International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Ingénierie des Systèmes d'Information
Vol. 30, No. 11, November, 2025, pp. 2825-2837
Journal homepage: http://iieta.org/journals/isi

Indonesian Sports Text Classification Modeling Based on Few-Shot NER

Nurchim[1,2*], Muljono[3], Edi Noersasongko[3], Ahmad Zainul Fanani[3], Deshinta Arrova Dewi[4]

[1] Doctoral Program of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[2] Faculty of Computer Science, Universitas Duta Bangsa Surakarta, Surakarta 57154, Indonesia
[3] Faculty of Computer Science, Universitas Dian Nuswantoro, Semarang 50131, Indonesia
[4] Faculty of Data Science and Information Technology, INTI International University, Nilai 71800, Malaysia

Corresponding Author Email: p41202000030@mhs.dinus.ac.id

https://doi.org/10.18280/isi.301102

Received: 14 August 2025
Revised: 29 October 2025
Accepted: 10 November 2025
Available online: 30 November 2025

Keywords:
Few-Shot learning, Named Entity Recognition, Indonesian NLP, process innovation, sports text classification, low-resource languages

ABSTRACT

This study investigates the development of a Few-Shot Named Entity Recognition (NER) model for Indonesian sports texts, addressing the persistent challenge of limited annotated resources. While conventional NER approaches such as SpaCy, BiLSTM-CRF, and transformer-based models typically require substantial training data, this work demonstrates that a Few-Shot learning strategy can achieve competitive and in many cases superior performance using only minimal labelled examples. By incorporating prototype-based adaptation and prompt-driven fine-tuning into a pre-trained SpaCy pipeline, the proposed model effectively distinguishes eight Indonesian sport categories despite severe token imbalance and low lexical diversity. This approach is particularly advantageous for this dataset, as Indonesian sports headlines are short, context-limited, and structurally compact, making Few-Shot methods better suited to capturing fine-grained entity distinctions than traditional high-data models. Experimental results show that the Few-Shot model outperforms all baselines, achieving a precision of 0.991 and an F1-score of 0.995, compared with the standard SpaCy model's F1-score of 0.983. These improvements highlight the model's enhanced generalization capability, reduced reliance on extensive annotation, and improved discrimination among closely related sport entities. The key contribution of this study lies in demonstrating the scalability and practical viability of Few-Shot NER for low-resource Indonesian NLP, offering an efficient and adaptable solution for domain-specific information extraction in scenarios where large annotated datasets are unavailable. The findings reinforce the potential of Few-Shot learning as a transformative method for sports analytics, media processing, and broader applications in under-resourced languages.

1. INTRODUCTION

Participation in sports plays a vital role in enhancing individual well-being and strengthening community engagement. Prior studies consistently show that involvement in both competitive and recreational sports improves mental resilience, fosters healthier coping strategies, and contributes to overall life satisfaction through structured social interaction [1, 2]. Participation in sports clubs further supports psychological well-being and nurtures social connectedness, which is especially important for adolescents and internal migrants adjusting to new environments [3]. Elite athletes also derive substantial meaning and purpose from their engagement in athletics, which has been linked to improved self-esteem and greater life satisfaction [4]. Additionally, the promotion of physical activity through social media and digital marketing has proven effective in reducing physical inactivity, demonstrating the increasingly important role of technology in supporting public health and general well-being [5]. Collectively, these findings underscore the multidimensional value of sports in promoting personal health, social cohesion, and life satisfaction across diverse communities.

At the same time, the global sports industry is undergoing rapid transformation driven by technological innovation, digitization, and shifting market dynamics. The emergence of eSports, for instance, has expanded the traditional sports ecosystem by introducing new forms of competition and digital fan engagement through professional virtual leagues [6]. Likewise, the development of the metaverse is reshaping audience experiences, enabling immersive virtual interactions that extend beyond physical venues [7]. Machine learning and data analytics have also become integral to optimizing athletic performance and informing sports betting strategies, where model calibration rather than accuracy alone plays a central role in maximizing financial returns [8]. These developments reflect the increasing interdependence between sports and technology, which continues to shape industry standards and create new commercial opportunities worldwide.

Natural Language Processing (NLP) contributes significantly to this transformation by enabling advanced methods for analysing, organizing, and enhancing sports-related content. NLP has been applied to mental training to help athletes manage anxiety and improve confidence [9]. It also powers automated sports commentary through Large Language Models (LLMs), generating real-time insights and narratives for football and other competitions [10]. In sports journalism, NLP assists in categorizing and structuring large volumes of textual data, including news articles and match reports, thus supporting efficient and timely information delivery [11, 12]. These innovations illustrate the growing potential of NLP in improving athlete preparation, streamlining media workflows, and enriching fan engagement.

Named Entity Recognition (NER) represents a core task within NLP, focusing on the extraction of structured entities from unstructured text. However, NER systems frequently encounter challenges when dealing with domain-specific corpora, ambiguous labels, or imprecise annotations. Noisy annotations in scientific information extraction, for example, can weaken model reliability and highlight the need for adaptive learning frameworks capable of distinguishing between clean and noisy samples [13]. Transformer-based architectures have substantially improved NER performance, yet they require rigorous evaluation using tools such as LangTest to ensure robustness, fairness, and compliance with ethical AI principles [14]. Moreover, recent research has shown the versatility of NER in tasks such as citation prediction, where entity extraction supports automated identification of citation contexts within academic writing [15]. In the context of Indonesian NLP, recent studies have shown that transformer-based architectures, particularly BERT, significantly outperform traditional sequence-labeling and machine-learning models, highlighting the importance of contextualized representations for processing complex Indonesian linguistic structures [16].

Developing NER systems for the Indonesian language, however, presents unique challenges due to the limited availability of annotated corpora across domains. Although resources such as NERSkill.Id provide domain-specific datasets, they lack comprehensive linguistic coverage [17]. Informal texts like tweets require specialized models, including Bidirectional LSTM-CRF architectures, to manage non-standard language and short text structures [18]. In religious contexts, Quran and Hadith translations have been analysed using HMM and SVM approaches, but data scarcity has constrained model effectiveness [19, 20]. Even within news domains where larger corpora exist, issues such as data imbalance and linguistic complexity continue to affect entity identification accuracy, necessitating advanced sampling and neural modeling techniques [21, 22].

In response to these constraints, Few-Shot NER has emerged as a promising solution for low-resource scenarios, offering the ability to learn effectively from only a small number of labelled examples. Few-Shot techniques such as metric learning and prototype networks enable models to classify entities based on similarity to learned prototypes, supporting stronger generalization under limited supervision [23, 24]. Multi-view learning approaches further enhance performance by integrating token-level and span-level features [25], while prompt-based strategies and GPT-driven data augmentation improve model adaptability to new domains [26]. Such methods have demonstrated competitive performance in specialized settings [23, 27], making Few-Shot NER particularly well suited for Indonesian sports texts a domain with limited annotated data yet significant practical relevance. Accordingly, this research aims to develop a Few-Shot NER model for sports classification in Indonesian texts, addressing challenges associated with low-resource environments while demonstrating the broader potential of Few-Shot learning for Indonesian NLP applications.

2. RELATED WORK

2.1 Sports technology innovation

Technology integration in sports has significantly transformed, with digital platforms and virtual environments becoming central to various sporting activities. For instance, platforms like Zwift offer mixed-reality environments where users' real-world performances are translated into virtual settings, enhancing motivation and physical performance through social interactions and immersive experiences [28]. Adopting technology-driven innovations like flywheel resistance training has gained traction in elite sports, providing neuromuscular and strength benefits through controlled, variable loading exercises [29]. These technologies reshape how sports are practiced and understood, pushing the boundaries of traditional sports concepts into the digital realm [30].

Despite these advancements, there are ongoing debates about how technology redefines the nature of sport. Scholars have highlighted that the terms used to describe these new forms, such as eSports, cybersport, and virtual sport, lack consistency, which creates challenges for comprehensive research [30]. However, the convergence of digital and traditional sports provides unique opportunities, especially for sports technology entrepreneurs, who utilize digital platforms to overcome gender biases and foster innovation [31]. As society continues to embrace digital transformation, sport will likely evolve, balancing tradition with cutting-edge technologies [28, 30].

The application of NLP in sports is still in its early stages, though it has shown significant potential across various domains. One example is using machine learning techniques for text classification in sports news [11]. The progress, however, is often constrained by factors such as limited data resources [12] and the complexity of language in sports commentary [10]. Despite the rise of LLMs, which have demonstrated success in sports commentary, the challenge of creating coherent and accurate automated commentary remains, as it requires handling domain-specific knowledge and diverse events.

Additionally, NLP techniques have been explored in the mental training of athletes, demonstrating their utility in improving psychosomatic skills like managing stress and enhancing self-confidence [9]. A study highlighted the benefits of NER-based interventions for managing anxiety and improving performance. This demonstrates that while NLP's application in sports is still developing, its impact is already evident in commentary generation and athlete mental training [9, 10]. These early advancements suggest that NLP could become a more integral part of sports.

