CSECU-Learners at CheckThat! 2025: Multilingual Transformer-based Approach for Subjectivity Detection in News Articles Across Multilingual and Zero-shot Settings

Monir Ahmad[1,*], Abu Nowshed Chy[1]

[1]Department of Computer Science and Engineering, University of Chittagong, Chattogram-4331, Bangladesh

Abstract

Subjectivity detection (SD) refers to identifying whether a sentence in a news article conveys the author's personal opinion or presents an impartial, factual statement. SD plays a significant role in fact-checking, sentiment analysis, and information extraction. However, detecting subjectivity remains challenging due to complex grammatical structures, the multifaceted nature of language, and the nuanced ways in which opinions can be expressed. To advance research in this area, the CheckThat! 2025 lab has launched a shared task aimed at developing automatic systems for subjectivity detection across monolingual, multilingual, and zero-shot scenarios. In this study, we present a multilingual transformer-based approach tailored to both multilingual and zero-shot subjectivity detection. Our method utilizes contextualized representations from pre-trained transformers and is fine-tuned using Focal Loss, which emphasizes harder-to-classify examples during training. Experimental results demonstrate the effectiveness of our approach, which achieved competitive performance in the shared task, most notably, securing first place in the zero-shot Ukrainian subjectivity detection track.

Keywords

multilingual subjectivity detection, zero-shot subjectivity detection, transformer, focal loss

1. Introduction

Subjectivity encompasses viewpoints, evaluations, or conclusions shaped by individual perceptions, emotions, opinions, or biases [1, 2]. Identifying whether a textual content conveys the author's subjective viewpoint has become a key area of research in natural language processing (NLP), due to its wide-ranging applications such as fact-checking [3, 4], claim detection [5], and sentiment analysis [6]. Although initial investigations in this domain were largely centered on the English language, recent efforts have increasingly expanded into multilingual contexts [7, 8]. Deep learning techniques have demonstrated superior performance in subjectivity detection tasks [2, 9] compared to traditional methods that rely on lexical or syntactic features [10, 11]. In multilingual settings, some approaches adopt a two-step pipeline, where texts are first translated into English and then processed using monolingual models trained on English data [2, 9]. Alternatively, other studies utilize multilingual models directly, enabling subjectivity detection without intermediate translation [12, 13].

To foster progress subjectivity detection in multiple languages, Ruggeri et al. present a shared task as part of CheckThat! 2025 at CLEF 2025 [14]. The task is organized into three distinct subtasks. Subtask 1 focuses on subjectivity detection in a monolingual setting, where both training and evaluation occur within the same language. This subtask includes five languages: Arabic, Bulgarian, English, German, and Italian. Subtask 2 addresses the multilingual scenario, requiring systems to be trained and tested on a combination of texts from different languages. In contrast, subtask 3 explores the zero-shot setting, where the model is trained on a subset of languages and evaluated on previously unseen ones. For this purpose, the organizers have added four additional test languages: Greek, Polish, Romanian, and Ukrainian. To demonstrate a clear view of the task definition, we articulate two examples in Table 1 for the English language.

CLEF 2025 Working Notes, 9 – 12 September 2025, Madrid, Spain

*Corresponding author.

ahmad.csecu@gmail.com (M. Ahmad); nowshed@cu.ac.bd (A. N. Chy)

Table 1
Examples of CheckThat! 2025 task 1

Sentence	Label
The Immigration Invasion symbolizes a lot about the present state of the immigration debate.	SUBJ
As has been pointed out elsewhere, the cost of rent is considerably greater than even the spiralling cost of energy bills.	OBJ

To tackle the problem of distinguishing between subjective and objective sentences across multilingual and zero-shot contexts, we propose a system in this paper. Our system harnesses a multilingual transformer to extract contextualized features from the given sentence. We utilize focal loss, which emphasizes harder-to-classify examples, helping the model focus on them during training.

