Sentiment Analysis of Google Maps Reviews of Hotel Grand Mansion Blitar Using Convolutional Neural Network Algorithm with Hyperparameter Optimization

Fatikhatul Trisna Ardiansyah[1], M.TaofikChulkamdi[2], Muhammad Wahyu Rizqi Pratama[3]

Informatics Engineering Study Program, Faculty of Engineering and Informatics, Universitas Islam Balitar, Blitar, Indonesia

The development of digital platforms such as Google Maps has made customer reviews an important source of information for the hotel industry in evaluating service quality. However, the high volume of unstructured reviews and the presence of imbalanced sentiment classes make it difficult for management to identify problems and evaluate customer satisfaction quickly, accurately, and objectively. Therefore, this study aims to apply the Convolutional Neural Network (CNN-1D) algorithm to classify sentiments in Google Maps reviews of Hotel Grand Mansion Blitar, compare model performance before and after hyperparameter optimization using Grid Search based on Confusion Matrix evaluation (accuracy, precision, recall, F1-score), and map service priorities using Importance-Performance Analysis (IPA). The data were collected through web scraping using Webscraper.io, resulting in 1,580 raw reviews. After validation, text preprocessing, automatic sentiment labeling, tokenization, padding, and additional selection, 723 reviews were used in the modelling stage. Text representation was performed using Word2Vec. The results show that based on the Confusion Matrix evaluation, the baseline model achieved an accuracy of 66.21%, precision of 76.77%, recall of 66.21%, and F1-score of 70.45%. After optimization, the best model was obtained with 32 filters, kernel size of 3, dropout rate of 0.3, and learning rate of 0.001, achieving an increased accuracy of 81.38%, precision of 84.80%, recall of 81.38%, and F1-score of 82.92%. The IPA results indicate that the room condition (interior) attribute falls into Quadrant I, making it the main priority for improvement. Therefore, hyperparameter optimization significantly improved the sentiment classification performance, while IPA provided more focused service evaluation recommendations for hotel management.

Keywords: Sentiment Analysis, CNN, Hyperparameter Optimization, Grid Search, IPA, Google Maps Review.

Corresponding Author:
Fatikhatul Trisna Ardiansyah
Informatics Engineering Study Program, Faculty of Engineering and Informatics, Universitas Islam Balitar, Blitar, Indonesia
ftardiansyah.net@gmail.com

1. Introduction

The hospitality industry in Blitar City is currently experiencing rapid growth, marked by the emergence of new competitors offering a variety of modern facilities. Grand Mansion Hotel Blitar, as one of the longest-established and leading hotels in the region, faces increasingly intense competition in maintaining its market share. In this competitive environment, service quality has become a crucial differentiating factor. In *Manajemen Pariwisata dan Perhotelan*, [1] emphasize that successful hotel management is not solely measured by the luxury of its physical facilities, but also by its ability to understand and fulfill guests' needs and satisfaction comprehensively. Customer satisfaction is a vital indicator, as a positive stay experience fosters customer loyalty and enhances the company's reputation, while negative experiences can quickly damage a hotel's image amid intense competition. In today's digital era, interactions between service providers and customers have shifted to online platforms. [2] explains that fierce competition in the hospitality industry compels hotels to continuously improve service quality to obtain positive reviews, as such reviews have become a primary reference for potential customers. One of the most influential platforms is the review

feature on Google Maps [3]. Reviews can be defined as written feedback provided by users regarding their experiences with a product or service. Unlike star ratings, which only represent numerical scores, textual reviews contain rich descriptive information explaining the reasons behind customer satisfaction or dissatisfaction, such as room cleanliness, food quality, and staff friendliness [4].

However, the large volume of reviews submitted daily presents a significant challenge for hotel management, namely the difficulty of processing textual data manually. Reading thousands of reviews individually is highly inefficient and prone to subjectivity. Therefore, an automated computational approach known as Sentiment Analysis is required. Sentiment analysis is a technique used to identify, extract, and measure the emotions or opinions (positive, negative, or neutral) expressed in textual data. By applying sentiment analysis, the management of Grand Mansion Hotel Blitar can transform large volumes of raw review data into strategic insights that support targeted improvements in hotel services and facilities [5]. To perform sentiment analysis with high accuracy, a reliable classification algorithm is required. In recent years, Deep Learning algorithms, particularly the Convolutional Neural Network (CNN), have demonstrated superior performance in text processing compared to traditional methods. This advantage was demonstrated [6], who applied CNN for hotel review sentiment analysis in Pekanbaru, achieving an accuracy rate of up to 98%. These findings are further supported [7], who showed that the CNN architecture is more effective than conventional methods such as Naïve Bayes and Support Vector Machine (SVM) in capturing local features and word patterns within review texts.

Nevertheless, sentiment classification results that merely categorize reviews as "Positive" or "Negative" are insufficient to provide actionable strategic guidance. Hotel management requires more specific information regarding which aspects such as service quality, cleanliness, or facilities should be prioritized for improvement. Therefore, this study integrates CNN-based sentiment classification with the Importance–Performance Analysis (IPA) method. IPA maps sentiment analysis results into priority quadrants, enabling management to identify facilities with low performance but high importance to customers, thereby highlighting areas that require immediate attention. Based on the urgent need for comprehensive hotel service evaluation and the opportunity to improve classification accuracy, this study is entitled "Sentiment Analysis of Google Maps Reviews of Grand Mansion Hotel Blitar Using the Convolutional Neural Network Algorithm with Hyperparameter Optimization."

2. Methods

Research Location

This study was conducted at Grand Mansion Hotel Blitar, located at Jl. Melati No. 90, Blitar City. The hotel was selected as the research location because it has a relatively higher volume of Google Maps reviews than other hotels in the region. This condition provides a sufficiently large dataset for sentiment analysis and enables the evaluation of customer perceptions of hotel services based on publicly available reviews.

Research Type

This research employs a quantitative experimental method supported by qualitative observations and interviews. The experimental component was conducted to examine the effect of CNN hyperparameter configurations, particularly the number of filters and kernel size, on sentiment classification performance. Qualitative observations and interviews were conducted during the initial stage to obtain contextual information regarding the hotel's services and to identify service attributes considered relevant for subsequent analysis. The results of the sentiment classification were then combined with Importance–Performance Analysis (IPA) to map the importance and performance of the identified service attributes.