2.2 Advances and constraints in domain-specific Named Entity Recognition

NER is an NLP technique that identifies and categorizes specific entities in a text, such as people, organizations, locations, and medical terms. NER plays a significant role in various domains, where it extracts critical information from electronic records and scientific documents [32, 33]. The application of NER has been widely adopted in machine learning and AI models, helping to structure unstructured data [13, 34]. Additionally, NER has been used to create knowledge graphs and supporting intelligent systems [35, 36]. One of NER's primary advantages is its ability to enhance the accuracy and efficiency of information extraction, which leads to improved decision-making [37]. Furthermore, domain adaptation in NER models allows them to handle specific linguistic patterns or other specialized domains, reducing errors in entity recognition and improving the performance of downstream tasks [33, 34].

NER systems can be implemented using different approaches, such as rule-based, machine learning-based, or hybrid techniques, each offering varying degrees of flexibility, accuracy, and resource dependence [23, 38]. Machine learning-based methods are popular due to their adaptability across domains and languages, though they require extensive annotated data. Combining rule-based and machine-learning techniques, hybrid models can enhance system performance, especially in complex domains [39]. Recent developments have introduced deep learning architectures, such as Long Short-Term Memory (LSTM), Gated Recurrent Units (GRU), and Convolutional Neural Networks (CNN), which improve NER performance by better capturing semantic and syntactic features [40]. These advanced methods significantly enhance NER tasks, particularly in under-resourced languages where traditional methods fall short [41].

NER has become an essential task in various specialized fields, but domain-specific challenges, data limitations, and the complexity of linguistic structures often hinder its effectiveness. For instance, the lack of publicly available annotated datasets hampers advancements in identifying key entities related to crop diseases and pests in agricultural domains [42]. Similarly, materials science faces challenges with overlapping entities in scientific literature, prompting Machine Reading Comprehension (MRC) techniques to enhance accuracy by converting entity extraction into a question-answering task [43]. In historical and low-resource domains, NER's complexity increases due to the lack of training data and the degradation of historical documents. In ancient Chinese texts, the concise nature of the language complicates the recognition process, while integrating external knowledge, such as dictionaries, can enhance the representation of entities [44]. In contrast, LLMs offer a potential solution for historical document NER but face challenges when handling noisy inputs, as their performance still lags behind dedicated NER systems [45].

Challenges in NER often arise from the limitations of data and the complexity of nested structures. One challenge is the recognition of nested entities, where entities overlap, making traditional models inadequate. Sequence labeling methods face performance issues due to the complexity of hierarchical structures, which results in poor accuracy and error propagation when external entities are misclassified after internal ones are recognized [44, 45]. While effective in theory, Span-based approaches require the enumeration of all possible spans, leading to high computational complexity and frequent misidentification of non-entities [46-48]. These issues highlight the difficulty in dealing with entities that have complex positional dependencies, further complicating the task of NER.

Another significant area for improvement stems from language-specific challenges. For example, Chinese NER is complicated by the need for word delimiters, making it difficult to identify entity boundaries, particularly in nested entities. Multi-feature fusion, which incorporates glyphs and pinyin, is essential but increases the complexity of the models [49]. Similarly, agglutinative languages such as Turkish present challenges in informal text settings, where abbreviations, irregular grammar, and orthographic errors are standard, leading to a substantial decrease in the performance of NER models [50]. These limitations underscore the need for more adaptable models capable of handling diverse linguistic features and overcoming the challenges posed by complex and noisy data.

2.3 Few-Shot learning

Few-Shot learning has become a crucial research subject aimed at tackling the issue of training models with few labelled data across several fields. Initial research primarily focused on meta-learning, which instructs a model on the learning process, using episodic training to improve generalization across tasks with few instances [49, 50]. Nonetheless, obstacles such as overfitting to the support set and meta-dataset persist, as seen in studies on ensemble meta-learning and adaptive transformers [51-53].

A notable strategy in Few-Shot learning is model-centric techniques, whereby researchers enhance designs or use ensemble learning to increase learning efficacy from limited datasets [52, 53]. Examples include the integration of multi-input and multi-output setups inside meta-learning to improve generalization performance by circumventing local minima during training [52]. The fine-tuning of pre-trained networks has proven effective, especially in image-based damage detection, where models such as ProtoNet exhibit encouraging outcomes after fine-tuning with a limited support set [52, 54].

Conversely, data-centric solutions emphasize enhancing data quality and addressing domain transitions, often hindering model performance in Few-Shot contexts [49, 51]. Cross-domain adaptation techniques seek to alleviate the effects of distribution discrepancies between the source and target domains. Moreover, self-supervised learning is often used to exploit unlabeled data successfully, as shown in recent initiatives for cross-domain threat intelligence extraction and segmentation tasks [23, 51].

Unsupervised and lifetime learning are becoming significant as academics investigate techniques to enhance model flexibility without extensive annotated datasets [55-57]. Unsupervised meta-learning networks use task-specific reconstruction techniques for domain adaptability, enhancing feature extraction across tasks with few labelled data. Lifelong ensemble learning in robotics utilizes many representations to adapt to new object categories dynamically, optimizing the balance between identification accuracy and processing efficiency.

LLMs and prompt-based fine-tuning have gained significance in text-based Few-Shot applications, including code review automation and named entity identification in cybersecurity. Few-Shot strategies, including personal-specific instructions and embedding-based retrieval methods, provide effective solutions for LLMs, especially when extensive fine-tuning is unfeasible [58]. These developments underscore the progression of Few-Shot learning towards more flexible, efficient, and scalable solutions relevant across many domains and workloads.

3. METHOD

This research employs a structured approach to develop a NER model using Few-Shot learning, as shown in Figure 1. The process begins with collecting relevant datasets for the specific NER problem domain. The system then processes the collected data by cleaning, transforming, and formatting it to ensure consistency and prepare it for annotation. Subsequently, domain experts conduct the data annotation phase, labelling entities in the dataset and creating a high-quality annotated corpus to serve as the foundation for the model.

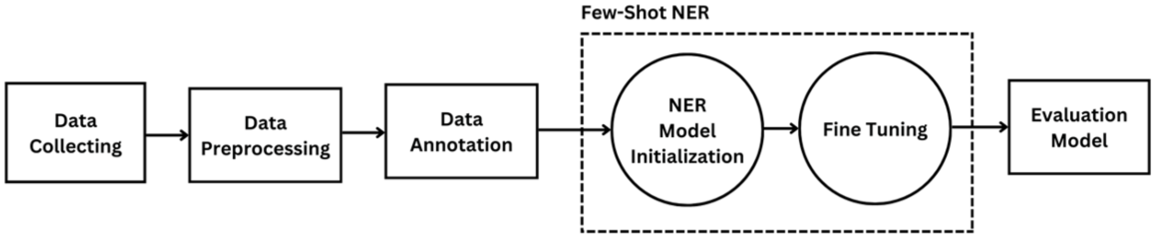


Figure 1. The proposed method Few-Shot NER

The Few-Shot NER framework integrates the annotated data through two core stages: initializing and fine-tuning the NER model. In the initialization phase, a pre-trained or baseline NER model leverages existing linguistic knowledge. The model is then fine-tuned with annotated data, using Few-Shot learning techniques to adapt effectively with minimal labelled data. The fine-tuned model is subjected to an evaluation phase to assess its performance based on precision, recall, and F1-score metrics. This research method ensures an efficient model development cycle by minimizing the need for large annotated datasets while maintaining high performance, demonstrating the effectiveness of Few-Shot learning in NER tasks.

Few-Shot learning offers several key advantages, making it highly beneficial in scenarios with limited data. Firstly, it allows models to achieve high performance even with minimal labeled data, which is crucial in fields where data collection is expensive or time-consuming [23]. Additionally, Few-Shot learning helps improve the generalization capabilities of models, allowing them to adapt to new tasks more efficiently without requiring extensive retraining [59]. Moreover, it significantly reduces the risk of overfitting, as the model is not trained on large, possibly redundant datasets [55, 58]. Lastly, its application in cross-domain tasks has been enhanced with techniques like adaptive transformers, which help bridge domain gaps and improve task adaptability [53].

4. RESULT

4.1 Data collecting

The news dataset was compiled through automated retrieval using Google News RSS feeds, which aggregate articles from a wide range of publishers. This approach enables efficient collection of diverse headlines while minimizing manual intervention. A sample of the extracted data showing the structured format containing the news title, publication date, and associated URL is presented in Figure 2. The final dataset comprises 1543 Indonesian sports-related news headlines, which were subsequently categorized according to the sport taxonomy listed in Table 1.

