Extracting Information on Aspects of Sustainable Tourism in ASEAN Using Named Entity Recognition (NER)

S Manalu[1*], Y T Wijaya[1]

[1] Politeknik Statistika STIS, Jakarta, Indonesia

*Corresponding author's email: 222112372@stis.ac.id

Abstract. Sustainable tourism is an important issue in the ASEAN region, which has experienced rapid growth in the tourism sector but faces challenges in maintaining a balance between economic, social, and environmental aspects. Information on sustainability practices is scattered across various forms of text, making it difficult to analyze manually. This study aims to extract information on aspects of sustainability in tourism using a transformer-based Named Entity Recognition (NER) approach. Three data sources were used: government websites, online news, and travel reviews on TripAdvisor. Five transformer models were compared, namely BERT, ALBERT, DistilBERT, ELECTRA, and RoBERTa, to evaluate entity extraction performance. The dataset was divided using an 80:10:10 ratio for training, validation, and testing. The results showed that DistilBERT provided the best performance with a balance of accuracy and computational efficiency. In addition, an analysis of the distribution of sustainability aspects in ASEAN countries and Indonesia in particular was conducted to identify practices that have already been implemented. These findings are expected to contribute to the development of more sustainable tourism policies and practices in the ASEAN region and Indonesia.

Keyword: ASEAN, Information Extraction, NER, Sustainable Tourism, Transformer Model.

1. Introduction

Tourism is a rapidly growing industry, contributing significantly to GDP and job creation [1]. The UNWTO projects that by 2030 there will be 2 billion international tourists with annual revenues of USD 2 billion [2]. Southeast Asia is recommended as one of the tourism-friendly regions [3]. with a projected 125.78 million visitors and becoming the region with the fastest growth in tourist arrivals [4]. Globally, the travel and tourism industry contribute $8,272.3 billion to GDP [5]. with a growth rate of 10.4–11.7% by 2028. In Southeast Asia, tourism contributes 12% of GDP and is expected to rise to 13% [6]. The sector also accounts for 20% of global employment and will create more than 100 million jobs in the coming decade [7] [8].

However, this rapid growth has had negative impacts. Economic dependence on tourism makes countries vulnerable to health and political crises [7] and triggers increases in property prices and living costs in popular destinations. Environmentally, tourism causes ecosystem degradation, carbon emissions, and marine and land pollution [9]. By 2020, tourism accounted for 8% of global greenhouse

gas emissions, with transportation contributing 75% and accommodation 20% [10] [11]. Studies show a correlation between tourism revenue and increased carbon emissions [12][13], underscoring the need for sustainable management.

Research in Southeast Asia has demonstrated that tourism has a positive impact on GDP, employment, and currency value. [1][14][15][16]. However, the potential for ecological damage is also significant [6] [17]. Reference [18] emphasizes the need for a proactive ASEAN strategy to reduce the negative impact of tourists. The ASCC 2025 policy is in line with the SDGs, emphasizing the development of sustainable tourism that reduces carbon emissions and promotes environmentally friendly cities [19]. The strategy includes cross-sector cooperation and the provision of clean public infrastructure [20].

Sustainable tourism aims to minimize negative impacts and maximize economic benefits [21]. The UNWTO designated 2017 as the year of sustainable tourism [9]. ASEAN also prioritizes safety, heritage protection, and climate change mitigation [3]. Although sustainability education programs have been implemented in various destinations [22], and encourages the provision of information for responsible tourist behaviour. In Indonesia, the main challenge is collecting and managing sustainable tourism information that is scattered in unstructured formats [23]. Previous studies tend to use conventional methods [24][25]. This study will examine data from the government, media, and online reviews to produce a more comprehensive analysis.

Advances in NLP through the Transformer architecture have revolutionized information extraction [26]. Models such as BERT, ALBERT, DistilBERT, ELECTRA, and RoBERTa show superior performance on the GLUE benchmark [27]. This study compares the effectiveness and efficiency of models in extracting sustainable tourism information, implements them in a dashboard, and compares Indonesian practices with other ASEAN countries to identify the best strategies.

The researchers will explore the balance between effectiveness, which is the model's ability to extract relevant information from text, and efficiency, which is the use of computational resources and the environmental impact caused by these models. As part of this research, the researchers aim to evaluate which of these models is the most effective and efficient in information extraction applications. In addition, researchers will also implement the results of model development into a dashboard that enables the process of extracting information directly from text, making it easier for users to obtain relevant data quickly and in a structured manner.

Researchers will also analyze the conditions of sustainable tourism in ASEAN and Indonesia and make comparisons between Indonesia and other ASEAN countries to learn about best practices that can be applied in Indonesia. The focus of this research is to understand how ASEAN countries manage tourism by considering sustainability aspects, especially in Indonesia. Thus, this research will not only provide insights into the application of Transformer-based models in information extraction but will also broaden the understanding of sustainability in the context of tourism in the ASEAN region.

The background of the research includes a description of the importance of the research that will be raised as a research plan in the thesis, the reasons underlying or motivating the research. In general, it can begin with an explanation of the research object. Then it continues with the problems found in the research object and how the methods or approaches have been used so far. Then, a solution design for these problems is presented, which also becomes the topic of this research.

2. Research Method

The research method applied in this study uses the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology. CRISP-DM is the most commonly used method in data mining and data

analysis. It is a step-based approach that provides a standard process to guide practitioners in extracting knowledge from data. CRISP-DM consists of six interconnected stages that are carried out iteratively. The stages of CRISP-DM consist of business understanding, data understanding, data preparation, modeling, evaluation, and deployment.

2.1. Business understanding

This study aims to analyze the condition of sustainable tourism in ASEAN with a focus on Indonesia, to identify best practices in environmentally and socially friendly management. Data was collected from government policies, news, and industry reviews, then processed to be more structured in order to overcome the limitations of the existing processing system. This study also developed an automated solution based on natural language processing to accelerate and improve the accuracy of information extraction. In addition, an evaluation of Transformer-based language models (BERT, ALBERT, DistilBERT, ELECTRA, RoBERTa) was conducted to determine the most effective and efficient ones, as well as an analysis of policy implementation in Indonesia and other ASEAN countries to provide strategic recommendations for tourism management that supports environmental sustainability and the welfare of local communities.

