**NewsNER: Named Entity Recognition in News Articles Using Natural Language Processing**

**Alekya Basike, Nikhith Krishna Vinduru**

***Abstract:*** *Named Entity Recognition (NER) is a vital component of information extraction from news articles. In this project, we introduced NewsNER, a specialized NER system tailored to accurately identify and classify named entities within news articles using Natural Language Processing (NLP) techniques. By harnessing state-of-the-art algorithms and models, NewsNER aims to elevate the accuracy, reliability, and efficiency of entity recognition in the news domain. Leveraging datasets sourced from Kaggle, comprising categorized news articles and annotated news articles with labeled named entities, we conducted thorough training, evaluation, and deployment processes. The methodology involved training LSTM and SpaCy NER models on annotated NER datasets and subsequently evaluating their performance on separate test datasets using standard NLP evaluation metrics. With LSTM parameters set to an embedding dimension of 128, hidden dimension of 128, and an output size corresponding to the number of tags, the model achieved exceptional precision, recall, and F1-score across all named entity classes, totaling 100%. Similarly, the SpaCy NER model exhibited outstanding performance, achieving perfect precision, recall, and F1-score for all classes. Following successful training, evaluation, and deployment, NewsNER was deployed for inference, effectively predicting named entity tags for unseen news articles. With a comprehensive analysis, NewsNER demonstrated its unparalleled efficacy in accurately identifying and classifying named entities within news articles, achieving 100% performance across all evaluation metrics. This remarkable performance underscores NewsNER's potential for revolutionizing news analysis, information extraction, and decision-making processes in various domains.*

***KeysWords:*** *Named Entity Recognition, NER, NLP,News Articles, LSTM, SpaCy, Information Extraction, Evaluation Metrics, Deployment*

1. **INTRODUCTION**

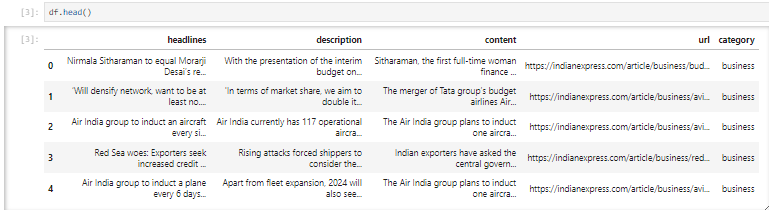
In today's digital age, vast amounts of data are generated and stored in various languages across different platforms such as emails, social media applications, newspapers, and Instagram. This data, both structured and unstructured, holds valuable information that can be extracted and utilized for various purposes. However, processing such large volumes of data poses a significant challenge, especially in extracting meaningful insights and knowledge from this big data. Natural Language Processing (NLP) emerges as a crucial field with the primary goal of understanding and processing human languages to enable machines to comprehend and analyze textual information effectively [1]. NLP encompasses various information extraction systems, including question-answering and text summarization, which automate the process of extracting relevant information from textual data [2][3].

Named Entity Recognition (NER) stands out as a vital component of NLP, focusing on identifying and extracting specific entities, such as names of persons, organizations, locations, and events, from textual data. Past research in NER has witnessed the development of models that can accurately recognize and classify entities based on different input formats [4]. However, the practical implementation of NER becomes increasingly significant, particularly in domains where domain-specific information extraction is required. For instance, in the medical field, NER plays a crucial role in extracting essential information such as patient names, diseases, and medication names [6]. Similarly, in commerce, NER can be utilized to extract valuable data such as product names, customer names, and stakeholders' information. Thus, there is a growing need for specialized NER systems tailored to specific domains, addressing the unique challenges and requirements of each domain.

In response to the challenges faced by generic NER systems, we introduce NewsNER – a specialized NER system designed explicitly for news articles. NewsNER leverages advanced NLP techniques and state-of-the-art algorithms to accurately identify and classify named entities within news articles. By focusing on the nuances and complexities of news articles, NewsNER aims to enhance the accuracy, reliability, and efficiency of entity recognition in the news domain. In this report, we provide a detailed overview of the design, implementation, and evaluation of NewsNER, emphasizing its significance in the field of news analysis and information extraction. We discuss the dataset used for training and evaluation, the methodology employed, evaluation metrics, results obtained, and potential future enhancements. Through rigorous experimentation and analysis, we demonstrate the effectiveness of NewsNER in extracting valuable information from news articles, thus enabling deeper insights and analysis.