No	News_Titles
1	Open Tournament Bola Voli Piala Panglima TNI 2024 Digelar di Solo
2	Tim bola voli putra Jabar pertahankan medali emas setelah hajar Jateng
3	Jatim tuntaskan dahaga juara sepak bola PON setelah penantian 16 tahun
...	...
...	...
1541	Pelatihan terpusat kunci tim golf Jateng raih emas perdana pada PON
1542	Jabar bangkit untuk curi emas pertama bulu tangkis PON 2024
1543	Kalahkan Jatim, Tim Bola Basket DKI Jakarta Sukses Raih Medali Emas PON XXI Aceh Sumut 2024

(a) Indonesia language

No	News_Titles
1	The 2024 Panglima TNI Cup Open Volleyball Tournament Held in Solo
2	West Java men's volleyball team retains the gold medal after defeating Central Java
3	East Java ends a 16-year championship drought in PON football
...	...
...	...
1541	Centralized training becomes the key for Central Java's golf team to win their first gold medal at PON
1542	West Java rises to seize their first badminton gold medal at PON 2024
1543	Defeating East Java, the DKI Jakarta basketball team successfully wins the gold medal at the 21st PON Aceh-North Sumatra 2024

(b) English language

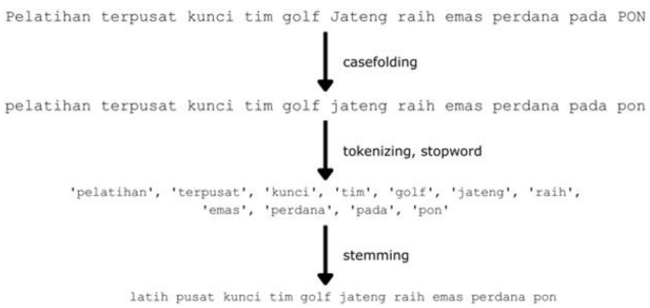
Figure 2. Data collecting result

Table 1. Indonesian sport entities

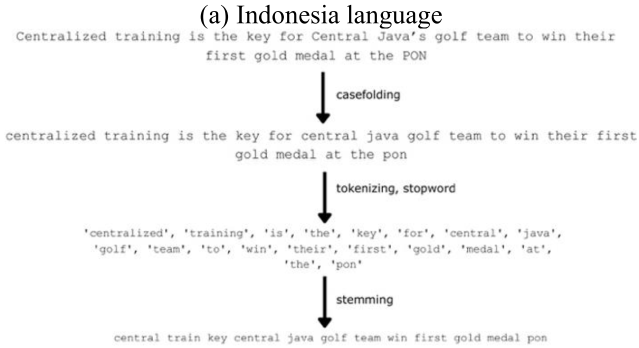
Sport Classification	Indonesian Sport		Entity	Total Number of News Titles Datasets
	Case-1	Case-2		
Combat	Tinju	Taekwondo	B-Sport_Combat	195
Individual aesthetic	Selancar Ombak	Diving	B-Sport_Individual-Aesthetic	190
Individual aiming	Golf	Panjat Tebing	B-Sport_Individual-aiming	191
Racing	Renang	Kano	B-Sport_Racing	193
Net/court	Bulutangkis	Voli	B-Sport_Net/Court	198
Invasion	Sepak Bola	Bola Basket	B-Sport_Invasion	196
Fielding	Baseball	Softball	B-Sport_Fielding	190
Target	Bowling	Panahan	B-Sport_Target	190

4.2 Pre-processing pipeline for Few-Shot NER

To prepare the text for Few-Shot NER, a comprehensive pre-processing pipeline was developed for the Indonesian language. This process includes lowercasing, punctuation and numeric removal, stop word elimination, tokenization, and stemming. The outcome of the Indonesian pipeline is shown in Figure 3(a), demonstrating how raw input is transformed into a cleaned and standardized token sequence. For comparison, Figure 3(b) presents the same pipeline applied to English text, highlighting cross-language consistency and generalizability of the pre-processing steps. These transformations ensure uniform input structure, thereby enhancing downstream model robustness and reducing noise-related variability.



(a) Indonesia language



(b) English language

Figure 3. Data collecting result

4.3 Data annotation

Annotation follows the widely adopted BIO tagging scheme, enabling explicit identification of entity boundaries within each sentence. An example of the annotated output is illustrated in Figure 4. In this example, tokens such as *golf* are labeled as B-Sport_Individual-aiming, marking the beginning of an entity span, while other non-entity tokens receive the O label. This annotation framework provides structural clarity and facilitates accurate supervised training of NER models, particularly in specialized domains such as sports news.

```
latih   O
pusat   O
kunci   O
tim     O
golf    B-Sport_Individual-aiming
jateng  O
raih    O
emas    O
perdana O
pon     O
```

Figure 4. Data annotation result

4.4 Dataset diagnostic analysis

A deeper evaluation of the dataset's structure was conducted to assess its suitability for Few-Shot NER and to identify potential sources of bias. Sentence length distributions for both training and test sets are presented in Figure 5. The boxplot reveals that the corpus is dominated by short headlines, with average lengths of 10.43 tokens (train) and 10.61 tokens (test). This characteristic reflects the typical structure of Indonesian news headlines, which tend to be concise and highly compressed.

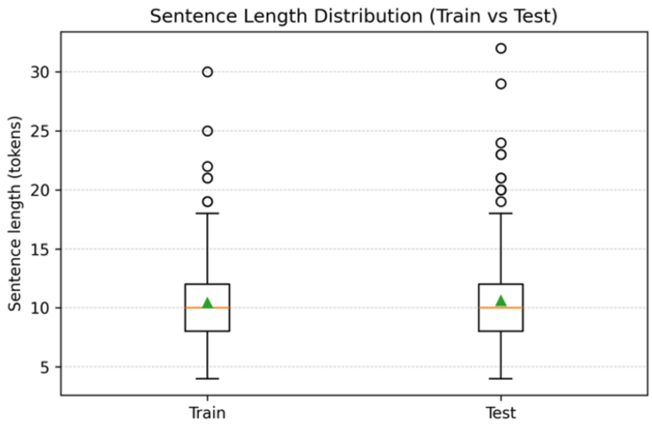


Figure 5. Sentence length distribution in train and test sets

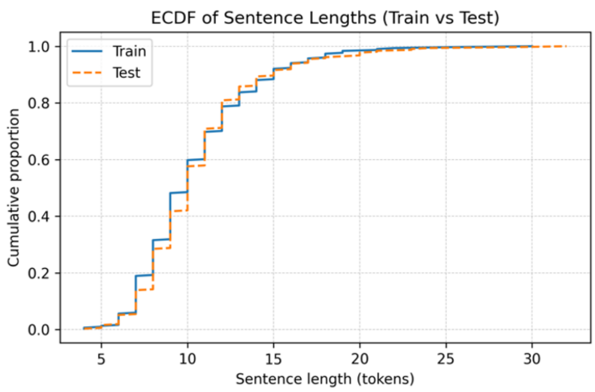


Figure 6. ECDF of sentence lengths

Further analysis using the Empirical Cumulative Distribution Function (ECDF) shown in Figure 6 demonstrates near-identical cumulative patterns between the two splits. This consistency indicates an effective and well-balanced data partition without structural drift.

Token-level label comparison, visualized in Figure 7, highlights a pronounced class imbalance. The O label accounts for approximately 89% of all tokens across both splits, while each sport-related entity label constitutes less than 1.3%. Although the distribution between splits is stable (points cluster closely along the diagonal), the extreme imbalance underscores the importance of compensatory strategies such as class weighting or data augmentation during model training.

A span-level analysis of entity categories is provided in Figure 8. All eight sport categories exhibit comparable proportions, each contributing between 9% and 14% of total spans, indicating balanced representation at the entity level. However, lexical diversity within categories remains limited some categories contain only one to three unique surface forms. This low diversity raises concerns regarding potential lexical bias, where the model may memorize specific entity words rather than learn generalized patterns.