2.2. Data understanding

This study uses data from three main sources: news portals, official government tourism websites, and travel reviews on TripAdvisor. News portals provide English-language articles with the keyword "sustainable tourism" from leading media outlets in ASEAN countries (January 1, 2024 – January 31, 2025) to gain insight into public perspectives. Official government tourism websites represent sustainability policies, programs, and initiatives from the government's perspective. Meanwhile, TripAdvisor reviews of 4–5 star hotels and restaurants were used to capture the perspective of tourism industry players. This approach combines the views of three key stakeholders—the government, the public, and industry players—who have different roles and perspectives in the development of sustainable tourism. The community perspective is obtained from news reports containing public opinion, social responses, and the impact of tourism on local communities. The industry perspective is reflected in service descriptions and tourist assessments of sustainability implementation. Through a combination of these three sources, the study builds a comprehensive and balanced picture of sustainable tourism practices in ASEAN as a basis for cross-country analysis and comparison.

2.3. Data preparation

2.3.1. Data Filtering. The filtering process is carried out to ensure that only text relevant to sustainable tourism is used. Three sources of data news, official tourism websites, and Tripadvisor reviews—are filtered using different methods. For news, the selection is carried out in two stages: selecting articles that contain economic, environmental, or social aspects, then separating and reselecting relevant sentences. On official websites, structured text is filtered directly based on its relevance to sustainable tourism. Meanwhile, concise Tripadvisor reviews are selected directly if they contain relevant travel experiences. This approach ensures that the NER model is built from targeted data to identify entities related to sustainable tourism in ASEAN. ensure that affiliations are as full and complete as possible and include the country. The addresses of the authors' affiliations follow the list of authors and should also be indented 25 mm to match the abstract. If the authors are at different addresses, numbered superscripts should be used after each surname to reference an author to his/her address. The numbered superscripts should *not* be inserted using Word's footnote command because this will place the reference in the wrong place at the bottom of the page (or end of the document) rather than next to the address. Ensure that any numbered superscripts used to link author names and addresses start at 1 and continue on to the number of affiliations.

***Website* Pemerintah**	
The government's investment in forest restoration helps create development of local tourism-based enterprises.	1
Visit our theme park for a day of fun and adventure, featuring thrilling rides, live performances, and delicious street food from around the world	0
Berita	
In places like George Town, Community Involvement has helped preserve the city's UNESCO-listed heritage while simultaneously promoting it as a major tourist attraction.	1
A new five-star hotel is set to open in the city center, offering luxury accommodations and exclusive services for international travelers.	0
Ulasan	
The hotel ensures Tourist Health with top-notch sanitization.	1
I loved staying at this eco-lodge! They use renewable energy, and the staff educates guests about conservation efforts.	0

Figure 1. Data Filter.

2.3.2. Data Labelling. NER Labeling The data labeling process in this study was carried out by applying the Named Entity Recognition (NER) approach, which focuses on three main aspects of sustainable tourism, namely economic, environmental, and social. To ensure accuracy and consistency in data annotation, labeling was done manually using Label Studio software in BIO (Begin, Inside, Outside) format. Each word or phrase in the text was analyzed and labeled based on the appropriate entity category:

- The Economic aspect includes information related to the economic impact in the tourism sector, such as revenue generation, job creation, fair wages, and investment in the tourism industry.
- Environmental Aspects include practices that contribute to ecological sustainability, such as eco-friendly accommodations, waste management policies, conservation efforts, and the use of renewable energy.
- Social aspects include the welfare of local communities, cultural preservation, and community involvement in tourism activities, such as cultural heritage preservation, community-based tourism, and the protection of workers in the tourism sector.

Each token in the text is labeled according to the BIO format, namely:

- B- (Begin) marks the beginning of an entity,
- I- (Inside) indicates the continuation of the same entity, and
- O- (Outside) is used for tokens that do not belong to the annotated entity category.

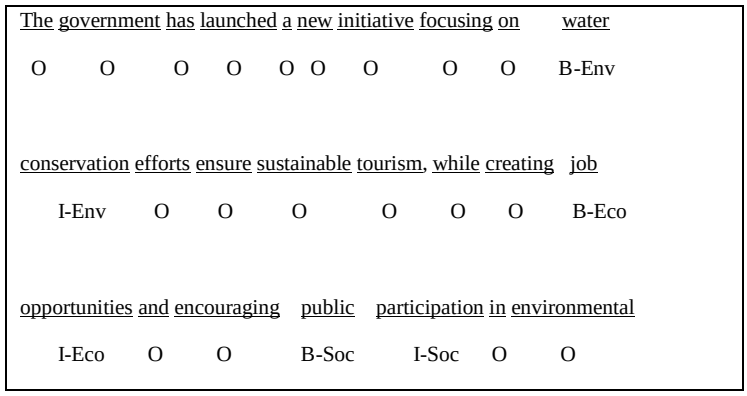

Figure 2. Data Labeling.

The labeling process was carried out by annotators based on labeling guidelines compiled by researchers. Data labeling was performed using Label Studio, a flexible open-source software for annotating various types of text. Each word or token could only have one entity label. To determine the final entity label, the majority vote principle is used, while the reliability of the labeling is evaluated using Cohen's Kappa coefficient to ensure the level of agreement between annotators. The annotators involved are students who have undergone labeling training.

2.3.3. Split Data. The filtered and labeled data is divided into 80% for training, 10% for validation, and 10% for testing. The training data is used to build the NER model's ability to recognize economic, environmental, and social entities. The validation data serves to optimize hyperparameters and prevent overfitting, while the test data objectively evaluates the final performance. This division ensures that the model has high accuracy and good generalization capabilities in identifying sustainable tourism entities from various sources.

2.4. Modelling

At the modeling stage, this research will apply various Transformer-based models in the context of extracting information related to sustainable tourism. The main objective of modeling is to develop a system that can automatically extract relevant data from scattered text sources, such as government policies, news articles, and tourism industry reviews. Several Transformer-based language models that will be evaluated and used include BERT, ALBERT, DistilBERT, ELECTRA, and RoBERTa.

2.5. Evaluation

Evaluation of the model in extracting aspects of sustainable tourism was conducted on testing data. The metrics used to evaluate the performance of the classification model in this study were Precision, Recall, F-1 score, and Accuracy.