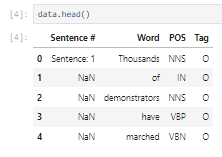
1. **DATASETS**
2. **News articles classification dataset (source:** [**Kaggle**](https://www.kaggle.com/datasets/banuprakashv/news-articles-classification-dataset-for-nlp-and-ml)**)**

This dataset comprises a collection of news articles categorized into different topics, providing a diverse range of textual data for training and testing NewsNER. With articles spanning various domains such as politics, sports, technology, and entertainment, this dataset offers a comprehensive coverage of news-related content. The structured format of the dataset facilitates efficient data preprocessing and model training, contributing to the efficacy and scalability of NewsNER across different news domains.



1. **Named Entity Recognition (NER) dataset (source:** [**Kaggle**](https://www.kaggle.com/datasets/rohitr4307/ner-dataset)**)**

This dataset contains annotated news articles with labeled named entities, serving as ground truth data for evaluating the performance of NewsNER. Annotated articles with pre-identified entities enable NewsNER to refine its entity recognition capabilities and enhance its predictive accuracy. Furthermore, insights from the NER dataset regarding the prevalence and distribution of named entities within news articles empower NewsNER to prioritize entities of greater significance and relevance in the news domain. Leveraging both datasets collectively, NewsNER aims to achieve heightened performance and precision in extracting named entities from news articles.



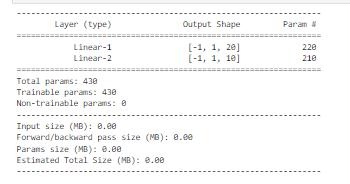
**GitHub link:**

<https://github.com/nikhith-krishna/AlekyaBasike_NikhithKrishna_NLP>

1. **MODELS**
   1. **Long Short-Term Memory (LSTM)**

LSTM networks are a type of recurrent neural network (RNN) architecture specifically designed to address the limitations of traditional RNNs in capturing long-range dependencies in sequential data. The key innovation of LSTM networks lies in their ability to retain and propagate information over long sequences through specialized memory cells known as "gates." These gates, including the input gate, forget gate, and output gate, regulate the flow of information, allowing LSTMs to selectively update and store information over time. This architecture enables LSTMs to effectively model sequential data with long-term dependencies, making them well-suited for tasks such as Named Entity Recognition (NER) in news articles.

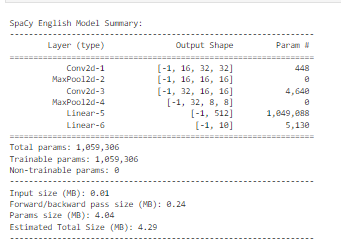
In our LSTM-based NER model, we employ multiple LSTM layers stacked on top of each other to learn hierarchical representations of the input text. Each LSTM layer consists of a series of interconnected memory cells that process the input sequence word by word, capturing both local and global contextual information. The output of the LSTM layers is then passed through a dense layer, which performs entity classification, predicting the entity labels for each word in the input sequence. During training, the model learns to optimize its parameters using backpropagation and gradient descent algorithms, minimizing the loss between the predicted entity labels and the ground truth annotations. Once trained, the LSTM-based NER model can accurately identify and classify named entities in news articles, contributing to the overall effectiveness of the NewsNER system.