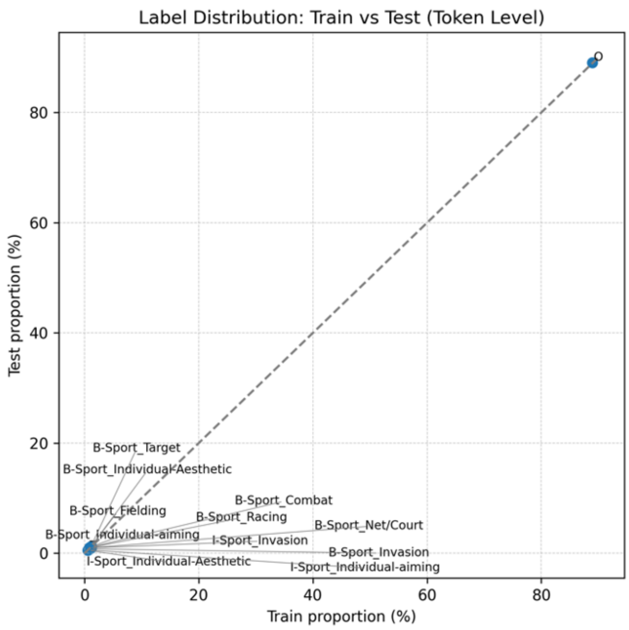
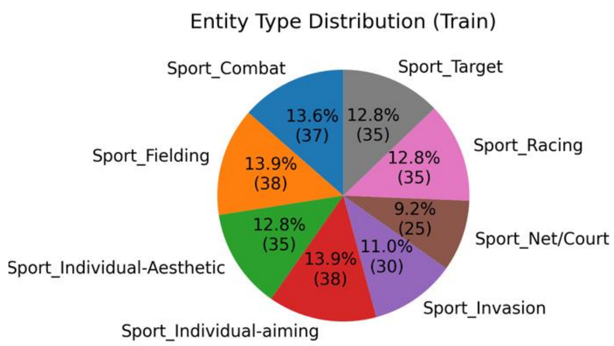
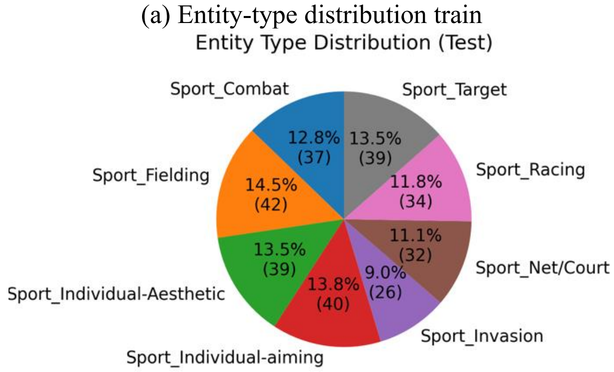


Figure 7. Token-level label distribution comparison



(a) Entity-type distribution train



(b) Entity-type distribution train

Figure 8. Entity-type distribution across train and test sets

Finally, Figure 9 compares entity-type proportions across the two splits. Points aligning closely with the diagonal confirm that the train-test partitioning maintained consistent entity distributions, ensuring fair evaluation conditions without category-specific skew.

Collectively, these analyses provide an in-depth understanding of dataset characteristics and highlight both strength such as well-balanced entity-type distribution and limitations, particularly in token-level imbalance and low surface-form diversity. These insights directly inform the design and expected performance of Few-Shot NER models applied to Indonesian sports news.

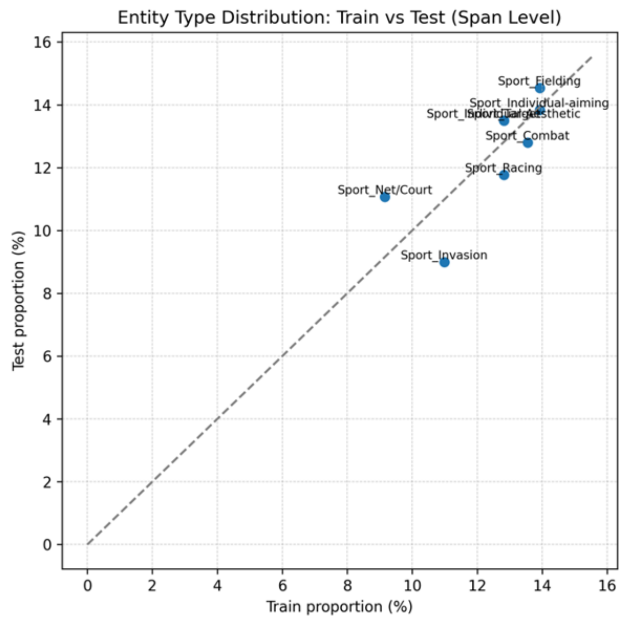


Figure 9. Span-level entity-type proportion comparison

Algorithm 1: Model Initialization Using SpaCy

Function train_ner_model(train_data_path, n_iter=5)
Create a blank nlp model for Indonesian (nlp)
Add a ner pipeline component to the model (ner)
Load training data from disk (train_data)
Convert data into a list of documents (train_data)
For each document in train_data:
For each entity in the document:
Add entity label to the ner pipeline (ner)
Initialize the model's optimizer
For i from 1 to n_iter:
Initialize an empty dictionary for losses
Create batches of training data with varying size (batches)
For each batch in batches:
For each document in the batch:
Convert document entities into example format
Update the model with the optimizer
Track the losses
Print the losses for the current iteration
Return the trained nlp model

4.5 Few-Shot NER fine tuning model

The construction of a few-Shot model comprises two essential phases: model initialization and fine-tuning. Model initialization involves loading a pre-trained pipeline that provides a foundation for customization, guaranteeing the model has essential linguistic characteristics such as tokenization, part-of-speech tagging, and named entity identification. The initialization stage is formally described in Algorithm 1, which summarizes the procedural pipeline for constructing the base Indonesian SpaCy model prior to downstream Few-Shot adaptation. Subsequently, fine-tuning entails adjusting or training the model using a limited, task-specific dataset by updating pertinent components, such as including new labels for text categorization or custom entities. When used with appropriate batching, the nlp.update ()

method in SpaCy modifies the model parameters according to the supplied Few-Shot data. Subsequently, following training, the optimized model is assessed for efficacy and preserved for deployment using nlp.to_disk () to guarantee its suitability for the designated application.

The function train_ner_model initializes a blank SpaCy model for Indonesian, adds an NER pipeline, and loads training data from a DocBin file. It registers entity labels from the training data and begins model training with an optimizer. The training loop runs for a specified number of iterations, where the data is divided into batches using minibatch and compounding. Each batch is used to update the model with Example objects that map entity spans, and the losses are tracked and printed for each iteration. Finally, the trained model is returned for further use.

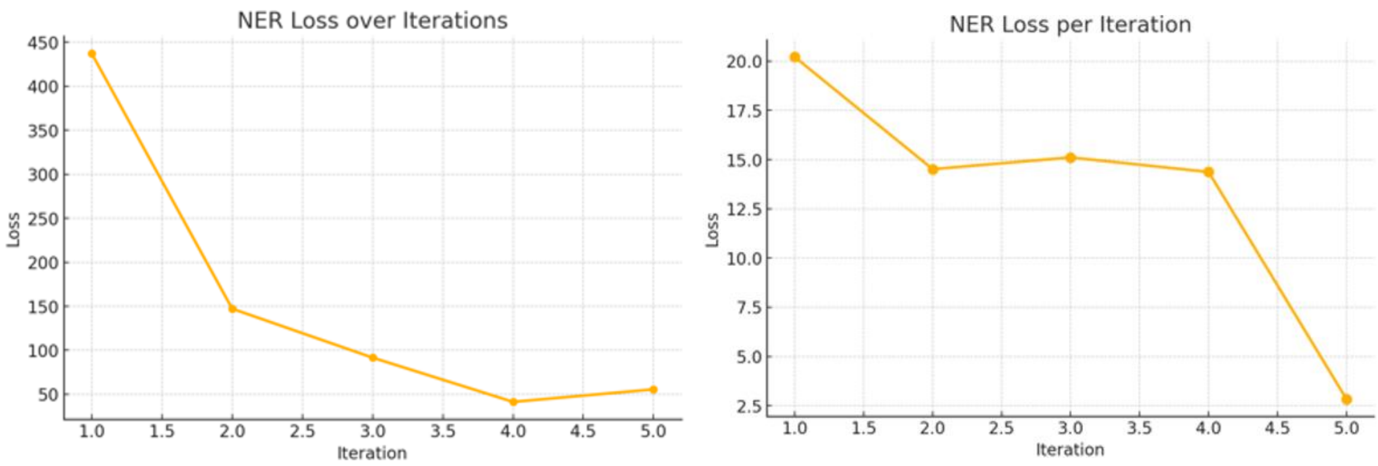
Algorithm 2: Fine Tuning Model Using Few-Shot Learning

```
Function fine_tune_ner_model(nlp, few_shot_data_path, n_iter=5):
    Load data from 'few_shot_data_path' into 'doc_bin'
    Extract documents from 'doc_bin' into 'few_shot_data'
    Retrieve the 'ner' pipeline component from 'nlp'
    For each 'doc' in 'few_shot_data':
    For each 'entity' in 'doc.ents':
    Add 'entity.label_' to the 'ner' component
    Initialize the optimizer using 'nlp.resume_training()'
    FOR 'i' in range from 0 to 'n_iter - 1':
    Initialize an empty dictionary 'losses'
    Generate 'batches' from 'few_shot_data'
    For Each 'batch' in 'batches':
    For each 'doc' in 'batch':
    Convert 'doc' into 'example' with entities labelled
    Update the model with the 'example' and track 'losses'
    Print "Iteration (i + 1), Losses: (losses)"
Return 'nlp'
```

The pseudocode Algorithm 2 describes the process of fine-tuning an existing NER model using Few-Shot data. It begins by loading the Few-Shot data from a DocBin file and extracting the documents with their annotated entities. The NER component of the NLP model is accessed, and any new entity labels found in the data are added to the model. The training resumes using an optimizer, running for a specified number of iterations (n_iter). In each iteration, the data is divided into batches of increasing size, and the model is updated with examples created from the documents and their annotations, while tracking the training losses. After all iterations have been completed, the refined model is returned for further application.

