extract topics from various kinds of corpora, including but not limited to microblogs. For example, Huang, Yang, Mahmood, and Wang (2012) applied LDA with web usage data; Xu, Zhang, and Yi (2018) and Lim and Buntine (2014) studied tweets with LDA modeling. Therefore, it is logical to follow that exploring topics from Weibo datasets with LDA topic modeling should also harvest satisfactory results.

2.2.8 Kansei Engineering

Kansei Engineering is defined as a consumer's emotion translation technology about upcoming products. By this definition, Kansei Engineering is trying to produce new products based on consumer feelings and requests. In the field of art and design, Kansei is one of the most important elements that bring about the will or power to create something. Kansei Engineering is a method to translate consumer images or consumer feelings into components of real design (Nagamichi & Mitsuo, 2011). Kensei Engineering was founded by Nagamichi at the Hiroshima University of ergonomics some 30 years ago. Kansei Engineering, as a consumer-oriented ergonomics technology, allows the ' 'consumer's image or feeling to be integrated into the new product design process.

CHAPTER III

RESEARCH METHODOLOGY

3.1 Research Object

The object of this research is a hotel located in Yogyakarta. The author selected Kangen Boutique Hotel as the research object. Kangen Boutique Hotel is a hotel located in Jalan Magelang Yogyakarta. This study will be focused on the factors that influence customer rating quality toward Kangen Boutique Hotel. Customer rating quality found in the customer's reviews stated on the Google review platform. Therefore, the reviews shared on Google Review Platform are used to identify what factors that influence customer rating quality. The data taken from google reviews were obtained by using the scrapping method on phyton. There are the total of 1100 reviews collected.

3.2 Literature Review

The literature reviews in this study are categorized into two categories, which are inductive review and deductive review. The deductive review is focused on the basic or general knowledge of the research's fundamental theory. It is aimed at giving a comprehension of the research conducted. In the deductive review, the fundamental theory of customer satisfaction, online reviews, latent Dirichlet allocation, and related methodologies used in this study were delivered with references. Meanwhile, inductive research is a study that discusses the topic related to the research conducted. It is aimed to deliver knowledge to readers or further researchers about the distinctive features of one research to another. In this review, research that contain scrapping, sentiment analysis, and other similar methodologies to this study were quoted as a reference.

3.3 Data Collection

This research applies primary data to conduct the research. The primary data are taken from the customer reviews section on the google reviews platform. The object of this study is the Kangen Boutique hotel. In order to gather the primary data from the source mentioned above, scrapping by using python is utilized. The data collection was done online, involving data taken from the google review section and the process of mining

the information of the customer reviews toward Kangen Boutique Hotel. Also, the text mining process with the help of a tool called python later on. After the primary data were collected, the next stage is conducting the text preprocessing by following these several steps, filtering, stopword and emoji removal, punctuation, and lemmatization which aimed to clean and prepare text to be the data input for the next step, Word Cloud. After the text preprocessing is done, the next step is applying topic modeling using Latent Dirichlet Allocation (LDA). It is for examining what word that is dominant in each review by using repeated words and extracting a set of reviews into the topic to cluster based on the subject of each word associated with the topic. LDA assumes each document is a distribution of topics, and topics is distributed over the word. The next step is a word cloud, to find the most frequent word related to the comments by the customer during their stays. Data visualization resulting from the work cloud is followed with the implementation of Kansei Engineering as the improvement guidelines. Therefore, the factor that influences the customer rating quality obtained from the work cloud is the reference to suggest the improvement.

3.4 Data Processing

Data processing will represent how the data is processed in order to provide improvement for the hotel. In this process, the method that will be used is Topic Modeling using latent Dirichlet allocation. Also, to find the most frequent word in each review is, use word cloud to give the illustration of represented words in each topic review. Then, the result of this data processing will be used for the interpretation of Kansei guidelines. The tool used in this step is python, LDA multicore, and word cloud. Below are the methodologies used for this research,

1. Topic Modelling is used to extract a set of documents into a topic. The topic represents a document within that topic. It uses repeated word in the customer review.
2. Latent Dirichlet Allocation is used to detect topics in documents along with the proportion of occurrences of these topics in certain documents. LDA is also able to associate words in documents and collections with certain topics. In this study, the author tries to use LDA to detect topics in a google review. The author can imagine that there is a number of reviews coming time by time. Therefore, the LDA piles various reviews from various topics into one without any separation. The

distribution of hidden topics in a number of reviews will be very useful for organizing, understanding, and even making a summary of what is “discussed” in customer review. It can be clustered according to the subject associated with the topic. Where it can explain the correlation of word that is hidden in the reviews stated by the customer on google review. Also, it will be used to determine the Kansei word. Kansei word will be utilized for constructing the guidelines for suggesting the improvement.

3. Work Cloud is used to visualize the most frequent word. The word is stated multiple times in the review section is most likely to appear bigger and bolder in this visualization. The visualized word of the word cloud is the factor that influences customer satisfaction the most.
4. Kansei Engineering is used as guidelines to develop products or services to fulfill customer satisfaction. In this study, the author proposes suggestions for Kangen boutique Hotels based on the factor that influences customer rating quality.

3.5 Discussion

The author conducts a series of data processing, then the result of the data processing will end up in a discussion part that will mainly discuss the problem and the way to solve it. The discussion will begin by identifying the problem using topic modeling until obtained the factors that result in customer rating quality toward Kangen Boutique Hotel.

3.6 Conclusion and Recommendation

This section will mainly be about answering the questions stated in the problem formulation in Chapter I. Also, this part gives some recommendations that might be used to improve for further research.

3.7 Research Flow

This research flowchart is to show how this research is conducted. The research flowchart explains and portrays how the author conducts the research from the beginning untill the end of the research, which the conclusion. The research flowchart is shown in Figure 3.1 below:

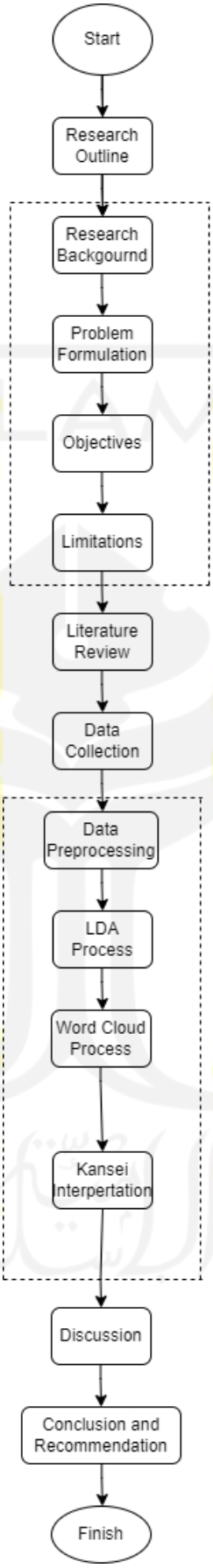


Figure 3. 1. Research Flowchart.