Precision

Precision is the ratio of the correct class calculations from all predicted classes. It measures how many correct predictions there are from the total positive predictions given by the model

$$Presisi = \frac{TP}{FP+TP} \quad (1)$$

Recall

Recall is the ratio of correct class predictions compared to all correct data. It measures how many entities that actually exist in the data are successfully detected by the model and ensures that all existing entities are successfully identified.

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

F-1 score

F1- Score is a calculation using the weight of the precision and recall values. The F-1 score provides an overall picture of the model's performance by considering the accuracy and completeness in detecting entities.

$$f1 - score = \frac{2\,(recall \times presisi)}{recall+presisi} \quad (3)$$

Accuracy

An evaluation metric that measures the extent to which the model makes correct predictions compared to the total number of predictions made. Accuracy is calculated by comparing the number of correct predictions to the entire test data.

$$Accuracy = \frac{TP+FN}{TP+TN+FP+FN} \quad (4)$$

Where TP (True Positive) is the correct entity that was successfully recognized, FP (False Positive) is not an entity but was recognized, FN (False Negative) is the correct entity that was not successfully recognized, and TN (True Negative) is not an entity that was not recognized by the model.

2.6. Deployment

The implementation phase of this research produced an NLP-based system for automating information extraction from sources such as policies, news, and tourism industry reviews. This system makes it easier for stakeholders to access structured data to support evidence-based decision making. The most effective and efficient Transformer model will be implemented to optimize data processing, accompanied by recommendations for best practices and sustainable tourism policies. This system is expected to improve policy effectiveness, minimize negative environmental and social impacts, and promote sustainable economic growth in ASEAN, particularly Indonesia.

3. Result and Discussion

3.1. Comparison of Transformer Models Dataset Development

Results of data collection using scraping on data sources. Entity labeling by annotators produced unbalanced data. Cohen's kappa for the dataset on aspects of sustainable tourism was 0.88. This value indicates a very high level of agreement among annotators in assigning labels. Furthermore, this value indicates that the labeling guidelines developed by the researchers were well interpreted by each annotator. The NER labeling guidelines developed by the researchers.

Table 1. Cohen's Kappa Interpretation.

Cohen's Kappa Index	Agreement
< 0,2	Very Low
0,2 – 0,4	Low
0,4 – 0,6	Moderate
0,6 – 0,8	High
0,8 - 1	Very High
< 0,2	Very Low

Table 2. Cohen's Kappa Value.

Label	Cohen's Kappa Value
B - ECONOMIC	0.87
I - ECONOMIC	0.91
B - SOCIAL	0.90
I - SOCIAL	0.88
B - ENVIRONMENT	0.89
I - ENVIRONMENT	0.93

Overall, most labels in this study showed a high level of agreement, with Cohen's Kappa values reflecting excellent consistency among observers. Labels such as B - ECONOMIC, I - ECONOMIC, and B - SOCIAL show a very high level of agreement, indicating strong consistency in the classification of data by annotators. However, the label I - SOCIAL requires attention.

Table 3. Comparison of entities.

Label	Cohen's Kappa Value
B - ECONOMIC	2044
I - ECONOMIC	2378
B - SOCIAL	2043
I - SOCIAL	2055
B - ENVIRONMENT	2004
I - ENVIRONMENT	2006
O	27570

Most tokens in the dataset, approximately 89.66%, have the label "O," which indicates that the word is outside the defined entity. In the BIO (Beginning, Inside, Outside) scheme, the label "O" is used for tokens that do not belong to any entity category. The analysis also shows that the majority of entities consist of two or more words, as evidenced by the number of "Inside" labels that are consistently higher than "Begin" labels for each entity type.

3.2. Comparison of Transformer Models

The application of the model is carried out in several stages, starting with converting the data into a file format compatible with the data format used in model development.

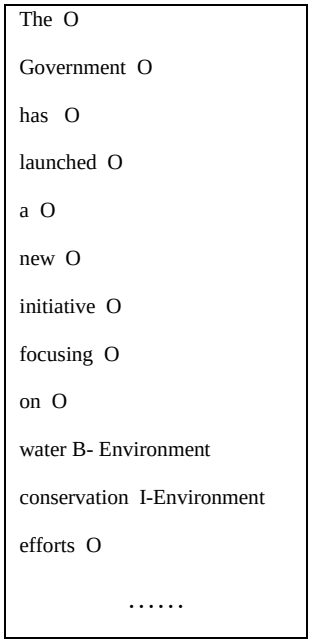

Figure 3. Data Format.

After labeling is complete, the data is divided into three parts: training (80%), validation (10%), and testing (10%) to ensure optimal training, validation during the process, and objective evaluation of new data. The next step is to fine-tune five transformer models—ALBERT, BERT, DistilBERT, ELECTRA, and RoBERTa—trained with the NER dataset specific to this research, taking into account the differences in size and efficiency of each model.

Table 4. Fine Tuned AlBERT.

Label	Precision	Recall	F-1 Score	Accuracy
B – ECONOMIC	0,80	0.94	0.86	
I – ECONOMIC	0.73	0.79	0.76	
B – SOCIAL	0.71	0.80	0.75	0.86
I – SOCIAL	0.89	0.94	0.91	
B – ENVIRONMENT	0.60	0.73	0.66	

I – ENVIRONMENT	0.88	0.91	0.89	
O	0.95	0.95	0.95	
macro avg	0.79	0.87	0.83	

Table 5. Fine Tuned BERT.

Label	Precision	Recall	F-1 Score	Accuracy
B – ECONOMIC	0.83	0.93	0.88	
I – ECONOMIC	0.77	0.80	0.78	
B – SOCIAL	0.84	0.87	0.85	
I – SOCIAL	0.90	0.89	0.89	
B – ENVIRONMENT	0.91	0.85	0.88	0.89
I – ENVIRONMENT	0.73	0.79	0.76	
O	0.92	0.95	0.93	
macro avg	0.84	0.88	0.86	

Table 6. Fine Tuned DistilBERT.