Figure 10 shows how loss decreased a NER model across five iterations. The first graph reveals a high loss of roughly 450, which decreases markedly by the second repetition, signifying substantial improvement. Nonetheless, after the fourth cycle, the loss exhibits a tiny rise, indicating a little variability. The second graph, by comparison, shows a smaller range of loss values, commencing at around 20 and progressively declining. Although the pattern seems more consistent in the first rounds, a significant decrease in loss transpires in the final iteration. These graphs illustrate the enhancement of the model's performance, with considerable diversity in optimization across the training cycles, particularly after fine-tuning.

Loss iteration in NER refers to the process of calculating and minimizing the error (loss) during the training of a machine learning model. The model predicts which words in a sentence correspond to specific named entities in each iteration. The model's predictions are compared to the actual labeled entities, and a loss function measures how far off the predictions are. This loss is then used to update the model's parameters via an optimization algorithm (like gradient descent), improving its accuracy over time. Each iteration refines the model, gradually reducing the loss until it reaches an optimal performance.



(a) SpaCy (b) Few-Shot fine tuning

Figure 10. Comparison of loss iteration between SpaCy and Few-Shot fine tuning

4.6 Evaluation model

The assessment of model performance must be conducted using the commonly used metrics based on accuracy. This study includes evaluation metrics such as precision (P), recall (R), and F1-score. The formulas for calculating precision (P), recall (R), and F1-score are as follows:

$$Precision\ (P) = \frac{True\ Positive}{(True\ Positive + False\ Positive)} \tag{1}$$

$$Recall\ (R) = \frac{True\ Positive}{True\ Positive + False\ Negatives} \tag{2}$$

$$F1 = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (3)$$

Table 2. Performance metrics of the Indonesian Few-Shot NER model across evaluation scenarios

Model	P	R	F1
SpaCy	0.967	1	0.983
Bi-LSTM-CRF	0.959	0.795	0.869
Transformer-IndoBERT	0.933	0.989	0.961
Few-Shot Fine tuning (proposed model)	0.991	1	0.995

The results presented in Table 2 demonstrate a clear performance advantage of the proposed Few-Shot Fine-Tuning model over the baseline methods. While the SpaCy model achieves a strong F1-score of 0.983 and perfect recall, its slightly lower precision indicates a modest tendency to over-predict entities. The Bi-LSTM-CRF model performs considerably weaker, particularly in recall (0.795), suggesting that it fails to identify many true entities likely due to its limited ability to capture contextual nuances in Indonesian sports headlines. The Transformer-IndoBERT model provides a substantial improvement, reaching an F1-score of 0.961 with a well-balanced precision and recall profile, reflecting the benefits of contextualized language modeling. However, the proposed Few-Shot Fine-Tuning approach surpasses all baselines with an F1-score of 0.995, supported by near-perfect precision (0.991) and perfect recall. This exceptional performance indicates that adapting a pre-trained model with a domain-specific Few-Shot dataset enables highly effective entity representation, allowing the model to generalize well within the sports domain while minimizing false predictions.

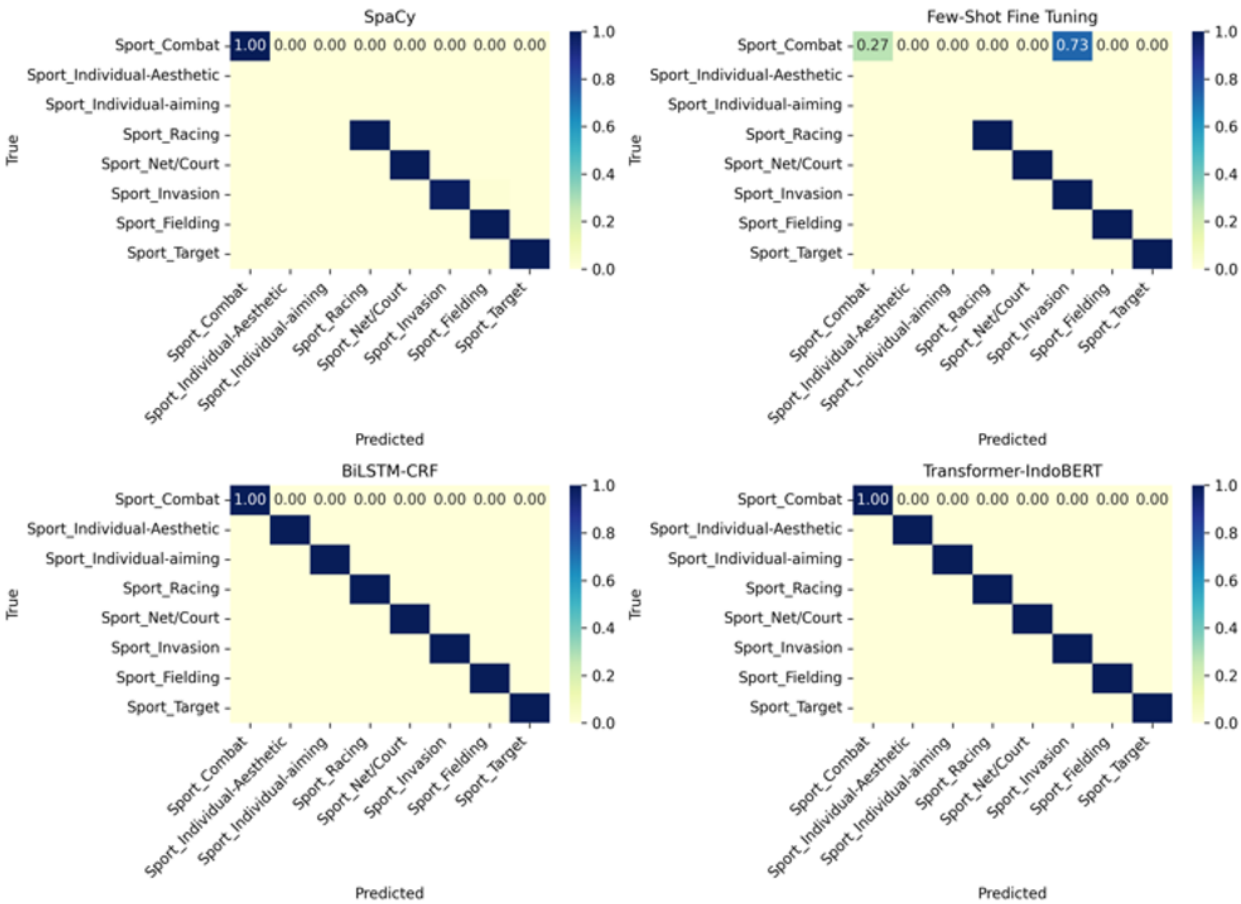


Figure 11. Entity-level, row-normalized confusion matrices for four Indonesian sports NER architectures

4.7 Error analysis and improvement methods

The entity-level confusion matrices across the four evaluated NER architectures SpaCy baseline, BiLSTM-CRF, Transformer-IndoBERT, and the proposed Few-Shot fine-tuned model provide a detailed view of misclassification behaviour among the eight Indonesian sport categories (Figure 11). While the SpaCy baseline, BiLSTM-CRF, and IndoBERT models display strong diagonal dominance with only negligible off-diagonal mass, indicating stable cross-category discrimination, the Few-Shot fine-tuned model exhibits marked deviations from this ideal pattern. In particular, a substantial proportion of *Sport_Combat* instances are confused with *Sport_Target* and, to a lesser extent, with *Sport_Individual-Aesthetic*, despite the generally high span-level F1 achieved in aggregate evaluation. This asymmetry is consistent with recent findings showing that NER models can appear robust at the corpus level while still exhibiting systematic, label-specific errors that only become visible through fine-grained confusion analysis and per-label diagnostics [59-62]. Given that Indonesian sports headlines are short and lexically compressed, with limited surface-form diversity within each category, the Few-Shot model is particularly prone to over-generalising from a small number of support examples and collapsing semantically adjacent classes into a single dominant prototype.

These error patterns point to several concrete directions for improving the robustness and generalisation capacity of the Few-Shot pipeline. First, an error-driven augmentation regime can be adopted, in which misclassified spans from categories involved in confusion (for example, *Sport_Combat* vs. *Sport_Target*) are selectively re-annotated and injected into the support set, sharpening local decision boundaries and mitigating prototype collapse. Related work on prototype-based and taxonomy-guided Few-Shot NER has shown that explicitly encoding label structure and class prototypes can substantially improve discrimination between closely related entity types [63]. Second, integrating prompt-based or continuous-prompt heads on top of the current transformer backbone may further align the model's latent space with label semantics; prompt-based metric learning and continuous prompt tuning have been reported to yield consistent gains in Few-Shot NER by enriching the interaction between label descriptions and contextual representations [61, 62]. Third, to alleviate the impact of low lexical variation and domain skew, the model could benefit from task-adaptive pre-training (TAPT) on a broader Indonesian sports corpus [63-66], followed by knowledge-guided instance generation that synthesises diverse yet semantically coherent entity mentions [65]. Together, these strategies offer a principled pathway to move from a strong yet narrowly adapted model towards a more stable and transferable NER system for low-resource Indonesian sports text.

