

Unveiling Risks through Machine Learning: Analyzing Indonesian User Feedback Dataset of Capsule Hotel Experiences

Yehezkiel Gunawan¹, Ford Lumban Gaol² and Tokuro Matsuo³ ¹Information System Management Department, BINUS Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia

²Information System Management Department, BINUS Graduate Program – Master of Information System Management, Bina Nusantara University, Jakarta, Indonesia

³Advanced Institute of Industrial Technology, Japan (Director of Research Center for AI and Service Science, Advanced Institute of Industrial Technology, Japan / Director, Professional Education Center, Advanced Institute of Industrial Technology, Japan)

E-mail: ¹*yehezkiel.gunawan@binus.ac.id,* ²*fgaol@binus.edu,* ³*matsuo@aiit.ac.jp*

Abstract

The rise in popularity of capsule hotels as a unique and affordable lodging alternative, especially in Indonesia, has highlighted the necessity of skillfully recognizing and controlling any potential risks connected with such unusual lodgings. This paper introduces the large collection of 700 data examples that includes priority scores, problem areas, and verbatim user comments. Furthermore, we conduct a two-phase experiment using the Random Forest algorithm to classify risks. In the first stage, a custom BERT model for word embedding is integrated, and in the second stage, the pre-trained Indo LEM (BERT) model is used. Our results clearly demonstrate the higher effectiveness of the second step, demonstrating how the addition of Indo LEM as word embedding considerably improves classification accuracy. This demonstrates the enormous potential of utilizing cutting-edge machine learning techniques to improve risk classification processes, providing players in the capsule hotel industry with priceless information to improve safety regulations and better the overall guest experience. At (https://github.com/yehezkielgunawan/thesis-risk-classification), we provide full access to all relevant coding scripts for reference and replication as an addition to the dataset.

Keywords: Risk Management, Risk Classification, Machine Learning, Natural Language Processing, Hospitality

1. Introduction

Travelers now have access to affordable and convenient sleeping quarters because of the growing popularity of capsule hotel [1]. Due to its unique idea of offering guests compact, cheap sleeping accommodations, capsule hotels have become very popular in the hospitality sector. For tourists on a tight budget who value comfort and usefulness, these places present a sensible solution. Capsule hotels respond to the needs of contemporary tourists looking for economical and effective lodging solutions by making the most of the limited space available [2].

But as these remote locations gain in popularity, there is a growing desire to comprehend and manage the possible risks related to capsule hotel experiences [3]. There are inevitable challenges regarding to safety, security, privacy, and overall guest pleasure when there are a variety of visitors and shared amenities. For guests to have a happy and secure experience as well as to sustain the reputation and



success of capsule hotels in the hospitality sector, it is essential to recognize and mitigate these hazards.

The ability to detect and categorize these risks through risk categorization enables hoteliers to implement the proper measures to minimize risk management [4]. Hotel operators can systematically identify and analyze the numerous risks associated with capsule hotel experiences through the core process of risk categorization. Operators can implement focused and efficient risk mitigation measures by precisely analyzing the nature and severity of these threats [5]. By taking a proactive approach, they may deal with potential issues before they become more serious, ensuring both the guests' safety, security, and satisfaction and the operational integrity of the capsule hotel [3].

In this study, we provide an extensive set of customer feedback consisting of 700 rows of data from capsule hotel experiences in Bahasa Indonesia, which is an invaluable source for machine learning risk classification. Using the dataset, we can analyze user input and derive insightful information that helps us classify and detect the dangers connected to capsule hotels. We are seeking to build a strong risk classification framework for capsule hotels that improves security measures and elevates the entire visitor experience by utilizing cutting-edge machine learning techniques like BERT and Random Forest. We hope that this research will contribute to ongoing efforts by Indonesian hotel operators to efficiently manage and reduce hazards in the sector.

2. Related Works

Capsule hotels are a unique and innovative accommodation concept that has gained significant popularity among both tourists and business travelers [6]. These hotels offer compact sleeping spaces, often resembling individual pods, providing a functional and minimalistic approach to lodging [7]. The trend of capsule hotels has emerged as a response to the increasing demand for affordable and convenient accommodation options in urban areas, catering to travelers seeking efficiency, affordability, and a novel lodging experience [8].