The remainder of this paper is structured as follows: Section 2 presents the architecture and components of our proposed subjectivity detection system. Section 3 outlines the experimental framework, including datasets, training configuration, and evaluation metrics. Section 4 provides an analysis and discussion of the results. Finally, Section 5 concludes the work and discusses possible directions for future research.

2. System Overview

This section outlines our approach for CheckThat! 2025 Task 1, which involves determining whether a given sentence reflects the subjective view of the author behind it or presents an objective perspective on the topic. The competition includes three distinct settings, we have participated in the latter two. The second setting focuses on detecting subjectivity in textual data across multiple languages, whereas the third requires training on multiple languages and evaluating performance on previously unseen ones. An overview of our proposed framework is depicted in Figure 1.

Given a sentence from a news article, our approach begins by generating contextual representations using a multilingual transformer encoder. Specifically, we fine-tune a multilingual variant of DeBERTa [15] to capture semantic features relevant to the task. The contextualized embedding corresponding to the `[CLS]` token is passed into a classification head to produce unnormalized output scores (logits). To compute the loss, we employ the Focal Loss function [16], which incorporates the predicted logits, the ground-truth labels, and a focusing parameter, γ that emphasizes harder-to-classify examples. The model parameters are subsequently updated through backpropagation to improve generalization on the training data. For zero-shot settings, where training is performed in some languages and evaluation occurs on previously unseen ones, we utilize our fine-tuned model in subtask 2.

2.1. Multilingual Transformer

In contrast to traditional sequence modeling approaches like LSTMs [17] and convolutional networks [18], transformer-based architectures are well-suited for modeling long-range relationships within sequences. Their use of multi-head self-attention mechanisms combined with positional encoding allows for richer token-level interactions and improved contextual embedding.

In our study, we evaluate four prominent multilingual transformer-based models: mBERT [19], XLM-RoBERTa [20], RemBERT [21], and multilingual DeBERTa [22]. These models are assessed based on their effectiveness in the multilingual setting, and the highest-performing model is further used for the zero-shot learning setup. Among the candidates, multilingual DeBERTa demonstrates the most competitive performance. Consequently, we adopt it as the encoder in our proposed framework. For implementation, we utilize the publicly available checkpoint from Hugging Face[1].