CHAPTER IV

DATA COLLECTION AND PROCESSING

4.1. Data Collection

The research data were taken from customers' reviews on google website reviews, that experience the products and services of the hotel. The object of this study is the Kangen Boutique Hotel. The customer reviews were collected by using python for the scrapping process. Figure 4.1 shows the screenshot of the reviews section on google reviews. There are a total of 1100 reviews posted by customers.

For collecting data, the researcher uses an API to gather data from google maps review API. Application Programming Interface is referred to as API. Application programming interfaces (APIs) expose services or data provided by a software application through a set of predefined resources, such as methods, objects, or URIs (Stylos, Faulring, Yang, & Myers, 2009). In other words, API is a way to communicate between computers. However, Google Maps Review does not explicitly provide API to the public. But, by inspecting the request on the browser when opening google map review, the API endpoint can be captured. The endpoint captured is shown in figure 4.1 below.

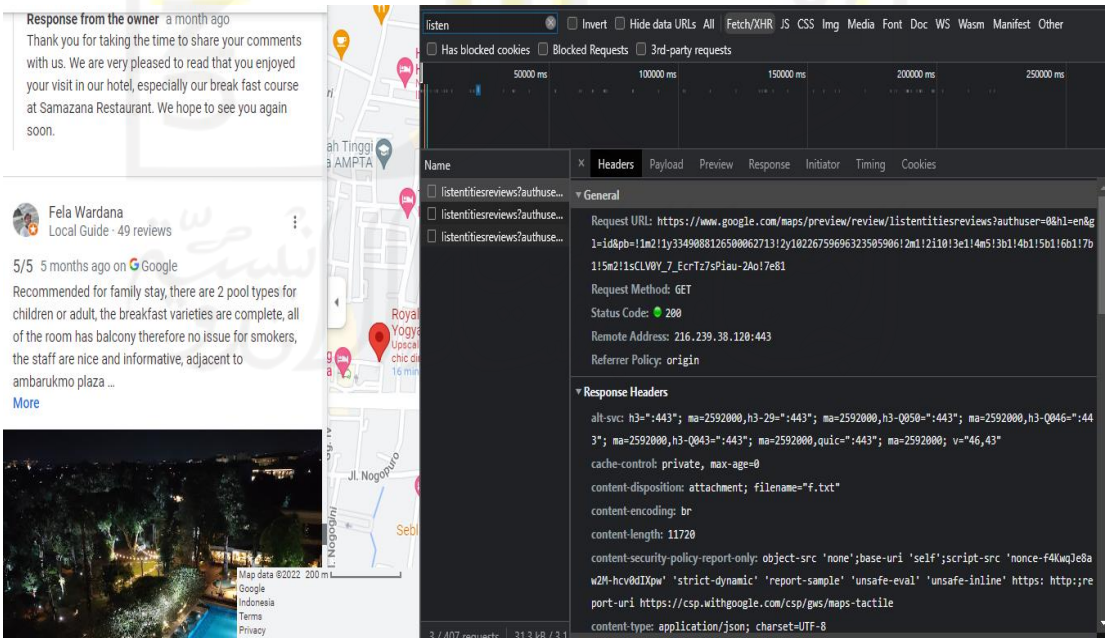


Figure 4. 1. Google Map Page Inspection.

It is shown in illustration figure 4.1 about the endpoint from google maps review. The behavior of this API, will load the next data when the user scrolls down to the next review.

```python
def get_data():
    n = 0
    page_size = 100
    url = "https://www.google.com/maps/preview/review/listentitiesreviews?authuser=0&hl=en&gl=id&pb=!1m2!1y334908812650006
    while n < 1000:
        modifed_url = re.sub('!1i\d+', '!1i' + str(n), url)
        print(modifed_url)
        captured_data = req.get(modifed_url)
        wait = randint(2, 10)
        time.sleep(wait)
        print ("wait for : ", wait)
        if captured_data.status_code == 200:
            print(captured_data.status_code)
            n += page_size
            cleaned_data = clean_data(captured_data)
            filename = str(n)
            save_data(cleaned_data, filename)
        else:
            print ("Error : ", captured_data.status_code)
```

Figure 4. 2. Scrapping Function on Python

Figure 4.2 shows the procedure of making a request to google maps review to gather specific data to get a review. In order to gather data, the researcher needs to modify each request to get review data. By modifying “!1i” to “!1i100” to get data from 1 to 100. In the next review data, the URL will change to “!1i200” to get the next data from 100 to 200. from this python code, the researcher can gather data from around 10.000 reviews. But somehow, the API is limiting the request to 9400 data and returning the error 429 code which is an error for the rate limit.

4.2. Data Preprocessing

The reviews taken from the google maps review section contain unnecessary things and wordings. Therefore, data cleaning is needed to get a clean data review before processing into the LDA process. By conducting the sequence of data processing is elaborated in chapter III, hence, the result of each data process is shown in the next subchapter.

4.2.1. Text Cleaning

1. Filter English Review

   The extracted data from the google maps review is contained Indonesia and Javanese review. For better accuracy in the next process, the researcher decided to use English translated Indonesian review.

```
def filter_english(data):
  if re.findall("(?<=\(Translated by Google\))[\s\S]*?(?=\(Original\))",data):
    english_only = re.findall("(?<=\(Translated by Google\))[\s\S]*?(?=\(Original\))",data)
    english_only = english_only[0]
  else:
    english_only = data
  return english_only
```

Figure 4. 3. Non-English Reviews Removal

2. Stopword Removal

   Removing stop words from the data frame is needed to eliminate less meaningful words that might make noise in the LDA process. This process uses the NLTK stopword. Some less-meaningful words such as this, that, it, is , for, etc. the researcher also extends the stopword by seeing the unnecessary word from the result. The function script can be seen in Figure 4.4.

```
def removeStopword(review):
  # when you need to extend the stop word to remove some unces word
  stopwords.extend(['...', "'s", '..', 'edu', 'use', 'The', 'also', 'i', 'I', 'a', '....', "n't", 'ok', 'OK'])

  review_tokens = create_token(review)
  # print (review_tokens)
  text_list=[]
  for word in review_tokens:
    # print (word)
    if word not in stopwords:
        text_list.append(word)
  return text_list
```

Figure 4. 4. Stopword Removal Function

3. Punctuation Removal

   Removing punctuation is needed to remove unnecessary punctuation from review which will disturb the dataset punctuation such as (! ()-[]{};:’”"\,<>./?@#$%^&*_~).

```
def removepunctuation(text):
  clean = []
  for word in text:
    if word not in string.punctuation:
      clean.append(word)
  return clean
```

Figure 4. 5. Punctuation Removal Function

Figure 4.5 represents the code for removing punctuation from the review. This process uses a string build-in python library.