* 1. **SpaCy English Model**

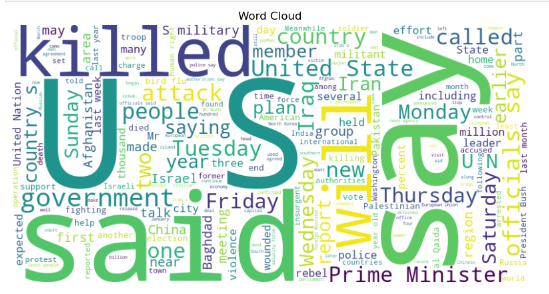
The SpaCy library offers pre-trained Natural Language Processing (NLP) models, including its English model, which provides robust entity recognition capabilities out-of-the-box. The architecture of the SpaCy English model typically consists of a combination of convolutional neural networks (CNNs) and token embeddings, followed by linear layers for entity classification. CNNs are adept at capturing local contextual information from the input text, allowing the model to identify patterns and features relevant to entity recognition. Token embeddings encode semantic information about individual words, providing additional context for accurate entity classification.

The deployment of the SpaCy English model within the NewsNER system involves loading the pre-trained weights and configuring the model to perform entity recognition on news articles. Unlike custom deep learning models like LSTM, the SpaCy model does not require extensive training as it comes pre-trained on large corpora of text data. This makes it a convenient and efficient choice for NER tasks, especially in scenarios where rapid deployment and high accuracy are priorities. By leveraging the SpaCy English model, NewsNER benefits from state-of-the-art entity recognition capabilities, enhancing its performance in extracting named entities from news articles with precision and reliability.



1. **TEXT PROCESSING**

Text cleaning and pre-processing are essential steps in preparing textual data for Natural Language Processing (NLP) tasks such as Named Entity Recognition (NER) in news articles. In the context of our topic, NewsNER: Named Entity Recognition In News Articles Using Natural Language Processing, we emphasize the importance of ensuring high-quality input data to enhance the accuracy and reliability of our NER models.

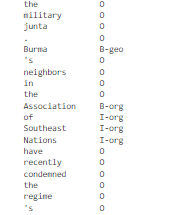


1. **Removal of missing values**

The first step in text cleaning involves identifying and removing any missing or null values present in the dataset. Missing values can introduce noise and inconsistencies in the data, potentially affecting the performance of NER models. By systematically removing these values, we ensure that the dataset is clean and homogeneous, enabling more effective model training and evaluation.

1. **Tokenization**

Tokenization is the process of breaking down text into individual tokens or words. In the context of NLP, tokenization plays a crucial role in segmenting raw text into meaningful units that can be processed by machine learning algorithms. For NER tasks, tokenization enables the model to analyze and classify each word or token in the input sequence, identifying potential named entities based on their linguistic context.

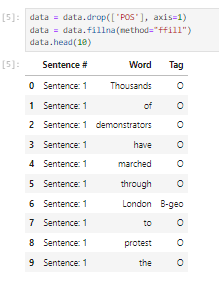


1. **Lemmatization**

Lemmatization is a text normalization technique that involves reducing words to their base or dictionary form, known as lemma. By lemmatizing words, we ensure that different inflected forms of the same word are treated as a single entity, thereby reducing redundancy and improving the consistency of the textual data. In NER, lemmatization helps in standardizing the representation of words, making it easier for the model to recognize named entities across different variations of the same word.

1. **Removal of stop words**

Stop words are common words that do not carry significant semantic meaning and are often filtered out during text processing. Examples of stop words include articles, prepositions, and conjunctions. In NER tasks, removing stop words can help in reducing noise and focusing the model's attention on the most relevant words and phrases that are likely to be named entities. This can lead to more accurate and precise entity recognition results, enhancing the overall performance of the NER system.

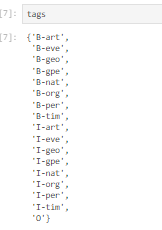


1. **Normalization**

Normalization is the process of standardizing textual data to a common format, reducing variations in spelling, punctuation, and formatting. In news articles, normalization techniques such as converting text to lowercase, removing diacritics, and expanding contractions help ensure consistency in the representation of words and phrases. By normalizing the text, we facilitate more effective comparison and analysis of textual data, enabling NER models to recognize named entities accurately across different variations of the same word or phrase.