[1] https://huggingface.co/microsoft/mdeberta-v3-base

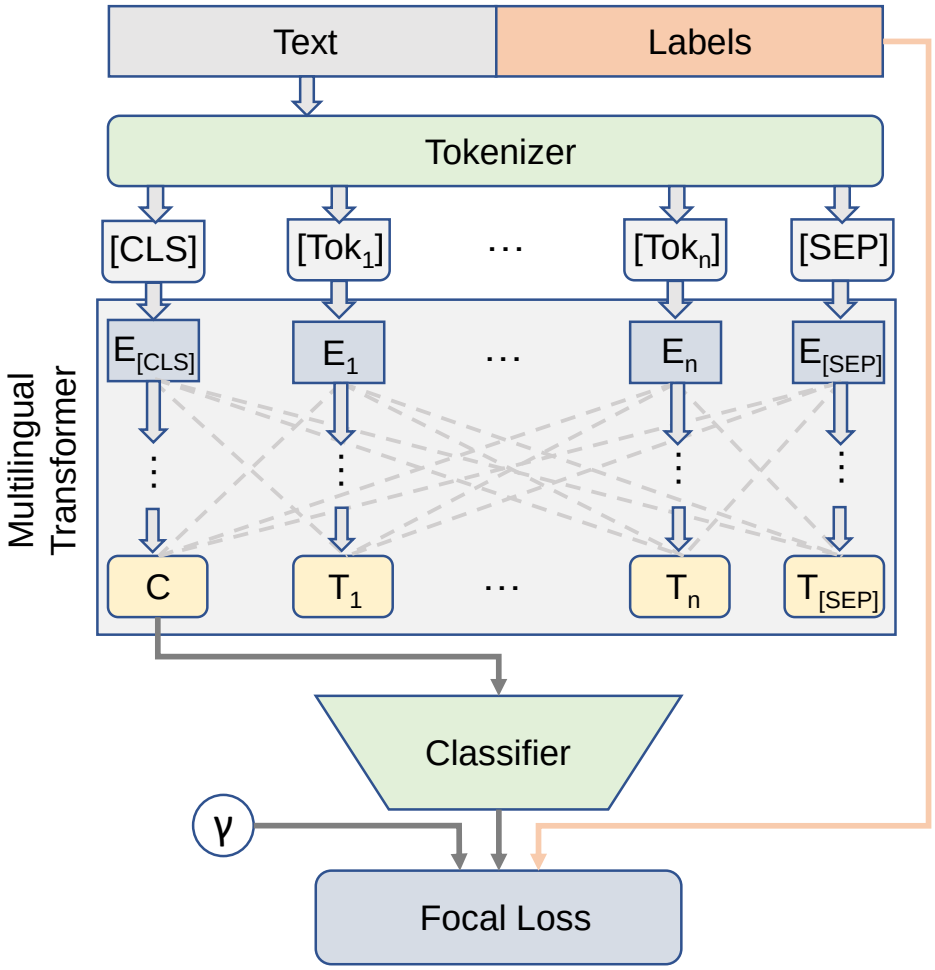

Figure 1: Overview diagram of our proposed system for CheckThat! 2025 Task 1: Identifying subjectivity in news articles

Multilingual DeBERTa is a variant of the DeBERTa [23] architecture designed for cross-lingual representation learning. DeBERTa, short for Decoding-enhanced BERT with Disentangled Attention, enhances the standard transformer design by decoupling the position and content embeddings and incorporating a relative position bias in the self-attention mechanism. The third version of the model introduces further improvements by integrating pre-layer normalization and optimized training objectives [15]. The multilingual version is pre-trained on a large-scale multilingual corpus covering over 100 languages, enabling it to capture semantic nuances across diverse linguistic contexts. We adapt this model to our classification task through fine-tuning, allowing it to learn task-specific representations.

2.2. Focal Loss

To address the challenge of class imbalance, we adopt Focal Loss [16] in this task. Traditional cross-entropy loss [24] tends to be dominated by easily classified examples, which can hinder the model's ability to learn from harder, misclassified instances. Focal Loss mitigates this by introducing a modulating factor that down-weights the contribution of well-classified samples, thereby encouraging the model to focus more on difficult cases.

Let the contextual embedding obtained from the encoder be denoted by c. The classifier then maps this embedding to unnormalized scores (logits) as follows:

$$\text{Logit} = cW^\top + b \tag{1}$$

where $W \in \mathbb{R}^{n \times d}$ and $b \in \mathbb{R}^n$ are the parameters of the classification layer. n and d represent the

number of target classes and the hidden size of the model respectively.

Let the t^{th} class be the true class label for a given input sample. Then, the predicted probability for the true class, denoted as p_t, is computed using the softmax function:

$$p_t = \frac{e^{\text{Logit}_t}}{\sum_{i=1}^{n} e^{\text{Logit}_i}} \tag{2}$$

We now compare the standard Cross-Entropy Loss and the Focal Loss for this class:

$$\mathcal{L}_{\text{CE}}(p_t) = -\log(p_t) \tag{3}$$

$$\mathcal{L}_{\text{FL}}(p_t) = -(1-p_t)^{\gamma}\log(p_t) \tag{4}$$

where γ is the focusing parameter. Therefore, the focal loss augments the standard cross-entropy formulation by introducing a factor $(1-p_t)^{\gamma}$, where p_t is the predicted probability for the correct class. The focusing parameter γ controls the degree to which well-classified examples are down-weighted. Higher values of γ reduce the relative loss assigned to correctly predicted examples, thus placing greater emphasis on learning from harder, misclassified samples.