Label	Precision	Recall	F-1 Score	Accuracy
B – ECONOMIC	0.91	0.94	0.87	
I – ECONOMIC	0.92	0.90	0.91	
B – SOCIAL	0.80	0.92	0.86	
I – SOCIAL	0.94	0.94	0.94	
B – ENVIRONMENT	0.86	0.73	0.79	0.90
I – ENVIRONMENT	0.80	0.91	0.85	
O	0.93	0.95	0.94	
macro avg	0.88	0.90	0.88	

Table 7. Fine Tuned Electra.

Label	Precision	Recall	F-1 Score	Accuracy
B – ECONOMIC	0.60	0.79	0.68	
I – ECONOMIC	0.89	0.80	0.84	0.85
B – SOCIAL	0.71	0.81	0.75	

I – SOCIAL	0.60	0.92	0.72	
B – ENVIRONMENT	0.81	0.90	0.85	
I – ENVIRONMENT	0.79	0.87	0.83	
O	0.96	0.95	0.95	
macro avg	0.76	0.86	0.80	

Table 8. Fine Tuned RoBERTa.

Label	Precision	Recall	F-1 Score	Accuracy
B – ECONOMIC	0.85	0.94	0.89	
I – ECONOMIC	0.95	0.88	0.91	
B – SOCIAL	0.98	0.86	0.92	
I – SOCIAL	0.85	0.97	0.91	0.91
B – ENVIRONMENT	0.89	0.81	0.85	
I – ENVIRONMENT	0.98	0.80	0.89	
O	0.91	0.97	0.94	
macro avg	0.92	0.89	0.90	

Table 9. Comparison Model.

Model	Paramters	Training	Testing	Accuracy
ALBERT	12 Juta	13m 17 s	8s	0.86
BERT	110 Juta	30m 0s	18s	0.88
DistilBERT	66 Juta	16m 38 s	10s	0.90
Electra	14 Juta	13m 17s	8s	0.85
RoBERTa	125 Juta	33m 18s	20s	0.91

These results show a trade-off between accuracy and computational efficiency: larger models are more accurate but require more time and resources. while smaller models are faster with a slight decrease in accuracy. DistilBERT is a balanced option because it approaches the performance of large models with shorter training times. The advantages of DistilBERT in this study are likely influenced by the relatively small dataset and short text. allowing lightweight models to process data quickly without losing much accuracy. Optimal training parameters also contributed to these results. Differences with other studies that favor larger models may be due to larger dataset sizes. more complex languages. or longer texts. Thus. the efficiency of DistilBERT is highly relevant in the context of this study but does not necessarily apply to all scenarios.

3.3. *Development of Information Extraction Dashboard*

Dashboard Link: https://website-personal-2-production.up.railway.app/

This information extraction dashboard was developed as a tool to analyze sustainable tourism practices in the ASEAN region. particularly Indonesia. with the aim of presenting previously scattered data in a structured and easily accessible format. Through the interactive map feature. users can see an overview of trends. policies. and the implementation of sustainable practices in each ASEAN country and province in Indonesia. For example. the ASEAN section displays initiatives such as the Bali Sustainable Tourism Program. while the Indonesia section includes practices such as community-based ecotourism in Aceh. This feature facilitates comparative analysis between countries or regions. while identifying potential. good practices. and challenges. In addition to the map. the dashboard has a Data section that contains information sources from three main categories. namely online news sites. official tourism websites. and TripAdvisor reviews. which were collected through a scraping process from January 2024 to February 2025. The Model section allows users to compare the performance of various NER models such as ALBERT. BERT. DistilBERT. Electra. and RoBERTa based on evaluation metrics including Precision. Recall. F1-score. and Accuracy. This feature helps users select the best model for their analysis needs.

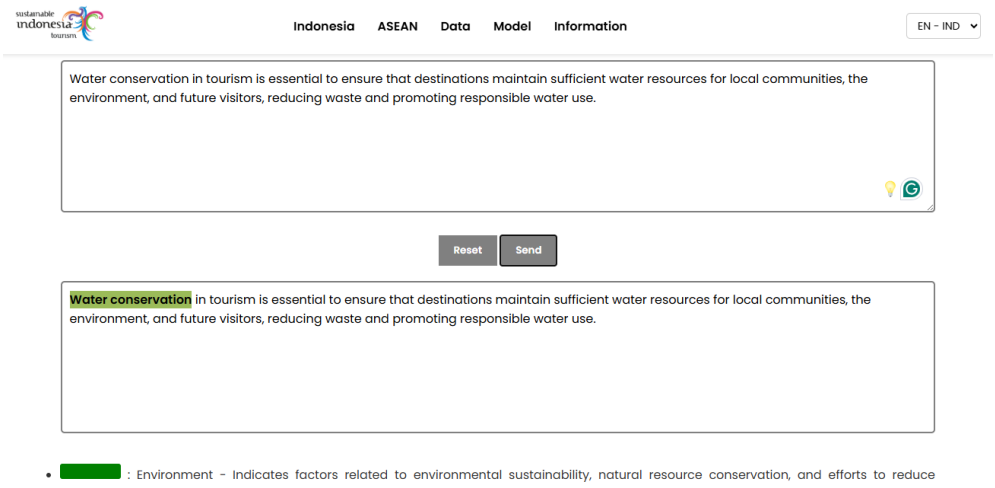

Figure 4. Information Extraction Dashboard Display.

The information extraction section allows users to enter text to be analyzed using the DistilBERT model. which will tag entities according to categories such as Environment. Economic. and Social. For example. the phrase “Water conservation” is recognized as belonging to the Environment category and displayed with a color marker. With this display. users can immediately see the entity classification results and understand the sustainability context contained in the text. so that this dashboard functions not only as a visualization tool but also as a means of interactive testing and exploration of tourism sustainability data.

3.4. *Analysis of Sustainable Tourism Aspects*

3.4.1. Sustainable Tourism Practices in ASEAN. Sustainable tourism in ASEAN has become an important issue as awareness of the wise management of natural and cultural resources increases. Social aspects play a major role. aiming to improve the welfare of local communities and ensure that the benefits of tourism are felt evenly. The government plays a role in creating supportive policies. the mass media disseminates educational information. while reviews from the community and tourists provide feedback for improving practices.