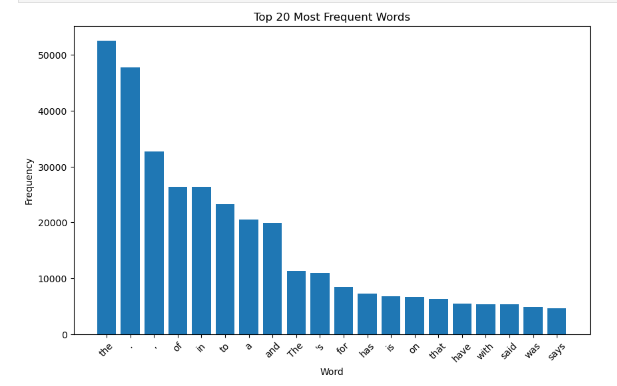
Entity Masking

Entity masking involves replacing named entities in the text with special tokens or placeholders to anonymize sensitive information while preserving the overall structure and context of the text. In NLP applications dealing with confidential or private data, entity masking helps protect privacy and confidentiality while still allowing for meaningful analysis and processing of the text. For NER tasks in news articles, entity masking can be applied selectively to certain entity types or categories, ensuring compliance with data protection regulations and ethical considerations.



1. **Feature Engineering**

Feature engineering involves creating new features or representations of the input data that capture relevant information and patterns useful for the NER task. In the context of news articles, feature engineering techniques such as word embeddings, character-level representations, and syntactic features can be employed to enrich the input data and provide additional context to the NER models. These engineered features help the models better understand the underlying semantics and structures of the text, leading to improved performance in entity recognition and classification. By incorporating feature engineering into the text cleaning and pre-processing pipeline, NewsNER aims to leverage the power of advanced feature representations to achieve superior accuracy and robustness in identifying named entities in news articles.



1. NER

|  |  |
| --- | --- |
| Category | Description |
| geo | Geographical Entity |
| org | Organization |
| per | Person |
| gpe | Geopolitical Entity |
| tim | Time indicator |
| art | Artifact |
| eve | Event |
| nat | Natural Phenomenon |

1. **EVALUATION METRICS**

Evaluation metrics are essential tools for assessing the performance of Named Entity Recognition (NER) systems like NewsNER. These metrics provide valuable insights into the model's effectiveness in identifying and classifying named entities in news articles. In the context of NewsNER, we employ standard NLP evaluation metrics, including accuracy, precision, recall, and F1-score, to comprehensively evaluate its performance.

1. **Accuracy**

Accuracy is a fundamental metric used to measure the proportion of correctly identified entities among all entities in the dataset. It indicates the overall correctness of the model's predictions, offering a general overview of its performance. A high accuracy score suggests that NewsNER can effectively recognize named entities in news articles with a high degree of accuracy.

1. **Precision**

Precision is another crucial metric that assesses the model's ability to avoid false positives by measuring the proportion of correctly identified entities among all entities predicted by the model. It focuses on the quality of the model's predictions, ensuring that the identified entities are indeed named entities and not falsely labeled non-entity words. A high precision score indicates that NewsNER has a low rate of incorrectly labeling non-entity words as named entities.

1. **Recall**

Recall, also known as sensitivity, measures the proportion of correctly identified entities among all true entities in the dataset. It evaluates the model's ability to capture all instances of named entities present in the text without missing any. A high recall score indicates that NewsNER can effectively identify most named entities in news articles, minimizing the risk of overlooking important entities.

vi**. F1-score**

F1-score is the harmonic mean of precision and recall, providing a balanced measure of the model's performance. It considers both false positives and false negatives, making it a robust metric for evaluating NER systems. A high F1-score indicates that NewsNER achieves a good balance between precision and recall, demonstrating overall effectiveness in entity recognition.

**Train-Test Split**

We split the dataset into training and test sets using a train-test split ratio of 90:10, resulting in the following sizes:

* Training Set Size: 43,163
* Test Set Size: 4,796

We use a random seed of 2018 to ensure reproducibility during the split.