3. Results

3.1. Dataset Overview

To assess the performance of submitted systems for subjectivity detection in the CheckThat! 2025 Task 1, the organizers provided an annotated benchmark dataset developed based on the annotation framework by Antici et al [25]. The dataset includes five languages for both mono- and multilingual settings, with an additional four languages incorporated for the zero-shot setting. The distribution of data samples across these configurations is summarized in Table 2. Since no dedicated multilingual dataset was available, we constructed one by aggregating the monolingual training and development sets from each language into a unified multilingual training set. The dev-test sets were similarly merged to form a multilingual development set. Our analysis of this aggregated dataset indicates a class distribution of 37.46% subjective instances and 62.54% objective instances, highlighting a significant class imbalance in the training data. For zero-shot scenarios, we employed the model trained on this combined multilingual dataset. During the evaluation phase, we combined the training and development sets to enhance model learning and evaluated its performance on the unseen test set provided in the CodaLab competition[2].

3.2. Parameter Settings

This section presents the system setup for our submission to the CheckThat! 2025 Task 1. We fine-tune the multilingual DeBERTa model available through the Hugging Face Transformers library [26]. All experiments are executed on Google Colab [27] using an NVIDIA T4 GPU. The random seed is fixed at 66 to ensure consistent results across runs.

We use the AdamW optimizer for training, which incorporates weight decay to improve generalization. To handle class imbalance, we adopt the Focal Loss function, setting the focusing parameter $\lambda = 1$. These and other hyperparameters were selected based on extensive experimentation across a range of values. In the final configuration, a batch size of 8, learning rate of 3×10^{-5}, and 3 training epochs provide optimal performance in our setting.

[2] https://codalab.lisn.upsaclay.fr/competitions/22756

Table 2
Statistics of the CheckThat! 2025 Task 1 dataset across different settings

Language	Label	Train	Dev	Dev-Test	Test
Subtask 1: Subjectivity detection in monolingual setting					
Arabic	SUBJ	1055	201	323	309
	OBJ	1391	266	425	727
	Total	2446	467	748	1036
Bulgarian	SUBJ	312	139	107	-
	OBJ	379	167	143	-
	Total	691	306	250	-
English	SUBJ	298	240	122	85
	OBJ	532	222	362	215
	Total	830	462	484	300
German	SUBJ	308	174	71	118
	OBJ	492	317	153	229
	Total	800	491	224	347
Italian	SUBJ	382	177	128	107
	OBJ	1231	490	334	192
	Total	1613	667	462	299
Subtask 2: Subjectivity detection in multilingual setting					
Multilingual	SUBJ	3286	751	-	619
	OBJ	5487	1417	-	1363
	Total	8773	2168	-	1982
Subtask 3: Subjectivity detection in zero-shot setting					
Greek	Total	-	-	-	284
Polish	Total	-	-	-	351
Romanian	Total	-	-	-	206
Ukrainian	Total	-	-	-	297

3.3. Evaluation Measures

To evaluate the performance of the participants' proposed systems, the organizers employ the macro-averaged F1 score [28], which is particularly suitable for datasets exhibiting a long-tail distribution. This metric provides a balanced assessment by computing the harmonic mean of precision and recall across all classes.

3.4. Results and Analysis

In this section, we present an evaluation of the CSECU-Learners system developed for the subjectivity detection task in news articles as part of CheckThat! 2025. Table 3 compares the performance of our approach with selected participant systems under the multilingual setting on the test data. Our system attained a macro-averaged F1 score of 0.7321, securing the 4th position in this task. Additionally, Table 3 includes the results for the zero-shot scenario, marked with the prefix "Zero-", where the system was tested on Greek, Polish, Romanian, and Ukrainian. The CSECU-Learners model consistently achieved competitive results across all evaluated languages. These outcomes underscore the robustness and generalization capability of our approach in both multilingual and zero-shot subjectivity detection settings.

