Jurnal Komputer, Informasi dan Teknologi Vol: 5, No 2, 2025, Page: 1-10

Sentiment Analysis of Public Reviews on the Internet for Gelora Sriwijaya International Stadium in Palembang Using Random Forest Classifier Title

Yoga Seprana[1]*, Herri Setiawan[2], Zaid Romegar Mair[3]

1,2,3 Department of Informatic Engineering, Indo Global Mandiri University, Indonesia

*Correspondence: Yoga Seprana
Email: yogaseprana@gmail.com

Received: 06-10-2025
Accepted: 17-11-2025
Published: 28-12-2025

Abstract: *This study was conducted with the aim of analyzing the sentiments contained in public reviews of the Gelora Sriwijaya International Stadium in Palembang. The analysis utilized the Random Forest Classifier algorithm as the main method for the classification process. The data source was obtained from Google Maps through web scraping using the Instant Data Scraper tool, resulting in a total of 1,000 reviews written between 2019 and 2025. These reviews then went through a series of preprocessing stages, including data cleaning to remove irrelevant information, case folding to standardize letter formats, stemming to return words to their root form, stopword removal to eliminate common words with little semantic value, and tokenization to split the text into individual word units. Subsequently, the text data was transformed into a numerical representation using the Term Frequency–Inverse Document Frequency (TF-IDF) method to make it suitable for processing by the classification algorithm. Sentiment classification was carried out by categorizing reviews into two main classes: positive sentiment and negative sentiment. The evaluation results showed that the model achieved an accuracy of 89% when tested using the K-Fold Cross Validation technique, which was higher than the 84% accuracy obtained without validation. These findings demonstrate that the Random Forest Classifier algorithm is highly capable of performing sentiment analysis and has the potential to serve as a useful tool for stadium management to understand public perceptions and improve the quality of facility management.*

Keywords: *Sentiment Analysis, Random Forest, Google Maps, Gelora Sriwijaya Stadium Public Reviews*

Introduction

A stadium is a sports facility or area specifically built to host various sporting activities, with a primary focus on football and athletics. Typically, a stadium has a centrally located main field surrounded by spectator seating, either in the form of open or covered stands. Stadiums are often located in areas with a large population, ensuring public use and participation in the activities held there [1]. Today, the use of football stadiums has expanded, extending beyond football clubs and their managers. Stadiums are also used for various other purposes, such as art performances, community gatherings, museums, shopping centers, and even sports facilities accessible to the surrounding community. In the design process, stadiums must adhere to applicable standards at the construction site, including provisions regarding structural design. Careful planning will result in a stadium that is safe for use, even when facing extreme conditions and certain external loads [2].

Sports are physical activities that are carried out regularly and systematically with the aim of maintaining and improving physical fitness and health. Furthermore, sports can

be a form of recreation, entertainment, and competition, involving certain physical abilities and technical skills. One such activity is soccer[3]. The construction of this stadium refers to soccer, a very popular sport with a large fan base around the world. The progress of this sport cannot be separated from the active participation of soccer clubs that regularly hold annual competitions. According to research by the multinational company Ipsos, Indonesia has the largest number of soccer fans in the world, namely 69%, the highest among dozens of other countries surveyed [4].

The city of Palembang is the administrative center of South Sumatra Province and is recognized as the oldest city in Indonesia, with a history dating back to the 7th century. The city has significant potential in the tourism sector, with one of its leading destinations being the Gelora Sriwijaya International Stadium. The Gelora Sriwijaya International Stadium is one of the largest multipurpose stadiums in Palembang, located in the Jakabaring area, namely Jakabaring Sport City. This area is known as an elite, modern area that is often visited by locals and tourists. Jakabaring Sport City itself is a sports complex built on 325 hectares of land, located about 5 kilometers from the center of Palembang, precisely in the Seberang Ulu area, with a capacity of up to 40,000 spectators [5]. Despite its great potential, the tourism sector in Palembang faces challenges in managing and monitoring destinations. The large number of tourist attractions makes it difficult for the Tourism and Culture Office to optimally monitor the development of each destination [6].

