CUET_KCRL at CheckThat!2025: EnsembleNet with RoBERTa-Large for Subjectivity detection in News Articles

Notebook for the CheckThat! Lab at CLEF 2025

Md. Tanvir Ahammed Shawon[1,*,†], Fariha Haq[1,†], Md. Ayon Mia[1], Golam Sarwar Md. Mursalin[1] and Muhammad Ibrahim Khan[1]

[1]*Department of Computer Science & Engineering, Chittagong University of Engineering and Technology, Chattogram,4349, Bangladesh*

Abstract

This study introduces EnsembleNet with RoBERTa-large , an innovative transformer-based architecture designed to detect subjectivity in News Article. Our approach EnsembleNet with RoBERTa-large introduce Multi-sample dropout for diverse feature representation, a Multi-head ensemble for efficient prediction stability, and Focal loss to handle minority class learning. We evaluated using traditional ML models and DL models, and transformer model. We found that EnsembleNet with RoBERTa-large achieves a weighted F1-score of 0.82 and a macro F1-score of 0.77. In the competition, we used a pre-trained BERT model and ranked 17th. After the competition, we explored different architectures and finalized our model using the EnsembleNet with RoBERTa-large. Despite encountering some false negatives that highlight areas for improvement, this work emphasizes the potential of EnsembleNet with RoBERTa-large as an efficient tool for handling imbalanced text classification.

Keywords

EnsembleNet, Multi-Sample Dropout, Multi-Head Ensemble, Focal Loss

1. Introduction

The News items influence people on how to perceive the world, but not all news is real[1]. Subjectivity in news articles refers to the inclusion of personal opinions, emotions, or bias instead of just facts. Detecting subjectivity is important to identify media bias, ensure fair reporting, and prevent misinformation[2][3]. Some news articles include opinions, feelings, or personal beliefs that making them subjective. However, some contain only clear, objective information. To improve news analysis, recognizing the difference between subjective (SUBJ) and objective (OBJ) is critical to recognize media biases and developing tools[4][5][6]. Objectivity can be measured, observed, and verified. It is polished through controlled experiments, established processes, and statistical analysis. It avoids personal bias or emotion, resulting in more trustworthy and replicable results[7][8]. On the other hand, subjectivity is influenced by personal experiences, ideas, and emotions. It is often regarded as as less reliable in science. For this reason, it is essential in subjects such as the humanities and social sciences, where personal understanding and context are important[7]. Several studies have investigated how to distinguish between subjective and objective news articles, with the majority of the methods focusing on widely spoken languages[1][9][10][11][12]. The main challenge in this detection is to negotiate the complicated, context-sensitive character of language, in which subjective texts frequently use subtle hints to indicate personal viewpoints. This paper makes numerous important contributions:

- Introduced an EnsembleNet with RoBERTa-Large architecture for Subjectivity detection in News Articles

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

*Corresponding author.

†These authors contributed equally.

u1904077@student.cuet.ac.bd (Md. T. A. Shawon); u1904051@student.cuet.ac.bd (F. Haq); u1804128@student.cuet.ac.bd (Md. A. Mia); sarwarmursalin1015@gmail.com (G. S. Md. Mursalin); muhammad_ikhan@cuet.ac.bd (M. I. Khan)

0009-0000-0966-4119 (Md. T. A. Shawon); 0009-0009-1650-6544 (F. Haq); 0009-0008-1693-6088 (Md. A. Mia); 0009-0006-1279-8840 (G. S. Md. Mursalin); 0000-0002-4228-6727 (M. I. Khan)

- Explored the performance of various ML, DL and transformer models to effectively detect Subjectivity in news articles

The following GitHub repository contains the complete implementation details: https://github.com/Ahammed-77/CheckThatLabCLEF2025_Sharedtask

2. Related Works

In recent years, the CLEF CheckThat! competition has showcased innovative approaches to claim detection. Previous study demonstrates diverse approaches to Subjectivity detection in English languages. Paran et al. [1] presented a method for identifying subjectivity and objectivity in news items written in Arabic and English that is based on LLM. The LLM (Llama-3-8b) achieved the greatest F1-scores of 50.36% (English) and 72.6% (Arabic), outperforming the other models. Dey et al. [9] applied XLM-RoBERTa which performed better than BERT and BERT-m. The model outperformed for the Multilingual dataset with an F1 score of 0.82, and it also performed better for the Arabic dataset with a macro F1 score of 0.79. Biswas et al. [10] fine-tuned the sentiment-based Transformer model 'MarieAngeA13/Sentiment-Analysis BERT'. Their model achieved the best performance on the German dataset, with an F1 Macro score of 0.79 and an accuracy of 0.81.