Data Collection Technique

The primary dataset for sentiment analysis consisted of Google Maps reviews of Grand Mansion Hotel Blitar. The reviews were collected through web crawling using Webscraper.io. The data collection process consisted of four stages: (1) accessing the Google Maps review page of Grand Mansion Hotel Blitar, (2) configuring Webscraper.io to identify and extract the relevant review elements, (3) conducting automated crawling by scrolling through the review page, and (4) exporting the extracted reviews into CSV format as raw data. In addition to review data, observations and interviews were conducted with hotel management to obtain contextual information regarding the services provided by the hotel. The interview results were used as supporting qualitative evidence in identifying and determining the service attributes included in the Importance–Performance Analysis (IPA). The selected attributes were based on information obtained from the interviews regarding service aspects considered relevant to hotel operations and customer service, and were subsequently aligned with the service dimensions identified in the research framework and customer review data.

Data Types

This study utilizes primary and secondary data. Primary data were obtained through direct field observations and in-depth interviews with hotel management. The interviews were conducted to obtain contextual information regarding hotel services, operational practices, and service attributes considered relevant to customer satisfaction. The number and characteristics of the interview informants, including their roles and relevance to hotel service management, are presented as part of the qualitative data collection procedure. Secondary data consisted of Google Maps reviews collected through web scraping. The population comprised 1,996 reviews, of which 1,580 reviews were successfully collected and subsequently used in developing the Convolutional Neural Network (CNN) sentiment classification model. To ensure data validity, the resulting CSV dataset was compared with the corresponding reviews displayed on Google Maps. This comparison was conducted as a form of source triangulation to verify that the data extracted using Webscraper.io accurately represented the original reviews.

Sentiment Labeling

Sentiment labels were determined using two sentiment lexicons. Each review was processed using both lexicons to obtain sentiment scores. The sentiment classification criteria were based on the predefined score thresholds established for each lexicon. Reviews whose scores met the positive threshold were assigned a positive label, whereas reviews meeting the negative threshold were assigned a negative label. Reviews falling within the defined neutral range were assigned a neutral label. When the sentiment results generated by the two lexicons were consistent, the corresponding sentiment label was directly assigned to the review. When the two lexicons produced different sentiment classifications, the final label was determined using the predetermined decision rule of the study rather than arbitrarily selecting one of the lexicons. This procedure was applied consistently to all reviews to maintain uniformity in the labeling process. The specific thresholds and conflict-resolution rules used in the labeling process should be explicitly reported in the methodological table to ensure that the sentiment-labeling procedure is reproducible.

Determination of Service Attributes for IPA

The service attributes used in the Importance–Performance Analysis (IPA) were determined through a combination of qualitative findings and the conceptual framework applied in the study. Interviews with hotel management were used to identify service aspects considered relevant to hotel operations and customer experience. These findings were then reviewed against the service dimensions and indicators established in the research framework. Attributes that were relevant to the hotel's service characteristics and could be evaluated in terms of importance and performance were subsequently selected as IPA indicators. Thus, the

interview results did not serve as the sole basis for determining the attributes. Rather, the interview findings provided contextual and empirical support for selecting service attributes, while the final indicators were aligned with the theoretical and methodological framework of the study. This approach was used to ensure that the attributes assessed in the IPA were both relevant to the actual hotel context and consistent with the objectives of the research.

3. Result and Discussion

Research Results

Data Collection

The primary data consisted of Grand Mansion Hotel Blitar customer reviews collected from Google Maps Reviews. Data were gathered through web scraping using the Webscraper.io browser extension, which enabled efficient extraction of large-scale public review data without requiring a paid API. The data collection (crawling) was conducted between December 25–27, 2025. From a total population of 1,996 Google Maps reviews, 1,580 public reviews were successfully extracted. The difference occurred because some reviews contained only ratings without textual content or could not be fully extracted due to Google Maps page-loading limitations. The raw dataset was saved in CSV format, including reviewer names, star ratings (1–5), and review text, before being processed in Microsoft Excel for further analysis.

Text Preprocessing

Prior to preprocessing, data validation and deduplication were performed by converting ratings into numeric format, removing missing ratings, eliminating reviews without meaningful text, and deleting duplicate entries. This process reduced the dataset from 1,580 to 855 valid textual reviews. The preprocessing stages included:

a. Cleaning: Lowercasing, removal of URLs, mentions, hashtags, numbers, punctuation, emojis, symbols, and extra spaces using Regular Expressions (RegEx).
b. Normalization: Converting non-standard Indonesian words and abbreviations into standardized forms using a custom slang dictionary.
c. Negation Handling: Combining negation words (e.g., *tidak*, *gak*, *kurang*) with subsequent terms (e.g., *tidak_bersih*) to preserve sentiment polarity.
d. Stopword Removal: Removing common Indonesian stopwords using PySastrawi StopWordRemoverFactory combined with a custom stopword list while preserving negation tokens.
e. Stemming & Tokenization: Applying the Nazief–Adriani stemming algorithm (PySastrawi) to reduce words to their root forms, followed by tokenization into individual words for Word2Vec feature extraction.
f. Labeling: Sentiment labels were generated automatically using a hybrid lexicon approach. InSet served as the primary lexicon, while VADER compound scores were used as a fallback for ambiguous reviews. Reviews were classified as Positive (1), Negative (0), or Neutral (2).

After preprocessing, 833 valid reviews remained for CNN-1D model training. The dataset was imbalanced, consisting of 647 positive, 153 negative, and 33 neutral reviews. To address this imbalance, class weighting was applied during CNN training to reduce model bias toward the majority class.