In several research fields, such as hotel evaluations and sentiment analysis, the availability of comprehensive datasets is essential. However, it becomes clear that there aren't any representative statistics for certain niche markets, such capsule hotels. This drawback is particularly apparent in the context of Indonesia's capsule hotels, where no existing dataset fully captures the distinctive experiences and feedback of visitors to this type of lodging [9].

Ghazi did a study on the motives and review components that drive guests to leave positive and negative reviews on Trip Advisor [10]. The study discovered variances in the motivations behind good and negative ratings by examining primary data gathered through an online poll from Trip Advisor users who assessed Egyptian 5-star hotel services. The research found that giving back to the hotel and receiving social advantages encouraged visitors to leave favorable ratings. Contrarily, submitting unfavorable evaluations was positively impacted by the ability to let out bad emotions, warn other customers, and reap social rewards. Additionally, the study found that review components, such as service quality, cleanliness, value for money, facilities, location, staff behavior, and food quality, had a positive influence on writing both positive and negative reviews.

ArCAR (Arabic text computer-aided recognition), on the other hand, is a revolutionary deep learning approach Muaad et al. devised for modeling and recognizing Arabic text at the character level [11]. Their system, ArCAR, created a 2D array for text recognition by using 1D vectors to represent Arabic characters. ArCAR showed remarkable performance in cross-validation testing and assessments on Arabic text document categorization and sentiment analysis. Using the

Alarabiya-balance dataset, ArCAR was able to classify documents with good levels of accuracy, recall, precision, and F1-score. Similar to this, ArCAR demonstrated impressive accuracy and F1-score in Arabic sentiment analysis utilizing the hotel Arabic reviews dataset (HARD) balance dataset.

Arabic reviews dataset (HARD) balance dataset. While Ghazi's research [10] provides insight into the factors and motivations that influence Trip Advisor hotel evaluations, Muaad et al.'s [11] ground-breaking ArCAR system advances the field of Arabic text classification. Nevertheless, despite these insightful studies, it is difficult to comprehend and analyze user feedback in this special lodging niche because there isn't a dataset that exclusively addresses capsule hotel experiences in Indonesia.

The creation of automatic text summarization systems is difficult, especially because large, publically available datasets are hard to come by and are generally sparse. Low resource languages like Indonesian make this issue much worse. Kurniawan and Louvan overcome this problem, nevertheless, by proposing INDOSUM, a brand-new benchmark dataset created especially for summarizing Indonesian literature [12]. INDOSUM, which includes news items and manually created summaries, is roughly 200 times larger than prior datasets in the same domain [13]. Using this dataset, the authors analyze various extractive summarization techniques. The results are encouraging and show the dataset's relevance while also offering helpful baselines for future research. INDOSUM is an important tool for developing research in Indonesian text summarizing because the code and dataset are made available under permissive licenses [14].

Despite being widely spoken, the Indonesian language has received relatively little focus in the study of natural language processing (NLP). This is due to difficulties such a lack of standardized resources, a lack of annotated datasets, and a lack of linguistic resources. Koto et al. present the INDOLEM dataset, which includes seven tasks covering morpho-syntax, semantics, and discourse for the Indonesian language [15], in order to solve these shortcomings. They also provide INDOBERT, a trained language model created exclusively for Indonesian NLP tasks. Modern performance on a range of tasks is shown in the tests utilizing INDOBERT on the INDOLEM dataset [16]. In addition to addressing the lack of resources, this extensive benchmark dataset and the INDOBERT language model offer an invaluable tool for scholars to examine the patterns and traits of the Indonesian language [17].

There's a study on the classification of emotions in Indonesian tweets was examined in the literature review [18]. It emphasised the benefits of extracting opinions from Twitter, but it also made note that the dataset utilised lacked specificity and was unrelated to customer reviews of capsule hotels. Although the dataset was good at categorising emotions based on overall sentiments, it might not accurately capture feelings unique to encounters at capsule hotels. As a result, there is a need for a more focused and industry-specific dataset that aims to gather customer opinions and feelings about capsule hotels. For this particular topic, a dataset like that would allow for more precise and contextually appropriate emotion classification.

In the meantime, sentiment analysis has been used in the field of digital banking to evaluate client happiness using Twitter data [19]. Various classifiers and ensemble approaches were used in the study, with SVM showing higher performance in predicting feelings. This study emphasises the value of sentiment analysis for determining customer satisfaction in the field of digital banking and demonstrates the potential of sentiment classification using machine learning techniques.