4.8 Discussion

The experimental results show that the proposed Few-Shot fine-tuning model achieves state-of-the-art performance for Indonesian sports NER under low-resource conditions. The model attains an F1-score of 0.995, clearly outperforming the SpaCy baseline (0.983), Transformer-IndoBERT (0.961), and Bi-LSTM-CRF (0.869). This combination of near-perfect recall and very high precision indicates that the model is able to capture almost all relevant entities while introducing very few false positives. In the context of Few-Shot NER, where achieving stable performance with limited annotation is notoriously challenging [67, 68], such results suggest that a carefully designed fine-tuning strategy on top of a strong pre-trained pipeline can be highly effective.

The dataset diagnostics provide important context for interpreting these findings. The corpus consists of short and syntactically compact Indonesian sports headlines with highly similar sentence-length distributions across training and test sets, which reduces structural distribution shift. At the token level, the label distribution is strongly skewed toward the O class (around 89%), while each sport-related entity label accounts for less than 1.3% of tokens. Despite this severe imbalance, span-level distributions across the eight sport categories remain relatively balanced, each contributing roughly 9-14% of all entities. This structure implies that the model must cope with extreme token-level skew but still receives a reasonably representative set of entity spans, a setting in which Few-Shot learning can be particularly effective when the label space is compact and well defined [66, 67].

Methodologically, the proposed approach complements recent advances that enrich Few-Shot NER with stronger label semantics and regularized adaptation. Taxonomy-guided prototype methods and prompt-based metric learning have shown that encoding label relationships and using prompts can substantially improve generalization from small support sets [60, 67]. Other work leverages semantics-induced optimal transport, knowledge-guided instance generation, and multi-scale feature extraction to stabilize fine-tuning and mitigate distribution drift in biomedical and domain-specific NER [67-69]. Compared to these more complex designs, our pipeline relies on standard gradient-based fine-tuning but is carefully aligned with the structure of the Indonesian sports corpus, demonstrating that a relatively lightweight strategy can still yield competitive gains when the domain is narrow and the support examples are well curated.

The findings also contribute to the broader body of work on NER for low-resource languages and specialized domains. Studies on Indonesian NER in news and legal text have primarily relied on transformer-based or hybrid neural architectures, yet continue to report challenges related to limited annotated corpora and domain shift [70, 71]. Parallel efforts in lower-resourced clinical and biomedical settings have shown that prompt-based and Few-Shot models can substantially improve performance when high-quality labeled data are scarce [68, 72, 73]. By showing that a Few-Shot fine-tuned SpaCy pipeline can match or exceed more complex architectures in a focused Indonesian sports domain, this study underscores the practicality of lightweight adaptation techniques for under-resourced languages.

Despite these strengths, several limitations remain. Lexical diversity within some sport categories is low, with only a few distinct surface forms per class, which means that part of the improvement may be due to memorization rather than fully generalizable representation learning. The current evaluation focuses on flat entities within a single domain and ontology; real-world applications may require nested entities, cross-domain transfer, or multilingual scenarios, where more advanced architectures such as nested or hierarchy-aware Few-Shot NER models may be necessary [74-78]. Future work should therefore explore integrating prototype-based or prompt-based heads into the existing pipeline, expanding the ontology to more fine-grained or cross-domain entities, and assessing performance in downstream tasks such as sports analytics, recommendation, and event monitoring.

5. CONCLUSION

This study presents a comprehensive investigation into the use of Few-Shot learning for NER in Indonesian sports news a domain that typically suffers from limited annotated resources. By integrating a domain-specific Few-Shot fine-tuning approach into a pre-trained SpaCy pipeline, the research demonstrates that high-quality entity extraction can be achieved even with minimal labelled data. The proposed model achieves superior performance compared to baseline architectures, including SpaCy, BiLSTM-CRF, and Transformer-IndoBERT, attaining a precision of 0.991 and an F1-score of 0.995. These results highlight the model's ability to generalize effectively within a corpus characterized by short, context-limited headlines and severe token-level class imbalance, as evidenced by the dataset analyses presented earlier.

Beyond outperforming existing baselines, the key innovation of this work lies in revealing the scalability and practical utility of Few-Shot learning for low-resource languages. The findings show that a lightweight fine-tuning strategy when coupled with well-curated examples can rival or surpass more complex architectures, making it especially suitable for rapid deployment in real-world scenarios where large annotated datasets are unavailable or costly to develop. This positions Few-Shot NER as a promising solution for applications such as sports analytics, media monitoring, and domain-specific information extraction across multilingual environments.

At the same time, the error analysis uncovers opportunities for improvement, particularly in addressing misclassification among semantically similar categories. Future work should explore error-driven data augmentation, prototype-based modelling, and prompt-guided regularization to further enhance model robustness. Expanding the approach to additional domains, incorporating hierarchical entity structures, and applying task-adaptive pretraining across broader Indonesian corpora also represent promising directions for improving scalability and cross-domain transfer. Collectively, this study demonstrates that Few-Shot learning offers not only technical efficiency but also strategic value for advancing NLP capabilities in low-resource settings, paving the way for practical and adaptable NER solutions.

REFERENCES

[1] Li, J., Leng, Z., Tang, K., Na, M., Li, Y., Alam, S.S. (2024). Multidimensional impact of sport types on the psychological well-being of student athletes: A multivariate investigation. Heliyon, 10(11): e32331. https://doi.org/10.1016/j.heliyon.2024.e32331

[2] Kan, B., Xie, Y. (2024). Impact of sports participation on life satisfaction among internal migrants in China: The chain mediating effect of social interaction and self-efficacy. Acta Psychologica, 243: 104139. https://doi.org/10.1016/j.actpsy.2024.104139

[3] Dings, A., Spinath, F.M. (2024). Sports club participation impacts life satisfaction in adolescence: A twin study. Psychology of Sport and Exercise, 73: 102639. https://doi.org/10.1016/j.psychsport.2024.102639

[4] Schmid, M.J., Oblinger-Peters, V., Örencik, M., Schmid, J., Conzelmann, A., Ronkainen, N.J. (2025). Meaning in life of elite athletes: A person-oriented study. Psychology of Sport and Exercise, 76: 102730. https://doi.org/10.1016/j.psychsport.2024.102730

[5] Zimmermann, S., Dauter, J., Gewald, H. (2024). Promoting sports and wellness with social media advertising–An analysis of different marketing channels. Procedia Computer Science, 231: 311-316. https://doi.org/10.1016/j.procs.2023.12.209

[6] López-Cabarcos, M.Á., Caby, J., Lugilde, S.A.S., Piñeiro-Chousa, J. (2024). An approach to innovative eSports from a business perspective. Journal of Innovation & Knowledge, 9(4): 100555. https://doi.org/10.1016/j.jik.2024.100555

[7] Kim, A., Kim, S.S. (2024). Engaging in sports via the metaverse? An examination through analysis of metaverse research trends in sports. Data Science and Management, 7(3): 181-188. https://doi.org/10.1016/j.dsm.2024.01.002

[8] Walsh, C., Joshi, A. (2024). Machine learning for sports betting: Should model selection be based on accuracy or calibration? Machine Learning with Applications, 16: 100539. https://doi.org/10.1016/j.mlwa.2024.100539

[9] Boughattas, W., Salha, M.B., Moella, N. (2022). Mental training for young athlete: A case of study of NLP practice. SSM-Mental Health, 2: 100076. https://doi.org/10.1016/j.ssmmh.2022.100076

[10] Cook, A., Karakuş, O. (2024). LLM-Commentator: Novel fine-tuning strategies of large language models for automatic commentary generation using football event data. Knowledge-Based Systems, 300: 112219. https://doi.org/10.1016/j.knosys.2024.112219

[11] Barua, A., Sharif, O., Hoque, M.M. (2021). Multi-class sports news categorization using machine learning techniques: Resource creation and evaluation. Procedia Computer Science, 193: 112-121. https://doi.org/10.1016/j.procs.2021.11.002

[12] Zakzouk, T.S., Mathkour, H.I. (2012). Comparing text classifiers for sports news. Procedia Technology, 1: 474-480. https://doi.org/10.1016/j.protcy.2012.02.104

[13] Bölücü, N., Rybinski, M., Dai, X., Wan, S. (2024). An adaptive approach to noisy annotations in scientific information extraction. Information Processing & Management, 61(6): 103857. https://doi.org/10.1016/j.ipm.2024.103857

[14] Nazir, A., Chakravarthy, T.K., Cecchini, D.A., Khajuria, R., et al. (2024). LangTest: A comprehensive evaluation library for custom LLM and NLP models. Software Impacts, 19: 100619. https://doi.org/10.1016/j.simpa.2024.100619