4. Lemmatization

   Lemmatization is a text normalization technique to switch a word into its base root word and group a word into a root form that has the same meanings. In this step, the process is done using nltk wordnet lemmatizer. Figure 4.6 shows python script for lemmatization.

```
def lemmatizer(data):
  word =[]
  wordnet_lemmatizer = WordNetLemmatizer()
  for w in data:
    lemma = wordnet_lemmatizer.lemmatize(w)
    word.append(lemma)
  return word
```

Figure 4. 6. Lemmatization Function

5. Main Function

Function in figure 4.7. is a function to run the process of data cleaning; the process is to start by loading the data from json, then will iterate for each dict on json list. The function grabs specific data to process, and the text review data will be processed first before appending to dict. The first process is removing non-English reviews; then the next process is removing stopwords and removing punctuation. Lastly, to make the text back to its root word, which uses a spacy lemmatizer to do the lemmatization process. Lastly the data will pass to the list on the cleaned list. Then will be loaded to the data frame using the pd.from_dict function.

```
# load data
data = load_data()

# Create empty dict
cleaned = []

for i in range(len(data)):
  review = data['body_review'][i]
  name = data['nama'][i]
  rating = data['rating'][i]

  reviews = filter_english(review)
  reviews = removeStopword(reviews)
  reviews = removepunctuation(reviews)

  reviews = lemmatizer(reviews)

  cleaned.append({
      "name" : name,
      "review" : reviews,
      "rating" : rating
  })

cleaned_data = pd.DataFrame.from_dict(cleaned)
```

Figure 4. 7. Main Python Function

4.3. Data Processing

The process in the data processing step is about how to conduct Topic modeling using Latent Dirichlet allocation. The gensim library will be used in the process. The gensim library has LDA multicore to process reviews to extract topics. LDA multicore uses all CPU cores to parallelize and speed up model training.

```
# Create Dictionary
id2word_1 = corpora.Dictionary(clean_brand_review)

# Create Corpus: Term Document Frequency
corpus_1 = [id2word_1.doc2bow(review) for review in clean_brand_review]

 # Build LDA model
ldamodel = LdaMulticore(corpus= corpus_1, num_topics =8, id2word=id2word_1,chunksize=2000, passes=50, per_word_topics=True)
```

Figure 4. 8. LDA Main Script

The review text must be converted to a dictionary using the corpora dictionary function before the lda multicore process can begin. Then, using the term document frequency method to generate a corpus. The library used in this process is doc2bow. Figure 4.8 depicts the Python script.

```
from pprint import pprint

pprint(ldamodel.show_topics(formatted=False))
x = ldamodel.show_topics(formatted=False)
```

Figure 4. 9. Show LDA Result

Figure 4.9 above is a function to show the topic from the lda process. The result is shown below in Figure 4.10

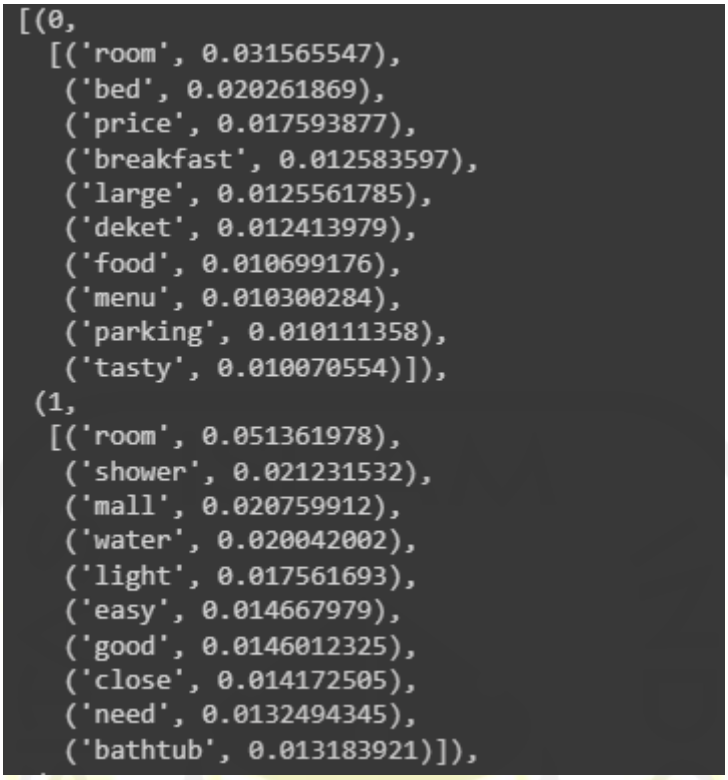


Figure 4. 10. Test Result

The result of LDA is a list of topics. The number of topics has been defined before. Each topic has terms and a score. The term in a topic has a score that indicates the frequency of the term appearing in the topic.

The researcher uses pyldaviz to create a visual of lda to make it easier to understand.