Modelling (CNN-1D)

The modeling stage is carried out after the review data has been pre-processed and sentiment labeled. At this stage, the cleaned text data is then prepared as input for the 1D Convolutional Neural Network (CNN-1D) model to classify sentiment into three classes: negative, positive, and neutral. The modeling process in

this study includes additional data filtering, training and test data formation, class imbalance handling, feature representation using Word2Vec, building a CNN-1D baseline model, and hyperparameter optimization using Grid Search.

a. Data Validation
   Before modeling, an additional validation process was performed to ensure data quality. Reviews with empty text, invalid labels (outside classes 0, 1, and 2), and fewer than two tokens were removed. As a result, 723 valid reviews were retained for model development.
b. Tokenization, Padding, and Data Splitting
   The validated reviews were converted into numerical sequences using the Keras Tokenizer. All sequences were standardized through padding to a maximum length of 100 tokens, resulting in an input shape of (723, 100).
   The dataset was then divided using stratified sampling to preserve class distribution:
   80% for training (train-full) and 20% for final testing.
   The training set was further split into train-grid and validation-grid for hyperparameter optimization.
c. Class Weighting
   To address class imbalance, class weighting was applied during model training. Higher weights were assigned to minority classes, particularly the neutral class, while the positive class received the smallest weight, reducing bias toward the majority class.
d. Word Embedding (Word2Vec)
   Word representations were generated using Word2Vec with the Skip-gram architecture and 100-dimensional vectors. The resulting embedding matrix initialized the CNN-1D embedding layer. The tokenizer vocabulary contained 1,452 words, while Word2Vec learned 1,455 words, successfully mapping 1,338 words (92.2% coverage) into the embedding matrix.
e. CNN Architecture (Baseline Model)
   A baseline CNN-1D model was developed to establish reference performance before optimization. The model was trained on the training set and validated using a separate validation set for 50 epochs.
f. Hyperparameter Optimization
   Model performance was optimized using Grid Search, evaluating combinations of filters, kernel size, dropout rate, and learning rate. Among 24 hyperparameter combinations, the best configuration achieved 86.21% validation accuracy, with 32 filters, kernel size 3, dropout 0.3, and learning rate 0.001.

Model Evaluation

Baseline Model

The baseline model is the initial CNN-1D model before hyperparameter optimization. The initial configuration of this model can be seen in Table 1below:

Table 1. Initial baseline configuration

Hyperparameter	Value
Filters	32
Kernel Size	3
Dropout Rate	0.3
Learning Rate	0.001
Embedding	Word2Vec

The baseline model was trained for up to 50 epochs and then tested using the final test data. The following is an excerpt from the baseline model evaluation program code:

ISSN 2828-9463


```
y_pred_bl = np.argmax(
    baseline_model.predict(X_test, verbose=0),
    axis=1
)
y_true = np.argmax(y_test, axis=1)
acc_bl  = accuracy_score(y_true, y_pred_bl)
prec_bl = precision_score(
    y_true, y_pred_bl,
    average='weighted',
    zero_division=0
)
rec_bl  = recall_score(
    y_true, y_pred_bl,
    average='weighted',
    zero_division=0
)
f1_bl   = f1_score(
    y_true, y_pred_bl,
    average='weighted',
    zero_division=0)
```

The results of the baseline model evaluation obtained performance values which can be seen in Figure 1 below:

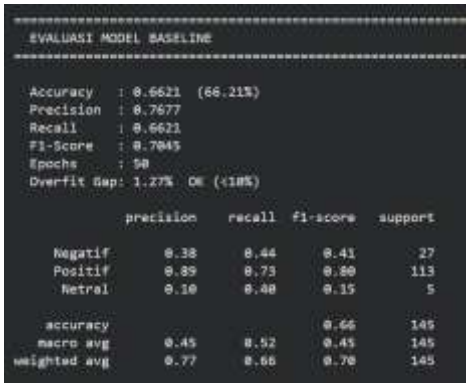

Figure 1. Baseline Model Evaluation

A visualization of the model evaluation can be seen in Figure 2 below:

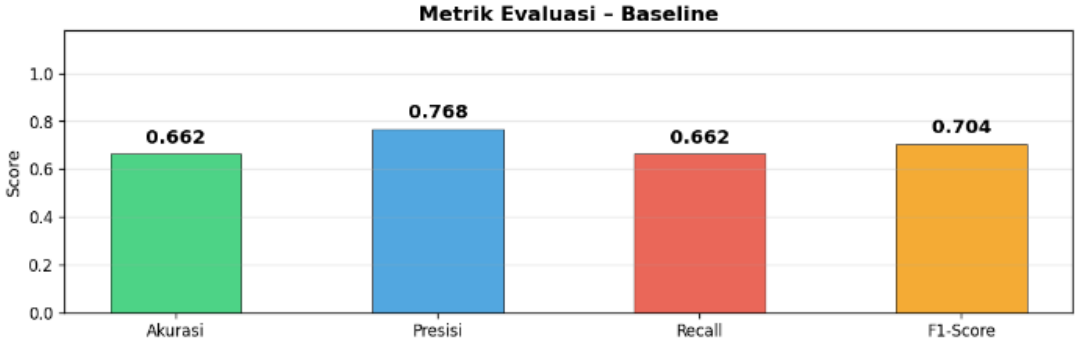

Figure 2. Visualization of the Baseline Model Evaluation

The visualization of the Baseline Model Confusion Matrix can be seen in Figure 3 below:

ISSN 2828-9463

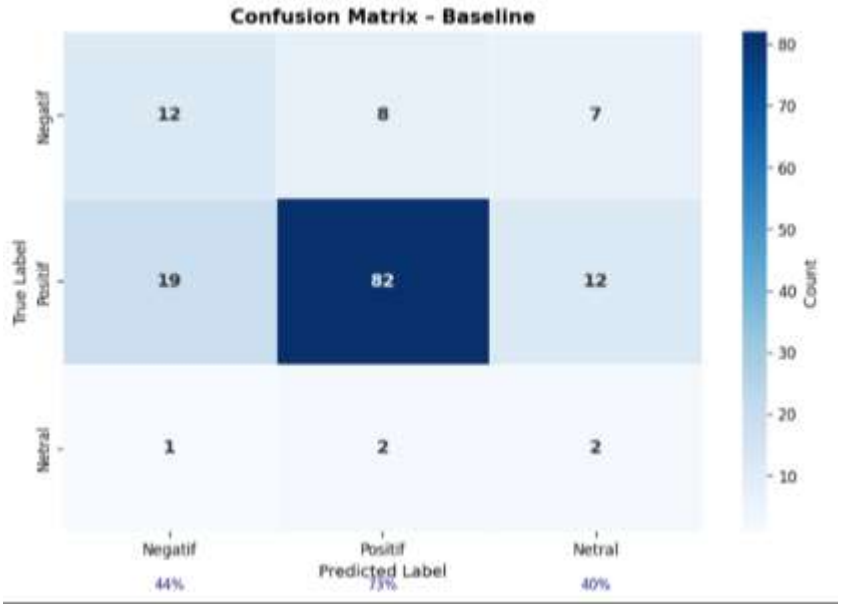

Figure 3. Visualization of Confusion in the Baseline MModel

The baseline model produced an accuracy of 66.21% and an F1-score of 70.45%. The precision value of 76.77% indicates that the model's predictions are generally quite accurate, especially for classes with a dominant amount of data. Meanwhile, the recall value of 66.21% indicates that there are still a number of reviews that the model has not correctly recognized. The baseline model performed best in the positive class, with an F1-score of 0.80. This may be due to the much larger number of positive data points in the dataset compared to other classes.