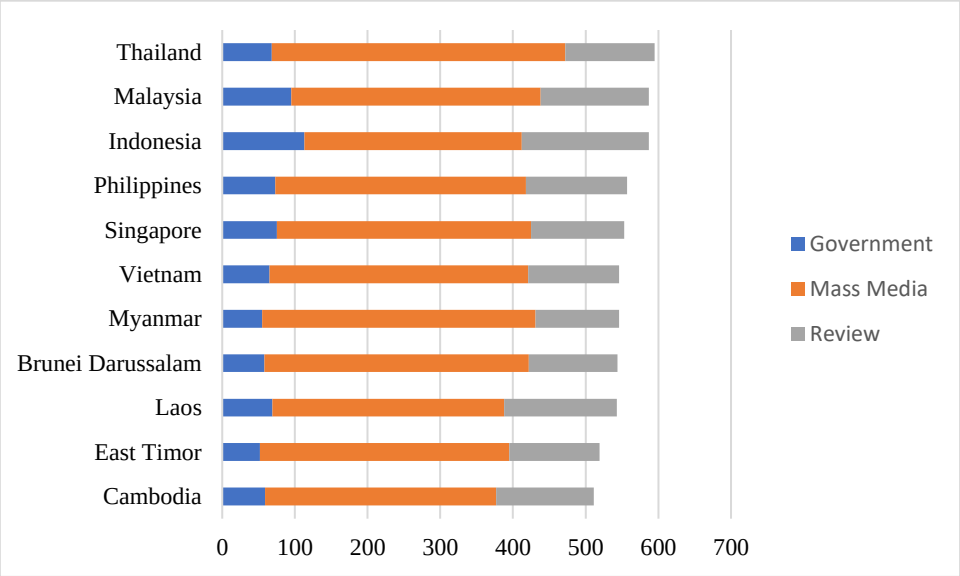

Figure 5. Comparison of the number of entities in each data source.

3.4.2. Economic Aspects of Sustainable Tourism. As tourism develops. this sector has become a major source of income for many ASEAN countries. accelerating economic growth and creating new jobs. In addition. sustainable tourism also supports better infrastructure development. such as transportation. public facilities. and accommodation for tourists. The following graph provides a clearer picture of the number of entities involved in the economic aspects of sustainable tourism in ASEAN countries. showing the role of the tourism sector in each country's economy. This comparison is important for understanding how the tourism sector can make a real contribution to a country's economy in a sustainable context.

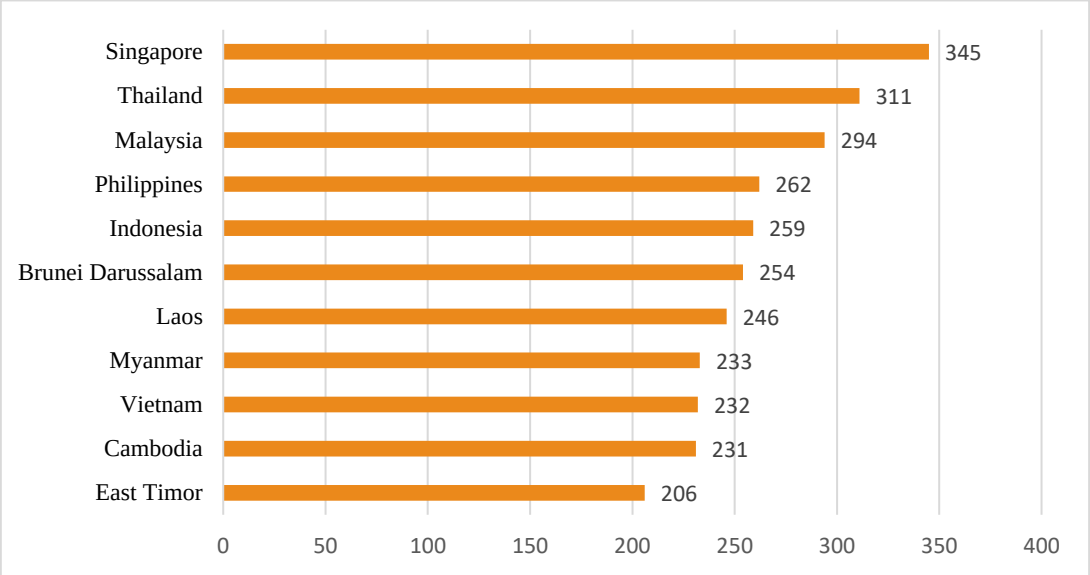

Figure 6. Number of entities in the economic aspect of each ASEAN country.

Overall. this graph provides a clear picture of the important role of the tourism sector in the economies of ASEAN countries.

Figure 7. ASEAN Economy Word Cloud.

The word cloud above provides an interesting visual representation of the economic aspects of sustainable tourism in ASEAN. highlighting the words that appear most frequently in discussions on this topic. Words such as “opportunities.” “economic.” “investment.” “livelihood.” and ‘tourism’ dominate. reflecting the main focus on the economic potential that can be explored through the

development of the tourism sector. For example. the word “opportunities” clearly shows the many opportunities available through the tourism sector. both in terms of creating jobs and encouraging sustainable investment.

3.4.3. Environmental Aspects of Sustainable Tourism. Environmental aspects are an integral part of the concept of sustainable tourism that needs to be considered in the development of tourist destinations in the ASEAN region. The tourism sector. despite its significant economic impact. often puts pressure on existing natural resources and ecosystems. Therefore. the application of sustainable tourism principles that take environmental sustainability into account is very important.