Gruman et al.[11] got a notable improvement by using their approach Googles pre-trained LLMs, Gemini. They achieved and F1 score of 0.445. Tran et al. [6] used BERT and RoBERTa models. They included an additional mean pooling and dropout layer on top of the model which help in reducing overfitting. For English, they achieved an accuracy of 0.696 with F1 weighted score of 0.687. Premnath et al. [13] applied RoBERTa model with additional POS tag features, achieved a macro-F1 score of 0.71. Rodriguez et al. [14] applied Zero-Shot Cross-Lingual transfer techniques using the datasets. Also fine-tuned two multilingual models, mDeBERTa v3 and XLM-RoBERTa. MDeBERTa v3 Base model that achieved with a score of 0.7372. Fariha et al. [15] evaluated multilingual transformer-based models and noticed that models trained in the multilingual setting achieved the best performance. Salas-Jimenez et al. [16] applied BERT-based classifiers and achieved a macro F1 score of 0.82 on the English dataset. Zehra et al. [17] applied ensemble approach and combined BERT-Base-Uncased and XLM-RoBERTa-Base. Their macro F1 score was 0.7081 for analyzing subjectivity. Antici et al. [18] set an innovative annotation guidelines for subjectivity detection which is applicable to any language.

3. Dataset and Task Description

Our study focuses on developing a system that can automatically determine if a sentence in a news story is subjective (SUBJ) or objective (OBJ).This task classifies texts into two types: subjective (SUBJ), which show opinions, and objective (OBJ), which give facts. We trained all models using the training set and evaluated the model's performance based on the dev set and predicted sentence labels using an unlabeled test set.

Table 1
Distribution of English text different categories in the training, validation, and test sets to detect Subjectivity in News Articles

Sets	Classes		Total
	SUBJ	OBJ	
Train	298	532	830
Dev	222	240	462
Dev_Test	122	362	484
Test	85	215	300

CLEF 2025- CheckThat! Lab[12][19][20] consists of Four Tasks. We participated in share task 1 (Subjectivity in News Articles). Table 1 shows the data set statistics for Task-1.

4. System Overview

Our framework introduces an innovative approach for subjectivity detection in News articles and differentiates between subjective (SUBJ) and objective (OBJ) sentences. We achieve reliable performance by combining a streamlined preprocessing pipeline, a custom dataset module, and a EnsembleNet transformer-based architecture. To address class imbalance and improve generalization, the system uses RoBERTa-large with innovative modifications such as multi-sample dropout, multi-head ensemble, and focal loss.

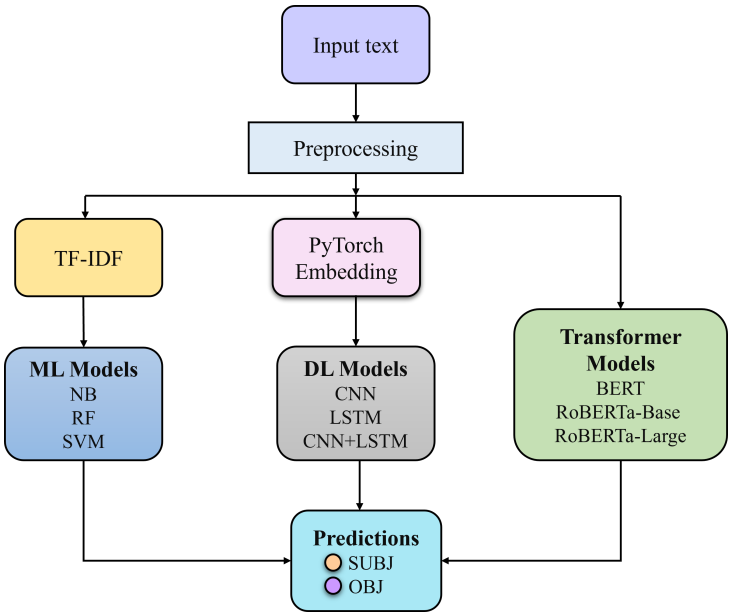

Figure 1: Overview of Subjectivity Detection.