Hyperparameter Optimization Model

The hyperparameter optimization model is a 1D CNN model obtained after a Grid Search process on 24 parameter combinations. Based on the optimization results, the configuration obtained is shown in Figure 4 below:

```
HYPERPARAMETER TERBAIK :
Filters       : 32
Kernel Size   : 3
Dropout Rate  : 0.3
Learning Rate : 0.001
Val Accuracy  : 0.8621
```

Figure 4. Best Hyperparameter Configuration

This configuration was chosen because it produced the highest validation accuracy of 86.21% in the Grid Search process. The optimization process was carried out systematically based on all predetermined parameter combinations. Evaluation of the optimized model was performed using the same final test data as the baseline model. The program code is as follows.

ISSN 2828-9463


```
y_pred_fin = np.argmax(
    final_model.predict(X_test, verbose=0),
    axis=1
)
acc_fin  = accuracy_score(y_true, y_pred_fin)
prec_fin = precision_score(
    y_true, y_pred_fin,
    average='weighted',
    zero_division=0
)
rec_fin  = recall_score(
    y_true, y_pred_fin,
average='weighted',
    zero_division=0
)
f1_fin   = f1_score(
    y_true, y_pred_fin,
    average='weighted',
    zero_division=0
)
```

The results of the Optimization Model evaluation can be seen in Figure 5 below:

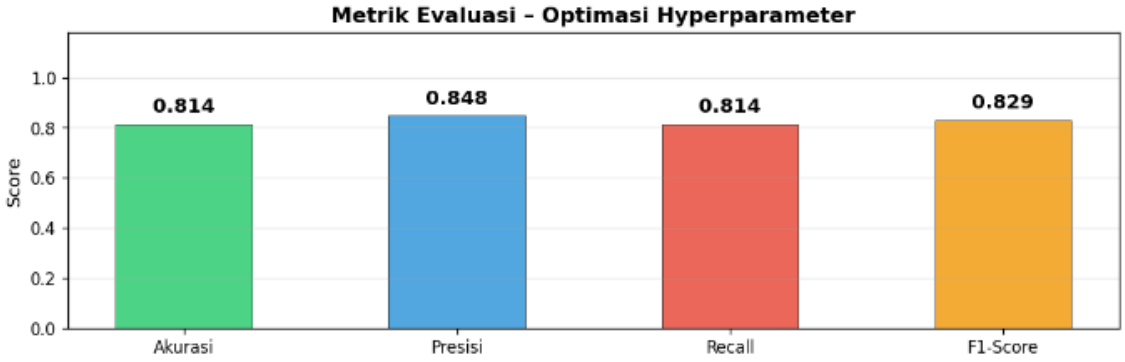

Figure 5. Results of Optimization Model Evaluation

The optimization model stopped at epoch 23, faster than the baseline model, which ran until epoch 50. This indicates that the Grid Search configuration achieved the best performance with a more efficient training process.

The optimization model produced an accuracy of 81.38% and an F1-score of 82.92%. These values indicate an improvement in classification ability compared to the baseline model. A precision of 84.80% indicates a significant increase in the proportion of correct predictions, while a recall of 81.38% indicates that the model has improved in recognizing actual data for each class.

The Classification Report for the Hyperparameter Optimization Model can be seen in Figure 6 below:

	precision	recall	f1-score	support
Negatif	0.70	0.59	0.64	27
Positif	0.92	0.89	0.91	113
Netral	0.08	0.20	0.12	5
accuracy			0.81	145
macro avg	0.57	0.56	0.55	145
weighted avg	0.85	0.81	0.83	145

Figure 6. Classification Report for the Hyperparameter Optimization Model

The Confusion Matrix visualization for the Hyperparameter Optimization Model can be seen in Figure 7 below:

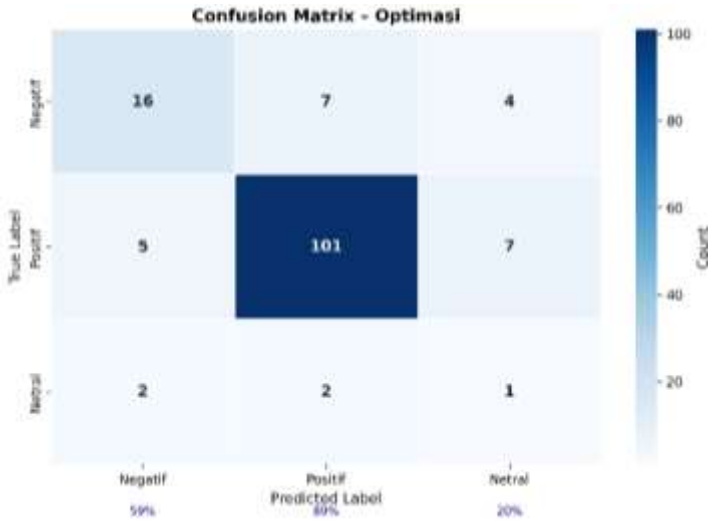

Figure 7. Confusion Matrix of the Hyperparameter Optimization Model

The optimized model showed significant improvements in both the negative and positive classes. In the negative class, the F1-score increased to 0.64, while in the positive class, it increased to 0.92. This indicates that hyperparameter optimization improved the model's ability to capture patterns related to negative and positive sentiment.