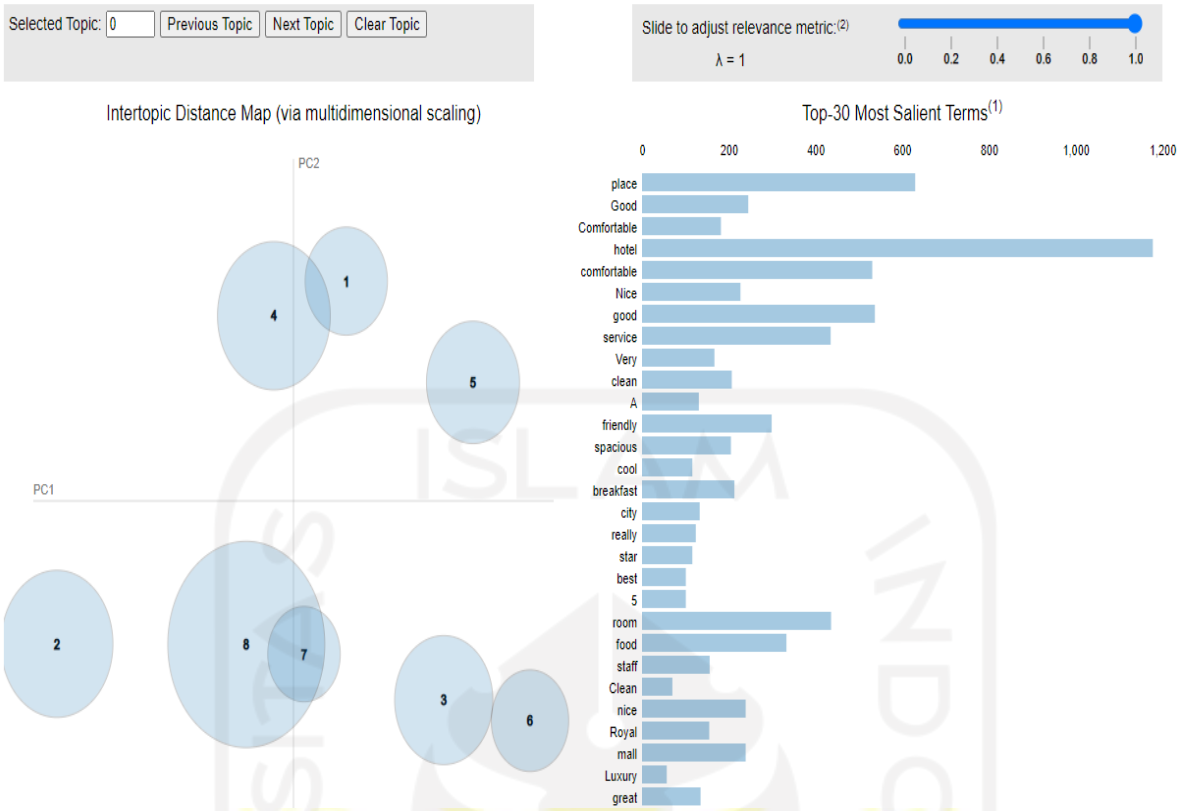


Figure 4. 11. Result Pyldavis

The pyldavis visual is shown above in figure 4.11. The topic from the LDA process is represented by the bubble in pyldavis. This visual's explanation will be continued in Chapter 5.

4.4. Data Visualization

Data visualization is a step in data processing that shows a chart, map, or graph to provide a better view of the resulting data. The goals of data visualization are to help readers identify patterns, trends, and outliers in large data sets. The function used in this process is shown below,

1. Create Chart

   The researcher constructed a function to get the frequency of top words in the review. The function can be seen in Figure 4.12. the function is converted to tokens using sklearn count vectorizer, and then a vocabulary of known words is built. Then the data from the frequency counter is passed to the drawing function. The drawing

function uses the matplotlib library. The data for drawing is provided from the get top word function. In Figure 4.13, it is shown the chart from the draw function.

```
def get_top_n_words(corpus, stp, n=None):
    wordlist =[]
    for i in range(len(corpus)):
      data = corpus[i]
      for w in data:
        wordlist.append(w)
    if stp != "":
      vec=CountVectorizer(stop_words=stp).fit(wordlist)
    else :
      vec=CountVectorizer().fit(wordlist)
    bag_of_words = vec.transform(wordlist)
    sum_words = bag_of_words.sum(axis=0)
    words_freq = [(word, sum_words[0, idx]) for word, idx in vec.vocabulary_.items()]
    words_freq =sorted(words_freq, key = lambda x: x[1], reverse=True)
    return words_freq[:n]
```

Figure 4. 12. Get Frequency Function

```
def DrawChart(titlename, pass_dataframe, by="count", x_axis="adjective", y_axis="count", color="red", savename=None):
  fig, ax = plt.subplots(figsize=(8, 8))

  pass_dataframe.sort_values(by=by).plot.barh(x=x_axis,
                    y=y_axis,
                    ax=ax,
                    color=color)
  ax.set_title(titlename)
  if savename != None:
    plt.savefig(path_picture + savename + ".png")
  return plt.show()
```

Figure 4. 13. Drawchart Function

2. Create Work Cloud

   In order to visualize the result of topic modeling research used word clouds and frequency charts to visualize the result of topic modeling. Figure 4.15 demonstrates the function of making a word cloud. For this method, the researcher must count each word and save the counter to the data frame. Then passing the data frame to the word cloud function. The word cloud represents the most frequent term that

appears in the data; the thicker the word cloud means, the more frequent word. The result of the word cloud function can be seen in Figure 4.15.

```
def get_wordcloud(dataframeAndColumn,wordcloudname,stp,saveornot,showornot):
  stop = stp
  text = " ".join(review for review in dataframeAndColumn.astype(str))
  print ("{} Word in ".format(len(text)) + wordcloudname)
  if stop != "":
    wordcloud = WordCloud(stopwords= stop, width=800, height=600, background_color='black',collocations=False).generate(text)
  else :
    wordcloud = WordCloud(width=800, height=600, background_color='black',collocations=False).generate(text)
  if showornot == True:
    plt.figure(figsize=(15, 15))
    plt.imshow(wordcloud, interpolation='bilinear')
    plt.axis("off")
    plt.show()
  if saveornot == True:
    wordcloud.to_file(path_picture + wordcloudname + ".png")
```

