

DOI: <https://doi.org/10.31933/dijemss.v5i5>Received: 30 Mayth 2024, Revised: 25 Juneth 2024, Publish: 30 Juneth 2024<https://creativecommons.org/licenses/by/4.0/>

Sentiment Analysis and Performance Forecasting Based on Online Review Hotel Harper M. T. Haryono Jakarta

Rizki Aulia Kusumawisanto¹, Athor Subroto²

¹Universitas Indonesia, Jakarta, Indonesia, rizky.aulia.k@gmail.com

²Universitas Indonesia, Jakarta, Indonesia, athor.subroto@ui.ac.id

Corresponding Author: rizky.aulia.k@gmail.com¹

Abstract: This thesis examines the sentiment of online reviews for Hotel Harper to identify key areas for service improvement, predict occupancy rates using sentiment analysis and forecasting models, and provide strategic recommendations for enhancing the hotel's image and performance. The research involves literature review, data collection, sentiment analysis, and forecasting model development. Data from online reviews of Hotel Harper M.T. Haryono from 2020-2023, spanning periods before, during, and after the COVID-19 pandemic, were collected. The Naive Bayes algorithm classifies sentiments into positive, negative, and neutral categories. Data visualization and classification performance evaluation are also conducted. Sentiment data is combined with occupancy rate data to develop an ARIMA forecasting model, evaluated using MAPE and RMSE. Results indicate that room quality and cleanliness significantly influence user evaluations, necessitating improvements in these areas. Negative reviews pointing to service-related issues suggest the need for enhanced staff training. Consequently, Hotel Harper M.T. Haryono should conduct regular training sessions for staff, especially those interacting with guests, to improve service quality. The occupancy rate predictions show an upward trend from 2020 to 2023 with low error rates, enabling Hotel Harper M.T. Haryono to use this model for strategic planning and informed decision-making.

Keyword: Sentiment Analysis, Forecasting, Hotel Reviews.

INTRODUCTION

Cawang area is a strategic location and a hub for traffic and multimodal integration such as Transjakarta Bus, Train, Halim Perdana Kusuma Airport, LRT, and Bandung - Jakarta High-Speed Train (Tambunan, 2022). With various modes of transportation available in the area, Cawang has become one of the main business and industrial districts in Jakarta. This has attracted many companies and businessmen to establish their offices in this area, consequently increasing the demand for accommodations such as hotels. The presence of hotels in this area also facilitates business professionals in conducting meetings and attending other business-related events. Hotel Harper M.T. Haryono Jakarta is a three-star hotel owned by PT HK Realtindo, located in the Cawang area, East Jakarta. The hotel is a modern minimalist establishment with 131 rooms and capacity for large-scale meetings accommodating up to 200 people. It has been operating since 2014 and has garnered a good

rating within the hotel environment in East Jakarta, as evidenced by numerous reviews on booking websites such as Traveloka. However, some customer dissatisfaction and areas for improvement have been identified in several reviews related to the hotel.

As an accommodation destination, Hotel Harper stands out for several aspects. Based on online reviews, one notable characteristic is the friendly and professional service provided by the hotel staff. Many guests highlight their positive experiences regarding the staff's hospitality and responsiveness to their needs. Additionally, the facilities provided by Hotel Harper M.T. Haryono Jakarta often receive praise in online reviews. From room cleanliness, bed comfort, to the variety and quality of food offered in the hotel restaurant, these are valued aspects appreciated by guests. Furthermore, the strategic location is also a major factor attracting visitors, facilitating accessibility to important places in Jakarta. In terms of performance forecast, this online review pattern also reflects that Hotel Harper M.T. Haryono Jakarta tends to maintain high service standards and continuously improve its facilities to meet guest expectations and ensure a satisfactory stay experience. The combination of friendly service, quality facilities, and strategic location makes this hotel an attractive choice for both leisure and business travelers seeking a satisfying stay experience in Jakarta.

In the hospitality industry, hotel reviews are valuable sources of information for prospective guests making decisions about their accommodations. These reviews can be found on various platforms, such as hotel booking websites, social media, travel forums, or booking sites like Traveloka. Consumer reviews in the form of hotel reviews can reflect customer needs to a certain extent and can be used to anticipate customer needs in the future.

Roy (2023) in his research mentioned that online customer reviews have become a valuable resource in the hospitality industry. The popularity of social media and reviews on websites has provided an opportunity for the hospitality industry to utilize this data to determine what factors can influence customer decision-making. Sentiment Analysis is one way to obtain sentiment from tourists and their emotions from online reviews (Wu et al., 2022). Customer sentiment plays a significant role in how hotel management can act to provide the best satisfaction for their customers. Sentiment analysis can also be used by hotel managers to encourage customer loyalty to the hotel. Effective sentiment analysis utilization can increase occupancy rates, profitability, and provide a better guest experience (Wu et al., 2022).

Sentiments obtained from customers can be influenced by lodging experiences. Jang and Moutinho (2019) in their research stated that customer-perceived experiences such as hotel security, service, location, price, and reputation can build customer sentiment and attitude toward the hotel. Some observable factors such as food, beds, lobby, and floors, as well as intangible factors such as stay experience, service, and hospitality, can influence customer sentiment. There are several types of sentiments or reviews given by customers about a product or service, namely positive, neutral, and negative. To classify these types of sentiments, data mining techniques involving information collection from available data sources are used. Data mining can be done using two main methods, namely classification and clustering. The sentiments expressed by guests or customers staying at the hotel are inseparable from the promises made by the hotel, which are the service level agreement to their guests. In a study conducted by Widyaningrum et al. (2023), Service level agreement (SLA) is an agreement formed based on the agreement that the parties involved need to follow some rules so that the relationship between the parties can run smoothly. In the context of hospitality, hotel management acts as a service provider and hotel guests represent clients in the SLA concept. Hotel management will strive to fulfill everything that has been conveyed to their guests so that satisfaction arises from both parties.