Comparison of Baseline and Optimization Models

After evaluating both models separately, a comparison was conducted to determine the impact of hyperparameter optimization on CNN-1D performance. The comparison was based on the main metrics: accuracy, precision, recall, and F1-score. The comparison between the baseline and optimization models can be seen in Figure 8 below:

TABEL PERBANDINGAN: BASELINE vs OPTIMASI HYPERPARAMETER

Metrik	Baseline	Optimasi Hyperparameter	Delta	Status
Accuracy	0.662069	0.813793	0.1517	Naik
Precision	0.767714	0.847957	0.0802	Naik
Recall	0.662069	0.813793	0.1517	Naik
F1-Score	0.704498	0.829152	0.1247	Naik

Figure 8. Comparison of Baseline & Optimization Models

Based on the comparison image, the application of hyperparameter optimization resulted in improved performance across all key metrics. Accuracy increased from 66.21% to 81.38%, a 15.17 percentage point increase. Precision increased by 8.03 percentage points, recall by 15.17 percentage points, and the F1-score increased by 12.47 percentage points. The improvement in the F1-score indicates that the optimization model is not only more accurate overall but also more balanced in combining prediction accuracy and the ability to recognize actual classes. This is important in the context of sentiment analysis, as research requires not only dominant predictions in the positive class but also better ability to recognize complaints or negative sentiment.

A clearer visualization of the comparison and improvements between the baseline and optimization models can be seen in Figure 9 below:

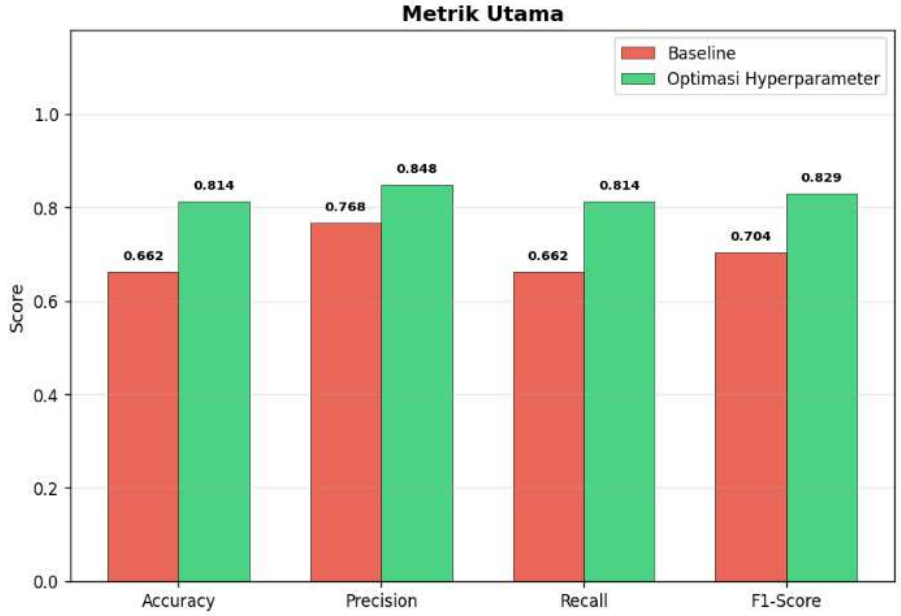

Figure 9. Visualization of Baseline vs. Optimization Comparison

Overall, the comparison results show that the hyperparameter-optimized model performed better than the baseline model. The optimized model was selected as the final model due to its higher accuracy and F1-score, particularly for the positive and negative classes. Based on the evaluation results, hyperparameter optimization through Grid Search significantly improved the performance of the CNN-1D model in classifying the sentiment of customer reviews at the Grand Mansion Blitar Hotel. The baseline model achieved an accuracy of 66.21%, while the optimized model increased to 81.38%. Similar improvements were also seen in the precision, recall, and F1-score values. Thus, hyperparameter optimization positively contributed to the model's classification quality. After the optimization model was selected as the final model based on the evaluation results, the next step was to compile a managerial analysis using Importance-Performance Analysis (IPA) to map the service priorities of the Grand Mansion Blitar Hotel.

Importance-Performance Analysis

Performance and Importance Calculation

The Performance and Importance values are calculated first using Google Collab modeling. Performance is calculated from the average rating of reviews containing keywords for each service indicator, while Importance is calculated from the number of times the indicator keywords appear in the entire corpus of reviews.

The following is the program code for calculating performance and importance in a Google Colab notebook.

```
# Performance: rata-rata rating ulasan relevan
perf = round(subset[COL_RATING].mean(), 2)
# Importance: total frekuensi kemunculan kata kunci
imp_raw = sum(
    sum(t.count(kw) for kw in keywords)
    for t in texts)
```

Once the Raw Importance value is obtained, it is normalized to a range of 1–5 using the Min-Max Scaling method to align with the rating scale for the Performance variable. The following is the program code.

```
scaler = MinMaxScaler(feature_range=(1, 5))
df_ipa['Importance_Likert'] = scaler.fit_transform(
    df_ipa[['Importance_Raw']]
).round(2)
```

The results of the calculation of the IPA variable can be seen in Figure 10 below:

```
======================================================================
  ANALISIS IPA - 10 VARIABEL LAYANAN HOTEL
======================================================================

  Variabel                          Perf       Raw   Likert
  ------------------------------------------------------------
  1. Kebersihan kamar               3.64       188     2.57
  2. Kecepatan pelayanan staf       4.07       123     1.95
  3. Kenyamanan kasur               2.26        56     1.30
  4. WiFi stabil                    3.17        25     1.00
  5. Variasi sarapan                3.69        95     1.68
  6. Kondisi kamar (interior)       3.15       439     5.00
  7. Area parkir                    4.22       262     3.29
  8. Lokasi hotel                   3.99       207     2.76
  9. Harga sesuai                   3.70       116     1.88
  10. Keamanan hotel                4.10       203     2.72

  Rata-rata Performance : 3.60
  Rata-rata Importance  : 2.42
```

Figure 10. IPA Variable Results

The indicator with the highest Performance score was Parking Area (4.22), while the indicator with the lowest Performance score was Mattress Comfort (2.26). The highest Importance score was achieved by the Room Condition (Interior) indicator (5.00), indicating that this attribute is the most frequently cited concern among customers in their reviews.