[15] Buscaldi, D., Dessí, D., Motta, E., Murgia, M., Osborne, F., Recupero, D.R. (2024). Citation prediction by leveraging transformers and natural language processing heuristics. Information Processing & Management, 61(1): 103583. https://doi.org/10.1016/j.ipm.2023.103583

[16] Solihin, F., Budi, I., Rochman, E., Mufarroha, F.A., Ramdlany, A.A., Dewi, D.A. (2025). Learning-based information extraction to obtain prominent named entities in Indonesian court decision documents. Mathematical Modelling of Engineering Problems, 12(5): 1627-1633. https://doi.org/10.18280/mmep.120517

[17] Tentua, M.N. (2024). NERSkill.Id: Annotated dataset of Indonesian's skill entity recognition. Data in Brief, 53: 110192. https://doi.org/10.1016/j.dib.2024.110192

[18] Wintaka, D.C., Bijaksana, M.A., Asror, I. (2019). Named-entity recognition on Indonesian tweets using bidirectional LSTM-CRF. Procedia Computer Science, 157: 221-228. https://doi.org/10.1016/j.procs.2019.08.161

[19] MAK, R.S., Bijaksana, M.A., Huda, A.F. (2019). Person entity recognition for the Indonesian Qur'an translation with the approach hidden Markov model-viterbi. Procedia Computer Science, 157: 214-220. https://doi.org/10.1016/j.procs.2019.08.160

[20] Yusup, F.A., Bijaksana, M.A., Huda, A.F. (2019). Narrator's name recognition with support vector machine for indexing Indonesian hadith translations. Procedia Computer Science, 157: 191-198. https://doi.org/10.1016/j.procs.2019.08.157

[21] Shidik, G.F., Saputra, F.O., Saraswati, G.W., Winarsih, N.A.S., et al. (2024). Indonesian disaster named entity recognition from multi source information using bidirectional LSTM (BiLSTM). Journal of Open Innovation: Technology, Market, and Complexity, 10(3): 100358. https://doi.org/10.1016/j.joitmc.2024.100358

[22] Gunawan, W., Suhartono, D., Purnomo, F., Ongko, A. (2018). Named entity recognition for Indonesian language using bidirectional LSTM-CNNs. Procedia Computer Science, 135: 425-432. https://doi.org/10.1016/j.procs.2018.08.193

[23] Wang, H., Yang, W., Feng, W., Zeng, L., Gu, Z. (2024). Threat intelligence named entity recognition techniques based on Few-Shot learning. Array, 23: 100364. https://doi.org/10.1016/j.array.2024.100364

[24] Ding, L., Ouyang, C., Liu, Y., Tao, Z., Wan, Y., Gao, Z. (2024). Few-shot named entity recognition via encoder and class intervention. AI Open, 5: 39-45. https://doi.org/10.1016/j.aiopen.2024.01.005

[25] Fang, W., Liu, Y., Ouyang, C., Ren, L., Li, J., Wan, Y. (2023). Joint span and token framework for few-shot named entity recognition. AI Open, 4: 111-119. https://doi.org/10.1016/j.aiopen.2023.08.009

[26] Cheng, Q., Chen, L., Hu, Z., Tang, J., Xu, Q., Ning, B. (2024). A novel prompting method for few-shot NER via LLMS. Natural Language Processing Journal, 8: 100099. https://doi.org/10.1016/j.nlp.2024.100099

[27] Liu, H., Song, J., Peng, W. (2024). GlyReShot: A glyph-aware model with label refinement for few-shot Chinese agricultural named entity recognition. Heliyon, 10(12): e32093. https://doi.org/10.1016/j.heliyon.2024.e32093

[28] Methlagl, M., Mairhofer, S., Michlmayr, F. (2025). Exploring technology acceptance patterns of users of the mixed-reality sport technology Zwift: Antecedents and consequences of technology acceptance. Universal Access in the Information Society, 24(1): 721-732. https://doi.org/10.1007/s10209-024-01094-2

[29] Beato, M., De Keijzer, K.L., Muñoz-Lopez, A., Raya-González, J., et al. (2024). Current guidelines for the implementation of flywheel resistance training technology in sports: A consensus statement. Sports Medicine, 54(3): 541-556. https://doi.org/10.1007/s40279-023-01979-x

[30] Lee, Y., Kang, J.H., Lee, S., Oh, T., Choi, S. (2024). The evolution of terminology: A scoping review of terms and concepts used to research sport in the digital realm. Quest, 76(4): 462-480. https://doi.org/10.1080/00336297.2024.2357370

[31] Ratten, V. (2024). Digital platform usage amongst female sport technology entrepreneurs. Journal of Small Business & Entrepreneurship, 36(5): 751-774. https://doi.org/10.1080/08276331.2022.2116678

[32] Ahaggach, H., Abrouk, L., Lebon, E. (2024). Enhancing car damage repair cost prediction: Integrating ontology reasoning with regression models. Intelligent Systems with Applications, 23: 200411. https://doi.org/10.1016/j.iswa.2024.200411

[33] Andrade, G.H.B., Nishiyama, T., Fujimaki, T., Yada, S., et al. (2024). Assessing domain adaptation in adverse drug event extraction on real-world breast cancer records. International Journal of Medical Informatics, 191: 105539. https://doi.org/10.1016/j.ijmedinf.2024.105539

[34] Gu, Z., He, X., Yu, P., Jia, W., et al. (2024). Automatic quantitative stroke severity assessment based on Chinese clinical named entity recognition with domain-adaptive pre-trained large language model. Artificial Intelligence in Medicine, 150: 102822. https://doi.org/10.1016/j.artmed.2024.102822

[35] Nessah, D., Hemam, S.M., Laouadi, M.L. (2024). Expanding the user's query to enhance semantic information retrieval using the reasoning mechanism based on homomorphism between semantic annotations. Ingénierie des Systèmes d'Information, 29(5): 1859-1866. https://doi.org/10.18280/isi.290518

[36] Yang, P., Wang, H., Huang, Y., Yang, S., et al. (2024). LMKG: A large-scale and multi-source medical knowledge graph for intelligent medicine applications. Knowledge-Based Systems, 284: 111323. https://doi.org/10.1016/j.knosys.2023.111323

[37] Nair, V., Pareek, J., Bhatt, S. (2024). A knowledge-based deep learning approach for automatic fake news detection using BERT on Twitter. Procedia Computer Science, 235: 1870-1882. https://doi.org/10.1016/j.procs.2024.04.178

[38] Badalotti, D., Agrawal, A., Pensato, U., Angelotti, G., Marcheselli, S. (2024). Development of a Natural Language Processing (NLP) model to automatically extract clinical data from electronic health records: Results from an Italian comprehensive stroke center. International Journal of Medical Informatics, 192: 105626. https://doi.org/10.1016/j.ijmedinf.2024.105626

[39] Rozo, A.Z., Campo-Archbold, A., Díaz-López, D., Gray, I., et al. (2024). Cyber democracy in the digital age: Characterizing hate networks in the 2022 US midterm elections. Information Fusion, 110: 102459. https://doi.org/10.1016/j.inffus.2024.102459

[40] Shelke, R., Vanjale, S. (2022). Recursive LSTM for the classification of named entity recognition for Hindi language. Ingénierie des Systèmes d'Information, 27(4): 679-684. https://doi.org/10.18280/isi.270420

[41] Narzary, S., Brahma, A., Nandi, S., Som, B. (2024). Deep learning based named entity recognition for the Bodo language. Procedia Computer Science, 235: 2405-2421. https://doi.org/10.1016/j.procs.2024.04.228

[42] Yao, X., Hao, X., Liu, R., Li, L., Guo, X. (2024). AgCNER, the first large-scale Chinese named entity recognition dataset for agricultural diseases and pests. Scientific Data, 11(1): 769. https://doi.org/10.1038/s41597-024-03578-5

[43] Huang, Z., He, L., Yang, Y., Li, A., et al. (2024). Application of machine reading comprehension techniques for named entity recognition in materials science. Journal of Cheminformatics, 16(1): 76. https://doi.org/10.1186/s13321-024-00874-5

[44] Xu, Y., Mao, C., Wang, Z., Jin, G., Zhong, L., Qian, T. (2024). Semantic-enhanced graph neural network for named entity recognition in ancient Chinese books. Scientific Reports, 14(1): 17488. https://doi.org/10.1038/s41598-024-68561-x

[45] González-Gallardo, C.E., Tran, H.T.H., Hamdi, A., Doucet, A. (2024). Leveraging open large language models for historical named entity recognition. In International Conference on Theory and Practice of Digital Libraries, pp. 379-395. https://doi.org/10.1007/978-3-031-72437-4_22

[46] Zhang, Y., Xiao, G. (2024). Named entity recognition datasets: A classification framework. International Journal of Computational Intelligence Systems, 17(1): 71. https://doi.org/10.1007/s44196-024-00456-1