In this study, the focus will be given to the classification method, using the Naive Bayes method. The Naive Bayes method is one of the common classification techniques used in sentiment analysis. This method is based on calculating the probability of an event

occurring based on the influence obtained from observations. In addition to the Naive Bayes method for sentiment analysis, the ARIMAX forecasting method can also be used to provide predictions about future behaviors or demands based on historical data. The ARIMAX method (Autoregressive Integrated Moving Average with Exogenous Variables) is a development of the ARIMA model that considers the influence of exogenous variables or external factors in the forecasting process. In the context of sentiment analysis, the results of Naive Bayes sentiment analysis can be used as exogenous variables in the ARIMAX model to improve forecasting accuracy. By incorporating exogenous variables from sentiment analysis results, we can take advantage of additional information revealed by customer sentiments in predicting future behaviors or demands. By combining the Naive Bayes method for sentiment analysis with the ARIMAX forecasting method, we can utilize information from both approaches to generate more accurate forecasts. The use of Naive Bayes sentiment analysis in the ARIMAX model allows us to capture the influence of customer emotions or evaluations on a product or service, which can impact future demand.

Research conducted by Wu et al. (2022) stated that sentiments generated by customers can be used to improve forecasting in the future. Research conducted by Nicolau et al. (2024) also found that sentiments provided by customers through online reviews can affect the performance of the hotel in the form of an increase in the ADR of the hotel. Research conducted by Roy (2023) found that hotel customer reviews are influenced not only by the hotel's services but also by factors such as facilities, location, and hotel type that affect customer reviews. The increasing number of negative reviews from Hotel Harper guests may affect the decrease in new guest interest to stay, potentially reducing the hotel's occupancy rate. On the other hand, these reviews contain valuable sentiment data for Hotel Harper to use as feedback for improvements and quality enhancements. Hotel Harper needs to utilize the sentiment analysis results from online reviews to identify priority areas and aspects of service improvement, such as room facilities, reception services, cleanliness, and others. Sentiment analysis can also provide insights into the goals or motivations of guests to visit Hotel Harper. By combining sentiment analysis results from online reviews and forecasting modeling techniques, Hotel Harper can predict future occupancy rates more accurately. Therefore, this research is important to analyze the relationship and influence of online review sentiments on Hotel Harper's weekly occupancy, so the hotel can develop more effective strategies to maintain and increase their occupancy rate.

Based on the background and problem formulation presented above, the objectives of this research are as analyze the sentiment of online reviews for Hotel Harper, identify priority areas and aspects of service improvement, determine the reasons for visiting Hotel Harper, predict occupancy rates using sentiment analysis and forecasting models, and investigate the relationship and influence of sentiment on occupancy rates. These objectives aim to provide Hotel Harper with valuable insights derived from sentiment analysis of online reviews, facilitating strategic decision-making to enhance guest satisfaction and improve occupancy rates.

METHOD

The research method encompasses a comprehensive array of elements vital for the systematic investigation undertaken. It delineates the type of research employed, which includes the qualitative exploration of literature pertinent to the concepts and methodologies used to address the issues at hand. Sources of literature review span diverse avenues, including academic journals and scholarly articles, providing in-depth insights into classification theories utilizing Naive Bayes and data mining techniques such as Arimax Model for performance forecasting (Wu et al., 2022). Additionally, the method elucidates the conceptual framework guiding the research endeavors, delineating key stages from selecting

hotel review sources to sentiment analysis using Naïve Bayes and subsequent forecasting utilizing ARIMAX.

Moreover, it outlines the meticulous process of data collection, featuring primary datasets sourced from platforms like Traveloka.com and Booking.com, subjected to scraping procedures. The collected data undergoes labeling, preprocessing, and classification to generate training and testing data sets, facilitating subsequent discussions and visualizations. Furthermore, the method section elaborates on the data sources, encompassing online reviews of Hotel Harper M. T. Haryono Jakarta, including occupancy rate analysis, chosen for their voluminous and diversified data. The temporal scope extends over four years, spanning the period before, during, and after the COVID-19 pandemic (2020-2023), allowing for a nuanced analysis of review patterns and hotel occupancy dynamics. Additionally, the method scrutinizes the impact of the COVID-19 pandemic on hotel occupancy rates, exploring fluctuations and identifying external and internal factors influencing occupancy levels. This entails visualizing trends and conducting detailed analyses to discern pandemic-induced shifts and hotel adaptation strategies. Subsequently, the sentiment analysis process is delineated, featuring the Naïve Bayes algorithm's application for classification, data labeling, preprocessing, and classification stages, evaluated using R Studio.

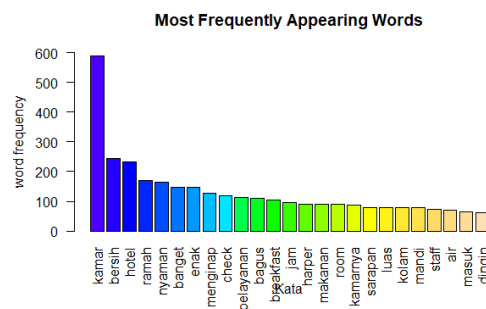
Furthermore, the forecasting model employs the Arimax Model, leveraging sentiment analysis outcomes to assign scores for subsequent analysis (Wu et al., 2022). This entails combining sentiment analysis data with occupancy rate data for forecasting, utilizing R Studio for analysis. Accuracy is evaluated using MAPE and RMSE metrics, gauging the model's forecasting precision. Additionally, Pearson correlation analysis and simple linear regression are employed to elucidate relationships between variables, facilitating a comprehensive understanding of the research domain (Ghozali, 2018).