Figure 4. 14. Create Work Cloud

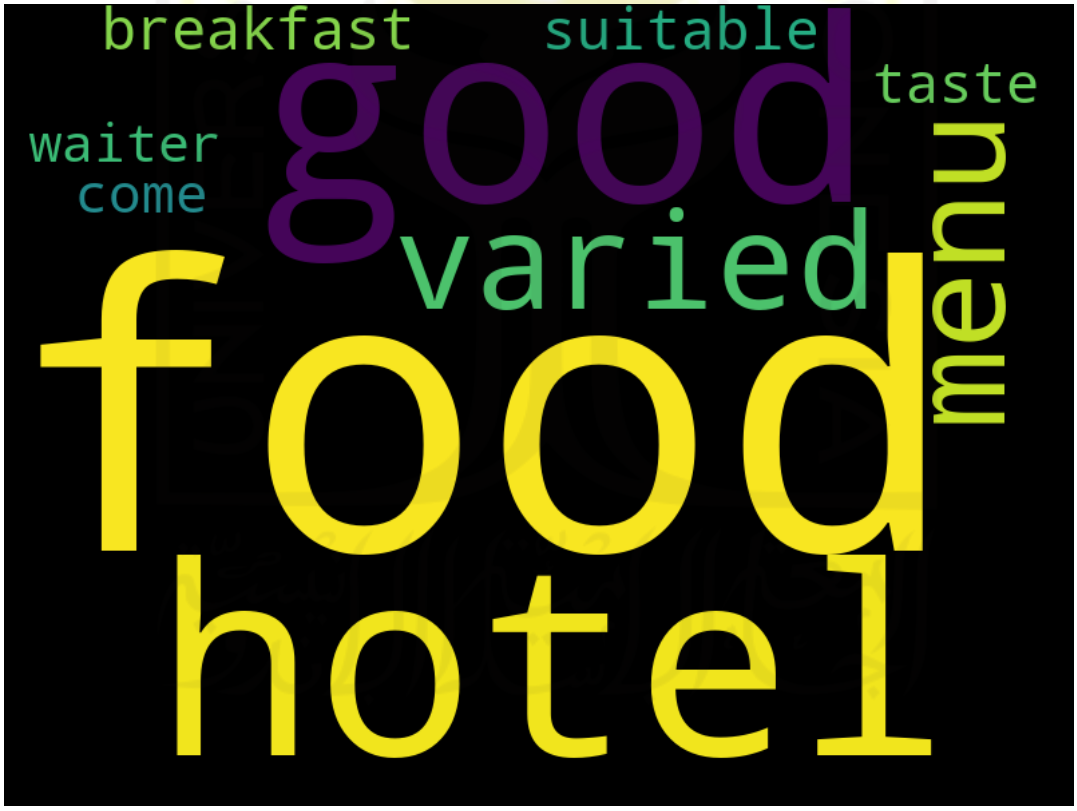


Figure 4. 15. Work Cloud Result

The next process of this research is the result and discussion. In order to make this research done, we must interpret the topic on topic modeling by human judging. The data visualization is needed to lead us to create Kansei product guidelines.

CHAPTER V

RESULT AND DISCUSSION

5.1. Analysis of Topic Modelling

The researcher will then delve deeper into each topic to investigate the findings in order to develop Kansei guidelines. To gain more insight from topic modeling, the researcher will investigate the word in each topic. There are 7 topics that found in this review.

Figure 5. 1. Word Cloud Topic 1

According to topic 1, The words "hotel," "room," "bathroom," "wall," "floor," "toilet," "smell," "dirty," "need," and "good" could potentially be assigned to a topic related to the overall cleanliness and condition of the hotel. The words could indicate that the reviewer did not like the hotel because the rooms and bathrooms were dirty, with a bad smell emanating from the toilet. The mention of the wall and floor may suggest that these areas also need cleaning or repair. Overall, the words indicate that the reviewer was dissatisfied with the hotel's condition and felt it was not up to a good standard.

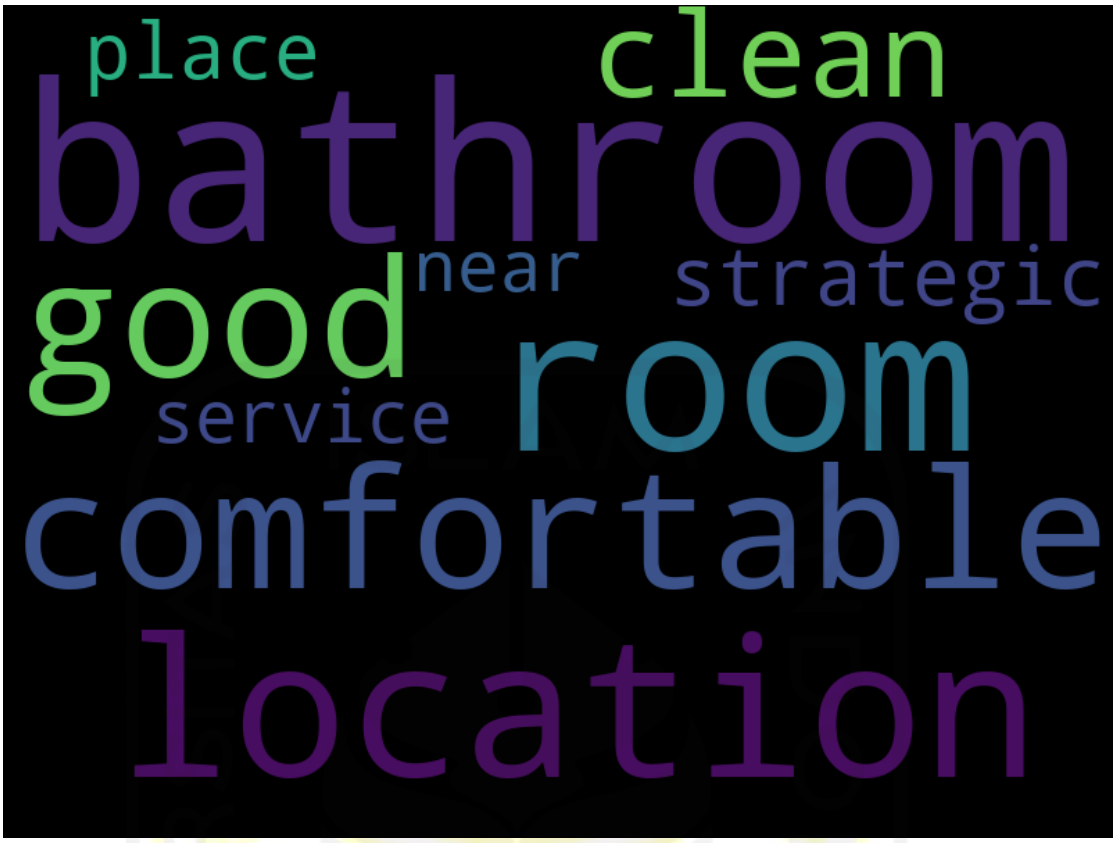


Figure 5. 2. Word Cloud Topic 2

According to topic 2, the words "place," "clean," "bathroom," "good," "near," "strategic," "service," "room," "comfortable," and "location" could potentially be assigned to a topic related to the overall location and cleanliness of the hotel. The words indicate that the reviewer did not like the hotel because it was not clean, the rooms and bathrooms were not comfortable, and the location was near popular attractions and amenities. The mention of service may suggest that the hotel did not offer good customer service. Overall, the words indicate that the reviewer was dissatisfied with the hotel's service and cleanliness.

Figure 5. 3. Word Cloud Topic 3

According to topic 3, the words "breakfast", "room", "service", and "hotel" could be grouped together as being related to the hotel's amenities and facilities. The words "okay", "good", "comfortable", and "friendly" could be grouped together as being related to the quality of the hotel and the staff. The word "nice" could potentially belong to either of these topics, depending on how it is used in the review.

Figure 5. 4. Word Cloud Topic 4

The customer specifically mentions the room, bathroom, floor, and facility in topic 4. In this topic, customers use adjectives such as dirty. It means they complain about the filthiness of the room, bathroom, or floor. It can be concluded from considering such reviews mentioned about "facility" that the hotel needs to perform renovation.