This also happened to Gelora Sriwijaya International Stadium, which received mixed reviews from the public. These differing views led to varying public sentiment, which ultimately affected the stadium's image and appeal as a tourist destination [7].Rapid technological developments and the increasingly widespread use of social media have transformed the way people communicate into a digital-based model. This change has had a significant impact on individual decision-making, both in the context of shopping and choosing tourist destinations. In the tourism sector, the phenomenon of Electronic Word of Mouth (e-WOM), which involves the dissemination of information and reviews about a tourist destination in the virtual world, plays a major role in increasing the popularity of that destination. Trending topics or discussions on social media can shape public perception and influence interest in visiting. Therefore, tourism businesses utilize text data from tourist reviews related to destinations, services, and supporting facilities as an important source of information. This data is then analyzed to understand tourists' needs, preferences, and expectations, thereby serving as a reference for adjusting strategies and improving service quality to keep pace with market developments and changing demands[8].

One method that can be used is sentiment analysis, which is a text data processing technique to identify and categorize sentiments into positive or negative categories. Review data from platforms such as Google Maps can be a rich source of information for measuring public perception. Through sentiment analysis, information about the quality, strengths, and weaknesses of a destination can be obtained more objectively [9]. This study utilizes the Random Forest algorithm, which is a machine learning method known to be effective in classifying large amounts of data. This algorithm works by building a number of decision trees simultaneously, then combining the results from each tree to produce more stable, consistent, and accurate predictions. The main advantage of Random Forest lies in its ability to reduce the risk of overfitting and handle data with many variables efficiently. Thanks to

its reliability, this method has been widely implemented in various sectors, including healthcare for disease diagnosis, banking for credit risk analysis, e-commerce in product recommendation systems, and market analysis to understand consumer behavior and demand trends [10].

Based on the background described above, this study focuses on analyzing public sentiment toward the Gelora Sriwijaya International Stadium. The goal is to facilitate the process of identifying the positive or negative meaning of a sentence, while also developing a system capable of performing effective sentiment analysis. This research is expected to provide more accurate information regarding public opinion on the Gelora Sriwijaya International Stadium as supporting material for decision making. In addition, this research also develops a prediction system that can process data in .xlsx or .csv format, both for individual and group analysis.

Methodology

This study applies the Knowledge Discovery from Data (KDD) method, which consists of five stages, namely data selection, preprocessing, transformation, data mining, and evaluation[11]. The KDD method is used to obtain information and discover patterns in the selected data. To clarify the research flow, a visual illustration is presented in Figure 1.

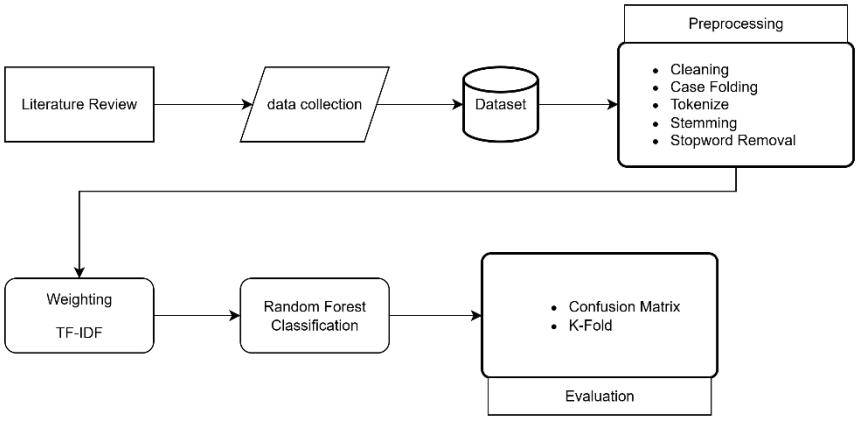

Figure 1. Research stages

The figure above shows the research flow starting from literature review and data collection to the formation of a dataset. The data was processed through preprocessing stages (cleaning, case folding, tokenizing, stemming, and stopword removal), then weighted using TF-IDF. Next, classification was performed using the Random Forest algorithm. The classification results were evaluated using Confusion Matrix and K-Fold Cross Validation to measure and ensure the reliability of the model.

Literatur Review

The literature study in this research was conducted by collecting, reading, and analyzing various relevant sources of information, such as books, scientific journals, articles, research reports, and other reliable sources. The purpose of the literature study was to obtain a strong theoretical basis, understand the latest research developments, identify

research gaps, and strengthen the framework used. Through this process, researchers can build a solid foundation before conducting further research and ensure that the research conducted has contribution and novelty value [12].