RESULTS AND DISCUSSION

RESULT

Sentiment Analysis of Harper MT Haryono Hotel Online Review

With the steps taken in the sentiment analysis process of Hotel Harper M.T. Haryono online reviews from Traveloka, Booking, and Agoda, the following results were obtained:



Source: Research Results (2024)

Figure 1. Chart of the Most Frequently Appearing Words at Harper MT Haryono Hotel

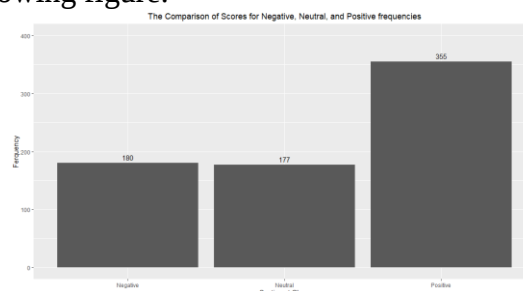
The results of the description show that the word room is the most popped-up word with a frequency of 590 followed by the words clean, hotel, and friendly by 243, 232, and 171. This shows that when assembled, most consumers leave reviews that "hotel rooms are clean" or "hotel friendly". To further find out the words that appear most often, it will then be displayed in the form of wordclouds.



Source: Research Results (2024)

Figure 2. Wordclouds Hotel Harper MT Haryono

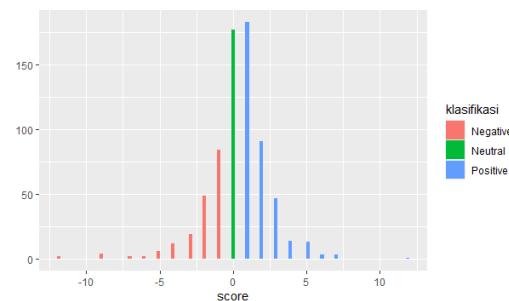
The results shown in wordclouds also display similar results where the most popping up words are room, hotel, and clean. Where these three words when assembled into "clean hotel room". Next we classify the data into 3 parts, namely negative, neutral, and positive. This result is shown in the following figure.



Source: Research Results (2024)

Figure 3. Harper MT Haryono Hotel Data Classification

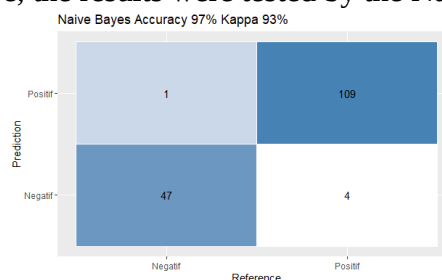
Based on figure 3, it can be seen that the sentiment class classification at Harper Traveloka Hotel is dominated by positive class of 355, neutral 177, and negative 180 reviews. With the following data dissemination:



Source: Research Results (2024)

Figure 4. Harper MT Haryono Hotel Data Classification Details

Based on figure 4, it can be seen that the spread of reviews tends to be more towards positive reviews. Furthermore, the results were tested by the Naive Bayes method as follows:



Source: Research Results (2024)

Figure 5. Confussion Matrix Naïve Bayes Hotel Harper MT Haryono

Based on figure 5 it is known that in the results of the evaluation of the model using the Confusion Matrix and related statistics, it can be seen that the prediction of the model is very satisfactory. The Confusion Matrix shows that there are 1 cases that are incorrectly predicted, namely negative-positive and 4 positive-negative, then categorized as negative by 47 negative categories, and positive by 109 successfully identified accurately. Overall, model accuracy reaches 97%, with a 93% confidence level interval, indicating a high level of confidence in model performance.

```

Accuracy : 0.9689
95% CI : (0.929, 0.9898)
No Information Rate : 0.6832
P-value [Acc > NIR] : <2e-16

Kappa : 0.9271

McNemar's Test P-Value : 0.3711

Sensitivity : 0.9216
Specificity : 0.9909
Pos Pred Value : 0.9792
Neg Pred Value : 0.9646
Prevalence : 0.3168
Detection Rate : 0.2919
Detection Prevalence : 0.2981
Balanced Accuracy : 0.9562

'Positive' Class : Negatif

```

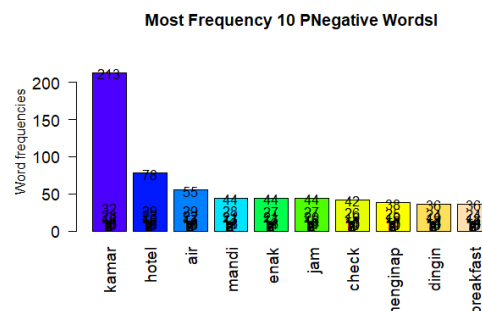
Source: Research Results (2024)

Figure 6. Overall Statistic Naïve Bayes Hotel Harper MT Haryono

Analysis of the classification model showed excellent results with an accuracy rate of 96.89%, which indicates that the model is able to classify the data correctly most of the time. The 95% confidence interval for accuracy is between 92.29% and 98.98%, indicating our level of confidence in the true value of accuracy. A high Kappa value, which is 0.9271, indicates an excellent level of agreement between the model's predictions and the actual label. The McNemar test showed no significant differences between the model and the reference method. Sensitivity of 92.16% and specificity of 99.09% indicate that the model is able to identify well both positive and negative classes (Wu *et al*, 2022). A positive prediction value of 95.83% and a negative prediction value of 96.46% indicate the reliability of the model in predicting both classes. Although the prevalence is only 31.68%, the detection rate of the model remains quite high with a detection rate of 28.57% and a prevalence detection rate of 29.19%. With a balanced accuracy of 95.62%, it can be concluded that the model is able to produce predictions that are balanced between sensitivity and specificity. Therefore, this model has excellent performance in classifying data with a class of "Negative" as a positive class (Wu *et al*, 2022).