Table 3
Comparative performance of our system against selected participants' systems

Language	Team	Rank	Macro F1
Multilingual	TIFIN India	1st	0.7550
	CSECU-Learners	4th	0.7321
	Baseline	14th	0.6390
	AI Wizards	17th	0.2380
Zero-Greek	AI Wizards	1st	0.5067
	CSECU-Learners	3rd	0.4919
	Baseline	9th	0.4159
	TIFIN India	15th	0.3337
Zero-Polish	CEA-LIST	1st	0.6922
	CSECU-Learners	3rd	0.6558
	Baseline	9th	0.5719
	TIFIN INDIA	15th	0.3811
Zero-Romanian	msmadi	1st	0.8126
	CSECU-Learners	2nd	0.7992
	Baseline	14th	0.6461
	TIFIN INDIA	15th	0.5181
Zero-Ukrainian	**CSECU-Learners**	1st	0.6424
	Investigators	2nd	0.6413
	Baseline	6th	0.6296
	TIFIN INDIA	15th	0.4731

4. Discussion

This section examines the influence of varying the focusing parameter γ in the Focal Loss function on our system's performance. As illustrated in Figure 2, we evaluate the model using different γ values from the set $\{0, 1, 2, 3\}$. The figure presents the F1 scores for both the Subjective (SUBJ) and Objective (OBJ) classes, along with their macro-averaged F1 score. When $\gamma = 0$, the modulating factor $(1 - p_t)^\gamma$ becomes 1, making the Focal Loss equivalent to the standard Cross-Entropy (CE) Loss. Under this setting, the model achieves F1 scores of 0.59 for the SUBJ class and 0.85 for the OBJ class, resulting in a macro F1 score of 0.72.

Among the tested values, the highest macro F1 score (0.73) is observed when $\gamma = 1$, while the lowest (0.70) occurs at $\gamma = 3$. These results suggest that setting $\gamma = 1$ provides the optimal balance, improving the F1 score for the subjective class by approximately 4% and the macro F1 score by 1% over the Cross-Entropy baseline. This improvement highlights the effectiveness of Focal Loss in guiding the model's attention toward more challenging examples—an aspect not adequately addressed by traditional loss functions.

5. Conclusion and Future Direction

In this paper, we introduce a method for identifying subjectivity in news articles across multilingual and zero-shot contexts by leveraging fine-tuned multilingual transformer models. We employ Focal Loss to focus the model's learning on hard-to-classify instances during training, thereby enhancing its generalizability. Empirical results validate the effectiveness of our approach. In the multilingual setting, our method achieved a competitive rank, placing 4th in the leaderboard. In the zero-shot scenario, the system demonstrated strong performance, securing 1st place for Greek, 2nd for both Polish and Romanian, and 3rd for Ukrainian.

For future directions, we intend to explore advanced multilingual transformer architectures and