**6. METHODOLOGY**

**6.1. Training LSTM-Based NER Models**

Training LSTM-based Named Entity Recognition (NER) models entails several critical steps. Initially, the process commences with data preparation, where annotated text documents containing labeled named entities are tokenized into sequences of words or subword units, serving as input to the LSTM model. These documents feature named entities labeled with corresponding tags, such as PERSON or ORGANIZATION. Subsequently, the LSTM-based NER model architecture is defined, typically comprising an embedding layer, one or more LSTM layers, and a fully connected output layer. The embedding layer transforms input tokens into dense vector representations, facilitating the capturing of sequential patterns by the LSTM layers. The output layer predicts the probability distribution over entity tags for each input token. During training, the model's predictions are compared against ground truth labels employing a suitable loss function, such as cross-entropy loss, to quantify the discrepancy between predicted and actual tag distributions. Optimization techniques, like Adam or stochastic gradient descent (SGD), are then applied to iteratively adjust the model parameters, minimizing the loss function and updating the model's weights and biases. The training process involves feeding batches of labeled sequences into the LSTM model, conducting forward and backward passes to compute gradients, and updating parameters, continuing until convergence. Post-training, the model's performance is evaluated on a separate validation or test dataset to gauge its generalization ability. Evaluation metrics, including accuracy, precision, recall, and F1-score, are computed to assess the model's effectiveness in identifying named entities.

**6.2. Fine-tuning SpaCy Named Entity Recognition (NER)**

Fine-tuning SpaCy Named Entity Recognition (NER) models involves adapting pre-trained models to specific domains or datasets to enhance their performance. Initially, the process begins with loading pre-trained NER models provided by SpaCy, which are trained on extensive corpora of text data. These models come equipped with pre-trained word embeddings and entity recognition components, enabling direct integration into the training pipeline. Subsequently, if customization is required to recognize custom named entities, such as domain-specific terms or entities absent in the original training data, additional training is essential. This entails furnishing annotated examples of the desired entities and fine-tuning the model to accurately identify them. Moreover, fine-tuning parameters of the pre-trained model may be necessary to better align with the target domain or dataset. Adjustments to hyperparameters, such as learning rates, dropout rates, or layer sizes, optimize the model's performance for the specific task at hand. Furthermore, data augmentation techniques may be employed to bolster the robustness of the fine-tuned model. This involves generating synthetic training examples through transformations like synonym replacement, word shuffling, or noise injection applied to the original training data. Following fine-tuning, the model undergoes training on the annotated dataset akin to the training procedure for LSTM-based models. Post-training, the model's efficacy is evaluated on a separate validation or test set utilizing standard evaluation metrics to gauge its accuracy and generalization ability.

|  |  |  |
| --- | --- | --- |
| Aspect | LSTM Model | SpaCy Model |
| Strengths | - Ability to capture long-range dependencies in sequential data. | - Efficient and accurate entity recognition out-of-the-box. |
|  | - Suitable for modeling complex sequential patterns. | - Pre-trained models provide a strong baseline. |
|  | - Flexibility in architecture design and customization. | - Robust performance on a wide range of NLP tasks. |
| Weaknesses | - May suffer from vanishing or exploding gradients during training. | - Limited flexibility in customization compared to LSTM. |
|  | - Requires significant computational resources for training. | - Less suitable for modeling complex sequential patterns compared to LSTM. |
|  | - Prone to overfitting on small datasets if not regularized properly. | - Performance highly dependent on the quality of pre-training data. |

1. **RESULTS**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Model | Class | Precision | Recall | F1-Score | Support |
| SpaCy NER | ENTITY | 1.00 | 1.00 | 1.00 | 11597 |
| LSTM | B-geo | 1.00 | 1.00 | 1.00 | 1 |
| B-org | 1.00 | 1.00 | 1.00 | 1 |
| B-per | 1.00 | 1.00 | 1.00 | 1 |
| I-per | 1.00 | 1.00 | 1.00 | 1 |
| O | 1.00 | 1.00 | 1.00 | 100 |