Identify Priority Service Areas and Aspects of Improvement

After sentiment analysis in the previous subchapter, the division of groups based on class classification was carried out in order to find out what needs to be improved on service priorities. First based on the following negative reviews:



Source: Research Results (2024)

Figure 7. Frequency 10 Negative Words Harper MT Haryono Hotel

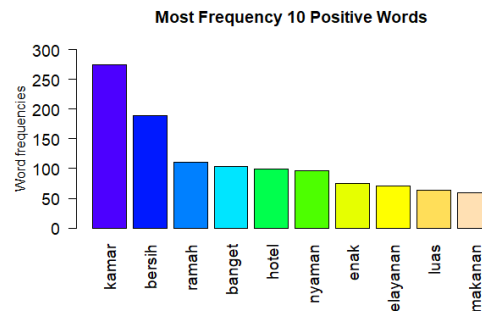
As a result of figure 7 there are 10 negative words that often appear including "room", "hotel", "water", "bath", "delicious", "hours", "check", "stay", "cold", "breakfast". For more details we use wordclouds in the image below as follows:



Source: Research Results (2024)

Figure 8. Wordclouds Negative Words Harper MT Haryono Hotel

Based on the results wordclouds also displays the same results as shown in the graph above where rooms are the main problem in negative reviews. Further based on positive reviews as follows:



Source: Research Results (2024)

Figure 9. Frequency 10 Positive Words Harper MT Haryono Hotel

As a result of figure 9 there are 10 positive words that often appear including "room", "clean", "friendly", "really", "hotel", "comfortable", "delicious", "service", "spacious", "food". For more details we use wordclouds in the image below as follows:



Source: Research Results (2024)

Figure 10. Wordclouds Positive Words Harper MT Haryono Hotel

Based on the results wordclouds also displays the same results as shown in the graph above where rooms are the most important problem in positive reviews.

Purpose of Visiting on Hotel Harper MT Haryono

The purpose of a person's stay in this study was carried out by testing the correlation of the words "Holiday" and "Work" to "Hotel". This result is shown in figure 11 as follows:

	Libur	Kerja	Hotel
Libur	1.000000000	-0.001990449	-0.01741921
Kerja	-0.001990449	1.000000000	0.08883156
Hotel	-0.017419206	0.088831556	1.000000000

Source: Research Results (2024)

Figure 11. Correlation of Visiting Visitor Destinations

Based on the results obtained in figure 11, spearman correlation results are obtained. Holiday relationship to hotel amounted to -0.01741921, this result is negative which indicates that the relationship of holiday to hotel is not linear or contrary to the number of visitors at the hotel. The working relationship to the hotel amounted to 0.08883156, this result shows that the working relationship to the hotel is linear so that the increase in hotel visitors is caused by the purpose of visitors to do work. To obtain more accurate results, regression tests were carried out in this study. The regression test results obtained in this study are as follows:

Coefficients:				
	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.32299	0.02629	12.284	<2e-16 ***
Libur	-0.32299	0.70062	-0.461	0.6449
Kerja	1.17701	0.49576	2.374	0.0179 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

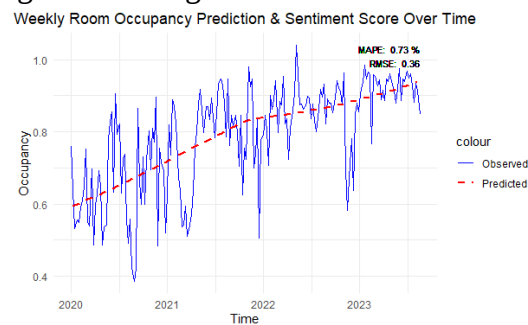
Source: Research Results (2024)

Figure 12. Regression of Visiting Destinations

Based on the results obtained in figure 12, it was obtained that the p-value in the holiday variable was $0.6942 > 0.05$ which showed that the holiday did not affect the interest in visiting the hotel. In the work variable, the p-value of $0.0119 < 0.05$ indicates that work affects the interest in visiting Harper M.T Haryono Hotel.

Occupancy Rate Prediction with Sentiment Analysis and Forecasting Model

Furthermore, namely predicting the weekly capacity of rooms at Harper MT Haryono hotels using the MAPE and RMSE Timeseries Forecasting methods. The following is the result of the data processing shown in figure 13 as follows:



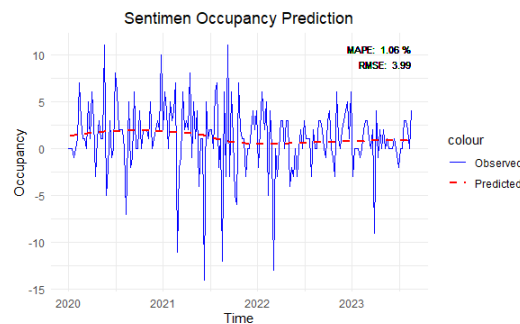
Source: Research Results (2024)

Figure 13. Weekly Room Occupancy Prediction Hotel Harper MT Haryono

Based on figure 13, it can be seen that the overall graph has increased relatively from 2020 – 2023, the result is due to the post-Covid-19 phase where there is a central government's determination in implementing PPKM. The increase is because it has returned to the normal phase so that hotel operations can work again as they should. The results in this study showed a MAPE value of 0.7332782% and an RMSE value of 0.3632695.