An further study investigating the effects of text pre-processing on sentiment analysis found that careful feature selection and representation significantly



increased the accuracy of SVM-based sentiment analysis [20]. Surprisingly, this accomplishment is noteworthy because sentiment analysis is frequently thought to be more difficult than subject categorization. These integrated research projects highlight the crucial function of domain-focused datasets and emphasise the applicability and importance of the dataset in determining consumer attitudes and perceptions in the context of capsule hotels.

3. Data Description

The customer input on capsule hotel experiences, coupled with the related labels for priority score and problem domain, make up the dataset used for this study. The labels has been verified using the expert judgement. Higher numbers denote more urgent issues. The priority_score label gives the feedback's level of urgency. The dataset's priority score labels are distributed as follows: Priority score 1 is assigned to 148 feedback instances, Priority_score 2 to 275 feedback instances, and Priority 3 to 277 feedback instances.

According to the problem_domain label, the feedback is divided into operational and technical domains. The operational domain is represented by a value of 0, whereas the technical domain is represented by a value of 1. The dataset's problem_domain labels are distributed as follows: 130 feedback incidents are classified as technical, while 570 are operational.

This dataset that contains user feedbacks in Bahasa Indonesia allows for the classification and analysis of user comments based on both priority score and problem domain, providing useful insights into user feedback on capsule hotel experiences. Researchers and practitioners can explore risk classification and mitigation strategies within the context of these particular domains thanks to the diverse distribution of priority_score and problem_domain labels (see Table 1), which provides a comprehensive representation of the various aspects and challenges associated with capsule hotels.

Table I. Dataset Label Distribution						
priority_score			problem_domain			
1	2	3	0	1		
148	275	277	570	130		

at Lahal Distrike D /

4. Results and Discussion

The purpose of this study is to demonstrate the merits of a dataset built especially for categorizing risks in the capsule hotel sector using various machine learning techniques. The dataset includes user reviews of capsule hotel experiences in Bahasa Indonesia together with labeling on risk indicators including priority scores and problem domains. This dataset will be used in the research to show how various machine learning algorithms can categorize and identify potential threats in the capsule hotel sector.

Two experiments will be run as part of the research for feedback analysis. In the first case, the dataset will be imported and preprocessed with the use of BERT, a potent language model, to create unique word embeddings that take contextual information into account. To ensure unbiased evaluation, the preprocessed dataset will be split into training and testing sets (70% training data and 30% test data). Based on priority rankings and problem areas, machine learning techniques, such as Random Forest, will be used to classify the data and detect potential threats. The outcomes of the categorization will be evaluated using measures including accuracy, precision, recall, and F1-score.



The project will use pre-trained BERT Indo LEM word embeddings for feedback analysis in the second scenario. The dataset will go through the same preprocessing procedures, then be split into training and testing sets. Once more, classification will be performed using machine learning techniques; such as Random Forest, and the effectiveness of the approach will be measured using evaluation criteria. The general risk classification flow can be seen in Figure 1.

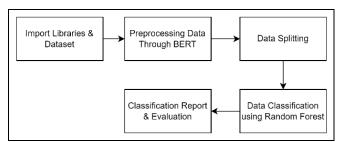


Figure 1. Experimental Risk Classification Flow

This study uses two different scenarios to categorise customer reviews of capsule hotels. In the first case, BERT is used to create unique word embeddings, and in the second, pre-trained BERT Indo LEM word embeddings are used. With consistent configurations for both circumstances, the experiment will employ Grid Search for hyperparameter tuning and Random Forest as the classification algorithm. Both of hyperparameter config can be seen in Table 2 and Table 3.

In the first scenario, illustrated in Figure 2, we will gather Play Store user feedback information on capsule hotels by scrapping the dataset using Python and Google Play Scrapper library. Following preprocessing to extract the textual content from the scraped user feedback, we will utilize BERT, a potent language model, to generate unique word embeddings for the words in the feedback. We will be able to capture the subtleties and sentiments stated by consumers in their feedback thanks to these bespoke word embeddings, which will contain contextual information. Our goal is to improve the comprehension and representation of the user feedback data using BERT-based word embeddings, which will make the process of risk assessment using machine learning algorithms like Random Forest even easier. Based on the mood and context provided in the user feedback, our method enables us to efficiently identify and categorize potential dangers related to capsule hotel experiences.