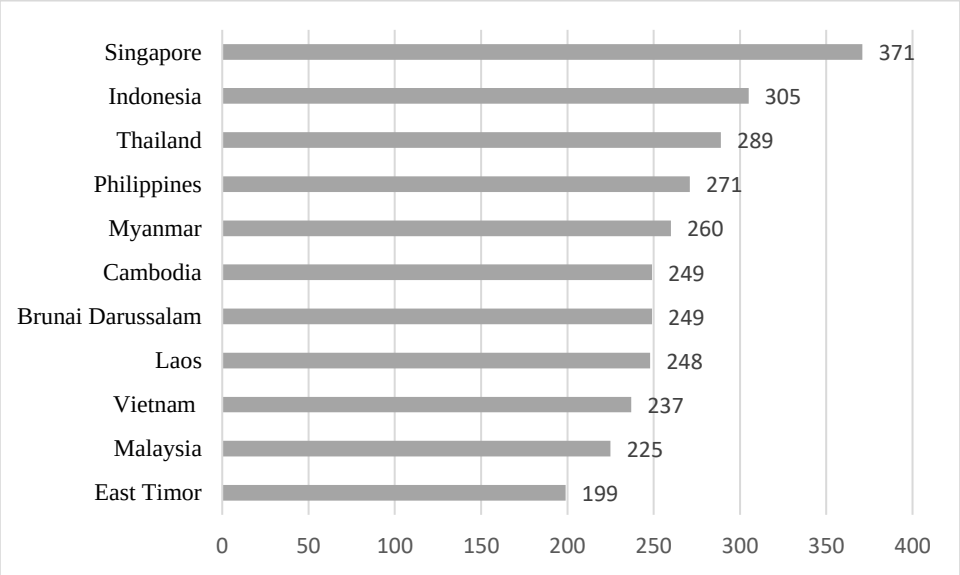

Figure 8. Number of entities in the environment aspect of each ASEAN country.

Singapore has the highest number of entities in the aspect of sustainable tourism with 374 entities. reflecting the country's commitment to sustainable environmental management in the tourism sector. Singapore has long been known as a country that prioritizes cleanliness and environmental sustainability. both in the urbanization and tourism sectors. The country has various policies and initiatives to preserve nature. such as tourist-friendly national park management and the implementation of green technology in tourism infrastructure.

Figure 9. Word cloud of environmental aspects in ASEAN.

The image is used in the discussion of sustainable tourism environmental aspects in the ASEAN region. Some of the most prominent words include “conservation.” “management.” “pollution.” “water.” “plastic.” “nature.” and “sustainability.” The dominance of the word “conservation” shows the importance of nature conservation efforts as part of sustainable tourism. which focuses on habitat protection. natural resource management. and ecosystem conservation. such as coral reefs and forests. In addition. the word “management” indicates the importance of effective management of waste. water. and pollution in the tourism sector to prevent environmental damage. Overall. this word cloud illustrates various practices related to environmental management in sustainable tourism in ASEAN.

3.4.4. Social Aspects of Sustainable Tourism. In the development of sustainable tourism. social aspects play a very important role in supporting the long-term success of this sector. Sustainable tourism is not only about preserving nature and optimizing economic potential. but also about how this sector can empower local communities.

These social aspects are closely related to improving access to education. health. and more equitable employment opportunities.

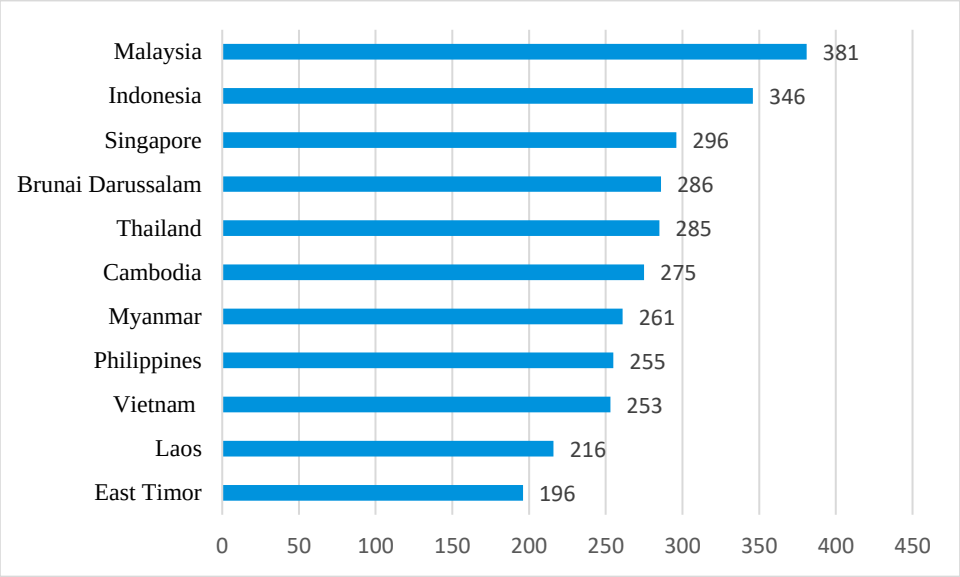

Figure 10. Number of social aspect entities in each ASEAN country.

From the data presented in 7. it can be seen that there are significant differences in the number of social entities involved in the tourism sector in various ASEAN countries. Malaysia is recorded as the country with the highest number of social entities. reaching 381. followed by Indonesia with 346 entities. These figures show that both countries have a strong structure for involving local communities in tourism-related social activities. Singapore and Brunei Darussalam. despite having slightly lower numbers of entities. with 296 and 286 respectively. still demonstrate their commitment to developing the social aspect of sustainable tourism.

Figure 11. Word cloud of social aspects in ASEAN.

The various keywords that appear in this word cloud describe the main values and principles that are the focus of tourism development. which is not only economically profitable. but also pays attention to social and environmental sustainability.

Words such as “Tourist Safety.” “Community.” “Public Participation.” and “Cultural Education” indicate that tourist safety. local community empowerment. and cultural and social education are the main focuses in sustainable tourism development. ‘Community’ and “Empowerment” stand out in this word cloud. emphasizing the importance of community involvement in every stage of tourism program planning and implementation.

3.4.5. Sustainable Tourism Practices in Indonesia. Sustainable tourism practices in Indonesia are efforts to develop the tourism sector in a balanced manner by considering economic. social. and environmental aspects. The government. community. and tourism industry players play an active role in promoting environmentally friendly destination management that is economically competitive and upports cultural preservation and the welfare of local communities.