The MAPE value of 0.7332782% and RMSE of 0.3632695 provide a positive picture related to model performance. MAPE, which measures the average relative error between predictions and actual values, shows that the model has a good accuracy rate, with an error rate of about 0.73% (Wu *et al.*, 2022). This indicates that the model's predictions are consistently close to their true values, making them a reliable tool in the context of analysis or prediction. On the other hand, a low RMSE (0.3632695) provides a clue that the mean

error between predictions and actual values has a small magnitude (Wu *et al.*, 2022). The model effectively reproduces actual variations in the data, producing predictions close to true values. Overall, these two metrics support the conclusion that the model performs well and reliably in the context of the analysis or prediction being performed. Furthermore, predictions were made using room *occupancy sentiment* at Harper MT Haryono hotels using the MAPE and RMSE Timeseries Forecasting methods. The following is the result of the data processing shown in figure 14 as follows:



Source: Research Results (2024)

Figure 14. Sentimen Occupancy Prediction Hotel Harper MT Haryono

Based on figure 14, it can be seen that the overall relatively stable graph is considered positive from 2020 – 2023 in the post-Covid-19 phase. The stability generated by the predicted line gives an impact to the increase in *Weekly Occupancy*. Where positive reviews also have a positive influence on the increase in the predicted trend in *Weekly Occupancy*. The results in this study showed a MAPE value of 1.055445% and an RMSE value of 3.99452. The MAPE value of 1.580946% and RMSE of 3.99452 give a positive picture related to model performance. MAPE, which measures the average relative error between predictions and actual values, shows that the model has a good accuracy rate, with an error rate of about 1.58% (Wu *et al.*, 2022). This indicates that the model's predictions are consistently close to their true values, making them a reliable tool in the context of analysis or prediction. On the other hand, the low RMSE (3.99452) provides a clue that the mean error between predictions and actual values has a small magnitude (Wu *et al.*, 2022). The model effectively reproduces actual variations in the data, producing predictions close to true values. Overall, these two metrics support the conclusion that the model performs well and reliably in the context of the analysis or prediction being performed.

Relationships and Influences Sentiment Hotel towards Weekly Occupancy

The final stage is to determine whether there is a relationship and influence between using the Sentiment Hotel delta value and the *Weekly Occupancy* delta. Before the correlation test, *hotel sentiment* data in week n was used to see the relationship in weekly occupancy in week $n + 1$. This is in accordance with tests conducted by Wu., *et al* (2022). To find out whether there is a relationship between *Hotel Sentiment* and *Weekly Occupancy*, in this study a pearson correlation test was conducted. The results of the pearson correlation test obtained are as follows:

```
Pearson's product-moment correlation
data: delta$delta_weekly and delta$delta_sentimen
t = 1.5874, df = 207, p-value = 0.114
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.02644255 0.24178173
sample estimates:
cor
0.1096656
```

Source: Research Results (2024)

Figure 15. Pearson correlation test results Sentimen Hotel towards Weekly Occupancy

Based on the test results in figure 15, it was obtained that the p-value was $0.114 > 0.05$ with a correlation coefficient of 0.1096656 (weak). These results explain that there is no relationship between *Hotel Sentiment* and *Weekly Occupancy*. Furthermore, a regression test was carried out to determine the magnitude of the influence caused by *Hotel Sentiment* on *Weekly Occupancy*.

```
Call:
lm(formula = delta_weekly ~ delta_sentimen, data = delta)

Residuals:
    Min       1Q   Median       3Q      Max
-0.64613 -0.07722 -0.01575  0.06325  1.13424

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.022866   0.013945   1.640   0.103
delta_sentimen 0.009567   0.006027   1.587   0.114

Residual standard error: 0.1947 on 207 degrees of freedom
Multiple R-squared:  0.01203,    Adjusted R-squared:  0.007254
F-statistic: 2.52 on 1 and 207 DF,  p-value: 0.114
```

Source: Research Results (2024)

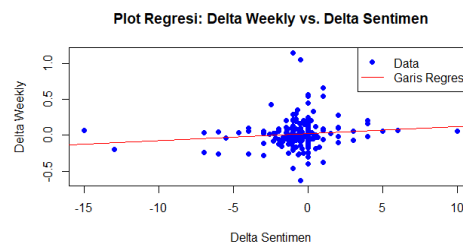
Figure 16. Regression Results Sentiment Hotel towards Weekly Occupancy

Based on the test results in figure 16, an estimate is obtained so that a regression model can be formed as follows:

$$Y = \alpha + \beta X + \epsilon$$

$$Y = 0.022866 + 0.009567 X + \epsilon$$

Based on the model, it can be explained that if the *Hotel Sentiment* value is constant (0), then the *Weekly Occupancy* value is 0.022866. If *Hotel Sentiment* increases by 1, then *Weekly Occupancy* increases by 0.009567 and vice versa. Based on the p-value obtained of $0.114 > 0.05$. These results explain that there is no significant influence between *Hotel Sentiment* and *Weekly Occupancy*. To strengthen the results of the analysis, plotting was carried out on the regression model. The results of regression plotting are as follows:



Source: Research Results (2024)

Figure 17. Plot Regression: Delta Weekly vs. Delta Sentiment

Based on the results obtained in figure 17, it can be explained that the hotel sentiment model can create an upward trend in weekly occupancy even though the results are not significant in influencing weekly occupancy.

DISCUSSION

Analysis Sentiment Online Reviews Hotel Harper MT Haryono

The sentiment analysis of the Harper M.T. Haryono Hotel's online reviews provides valuable insights into customer perceptions and experiences. By analyzing the most frequently mentioned words, such as "room," "hotel," and "clean," it's evident that users prioritize aspects like room quality and cleanliness when evaluating their stay. This aligns with common expectations for hotel experiences.