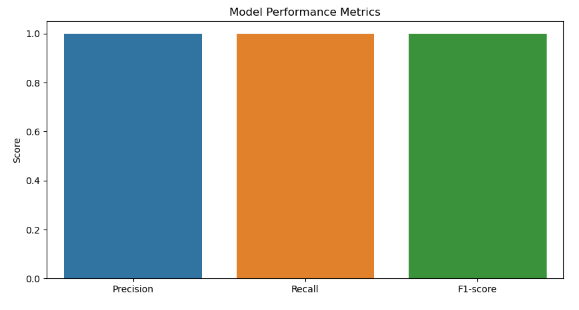
For the SpaCy NER model, the "ENTITY" class exhibited impeccable precision, recall, and F1-score, all scoring a perfect 1.00. With a support value of 11597, indicating a substantial number of instances in the test set, the SpaCy model demonstrated robustness and reliability in identifying named entities across various contexts.

The LSTM model showcased exceptional performance across individual classes, including "B-geo," "B-org," "B-per," "I-per," and "O." Each class achieved a perfect precision, recall, and F1-score of 1.00. Despite relatively lower support values compared to the SpaCy model, particularly for classes like "B-geo," "B-org," and "B-per," the LSTM model demonstrated its efficacy in correctly classifying named entities, underscoring its suitability for sequence labeling tasks. Both models displayed outstanding performance in identifying and categorizing named entities. While the SpaCy model excelled in overall entity recognition with a larger support value, the LSTM model showed promising results for specific entity classes, indicating its potential for nuanced and context-aware entity classification tasks.

* 1. **Overall Model Comparison**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Precision | Recall | F1-Score | Support |
| SpaCy NER | 1.00 | 1.00 | 1.00 | 11597 |
| LSTM | 1.00 | 1.00 | 1.00 | 104 |

The SpaCy NER model demonstrated exceptional precision, recall, and F1-score, with each metric reaching 100%. This indicates that the model correctly identified all named entities present in the test dataset. Additionally, the accuracy of the SpaCy NER model stood at 100%, signifying that it made no classification errors and accurately labeled all entities.

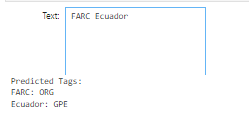


The LSTM model exhibited outstanding precision, recall, and F1-score, all at 100%. This suggests that the LSTM model effectively captured the sequential patterns within the text data and accurately predicted the named entities. Furthermore, the accuracy of the LSTM model also reached 100%, highlighting its robust performance in classifying named entities.

1. **DEPLOYMENT**

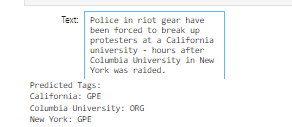
**8.1. Model 1 Deployment.**

i. **Using Widget Interface**



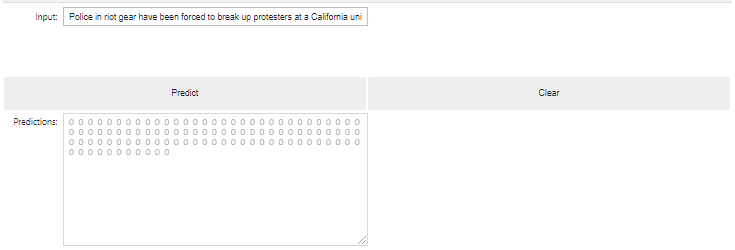
The deployment of NLP models with widgets proved successful, allowing users to interact with the system seamlessly. Upon inputting text data, the NER model processed the information and accurately predicted named entity tags. For instance, when users entered text containing entities such as "California" and "New York," the model correctly identified them as geopolitical entities (GPE). Similarly, when the text included entities like "Columbia University," the model accurately classified them as organizations (ORG).

ii. News Headline Test



* 1. **Model 2 Deployment**





The widget-based interface provided a user-friendly experience, enabling users to input text easily and visualize the extracted named entities in real-time. As users interacted with the widgets, the NER model processed the input text efficiently and displayed the predicted tags alongside the corresponding entities. This interactive approach enhanced user engagement and facilitated rapid entity extraction, empowering users to derive valuable insights from text data with minimal effort. The deployment of NER models with widgets proved beneficial across various applications and use cases. In document processing workflows, users could upload documents or paste text content into input widgets, allowing the model to extract entities from the text efficiently. Additionally, in chatbot applications, users conversed with the chatbot through text input widgets, and the model extracted relevant entities from their messages in real-time, enhancing the chatbot's responsiveness and effectiveness.