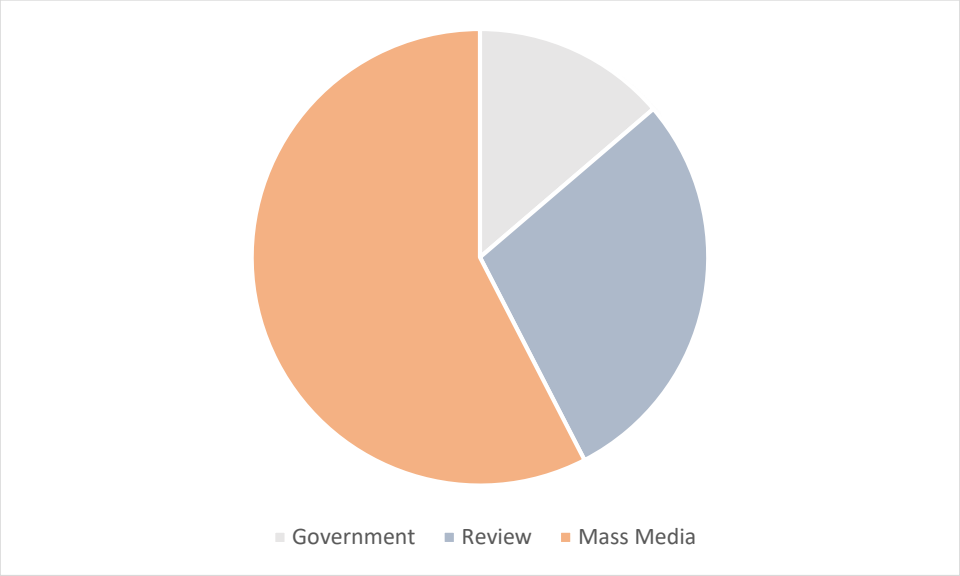

Figure 12. Number of entities based on data sources.

The figure shows the distribution of the number of entities successfully extracted based on data sources. namely from the government. user reviews. and online news. From this visualization. it can be seen that news sources are the largest contributors in generating entities. indicating that online media has a wide and diverse range of information related to tourism issues. including social. economic. and environmental aspects.

3.4.6. Economic Aspects of Sustainable Tourism in Indonesia. In Indonesia. this aspect is often associated with how tourism can drive the local economy through community empowerment and the involvement of micro. small. and medium enterprises (MSMEs). Before visualization in the form of a word cloud. the general understanding of the economic contribution of tourism covers three main aspects. namely increasing state revenue through foreign exchange and taxes. creating job opportunities

in various supporting sectors. and strengthening supporting infrastructure such as transportation. accommodation. and information technology.

Figure 13. Word cloud of Singapore's economic aspects.

Figure 14. Word cloud of Indonesia's economic aspects.

A word cloud visualization-based analysis of Indonesian tourism promotion documents shows the dominance of words such as Economic Growth. Job Opportunities. Tourism Sector. Empowerment. Labor Absorption. and Local-based enterprises. This indicates that Indonesia's economic narrative emphasizes community involvement in the tourism economic system. This approach is manifested through various programs. one of which is the Tourism Village program developed by the Ministry of Tourism and Creative Economy.

3.4.7. Environmental Aspects of Sustainable Tourism in Indonesia

The environmental aspect of sustainable tourism includes efforts to protect. preserve. and manage natural resources wisely so that environmental sustainability is maintained amid increasing tourism activities. A sustainable environment is not only a tourist attraction in itself. but also a key determinant of the long-term sustainability of a destination. Therefore. tourism policies and practices that take environmental conservation into account are important in designing strategies for the sustainable development of this sector.

Figure 15. Word cloud of environment aspects in Singapore.

Figure 16. Word cloud of environment aspects in Indonesia.

In comparison. Singapore has several advantages in managing the environmental aspects of tourism compared to Indonesia. First. environmental policies in Singapore are mandatory and integrated with business licensing. encouraging all tourism industry players to innovate in reducing their ecological footprint. Second. Singapore's capacity for utilizing green technology is very high. for example. the use of water recycling systems. automated solid waste management. and the use of solar energy in public facilities. Third. Singapore has a strong top-down approach. so that environmental regulations can be applied evenly and consistently across all tourism sectors.

3.4.8. Social Aspects of Sustainable Tourism. The social aspect of sustainable tourism focuses on improving the welfare of local communities and ensuring their active participation in the management and development of tourist destinations. The involvement of local communities not only enriches the tourist experience through the uniqueness of the local culture. but also plays an important role in maintaining the social sustainability of a destination.

Figure 17. Word cloud of social aspects in Malaysia.

Figure 18. Word cloud of social aspects in Indonesia.

In comparison. Indonesia has strengths in cultural diversity and broad local community participation. The community-based tourism model spread across various provinces is a distinct advantage because it promotes social balance and a more equitable distribution of benefits. especially in rural areas. However. the challenges are limited human resource capacity building and uneven social support infrastructure. such as access to education and tourism training. Meanwhile. Malaysia excels in the consistency of its cultural preservation programs and the integration of social security aspects. and has a strong institutional support system for the tourism community. Malaysia's approach is also relatively more structured in terms of training. certification of tourism communities. and service quality control.

4. Conclusion

This study confirms that the transformer-based Named Entity Recognition approach can make a significant contribution to identifying sustainable tourism practices in the ASEAN region. The results of the experiment show that RoBERTa achieved the highest accuracy of 91% with a macro F1 score of 0.90. while DistilBERT recorded an accuracy of 90% in only 5 minutes of training time. making it an efficient choice in the context of limited resources. Further analysis shows that each ASEAN country has a different focus on sustainability: Indonesia emphasizes the empowerment of local communities through tourism villages and conservation. Singapore prioritizes innovation and systemic policies that emphasize market efficiency. while Malaysia consistently maintains cultural preservation and social security through structured community programs. These findings prove that the integration of artificial intelligence in tourism research is not only effective for processing complex data. but also strategic in comparing sustainability approaches between countries. Thus. the implications of this research are highly relevant for policymakers. industry practitioners. and academics. particularly in formulating data-driven policies. strengthening community empowerment programs. and enhancing the competitiveness of sustainable tourism in ASEAN.

Acknowledgement

The author would like to express sincere gratitude to the academic advisors and lecturers of the Data Science program who have provided invaluable guidance and constructive feedback throughout the research process. Appreciation is also extended to fellow annotators who contributed their time and expertise in the data annotation process. ensuring the quality and reliability of the dataset. Special thanks are conveyed to the university research facilities and technical staff for their support in providing computational resources. Finally. the author acknowledges the encouragement and moral support from family and friends. which has been instrumental in the successful completion of this work.

References

[1] E. L. Yong. "Tourism Development and Economic Growth in BIMP-EAGA: A Conceptualization with Panel Data Evidence." *Tourism Planning and Development*. vol. 19. no. 6. pp. 500–525. 2022. doi: 10.1080/21568316.2021.2016930.