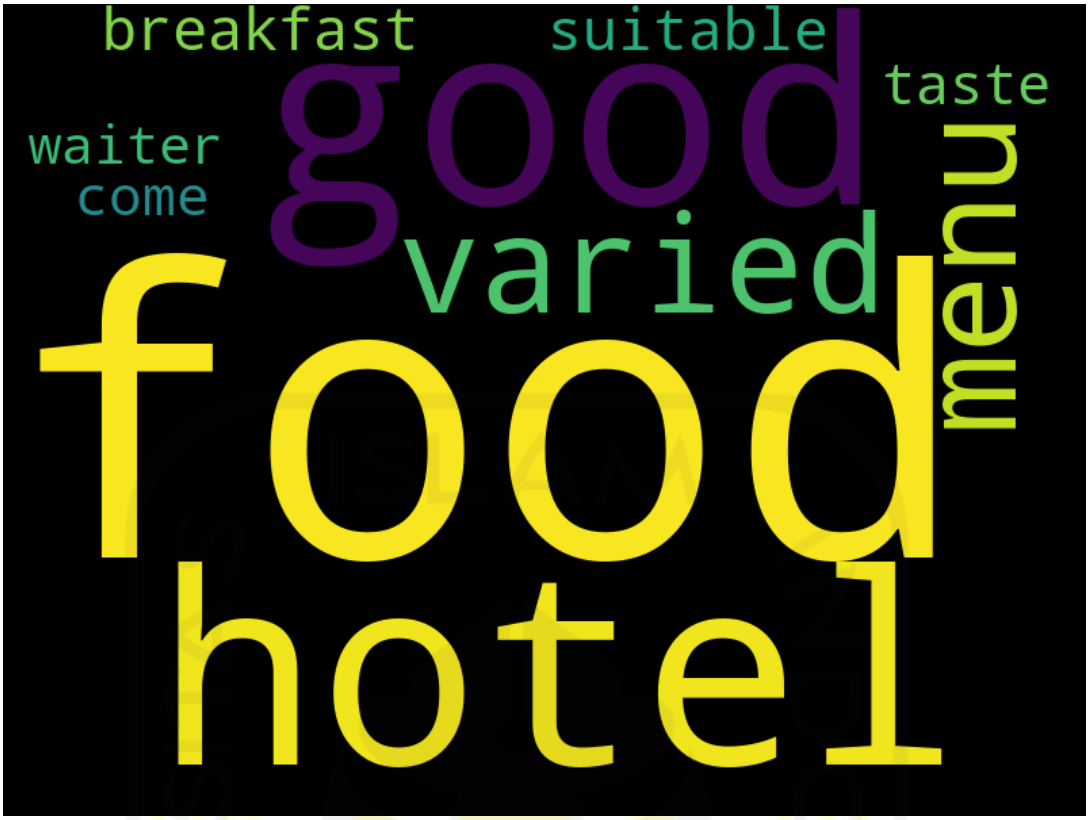


Figure 5. 5. Word Cloud Topic 5

The word "hotel" is about the overall experience at the hotel was not satisfactory. The word "food" may indicate that the quality and taste of the food at the hotel were not up to par. The mention of a "menu" may suggest that the options available were limited and not varied. The use of the word "good" may indicate that some aspects of the hotel, such as the breakfast, were satisfactory. However, the use of the word "suitable" may suggest that the food may not have been suitable for all guests. The mention of a "waiter" may indicate that the service was lacking, with guests having to constantly call for assistance. The use of the word "come" may suggest that the waiter did not come promptly to attend to the guests' needs. Overall, the review suggests a negative experience at the hotel in terms of food and service.

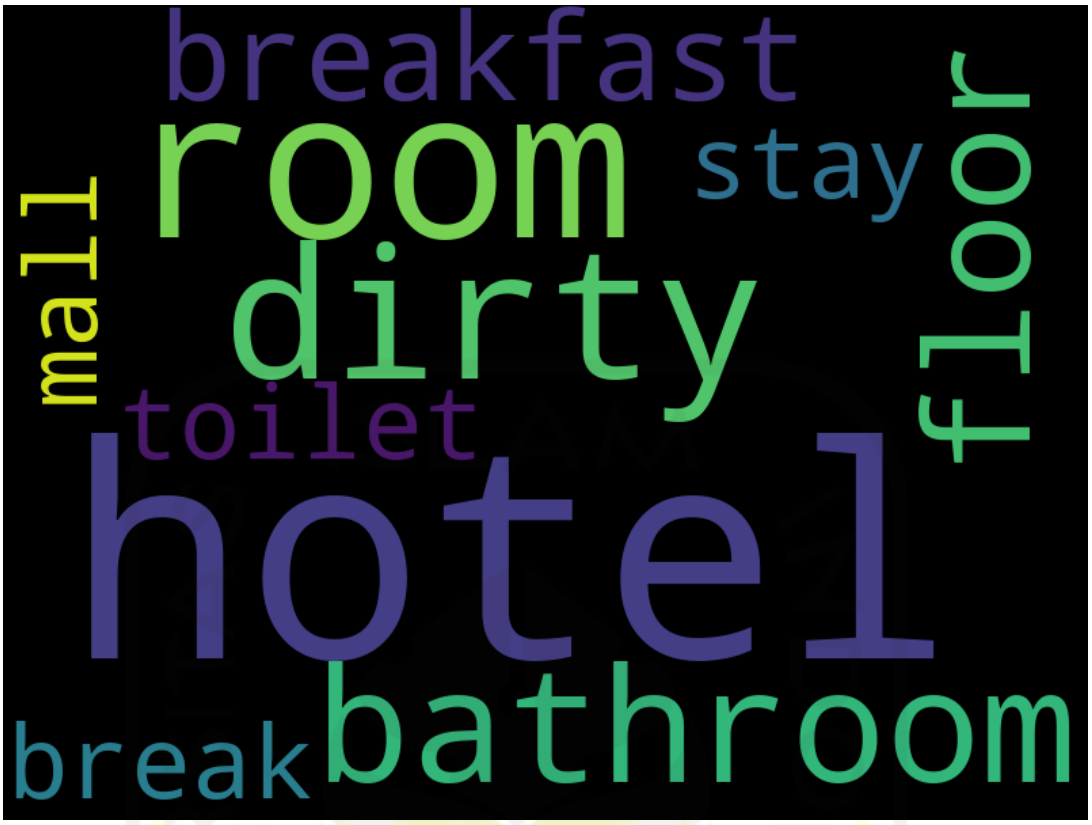


Figure 5. 6. Word Cloud Topic 6

The word "room" in this negative hotel review indicates that the condition of the rooms at the hotel was not satisfactory. The mention of the word "hotel" indicates that the overall experience at the hotel was not positive. The mention of "breakfast" indicate that the breakfast options at the hotel were not satisfactory. The mention of the word "mall" indicates that the hotel is located near a mall. The use of the word "floor" indicates that the condition of the floors in the hotel was not clean. The mention of the word "stay" indicate that the reviewer stayed at the hotel don't want to stay anymore in this hotel. The use of the word "dirty" indicate that the hotel was not well-maintained and clean. The mention of the word "break" indicate that the reviewer experienced some issue or problem during their stay at the hotel. The use of the word "toilet" indicates that the condition of the toilets at the hotel was not satisfactory. Overall, the review indicates a negative experience at the hotel regarding the condition of the rooms and cleanliness.