Visualizing these frequent words through word clouds further emphasizes the significance of room quality, hotel facilities, and cleanliness in user reviews, reinforcing their importance in shaping overall perceptions. The classification of reviews into negative, neutral, and positive categories reveals a predominance of positive sentiment, indicating that most users have had favorable experiences with the hotel. However, the presence of negative

and neutral reviews underscores the need for continuous improvement and attentiveness to diverse customer experiences. Employing the Naive Bayes method for sentiment prediction demonstrates high model accuracy, albeit with a single misclassification, suggesting room for refinement (Wu *et al*, 2022). Evaluating the model's performance through metrics like the Confusion Matrix highlights its overall effectiveness in predicting sentiment.

In summary, the sentiment analysis underscores the predominantly positive reception of Harper M.T. Haryono Hotel among users, with a focus on room quality and cleanliness. While acknowledging individual variations in experiences, ongoing evaluation and validation can further enhance service quality and customer satisfaction in the future.

Identify Priority Service Areas and Aspects of Improvement

The sentiment analysis provides valuable insights into user responses to Hotel Harper M.T. Haryono. To facilitate targeted improvements, it's crucial to identify priority areas for enhancement. This involves categorizing reviews into negative and positive classes. Ten frequently occurring negative words were identified, including "room," "hotel," "water," "bath," "delicious," "hours," "check," "stay," "cold," and "breakfast." Visualized through word clouds, it's evident that room-related issues dominate negative reviews, indicating that room improvements should be prioritized to enhance guest satisfaction. The correlation between these ten negative words is depicted in Figure 18.

skamar	deluxe	tangan	menunggu	benarbenar
	0.52	0.45	0.44	0.43
berfikir		disemprot	disinfektan	front
0.43	0.43	0.43	0.43	0.43
kinerja	membersihkan	office	tlp	
0.43	0.43	0.43	0.43	
tsb	menit	kalinya	pesan	
0.40	0.40	0.39	0.38	
shampo	antri	disana	superior	
0.38	0.37	0.36	0.36	
kunci	bikin	ngasih	luas	
0.36	0.36	0.36	0.35	
nunggu	app	brand	dial	
0.34	0.34	0.34	0.34	
dialami	diinfokan	dipahami	ditelpon	
0.34	0.34	0.34	0.34	
diturunkan	drastis	harinya	keesokan	
0.34	0.34	0.34	0.34	
kota	miring	operator	pengen	
0.34	0.34	0.34	0.34	
seminggu	suasana	sungguh	temanteman	
0.34	0.34	0.34	0.34	

Source: Research Results (2024)

Figure 18. Most Negative Word Correlation

The correlation analysis unveils significant insights regarding the word "room". Firstly, a strong correlation exists between "room" and "smoke", indicating many occupied rooms are by smokers, potentially discomforting other guests. Secondly, the correlation with "age" implies the available rooms might be old and in need of renovation. Additionally, the correlation with "next to each other" suggests adjacent rooms might compromise guests' privacy. Lastly, the correlation with "wait" indicates guests often experience delays during check-in or check-out, leading to inconvenience. These negative sentiments associated with "room" suggest potential failure to meet expectations or promised service levels by the hotel.

- 1 Guest Comfort and Privacy: Inconvenient room locations or smoke issues may compromise guest comfort and privacy.
- 2 Speed and Efficiency of Service: Long wait times for service or troubleshooting can lead to guest dissatisfaction.
- 3 Quality of Facilities: The condition of facilities that are old or poorly maintained can reduce the quality of the guest's stay.

Hotels need to act on these issues immediately to ensure SLA compliance and improve guest satisfaction:

- 1 Room Location: Consider offering room relocation options to guests who are bothered by their room location.
- 2 Ventilation and Hygiene: Improve ventilation and hygiene to prevent smoke and odor problems.

- 3 Speed of Service: Optimize service processes to reduce wait times and speed up troubleshooting.
- 4 Facility Maintenance: Perform regular maintenance and renovation if needed to ensure the condition of the facility remains good.

Addressing these issues can enhance guest satisfaction, bolster the hotel's reputation, and ensure compliance with promised service level agreements. Moving on, positive reviews were analyzed to identify areas of excellence. Ten frequently mentioned positive words include "room", "clean", "friendly", "really", "hotel", "comfortable", "delicious", "service", "spacious", and "food". While rooms continue to receive positive attention, other aspects like cleanliness, friendly service, and room quality were also highlighted positively by users.

skamar	dikamar	kekurangan	telepon	terdengar
	0.59	0.58	0.57	0.55
apartemennya	berbunyi	bermasalah	bnk	
	0.54	0.54	0.54	0.54
breafing	cape	cna	daftarkan	
	0.54	0.54	0.54	0.54
dengungan	dengungannya	dimana	dipindahkan	
	0.54	0.54	0.54	0.54
dkt	drttd	ganti	hrs	
	0.54	0.54	0.54	0.54
improve	kendaraannya	ketemu	krn	
	0.54	0.54	0.54	0.54
kualitas	lambat	lbh	mengambilkan	
	0.54	0.54	0.54	0.54
mengangkat	mengganggu	mengganggu	mineral	
	0.54	0.54	0.54	0.54
msh	mudahan	penghuni	penyelesaian	
	0.54	0.54	0.54	0.54
penyelesaiannya	pikir	plat	sandal	
	0.54	0.54	0.54	0.54
sanitizer	sekeras	sempet	seprot	
	0.54	0.54	0.54	0.54

Source: Research Results (2024)