[2] UNWTO. "International Tourists to Hit 1.8 Billion by 2030." Oct. 2011.

[3] World Economic Forum. "Travel & Tourism Development Index 2021: Rebuilding for a Sustainable and Resilient Future." 2022. Accessed: Mar. 28. 2025. [Online]. Available: https://www3.weforum.org/docs/WEF_Travel_Tourism_Development_2021.pdf

[4] UNWTO. *Tourism Highlights. 2012 Edition*. World Tourism Organization (UNWTO). 2012. doi: 10.18111/9789284414666.

[5] WTTC. "Economic Impact Research." 2019. Accessed: Mar. 28. 2025. [Online]. Available: https://wttc.org/research/economic-impact

[6] F. Ahmad. M. U. Draz. L. Su. and A. Rauf. "Taking the bad with the good: The nexus between tourism and environmental degradation in the lower middle-income Southeast Asian economies." *J Clean Prod*. vol. 233. pp. 1240–1249. Oct. 2019. doi: 10.1016/j.jclepro.2019.06.138.

[7] WTTC. "Travel & Tourism Economic Impact 2021." 2021. Accessed: Mar. 28. 2025. [Online]. Available: https://wttc.org/Portals/0/Documents/Reports/2021/Global%20Economic%20Impact%20and %20Trends%202021.pdf

[8] WTTC. "Travel & Tourism Economic Impact 2022." 2022. Accessed: Mar. 28. 2025. [Online]. Available: https://wttc.org/Portals/0/Documents/Reports/2022/EIR2022-Global%20Trends.pdf

[9] UNWTO. "2017 International Year of Sustainable Tourism for Development." UNTourism.

[10] Y. Jiaqi. S. Yang. Y. Ziqi. L. Tingting. and B. S. X. Teo. "The spillover of tourism development on CO2 emissions: a spatial econometric analysis." Apr. 01. 2022. *Springer Science and Business Media Deutschland GmbH*. doi: 10.1007/s11356-021-17026-z.

[11] Ottmar. Edenhofer. *Climate change 2014: mitigation of climate change: Working Group III contribution to the Fifth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press. 2015.

[12] M. Cetin. E. Ecevit. and A. G. Yucel. "The impact of economic growth. energy consumption. trade openness. and financial development on carbon emissions: empirical evidence from Turkey." *Environmental Science and Pollution Research*. vol. 25. no. 36. pp. 36589–36603. Dec. 2018. doi: 10.1007/s11356-018-3526-5.

[13] D. C. Bojanic and R. B. Warnick. "The Relationship between a Country's Level of Tourism and Environmental Performance." *J Travel Res*. vol. 59. no. 2. pp. 220–230. Feb. 2020. doi: 10.1177/0047287519827394.

[14] A. Benanav. "Demography and Dispossession: Explaining the Growth of the Global Informal Workforce. 1950-2000." *Soc Sci Hist*. vol. 43. no. 4. pp. 679–703. 2019. doi: 10.1017/ssh.2019.34.

[15] N. Benjamin. K. Beegle. F. Recanatini. and M. Santini. "Informal Economy and the World Bank." 2014.

[16] F. Manzoor. L. Wei. M. Asif. M. Z. Ul Haq. and H. Ur Rehman. "The contribution of sustainable tourism to economic growth and employment in Pakistan." *Int J Environ Res Public Health*. vol. 16. no. 19. Oct. 2019. doi: 10.3390/ijerph16193785.

[17] N. Fernandes. "Economic effects of coronavirus outbreak (COVID-19) on the world economy." 2020. [Online]. Available: https://ssrn.com/abstract=3557504

[18] J. W. Lee and T. Brahmasrene. "Investigating the influence of tourism on economic growth and carbon emissions: Evidence from panel analysis of the European Union." *Tour Manag*. vol. 38. pp. 69–76. 2013. doi: 10.1016/j.tourman.2013.02.016.

[19] The ASEAN Secretariat. “ASEAN Socio-Cultural Community Blueprint 2025.”

[20] H. Zhang. Q. Liang. Y. Li. and P. Gao. “Promoting eco-tourism for the green economic recovery in ASEAN.” *Economic Change and Restructuring*. vol. 56. no. 3. pp. 2021–2036. Jun. 2023. doi: 10.1007/s10644-023-09492-x.

[21] R. W. Butler. “Sustainable tourism: a state-of-the-art review.” 1999. Accessed: Jul. 14. 2025. [Online]. Available: https://www.unwto.org/tourism4development2017

[22] S. Pramana. D. Y. Paramartha. G. Y. Ermawan. N. F. Deli. and W. Srimulyani. “Impact of COVID-19 pandemic on tourism in Indonesia.” *Current Issues in Tourism*. vol. 25. no. 15. pp. 2422–2442. 2022. doi: 10.1080/13683500.2021.1968803.

[23] Moh. Fadli. E. Susilo. D. Puspitawati. A. M. Ridjal. D. P. Maharani. and A. Liemanto. “Sustainable Tourism as a Development Strategy in Indonesia.” *Journal of Indonesian Tourism and Development Studies*. vol. 10. no. 1. pp. 23–33. May 2022. doi: 10.21776/ub.jitode.2022.010.01.04.

[24] H. Basri. M. Rami. and Y. Birul Walid Sugandi. “Marketing Promotional Linguistic Contents: A Discourse Analysis On Touristic Magazine. Lombok-Indonesia.” vol. 5. no. 1. p. 2021.

[25] D. Vuković Vojnović. “Multimodal Discourse Analysis of Tourism Websites – the Promotion of Cultural Values Through the Language of Tourism.” 2020. pp. 449–469. doi: 10.18485/bells90.2020.1.ch27.

[26] G. Tucudean. M. Bucos. B. Dragulescu. and C. D. Caleanu. “Natural language processing with transformers: a review.” 2024. *PeerJ Inc.* doi: 10.7717/PEERJ-CS.2222.

[27] M. Escarda. C. Eiras-Franco. B. Cancela. B. Guijarro-Berdiñas. and A. Alonso-Betanzos. “Performance and sustainability of BERT derivatives in dyadic data.” *Expert Syst Appl*. vol. 262. Mar. 2025. doi: 10.1016/j.eswa.2024.125647.