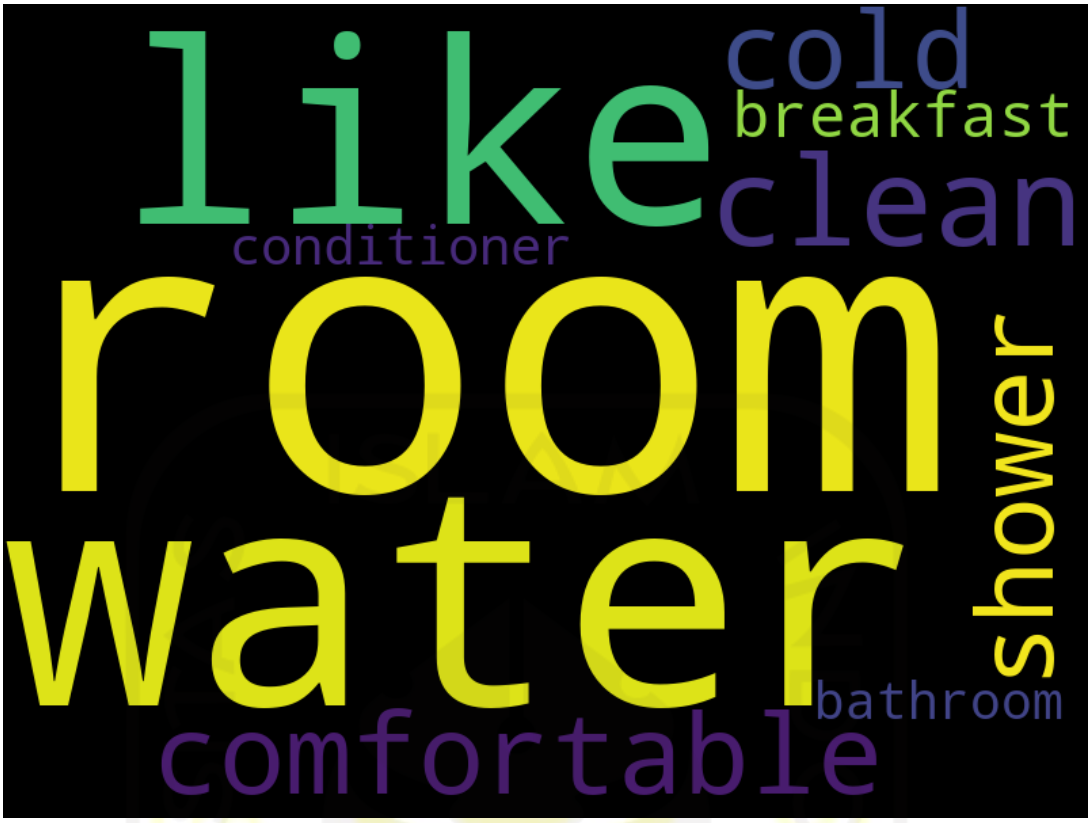


Figure 5. 7. Word Cloud Topic 7

According to topic 7, the word "room" in this negative hotel review indicates that the condition of the rooms at the hotel was not satisfactory. The mention of the word "bathroom" may indicate that the condition of the bathroom was not up to par. The mention of the word "shower" indicates that the shower was not functioning properly. The use of the word "air conditioner" may indicate that the air conditioner in the room was not functioning properly. The mention of the word "water" may suggest that the water in the room was not satisfactory. The use of the word "cold" may indicate that the air conditioner in the room was not cold. The mention of the word "clean" may suggest that the room was not clean. The use of the word "breakfast" may indicate that the breakfast options at the hotel were not satisfactory. Overall, the review suggests a negative experience at the hotel regarding the condition of the rooms and amenities.

5.2. Kansei Improvement Guidelines

This guideline results of the analysis of topic modeling that has been done. Kangen Boutique Hotel can use this guideline to improve the hotel service. There is a similarity on each topic. The researcher decides to summarize these topics into 4 aspects.

Table 5. 1. Improvement Guidelines

No	Aspect	Guideline
1	Service	Improve receptionist and waiter service. Train them to have a positive attitude and to be friendly to visitors.
2	Cleanliness	Renovate the room and bathroom because this word frequently appears in the review. Room cleanliness (floor and walls) must be improved.
3	Food Variant and Taste	Increase the menu to satisfy guests and improve food taste.
4	Facility on Room	Maintain the facilities in the rooms, such as the shower head, bathtub, and toilet, and ensure that the water is running. The light in the room must be checked on a regular basis to ensure that it is still bright.

The aspect found after topic model analysis, such as Service, Cleanliness, Food, and Room Facilities. Each topic represents a common word or sentiment expressed in the review. Consequently, there are many words in each topic that overlap or are repeated. Therefore, the researcher decides to simplify the aspect into these 4 aspects. The interpretation of this result can be varied using human judging, but it will not be the boundary of the topic. Because by using LDA, the result is grouped into each topic, which can make the understanding of the result better. Below is the priority of improvements aspect for the hotel in order. The topic interpretation can be seen below.

1. “Cleanliness” Topic, on the review term such as “dirty”, “bad smell”, “not clean” is often show up in reviews. The bad experience from customers is appeared together with the word “room”, “toilet”, “bathroom”, “wall”, and “floor”. From the review, they stated that the room, bathroom, and toilet are not in optimal condition, such as dirty, and has bad smell. On another side, the wall and floor are often found in imperfect condition, the wallpaper is peeled off and emitting a bad smell. The cleanliness of this hotel was a major concern for the customers during their stay at the

hotel. The recommendation for hotel management is to keep maintaining rooms to give the best experience for the customers during their stay at the hotel. The management must ensure the room is in perfect condition, such as well maintained, clean, and smells good.

2. “Room Facilities” Topic in the review, oftenly found described with certain terms, such as “break”, “broken”, “not cold”, “not well-functioned” and came up with the word “toilet”, “air conditioner”, “shower”, “water”, “bathroom”. It means that customers complain about the facility in the room, such as the bathroom, toilet, shower, and air conditioner is not in the best condition. From the review, it is found that the air conditioner is cold enough to satisfy the customer's expectation or the air conditioner is broken. The toilet is not functioning properly. The shower does not have enough water pressure to function properly. Those occurrences may give the bad experience to guests. The recommendation to hotel management is to give more attention to the room’s facility to give the best experience to customers. Regular maintenance should be held to make sure the room facility is in its best condition.