[47] Han, D., Wang, Z., Li, Y., Ma, X., Zhang, J. (2024). Segmentation-aware relational graph convolutional network with multi-layer CRF for nested Named Entity Recognition. Complex & Intelligent Systems, 10(6):

7893-7905. https://doi.org/10.1007/s40747-024-01551-8
[48] Zhao, H., Bai, X., Zeng, Q., Zhou, H., Bai, X. (2024). Nested entity recognition method based on multidimensional features and fuzzy localization. Neural Processing Letters, 56(3): 196. https://doi.org/10.1007/s11063-024-11657-2
[49] Ke, X., Wu, X., Ou, Z., Li, B. (2024). Chinese named entity recognition method based on multi-feature fusion and biaffine. Complex & Intelligent Systems, 10(5): 6305-6318. https://doi.org/10.1007/s40747-024-01464-6
[50] Yilmaz, S.F., Mutlu, F.B., Balaban, I., Kozat, S.S. (2024). TMD-NER: Turkish multi-domain named entity recognition for informal texts. Signal, Image and Video Processing, 18(3): 2255-2263. https://doi.org/10.1007/s11760-023-02898-0
[51] Chen, B., Han, Y., Yan, L. (2023). A few-shot learning approach for Monkeypox recognition from a cross-domain perspective. Journal of Biomedical Informatics, 144: 104449. https://doi.org/10.1016/j.jbi.2023.104449
[52] Ha, S., Yoon, Y., Lee, J. (2023). Meta-ensemble learning with a multi-headed model for few-shot problems. ICT Express, 9(5): 909-914. https://doi.org/10.1016/j.icte.2022.09.001
[53] Paeedeh, N., Pratama, M., Ma'sum, M.A., Mayer, W., Cao, Z., Kowlczyk, R. (2024). Cross-domain few-shot learning via adaptive transformer networks. Knowledge-Based Systems, 288: 111458. https://doi.org/10.1016/j.knosys.2024.111458
[54] Kasaei, H., Xiong, S. (2024). Lifelong ensemble learning based on multiple representations for few-shot object recognition. Robotics and Autonomous Systems, 174: 104615. https://doi.org/10.1016/j.robot.2023.104615
[55] Li, Y., Fu, Y., Gayo, I.J., Yang, Q., et al. (2023). Prototypical few-shot segmentation for cross-institution male pelvic structures with spatial registration. Medical Image Analysis, 90: 102935. https://doi.org/10.1016/j.media.2023.102935
[56] Gao, Y., Li, H., Fu, W. (2023). Few-Shot learning for image-based bridge damage detection. Engineering Applications of Artificial Intelligence, 126: 107078. https://doi.org/10.1016/j.engappai.2023.107078
[57] Liu, Y., Mu, C., Jiang, S., Liu, Y. (2025). Unsupervised meta-learning with domain adaptation based on a multi-task reconstruction-classification network for few-shot hyperspectral image classification. Journal of Information and Intelligence, 3(2): 103-112. https://doi.org/10.1016/j.jiixd.2024.06.001
[58] Pornprasit, C., Tantithamthavorn, C. (2024). Fine-tuning and prompt engineering for Large Language Models-based code review automation. Information and Software Technology, 175: 107523. https://doi.org/10.1016/j.infsof.2024.107523
[59] Li, X., Guo, Z., Zhu, R., Ma, Z., Guo, J., Xue, J.H. (2024). A simple scheme to amplify inter-class discrepancy for improving few-shot fine-grained image classification. Pattern Recognition, 156: 110736. https://doi.org/10.1016/j.patcog.2024.110736
[60] Deng, W., Xiong, W., Lu, Z., Yuan, X., Zhang, C., Wang, L. (2024). Few-shot power transformers fault diagnosis based on gaussian prototype network. International Journal of Electrical Power & Energy Systems, 160: 110146. https://doi.org/10.1016/j.ijepes.2024.110146
[61] Lubis, A.R., Lase, Y.Y., Rahman, D.A., Witarsyah, D. (2023). Improving spell checker performance for Bahasa Indonesia using text preprocessing techniques with deep learning models. Ingénierie des Systèmes d'Information, 28(5): 1335-1342. https://doi.org/10.18280/isi.280522
[62] Bhadauria, D., Sierra-Múnera, A., Krestel, R. (2024). The effects of data quality on Named Entity Recognition. In Proceedings of the Ninth Workshop on Noisy and User-Generated Text (W-NUT 2024), pp. 79-88.
[63] Feng, J., Xu, G., Wang, Q., Yang, Y., Huang, L. (2024). Note the hierarchy: Taxonomy-guided prototype for few-shot Named Entity Recognition. Information Processing & Management, 61(1): 103557. https://doi.org/10.1016/j.ipm.2023.103557
[64] Yu, Y., He, W., Kang, Y.M., Ji, Y.L. (2023). Leveraging continuous prompt for few-shot Named Entity Recognition in electric power domain with meta-learning. Data Intelligence, 5(2): 494-509. https://doi.org/10.1162/dint_a_00202
[65] Chen, Y., Zheng, Y., Yang, Z. (2023). Prompt-based metric learning for few-shot NER. In Findings of the Association for Computational Linguistics: ACL 2023, pp. 7199-7212.
[66] Gururangan, S., Marasović, A., Swayamdipta, S., Lo, K., Beltagy, I., Downey, D., Smith, N.A. (2020). Don't stop pretraining: Adapt language models to domains and tasks. arXiv preprint arXiv:2004.10964. https://doi.org/10.48550/arXiv.2004.10964
[67] Ge, Y., Guo, Y., Yang, Y.C., Al-Garadi, M.A., Sarker, A. (2022). A comparison of few-shot and traditional Named Entity Recognition models for medical text. In 2022 IEEE 10th International Conference on Healthcare Informatics (ICHI), pp. 84-89. https://doi.org/10.1109/ICHI54592.2022.00024
[68] Moscato, V., Postiglione, M., Sperlí, G. (2023). Few-shot Named Entity Recognition: Definition, taxonomy and research directions. ACM Transactions on Intelligent Systems and Technology, 14(5): 1-46. https://doi.org/10.1145/3609483
[69] Ullah, F., Gelbukh, A., Zamir, M.T., Riverón, E.M.F., Sidorov, G. (2024). Enhancement of Named Entity Recognition in low-resource languages with data augmentation and BERT models: A case study on Urdu. Computers, 13(10): 258. https://doi.org/10.3390/computers13100258
[70] Chen, A., Yu, Z., Yang, X., Guo, Y., Bian, J., Wu, Y. (2023). Contextualized medication information extraction using transformer-based deep learning architectures. Journal of Biomedical Informatics, 142: 104370. https://doi.org/10.1016/j.jbi.2023.104370
[71] Zhao, D., Mu, W., Jia, X., Liu, S., Chu, Y., Meng, J., Lin, H. (2025). Few-shot biomedical NER empowered by LLMs-assisted data augmentation and multi-scale feature extraction. BioData Mining, 18(1): 28. https://doi.org/10.1186/s13040-025-00443-y
[72] Zhou, D., Li, S., Chen, Q., Yao, H. (2024). Improving few-shot Named Entity Recognition via semantics induced optimal transport. Neurocomputing, 597: 127938. https://doi.org/10.1016/j.neucom.2024.127938
[73] Yulianti, E., Bhary, N., Abdurrohman, J., Dwitilas, F.W., Nuranti, E.Q., Husin, H.S. (2024). Named Entity Recognition on Indonesian legal documents: A dataset and study using transformer-based models. International Journal of Electrical & Computer Engineering (2088-8708), 14(5): 5489-5501.

https://doi.org/10.11591/ijece.v14i5.pp5489-5501
[74] Zainuddin, Z., Tahir, Z. (2025). Entity extraction in indonesian online news using Named Entity Recognition (NER) with hybrid method transformer, word2vec, attention and Bi-LSTM. JOIV: International Journal on Informatics Visualization, 9(3): 964-973. https://doi.org/10.62527/joiv.9.3.2902
[75] Xiao, Y., Zou, J., Yang, Q. (2025). Advancing few-shot Named Entity Recognition with large language model. Applied Sciences, 15(7): 3838. https://doi.org/10.3390/app15073838
[76] Richter-Pechanski, P., Wiesenbach, P., Schwab, D.M., Kiriakou, C., Geis, N., Dieterich, C., Frank, A. (2025). Clinical information extraction for lower-resource languages and domains with Few-Shot learning using pretrained language models and prompting. Natural Language Processing, 31(5): 1210-1233. https://doi.org/10.1017/nlp.2024.52
[77] Hwang, H., Jung, Y., Lee, C., Go, W. (2025). A nested Named Entity Recognition model robust in Few-Shot learning environments using label description information. Applied Sciences, 15(15): 8255. https://doi.org/10.3390/app15158255
[78] Li, S., Wu, Y., Fan, C., Ye, Y., Yao, H. (2025). Fine-grained geographic Named Entity Recognition with Few-Shot learning. International Journal of Geographical Information Science, 1-25. https://doi.org/10.1080/13658816.2025.2505046