Data Collection

The data source in this study comes from public reviews of Gelora Sriwijaya International Stadium obtained through Google Maps using web scraping techniques with Instant Data Scraper. A total of 1,000 reviews in Indonesian were collected. Next, the data will undergo preprocessing.

Pre-Processing

Preprocessing is the next stage in data processing that aims to prepare raw data so that it is ready for use in analysis or modeling processes. In text data, preprocessing is done to clean, tidy, and simplify the text so that it is easier for algorithms to process[13]. This process usually includes several steps, such as data cleaning to remove unnecessary characters or symbols, case folding to convert all letters to lowercase, (Tokenization) to separate text into words, stopword removal to eliminate common words that have no significant meaning, (Stemming) to return words to their basic form, and normalization to standardize the spelling of words. By performing preprocessing, the quality of the data improves, so that the results of the analysis or predictions can be more accurate and relevant [14].

Random Forest Classification

The classification process using the Random Forest algorithm is carried out in two main stages, namely training and testing. In the training stage, the model is formed using the available data, followed by the testing stage which considers the probability of results from the training data. The Random Forest Classifier is used to accurately group opinions, so that it can identify whether a comment is positive or negative towards the product or issue being discussed[15]. Next, all of these trees are combined to produce a more stable and accurate final prediction. The uniqueness of Random Forest lies in the application of the randomization process, not only in the selection of data samples (bootstrapping), but also in the selection of predictor variables used in the formation of each tree. This approach helps reduce the risk of overfitting while improving the model's ability to generalize to new data [16].

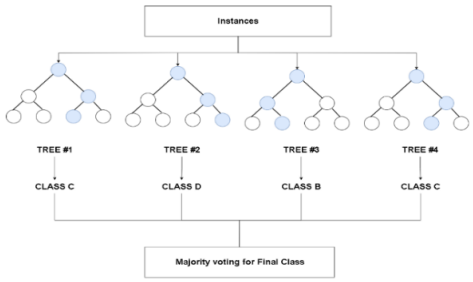

Figure 2. Process random forest clasification

a) Entropy is used to measure uncertainty in a data set
Formula:
Equation (1) is S = Data set, Pi = Proportion (probability) of class i

(1)

b) Information Gain is used to measure how much entropy is reduced when a feature is split.
Formula:
Equation (2) is S is a data set, A is a split feature, Si is a subset of S that is split based on A.

(2)

Evaluation

Evaluation is a step to measure the performance of a model that has been built in making predictions or classifications. The main objective is to ensure that the model not only works well on training data, but is also capable of providing accurate predictions on test data that has never been used before. To improve reliability and accuracy, this study applies the k-fold cross validation technique. This method divides the dataset into several folds, then training and testing are performed alternately on each fold. This approach produces a more consistent evaluation and helps reduce the risk of overfitting, so that the model's performance on new data is more reliable [17].

Result and Discussion

Scraping Results

	Ulasan
0	Stadion gelora Sriwijaya Jakabaring merupakan ...
1	Tempatnya rindang banyak pohon, bersih, aman d...
2	Stadion jakabaring merupakan stadion yang besa...
3	Gelora sriwijaya, pusat tempat olahraga di pal...
4	Tempatnya luas, asri, tertata. Cocok untuk aca...
..	...
995	Bagus dan rapih. Jadi salah satu tempat wisata...
996	Bagus dan indah
997	Closing Ceremonial Asian Games 2018, Sungguh F...
998	Cocok untuk anak anak PAS DEH BUAT ASIAN GAME...
999	T4 liburan kluarga

Figure 3.Scraping data results

In this study, data was collected by scraping public reviews on Google Maps about Gelora Sriwijaya International Stadium using Instant Data Scraper. A total of 1,000 reviews in Indonesian were collected.

Pre-Processing

Preprocessing is the initial stage in data processing that serves to prepare raw data so that it is cleaner, more structured, and ready for use in analysis and model building. Data obtained from original sources generally still contains noise (interference), empty data, duplicates, or inconsistent formats, so that if used directly, it can reduce the quality of the analysis results. Through preprocessing, data is cleaned, converted, and adjusted so that it

is more efficient and supports model performance to be more accurate. This stage includes data cleaning, data integration, data transformation, data reduction, and special processing for text data such as case folding, tokenization, stopword removal, and stemming. Thus, preprocessing is a crucial step that ensures data quality, reduces complexity, and optimizes the reliability of analysis results and models.