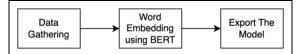


Figure 2. The Creation of Custom Word Embedding using BERT

Preprocessing of the text data will exclude stemming/lemmatization, stopword removal, and tokenization because of BERT usage can gather the text context without the traditional preprocessing. The preprocessed feedbacks will be used to train a unique word embedding model based on BERT, and the model's output will be used to generate the word vocabulary. The best hyperparameter setup for the Random Forest classifier will be discovered using Grid Search. On a different test dataset, the classifier's performance using the unique BERT word embeddings will be assessed.

We will apply the pre-trained Indo LEM model of BERT for the second case. The preparation of the text data will be similar to that of Scenario 1. Word embeddings for the preprocessed user feedbacks will be created using the pre-trained BERT

model. To identify the ideal hyperparameter setting for the Random Forest classifier, Grid Search will be used once more. On a different test dataset, the pre-trained Indo LEM word embeddings will be used to assess the classifier's performance

The hyperparameter configuration for the priority_score classification task is shown in Table 2. The hyperparameters describe various adjustments that can be made to the Random Forest algorithm's settings to improve its effectiveness when identifying the priority ratings of user feedback. The number of decision trees to be employed in the Random Forest ensemble is indicated by the parameter "n_estimators". We explore three alternative values: 100, 200, and 300. With options of None, 5, and 12, the "max_depth" parameter specifies the maximum depth of the decision trees, with "None" allowing the trees to grow until they have the fewest samples in each leaf node.

Last but not least, the "min_samples_split" parameter, which has possible values of 5, 7, and 10, specifies the minimum number of samples needed to split an internal node. By experimenting with different combinations of these hyperparameters, we hope to find the configuration that performs the best in accurately classifying the priority scores of user feedback. This will allow us to make decisions that are wellinformed and effectively address potential risks in the capsule hotel industry.

bie 2. Hyperparameter betup 1 of phoney_s				
Parameter	Possible Values			
n_estimator	[100, 200, 300]			
max_depth	[None, 5, 12]			
min_samples_split	[5, 7, 12]			

 Table 2. Hyperparameter Setup For priority_score

The hyperparameter configuration for the problem_domain classification job is shown in Table III. The Random Forest technique is used to classify the problem areas of user feedback, much like the priority_score classification. The "n_estimators" parameter, with values of 25, 75, and 125, indicates the number of decision trees utilized in the Random Forest ensemble. With options of None, 5, and 12, the "max_depth" parameter determines the maximum depth of the decision trees, with "None" allowing the trees to grow until they have the fewest samples in each leaf node.

The "min_samples_split" parameter, which can have values of 2, 5, and 10, defines the least number of samples necessary to split an internal node. The goal is to find the ideal configuration that performs better in accurately classifying the problem domains of user feedback, enabling the identification and effective mitigation of potential risks in the capsule hotel industry, by systematically varying these hyperparameters.

 te et hyperparameter betap i er problem_at				
Parameter	Possible Values			
n_estimator	[25, 75, 125]			
max_depth	[None, 5, 12]			
min_samples_split	[2, 5, 10]			

Table 3. Hyperparameter Setup For problem_domain

The performance of the classification model in the first scenario, categorizing priority_score using the hyperparameter setup 'max_depth': 5, 'min_samples_split': 7, and 'n_estimators': 200, is reflected in the precision, recall, accuracy, and F1-score values. The highest evaluation scores in this label, in the first scenario, are found in the priority_score label 2, which has an F1-score of 49%, providing quite good values for both precision and recall.



In the first scenario, the problem domain category that distinguishes between operational and technical domains uses the hyperparameter setup 'max_depth': 5, 'min_samples_split': 2, and 'n_estimators': 75. Relevant information can be seen in Table 4.3, which provides the results related to the Classification Report for the problem_domain label. Precision and recall generated lean more towards label 0. In contrast, when the model tries to predict label 1, it experiences a significant performance decrease, especially in precision and recall.

Switching to the second scenario, the prediction results for the priority_score category do not show a significant improvement; there is even an impression of a decrease in performance for predicting labels 1 and 2. It should be noted that Scenario 2 uses IndoBERT, which employs BERT as word embeddings and a pre-trained model. The hyperparameter setup used in this case is 'max_depth': 5, 'min_samples_split': 5, and 'n_estimators': 300. Performance improvement in the priority_score category is observed in the prediction results for label 3, which, in terms of distribution, has the highest number of instances. Therefore, it is not surprising that the classification model created becomes more sensitive to label 3. A precision value of 41% and a recall value of 48% for label 3 show a promising performance improvement, resulting in an F1-Score of 47%.