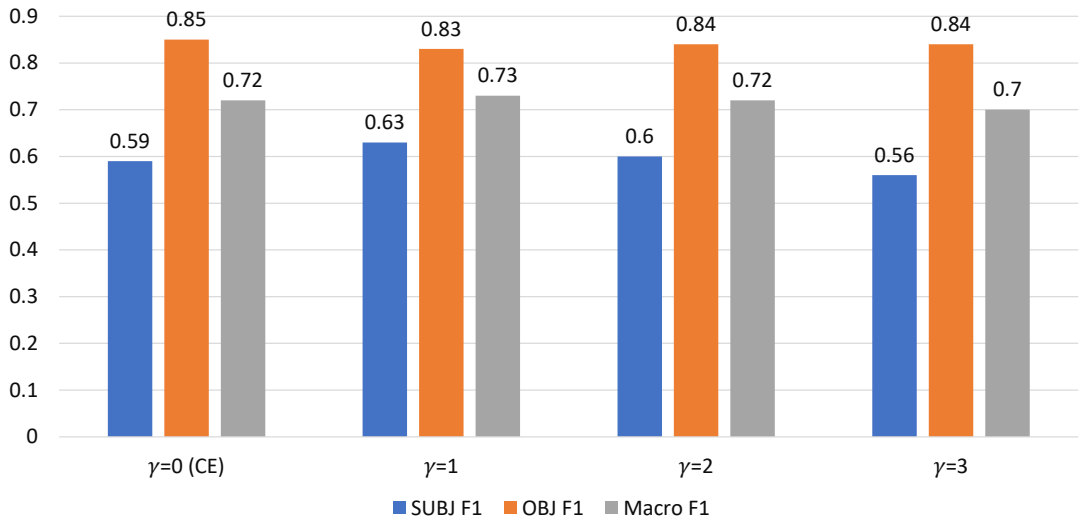

Figure 2: Effect of focusing parameter γ in Focal Loss

evaluate ensemble strategies that combine multiple transformer models. Given the class imbalance in the dataset, we also plan to implement data augmentation techniques to enhance representation across categories, intending to further improve system performance in both multilingual and zero-shot subjectivity detection tasks.

Declaration on Generative AI

During the preparation of this work, the authors utilized ChatGPT and Grammarly for grammar and spelling checks, paraphrasing, and rewording. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] J. Kocoń, M. Gruza, J. Bielaniewicz, D. Grimling, K. Kanclerz, P. Miłkowski, P. Kazienko, Learning personal human biases and representations for subjective tasks in natural language processing, in: 2021 IEEE International Conference on Data Mining (ICDM), 2021, pp. 1168–1173. doi:10.1109/ICDM51629.2021.00140.

[2] M. R. Biswas, A. T. Abir, W. Zaghouani, Nullpointer at checkthat! 2024: Identifying subjectivity from multilingual text sequence, 2024. URL: https://arxiv.org/abs/2407.10252. arXiv:2407.10252.

[3] L. L. Vieira, C. L. M. Jerônimo, C. E. Campelo, L. B. Marinho, Analysis of the subjectivity level in fake news fragments, in: Proceedings of the Brazilian Symposium on Multimedia and the Web, 2020, pp. 233–240.

[4] C. L. M. Jeronimo, L. B. Marinho, C. E. Campelo, A. Veloso, A. S. da Costa Melo, Fake news classification based on subjective language, in: Proceedings of the 21st International Conference on Information Integration and Web-based Applications & Services, 2019, pp. 15–24.

[5] E. Riloff, J. Wiebe, Learning extraction patterns for subjective expressions, in: Proceedings of the 2003 conference on Empirical methods in natural language processing, 2003, pp. 105–112.

[6] T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in: Proceedings of human language technology conference and conference on empirical methods in natural language processing, 2005, pp. 347–354.

[7] A. Galassi, F. Ruggeri, A. Barrón-Cedeño, F. Alam, T. Caselli, M. Kutlu, J. M. Struß, F. Antici,