The following is an example of the preprocessing process, which is the initial stage of data processing that aims to prepare raw data so that it is cleaner, more structured, and ready for use in analysis and model development.

Table 1. Pre-processing

Process	Before	After
Cleaning	Stadion gelora Sriwijaya Jakabaring merupakan salah satu stadion sepabkola ada di Palembang... 👏👏😂😂	Stadion gelora Sriwijaya Jakabaring merupakan salah satu stadion sepabkola ada di Palembang
Case Folding	Stadion gelora Sriwijaya Jakabaring merupakan salah satu stadion sepabkola ada di Palembang... 👏👏😂😂	stadion gelora sriwijaya jakabaring rupa salah satu stadion sepak bola ada di palembang
Stemming	Stadion gelora Sriwijaya Jakabaring merupakan salah satu stadion sepakbola di Palembang... 👏👏😂😂	stadion gelora sriwijaya jakabaring rupa salah satu stadion sepak bola ada di palembang
Stopword Removal	Stadion gelora Sriwijaya Jakabaring merupakan salah satu stadion sepakbola di Palembang... 👏👏😂😂	stadion gelora sriwijaya jakabaring stadion sepakbola palembang
Tokenize	Stadion gelora Sriwijaya Jakabaring merupakan salah satu stadion sepakbola di Palembang... 👏👏😂😂	'stadion', 'gelora', 'sriwijaya', 'jakabaring', 'rupa', 'salah', 'satu', 'stadion', 'sepak', 'bola', 'ada', 'di', 'palembang'

Weighting TF-IDF

Weighting of words in reviews is performed using the Term Frequency–Inverse Document Frequency (TF-IDF) method. This method assigns a weight value to each word according to its level of importance in a document. After the weighting stage is complete, the dataset is separated into two parts: training data for building the model and test data for evaluating its performance.

Classification Random Forest

Classification using the Random Forest algorithm is carried out in two main stages: training to build a model, and testing, which refers to the probability of results from the training data. The Random Forest Classifier method has proven effective in separating opinions and determining whether a comment is positive or negative regarding tourist destinations and trending issues.

```
Hasil Klasifikasi:
              precision    recall  f1-score   support

     POSITIF       0.93      0.71      0.81        91
     NEGATIF       0.80      0.95      0.87       109

    accuracy                           0.84       200
   macro avg       0.86      0.83      0.84       200
weighted avg       0.86      0.84      0.84       200

Persentase Positif: 65.00%
Persentase Negatif: 35.00%
```

Figure 4.Classification results

As shown in Figure 4, the classification results were obtained by searching for the highest accuracy using a test scenario of 800 test data and 200 training data. Of the 200 data, there were 91 data with positive sentiment and 109 data with negative sentiment. The classification model achieved an accuracy of 84%, with a precision value of 93%, recall of 95%, and an F1-score of 87%. Based on the sentiment classification results in the test with the highest accuracy, 65% of the data was categorized as positive sentiment, while 35% of the data was categorized as negative sentiment .

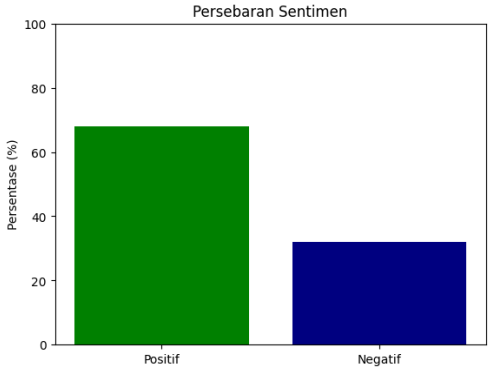

Figure 5.Classification results chart

Evaluation

In this study, the confusion matrix is used as one of the main tools to evaluate the performance of the classification model that has been built. The confusion matrix itself is an evaluation method that displays the model's prediction results in a table, where each row represents the actual class and each column shows the predicted class. With this table, researchers can see in detail the number of correct and incorrect predictions for each class. Through confusion matrix analysis, not only can the overall accuracy be measured, but also other evaluation metrics such as precision, recall, and F1-score. This information helps in understanding the extent to which the model is able to accurately recognize data from various classes, while also identifying patterns of errors that occur. Thus, the use of a confusion matrix provides a more comprehensive picture of the model's performance, making it easier for researchers to make improvements or adjustments to enhance the quality of the classification results.