Conversely, the indicator with the lowest Importance score was Stable Wi-Fi (1.00). This indicates that although Wi-Fi remains an important aspect of hotel service, its frequency of discussion in this research dataset is lower than other indicators.

Inputting IPA Data into SPSS

After obtaining the numerical results from the modeling notebook, the data was further processed using IBM SPSS Statistics software to generate a Cartesian diagram visualization. The first step was to define three main variables:

1. "Variable" as the name of the service indicator.
2. "Performance" as the average value of service performance.
3. "Importance" as the normalized importance value.

The definition of variable types in SPSS can be seen in Figure 11 below:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	VARIABEL	String	64	0		None	None	24	Left	Nominal	Input
2	PERFORM...	Numeric	8	2		None	None	13	Right	Scale	Input
3	IMPORTAN...	Numeric	8	2		None	None	12	Right	Scale	Input

Figure 11. Defining Variable Types

In the next step, the numerical data for each service indicator is entered into SPSS Data View. The data entry includes ten hotel service indicators along with their Performance and Importance values. Entering the IPA variable data in SPSS can be seen in Figure 12 below:

	VARIABEL	PERFORMANCE	IMPORTANCE
1	Kebersihan kamar	3,64	2,57
2	Kecepatan pelayanan staf	4,07	1,95
3	Kenyamanan kasur	2,26	1,30
4	WiFi stabil	3,17	1,00
5	Variasi sarapan	3,69	1,68
6	Kondisi kamar (interior)	3,15	5,00
7	Area parkir	4,22	3,29
8	Lokasi hotel	3,99	2,76
9	Harga sesuai	3,70	1,88
10	Keamanan hotel	4,10	2,72

Figure 12. Filling in the Science Variables in SPSS

The use of SPSS at this stage aims to facilitate the process of mapping the coordinates of each service indicator into a Cartesian diagram visually and systematically. This aligns with the research method design in Chapter III, which states that SPSS is used to map the results of the Science analysis into four priority quadrants.

Science Quadrant Mapping Results

The results of mapping the service variables into the Science quadrants can be seen in Table 2 below.

Table 2. Science Quadrant Mapping

Quadrant	Category	Variables
Quadrant I	Concentrate Here	Room condition (interior)
Quadrant II	Keep Up the Good Work	Room cleanliness, Parking area, Hotel location, Hotel security
Quadrant III	Low Priority	Bed comfort, Stable Wi-Fi connection
Quadrant IV	Possible Overkill	Staff service speed, Breakfast variety, Price appropriateness

Furthermore, the visualization of the results of the quadrant mapping presented in the Cartesian diagram of the SPSS application can be seen in Figure 13 below.

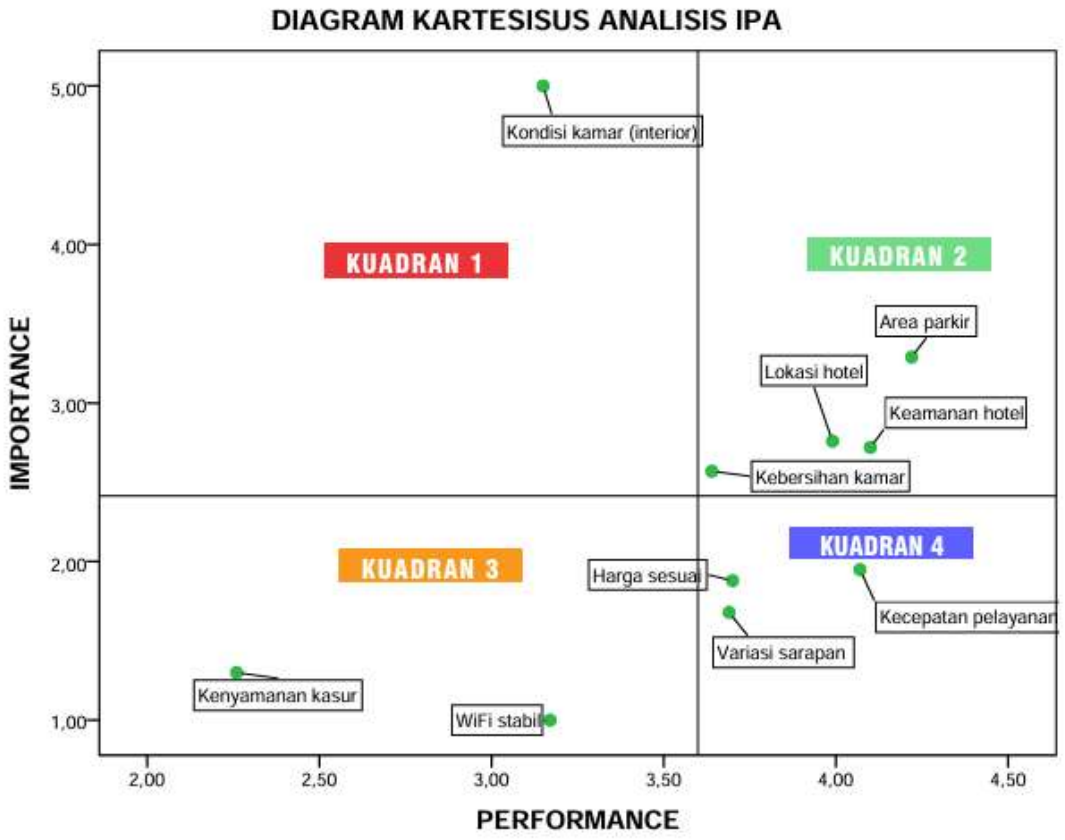

Figure 13. Cartesian Diagram of Science Analysis

Interpretation of Science Quadrants

1. Quadrant I (Top Priority)
   Quadrant I shows indicators with a high level of importance, but below-average performance. The indicator in Quadrant I is room condition (interior). This indicator has a Performance value of 3.15 and an Importance value of 5.00. These values indicate that room condition is the aspect most frequently discussed by customers, but its performance level has not yet reached the expected average.
   This finding indicates that the management of the Grand Mansion Hotel Blitar needs to prioritize evaluation of aspects of the physical condition of the rooms, such as interior cleanliness, the suitability of room facilities, air conditioning, lighting, television, bathroom water, and elements of renovation or room maintenance. Because this indicator is a top priority, improvements to room condition are expected to have a direct impact on improving customer perceptions.
2. Quadrant II (Maintain Performance)
   Quadrant II contains indicators with a high level of importance and high performance. The indicators in this quadrant are (1) room cleanliness, (2) parking area, (3) hotel location, and (4) hotel security.
   These four indicators have above-average Performance and Importance scores. This indicates that these attributes are considered important by customers and have been assessed as having relatively good performance.
3. Quadrant III (Low Priority)
   Quadrant III indicates indicators with low importance and low performance. In this study, the indicators in Quadrant III are (1) mattress comfort and (2) stable Wi-Fi. The mattress comfort indicator has a Performance score of 2.26 and an Importance score of 1.30, while stable Wi-Fi has a Performance score of 3.17 and an Importance score of 1.00.