M. Hasanain, J. Köhler, et al., Overview of the clef-2023 checkthat! lab: Task 2 on subjectivity in news articles, in: 24th Working Notes of the Conference and Labs of the Evaluation Forum, CLEF-WN 2023, CEUR Workshop Proceedings (CEUR-WS. org), 2023, pp. 236–249.
[8] J. M. Struß, F. Ruggeri, A. Barrón-Cedeño, F. Alam, D. Dimitrov, A. Galassi, G. Pachov, I. Koychev, P. Nakov, M. Siegel, et al., Overview of the clef-2024 checkthat! lab task 2 on subjectivity in news articles, in: CEUR Workshop Proceedings, volume 3740, CEUR-WS, 2024, pp. 287–298.
[9] M. Casanova, J. Chanson, B. Icard, G. Faye, G. Gadek, G. Gravier, P. Égré, Hybrinfox at checkthat! 2024-task 2: Enriching bert models with the expert system vago for subjectivity detection (2024).
[10] E. Gajewska, Eevvgg at checkthat! 2024: evaluative terms, pronouns and modal verbs as markers of subjectivity in text, Faggioli et al.[22] (2024).
[11] P. Premnath, P. Subramani, N. R. Salim, B. Bharathi, Ssn-nlp at checkthat! 2024: From feature-based algorithms to transformers: A study on detecting subjectivity, in: Conference and Labs of the Evaluation Forum, 2024. URL: https://api.semanticscholar.org/CorpusID:271793774.
[12] A. Rodríguez, E. Golobardes, J. Suau, Tonirodriguez at checkthat!2024: Is it possible to use zero-shot cross-lingual methods for subjectivity detection in low-resources languages?, in: Conference and Labs of the Evaluation Forum, 2024. URL: https://api.semanticscholar.org/CorpusID:271851634.
[13] F. Leistra, T. Caselli, Thesis titan at checkthat!-2023: Language-specific fine-tuning of mdebertav3 for subjectivity detection, in: Conference and Labs of the Evaluation Forum, 2023. URL: https://api.semanticscholar.org/CorpusID:264441796.
[14] F. Ruggeri, A. Muti, K. Korre, J. M. Struß, M. Siegel, M. Wiegand, F. Alam, R. Biswas, W. Zaghouani, M. Nawrocka, B. Ivasiuk, G. Razvan, A. Mihail, Overview of the CLEF-2025 CheckThat! lab task 1 on subjectivity in news article, in: G. Faggioli, N. Ferro, P. Rosso, D. Spina (Eds.), Working Notes of CLEF 2025 - Conference and Labs of the Evaluation Forum, CLEF 2025, Madrid, Spain, 2025.
[15] P. He, J. Gao, W. Chen, Debertav3: Improving deberta using electra-style pre-training with gradient-disentangled embedding sharing, 2021. arXiv:2111.09543.
[16] T.-Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollar, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017.
[17] M. Schuster, K. K. Paliwal, Bidirectional recurrent neural networks, IEEE transactions on Signal Processing 45 (1997) 2673–2681.
[18] I. Goodfellow, Y. Bengio, A. Courville, Y. Bengio, Deep learning, volume 1, 2016.
[19] J. Devlin, Multilingual bert readme document, https://github.com/google-research/bert/blob/master/multilingual.md, 2018.
[20] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, E. Grave, M. Ott, L. Zettlemoyer, V. Stoyanov, Unsupervised cross-lingual representation learning at scale, in: D. Jurafsky, J. Chai, N. Schluter, J. Tetreault (Eds.), Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, Association for Computational Linguistics, Online, 2020, pp. 8440–8451. URL: https://aclanthology.org/2020.acl-main.747/. doi:10.18653/v1/2020.acl-main.747.
[21] H. W. Chung, T. Févry, H. Tsai, M. Johnson, S. Ruder, Rethinking embedding coupling in pre-trained language models, 2020. URL: https://arxiv.org/abs/2010.12821. arXiv:2010.12821.
[22] P. He, J. Gao, W. Chen, mdeberta-v3 - multilingual deberta v3 model, https://huggingface.co/microsoft/mdeberta-v3-base, 2023.
[23] P. He, X. Liu, J. Gao, W. Chen, Deberta: Decoding-enhanced bert with disentangled attention, arXiv preprint arXiv:2006.03654 (2020).
[24] A. Mao, M. Mohri, Y. Zhong, Cross-entropy loss functions: Theoretical analysis and applications, in: International conference on Machine learning, PMLR, 2023, pp. 23803–23828.
[25] F. Antici, F. Ruggeri, A. Galassi, K. Korre, A. Muti, A. Bardi, A. Fedotova, A. Barrón-Cedeño, A corpus for sentence-level subjectivity detection on English news articles, in: N. Calzolari, M.-Y. Kan, V. Hoste, A. Lenci, S. Sakti, N. Xue (Eds.), Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024), ELRA and ICCL, Torino, Italia, 2024, pp. 273–285. URL: https://aclanthology.org/2024.lrec-main.25/.
[26] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Fun-

towicz, et al., Huggingface's transformers: State-of-the-art natural language processing, arXiv preprint arXiv:1910.03771 (2019).
[27] E. Bisong, Google colaboratory, in: Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners, Springer, 2019, pp. 59–64.
[28] M. Sokolova, G. Lapalme, A systematic analysis of performance measures for classification tasks, Information processing & management 45 (2009) 427–437.