3. “Service” Topic on the review, is shown up with the terms “staff”, “waiter”, “room service” with the adjectives word “not helpful”, “not friendly”. It means the staff and waiter service need improvement due to customers’ complaints. The recommendation for hotel management is to provide the training program to staff to improve their performance.

4. “Food” Topic, on the review term such as “menu”, “variant”, “taste”, “not good”, “not suitable”. It means that the hotel's food menu did not offer a lot of variety, and the available options did not suit customer tastes. The taste of the food was not great. That makes the customer disappointed and impacts the customers' experience during their stay at the hotel. The recommendation to hotel management is to improve their food taste by providing training to kitchen staff and making sure the ingredients are in good condition. The food should be varied to address customer satisfaction.

The topics evaluated above are the priority of the hotel for improvement because it is derived from the customer reviews. The hotel management should aware and concern

about these kind of comments from customer during their visits. thus, the author confirm to the hotel management regarding these issues. The management always seek for greatness in term of keeping their business. they do listen to the comments of customer either online and verbal. Improvement toward the room condition and cleanliness are being their priority since the most negative comments are coming from these aspects.. It has been renovated room by room and checked the condition of the room. Although, the management has done the improvements, it is not meant that the customers are satisfied with their products or services. By looking at their recent comments, they still fail to impress their customers. There are still customers are complaining about the staff services, room facility not working and the hotel cleanliness. Hence, the author suggest that it is key to have detailed improvement on each factors. The author already stated the factors that influenced by customer in this research. The factors are coming from the most stated comments from the customer visit.

The result and discussion of the research have been delivered in this chapter. The results are the output of the chosen methodology by the author. The methodology used primary data. All the data were taken online from the google review platform. Due to limited time, this research excludes observation and survey for data collecting. The author realizes that this research is not flawless that still contains shortcomings. Additional methodology, such as service quality (SERVQUAL), can be implemented in order to portray the real condition of Kangen Boutique Hotel. The method involves five dimensions, tangible, responsiveness, assurance, reliability, and empathy. This additional method could offer a different solution for improving the service quality since it has attributes that is not mentioned in this research.

CHAPTER VI

CONCLUSION AND RECOMMENDATION

6.1 Conclusions

1. Based on the research result and discussions after the completion of all chapters of the research, a conclusion can be drawn. In this research, Topic Modeling based on Latent Dirichlet Allocation is used to analyze Kangen Boutique Hotel. The data collected for this research is 1470 reviews. The researchers used LDA to identify the underlying topics in the reviews, and the resulting topics represented the key factors that influenced the rating quality. The topics include:
    1. Staff Service
    2. Cleanliness
    3. Food
    4. Room Facilities

    Each of these topics represents a common word or sentiment expressed in the review. One of the topic is the “Room Conditions” topic might include words and phrases like “bathroom”, “floor”, “room”, “wall”. From that topic, also get sentiments such as “smelly”, “dirty”, “break”, “broken”.

2. By identifying these topics, the researchers can gain a better understanding of the factors that influence the rating quality of Kangen Boutique Hotel. This information can be used to improve the hotel and enhance the customer experience. The recommendations suggested by the author are the hotel might focus on maintenance and improving the comfort of the rooms, giving suitable training programs to staff to improve their performance, adding more menu and making sure ingredients are fresh to address customer satisfaction, and scheduling regular maintenance should be held to make sure the room facility in the best condition.

6.2 Recommendations

Based on the research results, the researcher offers recommendations from this scope of the study, which can be used for improvement later on. The recommendations for further

research are implementing other research methodologies, such as service quality (SERVQUAL), to gain a better comprehension of the real condition of the object. Therefore, it offers a comparison between the two methodologies in order to reach customer satisfaction.

REFERENCES

Aggarwal, C.C., Zhai, C., 2012. Mining text data. Springer, Boston, MA

Barua, A., Thomas, S. W., & Hassan, A. E. (2014). What are developers talking about? An analysis of topics and trends in Stack Overflow. Empirical Software Engineering, 19(3), 619–654.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. Journal of Machine Learning Research, 3(Jan), 993–1022.

Chung, J., Lee, J., Yoon, J. 2022. Understanding music streaming services via text mining of online customer reviews. *Electronic Commerce Research and Applications*, Vol. 53.

Fu, W., Choi, E. 2022. Text-mining with network analysis of online reviews and consumer's satisfaction. *Information*, 127.

Gueriro, J., Rita, P. 2020. How to predict explicit recommendations in online reviews using text mining and sentiment analysis. *Journal of Hospitality and Tourism Management*, 43.

Hu, X., Choi, K., Hao, Y., Cunningham, S. J., Lee, J. H., Laplante, A., ... & Downie, J. S. (2017). Exploring the music library association mailing list: A text mining approach. In X. Hu, S. J. Cunningham, D. Turnbull, & Z. Duan (Eds.), Proceeding of 18th International Society for Music Information Retrieval Conference, 302–308.

Jung, Y., Suh, Y. 2019. A text mining approach to identifying and analayzing job satisfaction factors from online employee reviews. *Decision Support Systems*, Vol. 123.

Kumar, S., Ilavarsan, V. 2021. Application of text mining in service management. *International Journal of Information Management Data Insights.*

Li, H., Liu, H. 2020. Online persuasion of review emotional intensity on restaurant reviews. *Journal of Hospitality and Tourism Management*, 89.

Lim, K. W., & Buntine, W. (2014). Twitter opinion topic model: Extracting product opinions from tweets by leveraging hashtags and sentiment lexicon. Proceedings of the 23rd ACM international conference on conference on information and knowledge management, 1319–1328.

Lucini, F., Tonetto, L. 2020. Text mining approach to explore dimensions of airline customer satisfaction using online customer reviews. *Journal of Air Transport Management*, Vol. 83.

Madhumita, M. 2022. Perceived useefulness of online customer reviews: A review mining approach using machine learning and exploratory data analysis. *Journal of Business Research*, 147-164.

Verkjika, S., Neneh, B. 2021. A text-mining study on the recommendation of mobile payment apps. *Journal of Retailing and Consumer Services*, Vol. 63.

Villenuve, H., Obrien, W. 2020. Text-mining Airbnb reviews to explore indoor enviromental quality. *Building and Environmet*, Vol. 169.

Xu, X. 2020. What are customers commenting on, and how is their satisfaction affected?. *Decision support systems*, Vol. 142.

Zhao, F., Ren, X., Zhao P. 2020. Latent Dirichlet Allocation model training with differential privacy. IEEE Transactions of Information Forensic an Security. 16. 1290-1305.