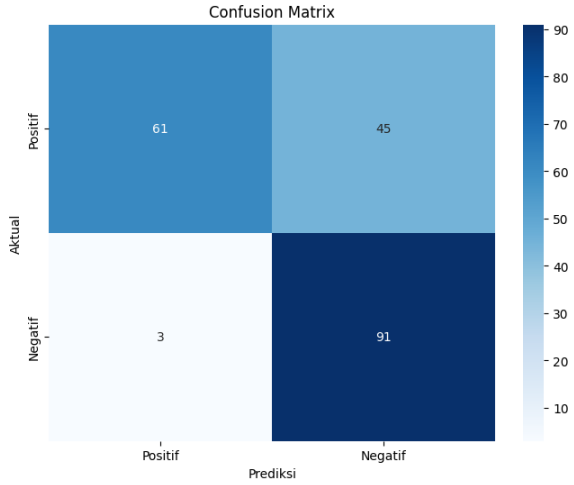

Figure 6. Confusion matrix

From Figure 6, several conclusions can be drawn as follows:

1. TP (True Positive) = Positive labeled data that is correctly predicted as a positive label is 61
2. FP (False Positive) = Positive labeled data that is incorrectly predicted as a positive label is 45
3. TN (True Negative) = Negative labeled data correctly predicted as negative labels 91
4. FN (False Negative) = Negative labeled data incorrectly predicted as negative labels 3

Therefore, this study uses the k-fold cross validation method with 7 folds in an effort to improve model accuracy. From this process, the highest accuracy result will be selected, as shown in Figure 7.

Table 2.*K-fold results*

Fold	Accuracy	Precision	Recall	F1-Score
1	0.81	0.85	0.81	0.80
2	0.89	0.91	0.89	0.89
3	0.78	0.81	0.78	0.77
4	0.76	0.81	0.76	0.75
5	0.74	0.7	0.74	0.73
6	0.81	0.82	0.81	0.80
7	0.82	0.86	0.82	0.81
Best Fold =2	**0.89**	**0.91**	**0.89**	**0.89**

In the Results section, summarize the collected data and the analysis performed on those data relevant to the issue that is to follow. The Results should be clear and concise. It should be written objectively and factually, and without expressing personal opinion. It includes numbers, tables, and figures (e.g., charts and graphs). Number tables and figures consecutively in accordance with their appearance in the text.

Conclusion

Based on the results of sentiment analysis of public reviews of the Gelora Sriwijaya International Stadium in Palembang, collected through scraping 1,000 Indonesian-language reviews from Google Maps, the distribution of positive sentiment was 650 data points (65%) and negative sentiment was 350 data points (35%). The data was processed through the stages of cleaning, case folding, stemming, stopword removal, tokenization, and TF-IDF weighting, then classified using the Random Forest Classifier algorithm. The test results showed that the model was able to achieve 84% accuracy in the initial test scenario, increasing to 89% after applying K-Fold Cross Validation (7 fold), with 91% precision, 89% recall, and an F1-score of 89%. Words that frequently appear in negative reviews include "parking," "traffic jam," "dirty," and "broken," which indicate public complaints related to stadium facilities and management. These findings show that the Random Forest algorithm is effective for analyzing the sentiment of tourist destination reviews, while also providing important information for managers to improve the quality of stadium services and facilities.

References

Alam, A. (2010). Perpustakaan Tempat Belajar Sepanjang Hayat. Media Indonesia, Jakarta: Kamis, 7 Oktober: hlm.1, kolom 2.

Andika T Mentaruk, Banu D. Handono, and Ronny E. Pandaleke, "Perencanaan Struktur Stadion Di Kota Manado," 2024.

C. Aluyah and F. Quarist, "Motivasi Dan Persepsi Pengunjung Terhadap Ruang Terbuka Hijau Jakabaring Sport City Palembang Provinsi Sumatera Selatan," 2024.