Although these two indicators are in the low priority category based on the IPA mapping, their scores still warrant attention because their performance is still below average. Specifically, mattress comfort has the lowest performance score among all indicators, so it still deserves attention as part of efforts to improve the guest experience.

4. Quadrant IV (Excessive)
   Quadrant IV contains indicators with low importance but high performance. Indicators in this quadrant are (1) staff service speed, (2) breakfast variety, and (3) reasonable pricing. These three indicators have above-average Performance scores, but below-average Importance scores. This indicates that, from a customer perspective, the hotel's performance in these aspects is considered good, but the frequency of discussion is not as high as other indicators.

Discussion

This study demonstrates that Google Maps reviews can serve as a valuable digital data source for understanding customer perceptions of Grand Mansion Hotel Blitar. From 1,580 scraped reviews (out of 1,996 available reviews), preprocessing and validation reduced the dataset to 723 valid reviews for modeling. This confirms that scraped data require cleaning and validation before analysis, consistent with [8].

The preprocessing stage including cleaning, case folding, normalization, stopword removal, stemming, tokenization, and negation handling significantly improved data quality by reducing noise and preserving sentiment context. These findings support [9], who emphasized preprocessing as a critical factor in sentiment analysis accuracy.

Sentiment labeling was performed automatically using a hybrid lexicon-based approach combining InSet and VADER, enabling efficient annotation of large datasets. This approach is consistent with [10], although the CNN model learned from automatically generated rather than manually annotated labels.

For sentiment classification, Word2Vec (100-dimensional embeddings) and CNN-1D were employed. Word2Vec achieved 92.2% embedding coverage, supporting effective feature extraction, consistent with [11]. The baseline CNN-1D achieved 66.21% accuracy, 76.77% precision, 66.21% recall, and 70.45% F1-score. After Grid Search optimization, performance improved to 81.38% accuracy, 84.80% precision, 81.38% recall, and 82.92% F1-score, confirming the importance of hyperparameter optimization as reported by [12]. However, class imbalance continued to limit neutral sentiment recognition, highlighting the importance of evaluating F1-score alongside accuracy [13].

The best-performing model was integrated with Importance–Performance Analysis (IPA). Room condition (interior) was identified as the highest priority (Quadrant I; Performance = 3.15, Importance = 5.00), indicating that it is highly important to customers but underperforms. This finding aligns with [14], who suggested that Quadrant I attributes should receive immediate managerial attention.

Other attributes including room cleanliness, parking area, hotel location, and security were classified into Quadrant II (maintain performance), while bed comfort and Wi-Fi stability fell into Quadrant III, and staff service speed, breakfast variety, and price fairness into Quadrant IV. These findings support [15], who argued that textual reviews provide richer managerial insights than ratings alone.

Overall, the study successfully applied CNN-1D for sentiment classification, demonstrated that Grid Search significantly improved model performance, and showed that integrating deep learning with IPA produces practical recommendations for service improvement [16]. Nevertheless, limitations remain, including reliance on automatic lexicon-based labeling, class imbalance, and keyword-based IPA mapping. Future research should incorporate manual annotation, larger datasets, and aspect-based or contextual language models to improve sentiment classification and service attribute analysis [17].

4. Conclusion

Based on the findings of this study on Google Maps review sentiment analysis of Grand Mansion Hotel Blitar using a Convolutional Neural Network (CNN) with hyperparameter optimization, the following conclusions were drawn:

1. CNN-1D was successfully implemented to classify customer reviews into positive, negative, and neutral sentiments. The workflow included web scraping, data validation, text preprocessing, automatic sentiment labeling, tokenization, padding, Word2Vec embedding, model training, and evaluation. From 1,580 raw reviews, 723 valid reviews were processed for sentiment classification.
2. Grid Search hyperparameter optimization significantly improved model performance. The baseline model achieved 66.21% accuracy, 76.77% precision, 66.21% recall, and 70.45% F1-score. After optimization (32 filters, kernel size = 3, dropout = 0.3, learning rate = 0.001), performance increased to 81.38% accuracy, 84.80% precision, 81.38% recall, and 82.92% F1-score, demonstrating better handling of class imbalance.
3. The integration of Importance–Performance Analysis (IPA) provided practical managerial recommendations. The IPA results identified:
   a. Quadrant I (Priority for Improvement): Room condition (interior).
   b. Quadrant II (Maintain Performance): Room cleanliness, parking area, hotel location, and security.
   c. Quadrant III (Low Priority): Bed comfort and Wi-Fi stability.
   d. Quadrant IV (Possible Overinvestment): Staff service speed, breakfast variety, and price suitability.

5. References

[1] Ni Wayan Yunita, Syamsuddin R, And Fadila Almahdali, "Pengaruh Kompensasi Dan Gaya Kepemimpinan Terhadap Kinerja Karyawan Pada Hotel Jazz Di Kota Palu," *Jurnal Ekonomi, Manajemen Pariwisata Dan Perhotelan*, Vol. 3, No. 1, Pp. 88–100, Jan. 2024, Doi: 10.55606/Jempper.V3i1.2658.

[2] D. W. Pardede, Y. K. S. Sitepu, R. Juni, T. Sitio, M. Silalahi, And R. Simbolon, "Partisipasi Kelompok Sadar Wisata Dalam Pengembangan Desa Wisata Meat Kecamatan Tampahan Kabupaten Toba," *Jurnal Manajemen Pariwisata Dan Perhotelan*, Vol. 1, No. 4, Pp. 159–171, 2023, Doi: 10.59581/Jmpp-Widyakarya.V1i4.1469.