4.1. Data Preprocess

For standardizing English News articles, we implement a systematic preprocessing pipeline. The raw text is subjected to multiple cleaning processes, which includes removing URLs, handling emojis, eliminating hashtags and mentions, and normalizing sequential punctuation. Labels are converted to binary values with SUBJ = 1 and OBJ = 0. A custom SubjectivityDataset class is intended to hold sentences and their binary labels. The RoBERTa-large tokenizer generates input IDs and attention masks with a fixed sequence length of 256 for each sentence. This ensures consistent text representation before feeding into our models.

4.2. ML Models

For the subjectivity detection in News articles, we employed several classical machine learning models: Multinomial Naive Bayes (NB), Support Vector Machine (SVM), and Random Forest. We used a TF-IDF vectorizer to convert input sentence into feature vectors, with a vocabulary size of 5000. We also used some preprocessing technique. The SVM classifier model used a linear kernel, and Random Forest was build up with 100 estimators and fixed random seed for consistency in performance. The Naive Bayes model build a MultinomialNB implementation, which had default settings. We trained all models on the training data and tested on the labeled test data using Macro F1-score metrics.

4.3. DL Models

We implemented three deep learning models—CNN, LSTM, and CNN+LSTM—for subjectivity detection in News Articles. In CNN Model, we use 1D convolutions with different filter size to capture local

n-gram patterns from 128-dimensional word embeddings. We also used max-pooling extracts key features which is followed by dense layers with 0.5 dropout. In LSTM Model, we used a bidirectional LSTM with two layers (128 hidden units, 0.3 dropout) processes 128-dimensional embeddings to model long-range dependencies. We used a hybrid model by combining CNN's local feature extraction (filters of sizes 3 and 5) with bidirectional LSTM. The features are merged and passed through dense layers (256 and 128 units, 0.5 dropout), to balance the local and global context and enhance performance.

4.4. Transformer models

We explored transformer-based techniques using three powerful pre-trained models: RoBERTa-large, RoBERTa-base [21] (Liu et al., 2019,) and BERT-base-uncased [22] (Devlin et al., 2018.) accessed via the Hugging Face platform [23] (Wolf et al., 2019) and implemented in PyTorch. We fine-tuned each on our dataset using the AdamW optimizer with a batch size of 16 across four epochs, integrating early stopping linked to the validation F1 score to avoid overfitting and improve classification accuracy.