Conyers, D. (1994). Perencanaan Sosial di Dunia Ketiga: Suatu Pengantar, Yogyakarta: Gadjah Mada University Press.

D. Nishfi, I. Huda, and C. Prianto, "Analisis Sentimen Layanan Jasa Pengiriman Pada Ulasan Play Store: Systematic Literature Review," 2023, [Online]. Available: https://ejurnalunsam.id/index.php/jicom/

F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," 2022. [Online]. Available: http://j-ptiik.ub.ac.id

G. J. V Turang, S. Sambiran, and D. K. Monintja, "Strategi Dinas Kepemudaan dan Olahraga Dalam Pembinaan Olahraga (Studi di Dinas Kepemudaan dan Olahraga Kota Kotamobagu)," JURNAL GOVERNANCE, vol. 1, no. 2, p. 2021, 2021.

H. M. Falah, M. R. Jamil, A. Taufik, M. Botha, and N. Agustina, "Analysis Sentiment Terhadap Ginjal Akut pada Twitter Menggunakan Algoritma Random Forest," Jurnal Ilmu Komputer dan Informatika, vol. 3, no. 2, pp. 99–106, Jan. 2024, doi: 10.54082/jiki.65.

H. Setiawan, M. Agus Munandar, and L. W. Astuti, "Penggunaan Metode Signatured Based Dalam Pengenalan Pola Serangan di Jaringan Komputer," vol. 8, no. 3, pp. 517–524, 2021, doi: 10.25126/jtiik.202184200.

M. Anggraini, P. Geografi, and U. PGRI Palembang, "Kajian Potensi Pengembangan Wisata Buatan Jakabaring Sport City Kota Palembang," 2021.

M. Tsaqib, Z. R. Mair, Gasim, R. Heriansyah, and H. Setiawan, "Detection and Classification Disease Citrus Nobilis Lour Using Convolutional Neural Network Method," 2024.

Muhammad Yusril Aldean, Paradise, and Novanda Alim Setya Nugraha, "Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac)," 2022.

P. Pratama, E. Indarbensyah, and N. Rochmawati, "Penerapan N-Gram menggunakan Algoritma Random Forest dan Naïve Bayes Classifier pada Analisis Sentimen Kebijakan PPKM 2021," Journal of Informatics and Computer Science, vol. 02, 2021.

Subroto, E., Tensiska, dan Indiarto. R. (2014). Peningkatan Pengetahuan dan Keterampilan dalam upaya Mendukung Ketahanan Pangan di Desa Girijaya dan Mekarjaya, Kecamatan Cikajang, Kabupaten Garut. Dharmakarya. 13 (1) 1-4.

Suwahyono, N. (2004). Pedoman Penampilan Majalah Ilmiah Indonesia. Jakarta: Pusat Dokumentasi dan Informasi Ilmiah, LIPI.

T. Ariaji, S. N. Wahyuni, and M. Ikhsan, "Perbandingan Algoritma SVM dan RF pada Analisis Sentimen menggunakan Pendekatan Machine Learning," Journal Automation Computer Information System, vol. 5, no. 1, pp. 80–92, May 2025, doi: 10.47134/jacis.v5i1.107.

T. Tasmi, F. Antony, D. Dhamyanti, H. Setiawan, and F. Oklilas, "Pengenalan Pola Serangan pada Internet of Thing (IoT) Menggunakan Support Vector Mechine (SVM) dengan Tiga Kernel," Jurnal PROCESSOR, vol. 18, no. 2, Nov. 2023, doi: 10.33998/processor.2023.18.2.1457.

Ully Aulina and Nyanyu, "Peran Promosi Dalam Meningkatkan Wisata Di Pendestrian Kota Palembang," 2021.

Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," Article in Journal of Information Systems and Informatics, 2021, doi: 10.33557/journalisi.v3i3.171.

Y. Pratama, "Penerapan Konsep Green Architecture pada Perancangan Stadion Sepak Bola Standar Internasional di Deli Serdang," 2023.

Z. A. Zainuddin, I. Nurjannah, ; I Made, and K. A. Dahrma, "Stadion Sepak Bola Berstandar Nasional Dengan Pendekatan Arsitektur Post Modern Di Kabupaten Kolaka," 2024.