[3] R. Kusumaningrum, I. Z. Nisa, R. P. Nawangsari, And A. Wibowo, "Sentiment Analysis Of Indonesian Hotel Reviews: From Classical Machine Learning To Deep Learning," *International Journal Of Advances In Intelligent Informatics*, Vol. 7, No. 3, P. 292, Nov. 2021, Doi: 10.26555/Ijain.V7i3.737.

[4] S. Melamane *Et Al.*, "Artificial Neural Network–Based Inference Of Drug–Target Interactions," In *Nanotechnology Principles In Drug Targeting And Diagnosis*, Elsevier, 2023, Pp. 35–62. Doi: 10.1016/B978-0-323-91763-6.00015-1.

[5] Retno Kusumaningrum, Asyraf Humam Arafifin, Selvi Fitria Khoerunnisa, Priyo Sidik Sasongko, Panji Wisnu Wirawan, And Muhammad Syafrudin, "Hyperparameter Optimization For Convolutional Neural Network-Based Sentiment Analysis," *Journal Of Advanced Research In Applied Sciences And Engineering Technology*, Vol. 53, No. 1, Pp. 44–56, Nov. 2024, Doi: 10.37934/Araset.53.1.4456.

[6] I. Irmayanty And M. Hendri, "PERENCANAAN BISNIS PENGEMBANGAN HOTEL PURI MUTIARA SUMEDANG," *Jurnal Ilmiah Manajemen, Ekonomi, & Akuntansi (MEA)*, Vol. 9, No. 2, Pp. 3435–3458, Aug. 2025, Doi: 10.31955/Mea.V9i2.6165.

[7] M. Kyei-Frimpong *Et Al.*, "Employee Empowerment And Organizational Commitment Among Employees Of Star-Rated Hotels In Ghana: Does Perceived Supervisor Support Matter?," *Journal Of*

Work-Applied Management, Vol. 16, No. 1, Pp. 65–83, Apr. 2024, Doi: 10.1108/JWAM-05-2023-0038.

[8] R. Karisma, S. Lestanti, And M. T. Chulkamdi, "APLIKASI KLASIFIKASI SENTIMEN PADA ULASAN SMARTPHONE DI SITUS JUAL BELI ONLINE BERBASIS WEB MENGGUNAKAN NAIVE BAYES DENGAN TF-IDF," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 6, No. 1, Pp. 31–37, Dec. 2021, Doi: 10.36040/Jati.V6i1.4365.

[9] E. U. H. Qazi, A. Almorjan, And T. Zia, "A One-Dimensional Convolutional Neural Network (1D-CNN) Based Deep Learning System For Network Intrusion Detection," *Applied Sciences*, Vol. 12, No. 16, P. 7986, Aug. 2022, Doi: 10.3390/App12167986.

[10] F. Nugraha, "ANALISIS EFEKTIVITAS PENERAPAN WEBSITE KPRO PADA UNIT ACCSESS SERVICE OPERATION MENGGUNAKAN METODE IMPORTANCE PERFORMANCE ANALYSIS (IPA) PADA PT TELKOM WITEL CIREBON," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 8, No. 3, Pp. 3581–3586, Jun. 2024, Doi: 10.36040/Jati.V8i3.9495.

[11] M. D. Ibadurahman And E. Nurhayaty, "Pengaruh Harga Dan Kualitas Pelayanan Terhadap Kepuasaan Pelanggan Ekspedisi J&Amp;T Cargo Di Jakarta," *Benefit: Journal Of Bussiness, Economics, And Finance*, Vol. 3, No. 2, Pp. 902–925, Jul. 2025, Doi: 10.70437/Benefit.V3i2.1334.

[12] S. Mahesa Putra And D. Sundari, "KLASIFIKASI SENTIMEN ULASAN PENGGUNA APLIKASI SHOPEE DENGAN ALGORITMA NAÏVE BAYES DAN ID3," *JATI (Jurnal Mahasiswa Teknik Informatika)*, Vol. 9, No. 6, Pp. 9638–9643, Nov. 2025, Doi: 10.36040/Jati.V9i6.15700.

[13] S. W. Ritonga, . Y., M. Fikry, And E. P. Cynthia, "Klasifikasi Sentimen Masyarakat Di Twitter Terhadap Ganjar Pranowo Dengan Metode Naïve Bayes Classifier," *Building Of Informatics, Technology And Science (BITS)*, Vol. 5, No. 1, Jun. 2023, Doi: 10.47065/Bits.V5i1.3535.

[14] Ahmad Mansur, Tonny Hendratono, And Sugiarto Sugiarto, "Tourism Village Management In Building The Local Economy Through Community Partnerships," *Proceeding Of The International Global Tourism Science And Vocational Education*, Vol. 1, No. 2, Pp. 01–08, Aug. 2024, Doi: 10.62951/Icgtsave.V1i2.24.

[15] G. M. Ramopolii, T. Runtu, And L. D. Latjandu, "Analisis Kinerja Keuangan Menggunakan Metode Economic Value Added Dan Market Value Added Pada Perusahaan Sub Sektor Perdagangan Ritel Barang Primer Yang Terdaftar Di Bursa Efek Indonesia," *Riset Akuntansi Dan Portofolio Investasi*, Vol. 2, No. 2, Pp. 458–465, Nov. 2024, Doi: 10.58784/Rapi.232.

[16] A. Zakharia, H. Hasanah, And A. Srirahayu, "ANALISIS SENTIMEN PUBLIK DI TWITTER TENTANG PEMANFAATAN AI DALAM SENI DIGITAL MENGGUNAKAN CNN," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 13, No. 3S1, Oct. 2025, Doi: 10.23960/Jitet.V13i3s1.7701.

[17] N. C. Ramadani, I. Tahyudin, And A. Shouni Barkah, "Perbandingan Algoritma Support Vector Machine, Decision Tree, Dan Logistic Regresion Pada Analisis Sentimen Ulasan Aplikasi Netflix," *Jurnal Nasional Teknologi Dan Sistem Informasi*, Vol. 10, No. 2, Pp. 110–117, Aug. 2024, Doi: 10.25077/TEKNOSI.V10i2.2024.110-117.