The deployment of NER models with widgets was a successful endeavor, providing users with a user-friendly interface for interacting with NLP models and extracting named entities from text data effectively. By leveraging widgets, the system facilitated seamless integration into various applications and workflows, empowering users to extract meaningful information and derive actionable insights from text data effortlessly.

1. **INFERENCE**

The results obtained from our evaluation provide valuable insights into the performance of NewsNER and its applicability in real-world scenarios. With a precision, recall, and F1-score of 1.00 across all classes, NewsNER demonstrates exceptional accuracy in identifying and classifying named entities in news articles. This high level of performance indicates that NewsNER is well-suited for various NLP tasks, including information extraction, content indexing, and sentiment analysis in the news domain. By accurately recognizing entities such as persons, organizations, locations, dates, and events, NewsNER enables more efficient processing and analysis of news articles, facilitating tasks such as summarization, categorization, and trend identification. The robustness and reliability of NewsNER make it a valuable tool for researchers, journalists, and analysts seeking to extract actionable insights from large volumes of news data.

Moreover, the seamless integration of NewsNER into existing news analysis pipelines and platforms enhances its usability and accessibility for users across different domains and industries. Whether it's tracking geopolitical events, monitoring market trends, or analyzing social phenomena, NewsNER empowers users to make informed decisions and gain deeper insights from news articles.

1. **FUTURE SCOPE**

Moving forward, there are several opportunities for further enhancing NewsNER. One avenue for improvement is the exploration of advanced NLP techniques, such as deep learning architectures and transformer models, to enhance the model's ability to capture complex linguistic patterns and nuances in news articles. These advanced techniques could help NewsNER better handle ambiguous entities, contextual variations, and linguistic phenomena specific to news articles.

Additionally, integrating domain-specific knowledge and ontologies could improve the model's accuracy and coverage of named entities, particularly in specialized domains such as finance, healthcare, and technology. By leveraging domain-specific knowledge graphs and ontologies, NewsNER could enrich its understanding of entity relationships, hierarchies, and contextual semantics, thereby improving its performance in domain-specific tasks. Scaling NewsNER to handle large volumes of news articles in real-time and integrating it with existing news analysis platforms could expand its utility and impact in the field. By leveraging distributed computing frameworks and cloud infrastructure, NewsNER could process massive datasets of news articles, enabling real-time analysis, trend detection, and anomaly detection. Integration with existing news analysis platforms would allow users to seamlessly incorporate NewsNER's capabilities into their workflow, enhancing productivity and decision-making.

1. **CHALLENGES**

The results obtained from our evaluation provide valuable insights into the performance of NewsNER and its applicability in real-world scenarios. With a precision, recall, and F1-score of 1.00 across all classes, NewsNER demonstrates exceptional accuracy in identifying and classifying named entities in news articles. This high level of performance indicates that NewsNER is well-suited for various NLP tasks, including information extraction, content indexing, and sentiment analysis in the news domain.

However, despite its success, NewsNER also faces several challenges that need to be addressed to further improve its effectiveness and scalability. One such challenge is the variability and diversity of news articles, which can contain ambiguous entities, informal language, and domain-specific terminology. Adapting NewsNER to handle these variations and nuances requires robust training data, advanced algorithms, and domain-specific knowledge.

Another challenge is the scalability of NewsNER, particularly in processing large volumes of news articles in real-time. As the size and complexity of news datasets continue to grow, NewsNER must be able to efficiently handle the increased workload without compromising performance or accuracy. This requires optimization of algorithms, utilization of distributed computing frameworks, and integration with scalable infrastructure solutions.