Still in the second scenario, the problem_domain category also experiences a similar situation to the priority_score category. The optimal hyperparameters used in this case are 'max_depth': 5, 'min_samples_split': 10, and 'n_estimators': 25, resulting in improved prediction performance, especially in terms of precision and recall. Label 0, which represents operational-related reviews, shows an improvement in prediction results reflected in precision, recall, and F1-score. This performance improvement indicates that the model can better predict user reviews of capsule hotels, categorizing them into operational or technical domains.

The overall performance of the model does show improvement after being combined with pre-trained IndoBERT. However, on the flip side, the imbalanced nature of the dataset has a significant impact on the developed model. The results are illustrated in Figure 3. Predictions for labels that are minorities tend to experience mispredictions. Nevertheless, with the combination of IndoBERT and Random Forest, there is a slight improvement in the accuracy of the developed model's predictions.

In general, there is a finding that the combination of IndoBERT as a pre-trained model with Random Forest appears to be quite effective in improving the risk classification performance. This is evident in metrics such as precision, recall, F1-Score, and accuracy. The Indonesian vocabulary from IndoBERT used in the pre-trained model proves to be effective in recognizing character or string inputs from user reviews of capsule hotels written in Indonesian (Koto et al., 2020). With this, the results of machine learning models for risk classification become better, which is undoubtedly very helpful for hotel owners in the decision-making process.

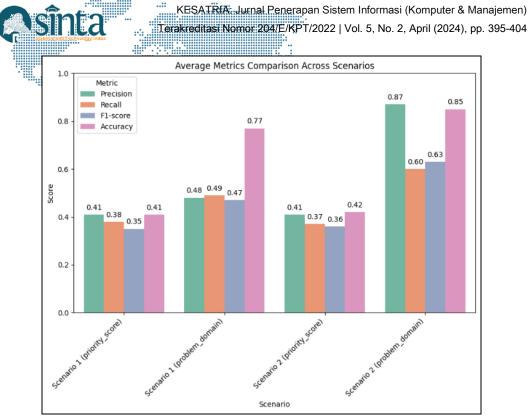


Figure 3. Average Metrics Comparison Across Scenarios

5. Conclusion

There are several conclusions that can be drawn from this research. The machine learning model created tends to be sensitive to labels that are in the majority. The pre-trained model from IndoLEM is quite helpful in improving the model's performance when combined with other machine learning algorithms such as Random Forest.

Although the dataset in this research is relatively small, consisting of only 700 rows of data, it can still be used for implementation and the creation of machine learning models, especially in the field of risk classification. It is hoped that with this dataset, it can contribute to further research related to risk classification that utilizes machine learning in the hospitality industry in Indonesia.

A suggestion for further research is to explore different machine learning approaches to enrich the perspectives of the implementation results. Additionally, there is a need for an increase in the number of data rows, making the dataset richer, and enabling the developed machine learning models to achieve better performance. With more data, it is expected that the model results will have a higher level of prediction accuracy.

References

- [1] Amornpornwiwat, N., And Kapasuwan, S. 2018. *Chapter 5: Tourists*' Perceptions Of And Intentions-To-Stay At A Capsule Hotel In Bangkok. (Https://Doi.Org/10.1108/S1871-317320180000015010).
- [2] Andrian, B., Simanungkalit, T., Budi, I., And Wicaksono, A. F. 2022. "Sentiment Analysis On Customer Satisfaction Of Digital Banking In Indonesia," International Journal Of Advanced Computer Science And Applications (13:3). (Https://Doi.Org/10.14569/Ijacsa.2022.0130356).
- [3] Anita, T. L., Pratomo, A., And Subakti, A. G. 2019. Effects Of Product Uniqueness On Re-Purchase Intention Case Study At Kini Capsule Hotel Jakarta. (Https://Doi.Org/10.2991/Isot-18.2019.75).