Figure 19. Most Positive Word Correlation

Based on the correlation analysis, the word "room" demonstrates a strong correlation with "children," suggesting many occupied rooms are by families, making it a child-friendly or family-friendly hotel unit. Additionally, the correlation with "people" indicates the hotel has staff who are friendly, helpful, and responsive to guest needs. The correlation with "advice" suggests the hotel welcomes suggestions to improve room quality. Moreover, correlations with "application" and "booking" imply the hotel offers convenient booking options through an application, facilitating access to other hotel information. These positive sentiments suggest the hotel has met or exceeded guest expectations in various aspects of the service level agreement:

- 1 Family-Friendly Facilities and Services: The hotel provides adequate facilities and services for families with children, in accordance with SLAs that promise a comfortable and enjoyable stay for all guests.
- 2 Staff Friendliness and Professionalism: The hotel's friendly, helpful, and responsive staff demonstrate the hotel's commitment to high-quality service, in accordance with SLAs that guarantee friendly and professional service.
- 3 Ease of Booking: The mobile app and easy booking process show that the hotel has met SLAs that ensure ease and efficiency in the booking process.

Based on the findings, it's evident that room improvements should be the primary focus, necessitating actions such as revamping facilities, enhancing hygiene standards, upgrading bathing facilities, and addressing potential odor issues. Additionally, attention should be given to other positively noted aspects, such as enhancing friendly service, maintaining cleanliness, and ensuring room quality. To tackle these improvement challenges, the management of Harper M.T. Haryono Hotel can implement concrete measures like facility renovations, staff training for better service delivery, and enhancing the overall quality management system. By prioritizing room enhancements and taking decisive actions to improve other service aspects, the hotel aims to significantly boost user satisfaction, fortify its reputation, foster customer loyalty, and ensure long-term business viability.

Purpose of Visiting Harper M.T. Haryono Hotel

In this study, research was conducted to find out the purpose of visiting Harper MT Haryono Hotel by considering keywords such as "Holiday" and "Work". The data is analyzed

through correlation and regression methods to gain a deeper understanding. First, it is necessary to pay attention to the correlation results between these variables, as shown in Figure 11. Spearman's correlation shows that the relationship between the keywords "Holidays" and "Hotels" has a coefficient of -0.01741921. This negative value indicates that there is no linear relationship between holidays and the number of hotel visitors. That is, the presence of guests in the hotel is not directly influenced by the holiday factor. On the other hand, the relationship between the keywords "Work" and "Hotel" shows a coefficient of 0.08883156. This positive value indicates a linear relationship between work and the number of hotel visitors. This suggests that a portion of hotel visitors have a purpose for work, which is most likely influencing the increase in the number of visitors. However, to ensure the accuracy of the results, regression tests are performed as listed in Figure 12. The results showed that the variable "Holiday" did not have a significant influence on interest in visiting hotels, with a p-value of 0.6942 which is greater than 0.05. That is, holidays do not significantly affect the interest in visiting hotels. It can be concluded that hotel guests at the M.T.Haryono hotel do not stay at the hotel for tourist purposes or on vacation, this may be due to the location of the hotel is not close to tourist attractions.

Meanwhile, the variable "Work" has a significant influence on interest in visiting hotels, with a p-value of 0.0119 which is smaller than 0.05. This shows that the purpose for work affects the interest in visiting the hotel significantly. This is in line with Tambunan E's explanation (2022) which states that the Cawang Area is the main business and industrial area in Jakarta. With many offices and business centers in the Cawang area, making Harper M.T Haryono Hotel one of the destinations to stay because of its strategic location and easy access. Thus, from the overall results of the analysis, it can be concluded that the main purpose of people visiting Harper MT Haryono Hotel is for work. Although some visitors may visit for work reasons, the effect is not as great as the purpose for staying. This demonstrates the importance of understanding visitor motives in the context of a particular hotel to inform more effective marketing strategies and service offerings.

Occupancy Prediction with Sentiment Analysis and Forecasting Model

In this analysis, we will discuss the results of the prediction of weekly room capacity and *Hotel Sentiment* at Harper MT Haryono Hotel using the Timeseries Forecasting method, focusing on the Average Percentage Error (MAPE) and Root Mean Square Error (RMSE) Methods as evaluation metrics. First, the results of the prediction of weekly room capacity at Harper MT Haryono Hotel as shown in Figure 13. The graph shows the overall increase from 2020 to 2023. This increase occurred after the Covid-19 period, where the central government implemented Restrictions on the Implementation of Community Activities (PPKM). This increase indicates that the hotel has returned to its normal phase of operations, so the number of visitors has increased again. The results of this prediction show that MAPE has a value of 0.7332782% and RMSE of 0.3632695. A low MAPE indicates a good level of accuracy of the prediction model. With an error rate of about 0.73%, the model is reliable in predicting weekly room capacity well (Wu et al., 2022). In addition, a low RMSE indicates that the model can reproduce actual variations in the data well, so that its predictions are effectively close to its true values.

Furthermore, the predicted results of *Occupancy Sentiment* at Harper MT Haryono Hotel are reviewed as shown in Figure 1 The graph shows overall stability from 2020 to 2023, with a trend towards positive values post Covid-19. This stability is influenced by positive sentiment which has a positive impact on the increase in the weekly occupancy prediction trend. The results of this prediction show MAPE of 1.055445% and RMSE of 3.99452. Although the MAPE and RMSE values are slightly higher compared to the predicted room capacity, they are still within acceptable ranges. A MAPE of 1.06% indicates that the model's predictions are consistently close to their true values with a low error rate.

While the relatively low RMSE indicates that the model can produce predictions close to the actual value well. Overall, the results of both predictions show good performance from the Timeseries Forecasting model used in predicting room capacity and *Hotel Sentiment* at Harper MT Haryono Hotel. Although there are some prediction errors, the high level of accuracy indicates that this model can be a useful tool in future analysis and prediction.