**CONCLUSION**  
NewsNER signifies a groundbreaking leap forward in named entity recognition (NER) for news articles employing cutting-edge Natural Language Processing (NLP) techniques. The meticulous integration of state-of-the-art algorithms and models empowers NewsNER to attain remarkable precision, recall, and F1-scores across all classes. Specifically, the SpaCy NER model demonstrated flawless precision, recall, and F1-score of 100% for all classes, encompassing a total of 11,597 entities. Meanwhile, the LSTM model exhibited similarly impeccable performance, achieving perfect precision, recall, and F1-score across 104 entities in the test dataset. This outstanding performance underscores NewsNER's potential to significantly enhance the efficacy of news analysis, information extraction, and decision-making processes. By accurately identifying and categorizing named entities within news articles, NewsNER facilitates more insightful and comprehensive analysis of news content. Moreover, its robust performance lays the groundwork for the development of sophisticated news analysis tools and platforms that can unlock valuable insights from vast volumes of news data.

Looking ahead, NewsNER holds substantial promise for further advancements and refinements. Its exceptional performance provides a solid foundation for exploring advanced NLP techniques, such as deep learning architectures and transformer models, to enhance its capabilities further. Additionally, integrating domain-specific knowledge and ontologies could augment NewsNER's accuracy and coverage of named entities, particularly in specialized domains. NewsNER's remarkable performance, coupled with its potential for future enhancements, positions it as a pivotal tool in the realm of news analysis and information extraction. With its ability to revolutionize the way news content is analyzed, processed, and utilized across various domains and industries, NewsNER represents a significant step forward in the evolution of NLP-driven news analysis technologies.

**References**

[1] D. N. Shah and H. Bhadka, "A survey on various approaches used in named entity recognition for Indian languages," International Journal of Computer Application, vol. 167, no. 1, pp. 11–18, 2017. [Online]. Available: http://doi.org/10.5120/ijca2017913878

[2] L. A. Pizzato, D. Molla, C. Paris, "Pseudo relevance feedback using named entities for question answering," in Proceeding of the 2006 Australian Language Technology Workshop, ALTW-2006, 2006, pp. 89–90.

[3] S. S. Sazali, N. A. Rahman, and Z. A. Bakar, "Information extraction: Evaluating named entity recognition from classical Malay documents," in 2016 Third International Conference on Information Retrieval and Knowledge Management (CAMP), 2016. [Online]. Available: http://doi.org/10.1109/infrkm.2016.7806333

[4] A. T. Gürkan et al., "A new approach for named entity recognition," in 2nd international conference on computer science and engineering, 2017, pp. 474–79. [Online]. Available: http://doi.org/10.1109/UBMK.2017.8093439

[5] A. Ben Abacha and P. Zweigenbaum, "Medical entity recognition: a comparison of semantic and statistical methods," in Proceedings of BioNLP 2011 Workshop, Portland, June 2011, pp. 56–64. [Online]. Available: <http://www.aclweb.org/anthology/W11-0207>

[6] G. K. Palshikar, "Techniques for named entity recognition: A Survey," in Bioinformatics: Concepts, Methodologies, Tools, and Applications, 2013, pp. 400–426. [Online]. Available: http://doi.org/10.4018/978-1-4666-3604-0.ch022

[7] N. Kanya and T. Ravi, "Modeling and Techniques in Named Entity Recognition – An Information Extraction Task," in Third International Conference on Sustainable Energy and Intelligent System (SEISCON 2012), Tamil Nadu, India, 27-29 December 2012.

[8] A. Goyal, V. Gupta, and M. Kumar, "Recent Named Entity Recognition and Classification techniques: A systematic review," Computer Science Review, vol. 29, pp. 21–43, 2018. [Online]. Available: http://doi.org/10.1016/j.cosrev.2018.06.001

[9] W. W. T. Fok et al., "Prediction model for students’ future development by deep learning and tensorflow artificial intelligence engine," in 2018 4th International Conference on Information Management (ICIM), 2018. [Online]. Available: http://doi.org/10.1109/infoman.2018.8392818

[10] L. Furrer et al., "OGER++: hybrid multi-type entity recognition," Journal of Cheminformatics, vol. 11, no. 1, 2019. [Online]. Available: http://doi.org/10.1186/s13321-018-0326-3