- [4] Chen, H. J., Wong, S. W., Bilgihan, A., And Okumus, F. 2020. "Capsule Hotels: Offering Experiential Value Or Perceived As Risky By Tourists? An Optimum Stimulation Level Model," International Journal Of Hospitality Management (86). (Https://Doi.Org/10.1016/J.Ijhm.2019.102434).
- [5] Chen, S., And Wei, H. 2022. "Minimalism In Capsule Hotels: Enhancing Tourist Responses By Using Minimalistic Lifestyle Appeals Congruent With Brand Personality," Tourism Management (93). (Https://Doi.Org/10.1016/J.Tourman.2022.104579).
- [6] Ghazi, K. M. 2017. "Guests' Motives To Write Positive And Negative Hotel Reviews On Trip Advisor," Journal Of Tourism & Hospitality (06:03). (Https://Doi.Org/10.4172/2167-0269.1000283).
- [7] Haddi, E., Liu, X., And Shi, Y. 2013. "The Role Of Text Pre-Processing In Sentiment Analysis," In Procedia Computer Science (Vol. 17). (Https://Doi.Org/10.1016/J.Procs.2013.05.005).
- [8] Hidayat, I. R., And Maharani, W. 2022. "General Depression Detection Analysis Using Indobert Method," International Journal On Information And Communication Technology (Ijoict) (8:1). (Https://Doi.Org/10.21108/Ijoict.V8i1.634).
- [9] Isa, S. M., Nico, G., And Permana, M. 2022. "Indobert For Indonesian Fake News Detection," Icic Express Letters (16:3). (Https://Doi.Org/10.24507/Icicel.16.03.289).
- [10] Jablonska, J., Tarczewski, R., And Trocka-Leszczynska, E. 2018. "Ergonomic Solutions In Capsule Hotels?," In Advances In Intelligent Systems And Computing (Vol. 600). (Https://Doi.Org/10.1007/978-3-319-60450-3_23).
- [11] Khotimah, N., And Girsang, A. S. 2022. "Indonesian News Articles Summarization Using Genetic Algorithm," Engineering Letters (30:1).
- [12] Koto, F., Rahimi, A., Lau, J. H., And Baldwin, T. 2020. "Indolem And Indobert: A Benchmark Dataset And Pre-Trained Language Model For Indonesian Nlp," In Coling 2020 - 28th International Conference On Computational Linguistics, Proceedings Of The Conference. (Https://Doi.Org/10.18653/V1/2020.Coling-Main.66).
- [13] Kurniawan, K., And Louvan, S. 2019. "Indosum: A New Benchmark Dataset For Indonesian Text Summarization," In Proceedings Of The 2018 International Conference On Asian Language Processing, Ialp 2018. (Https://Doi.Org/10.1109/Ialp.2018.8629109).
- [14] Leftwich, B., Opoku, S., Yin, J., And Adhikari, A. 2021. "Assessing Hotel Employee Knowledge On Risk Factors And Risk Management Procedures For Microbial Contamination Of Hotel Water Distribution Systems," International Journal Of Environmental Research And Public Health (18:7). (Https://Doi.Org/10.3390/Ijerph18073539).
- [15] Muaad, A. Y., Jayappa, H., Al-Antari, M. A., And Lee, S. 2021. "Arcar: A Novel Deep Learning Computer-Aided Recognition For Character-Level Arabic Text Representation And Recognition," Algorithms (14:7). (Https://Doi.Org/10.3390/A14070216).
- [16] Peterson, K. E. 2019. "What Is Risk Management?," In The Professional Protection Officer: Practical Security Strategies And Emerging Trends. (Https://Doi.Org/10.1016/B978-0-12-817748-8.00053-5).
- [17] Riccosan, Saputra, K. E., Pratama, G. D., And Chowanda, A. 2022. "Emotion Dataset From Indonesian Public Opinion," Data In Brief (43). (Https://Doi.Org/10.1016/J.Dib.2022.108465).
- [18] Subakti, A. G., Anita, T. L., And Triana, I. 2020. "The Impact Of Consumer Perceptions To Technology-Based Facilities At Bobobox Capsule Hotel, Jakarta," In Proceedings Of 2020 International Conference On Information



- [19] Wilson, A., And Zahra, A. 2022, "Improving Text Features In Text Summarization Using Harris Hawks Optimization," Icic Express Letters (16:9). (Https://Doi.Org/10.24507/Icicel.16.09:923).
- [20] Zaimah, N. F., And Wardhana, A. 2020. Customer Experience Pada Bobobox Di Kota Bandung Customer Experience Analysis Of Bobobox In Bandung City," E-Proceeding Of Management (7:2).