Relationships and Influences Hotel Sentiment towards Weekly Occupancy

The final stage in this analysis is to determine whether there is a relationship or influence between *Hotel Sentiment* and *Weekly Occupancy*. To answer this question, a Pearson correlation test and a regression test were performed. The results of the Pearson correlation test show a p-value of 0.114, which is greater than 0.05. This result indicates that there is no significant relationship between *Hotel Sentiment* and *Weekly Occupancy*. The Pearson correlation measures the strength and direction of the linear relationship between two variables. In this context, the results obtained show that there is no significant linear relationship between *Hotel Sentiment* and *Weekly Occupancy*. After the correlation test, a regression test was carried out to determine how much influence *Hotel Sentiment* had on *Weekly Occupancy*. The results of the regression test show that the estimated regression coefficient (β) is 0.009567. From the regression model formed, it can be concluded that when *Hotel Sentiment* increases by 1-unit, *Weekly Occupancy* will increase by 0.009567 units, and vice versa. However, the p-value of the regression test is also 0.114, which is greater than 0.05. This shows that there is no significant influence between *Hotel Sentiment* and *Weekly Occupancy* based on the regression model created.

Based on the results of the Pearson correlation test and regression test, it can be concluded that there is no significant relationship or influence between *Hotel Sentiment* and *Weekly Occupancy*. This suggests that other factors may have a more dominant role in influencing *Weekly Occupancy*, and *Sentiment Hotel* may not be a strong predictor in determining *Weekly Occupancy* rates. There is no relationship between *Sentiment Hotel* and *Weekly Occupancy* because guests staying at Hotel M.T Haryono who stayed at the hotel did not make reservations through the website. In addition, the tendency of existing hotel guests rarely gives reviews on hotels that have been occupied. In addition, according to research from Syah and Indriani (2020), it is stated that consumers assume that reviewers are not completely honest if they only make predominantly positive or negative comments. Therefore, to understand the factors that influence *Weekly Occupancy* better, it is necessary to conduct further research by considering other variables that may contribute.

CONCLUSION

The online review sentiment analysis of Harper M.T. Haryono Hotel shows a positive response from users towards the hotel, with key findings highlighting users' attention to room quality and hotel cleanliness as key factors in accommodation ratings. Although the majority of reviews tend to be positive, the presence of negative and neutral reviews indicates room for service improvement. The identification of priority service areas and aspects to be improved for Hotel Harper M.T. Haryono is based on sentiment analysis from online reviews, covering several key aspects, particularly room quality and cleanliness. Concrete steps that the Harper Hotel management can take include renovating or repairing damaged facilities, enhancing staff training to provide better service, and improving the overall quality management system. Additionally, this research highlights the importance of predicting weekly room capacity and sentiment analysis of reviews in hotel management. Through the Timeseries Forecasting method, the trend of increasing room capacity of Harper MT Haryono Hotel from 2020 to 2023 was successfully predicted with good accuracy. The analysis also revealed a significant correlation between review sentiment and weekly room capacity, confirming the important role of sentiment analysis in understanding consumer behavior and

planning more effective marketing strategies. However, no significant relationship was found between Hotel Sentiment and Weekly Occupancy. This suggests that other factors may have a more dominant role in determining the level of Weekly Occupancy, while Hotel Sentiment may not be a strong predictor in this context. Therefore, further research considering other variables that may contribute to understanding the factors affecting Weekly Occupancy is recommended.

REFERENSI

- Ghozali, I. (2018). *Aplikasi Analisis Multivariate dengan Program IBM SPSS 25*. Semarang: Badan Penerbit Universitas Diponegoro.
- Jang, S., & Moutinho, L. (2019). Do price promotions drive consumer spending on luxury hotel services? The moderating roles of room price and user-generated content. *International Journal of Hospitality Management*, 78, 27–35. <https://doi.org/10.1016/j.ijhm.2018.11.010>
- Nicolau, J. L., Xiang, Z., & Wang, D. (2024). Daily online review sentiment and hotel performance. *International Journal of Contemporary Hospitality Management*, 36(3), 790–811. <https://doi.org/10.1108/IJCHM-05-2022-0594>
- Roy, G. (2023). Travelers' online review on hotel performance – Analyzing facts with the Theory of Lodging and sentiment analysis. *International Journal of Hospitality Management*, 111, 103459. <https://doi.org/10.1016/j.ijhm.2023.103459>
- Syah, M. A., & Indriani, F. (2020). Analisis Pengaruh Online Customer Review terhadap Minat Beli Konsumen pada Produk Kosmetik (Studi Kasus pada Pengguna Famaledaily Indonesia). *Diponegoro Journal of Management*, 9(3), 1–9.
- Tambunan, E. (2022). Analisis Kondidi Lalu Lintas Jalan Mayjend Sutoyo Cawang. *Jurnal Rekayasa Teknik Sipil dan Lingkungan-CENTECH*, 3(2), 99–107.
- Widyaningrum, B., Helbawanti, O., Triyanto, S. A., Nuraini, C., & Mutolib, A. (2023). Permodelan Service Level Agreement (SLA): Upaya Mendukung Kelulusan Tepat Waktu Mahasiswa di Perguruan Tinggi. *Prosiding SEMDIKJAR (Seminar Nasional Pendidikan dan Pembelajaran)*, 6, 648–661.
- Wu, D. C., Zhong, S., Qiu, R. T. R., & Wu, J. (2022). Are customer reviews just reviews? Hotel forecasting using sentiment analysis. *Tourism Economics*, 28(3), 795–816. <https://doi.org/10.1177/13548166211049865>