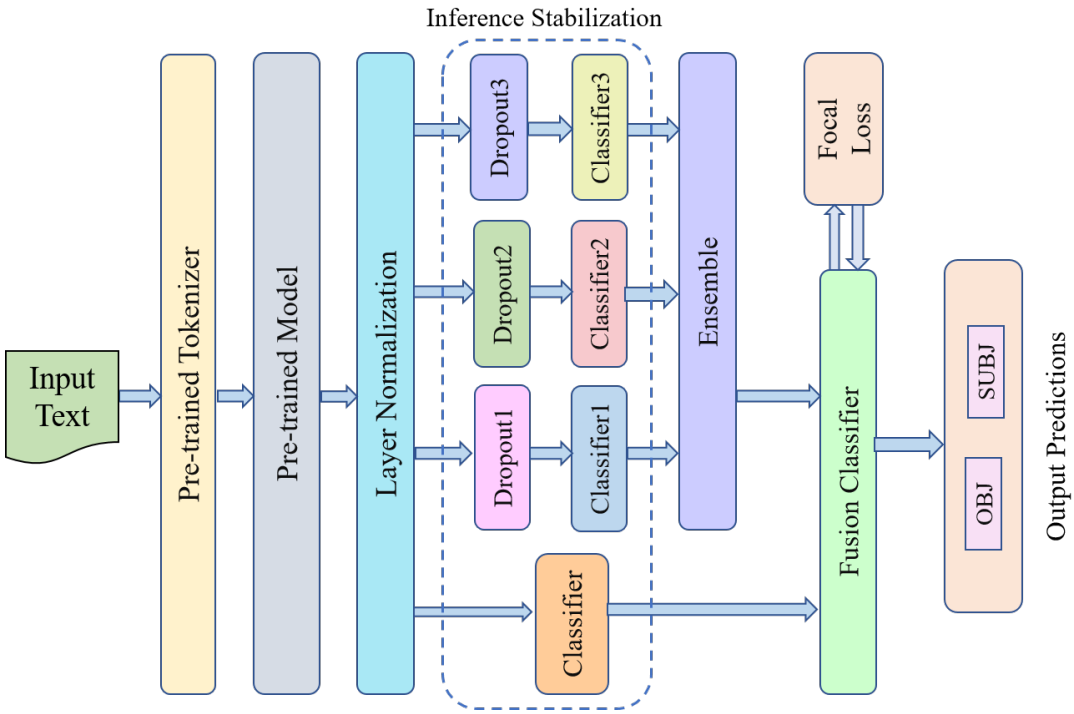

Figure 2: Architecture of EnsembleNet with RoBERTa-large for Subjectivity Detection.

4.5. EnsembleNet with RoBERTa-Large Architecture

Our core model, named EnsembleNet, is built upon the pre-trained RoBERTa-large model, which provides efficient contextual embeddings for English text. For an input sequence $S = \{s_1, s_2, \ldots, s_n\}$. RoBERTa produces a pooled output $h_p \in \mathbb{R}^{1024}$:

$$h_p = \text{RoBERTa}(S).$$

EnsembleNet with incorporates some innovative techniques:

- **Multi-sample Dropout:** We apply three dropout layers ($p = 0.3$) to produce diverse representations: h_{d1}, h_{d2}, h_{d3}, which enhances feature diversity, reducing overfitting and improving generalization.

- **Multi-head Ensemble:** The model has three linear classification heads, each mapping the RoBERTa pooled output to binary classes (SUBJ and OBJ). Three linear classifiers map h_{di} to logits:

$$l_i = W_i h_{di} + b_i, \quad i \in \{1, 2, 3\},$$

where $W_i \in \mathbb{R}^{2\times 1024},\ b_i \in \mathbb{R}^2$. The ensemble logits are averaged:

$$l_e = \frac{1}{3}\sum_{i=1}^{3} l_i.$$

A parallel path projects h_p to a reduced space (512 dimensions) with ReLU activation:

$$h_f = \text{ReLU}(W_f h_p + b_f), \quad l_m = W_m h_f + b_m,$$

where $W_f \in \mathbb{R}^{512\times 1024},\ W_m \in \mathbb{R}^{2\times 512}$. Final logits are:

$$l_{\text{final}} = \frac{l_e + l_m}{2}.$$

- **Inference Stabilization:** During inference, five forward passes with dropout are averaged to produce stable predictions:

$$l_{\text{inf}} = \frac{1}{5}\sum_{k=1}^{5} l_{\text{final}}^{(k)}.$$

- **Focal Loss:** To handle potential class imbalance problem, we introduce focal loss in EnsembleNet with RoBERTa-Large.

$$\text{FL}(p_t) = -\alpha(1-p_t)^{\gamma}\log(p_t),$$

with $\alpha = 1,\ \gamma = 2$, which prioritizes the probability of true class. This loss prioritizes hard-to-classify examples, enhancing performance on underrepresented classes. Layer normalization is applied to h_p to stabilize training:

$$h_p \leftarrow \text{LayerNorm}(h_p).]$$

EnsembleNet is trained over 5 epochs using the AdamW optimizer, with learning rates of 2×10^{-5} for general parameters and 4×10^{-5} for bias and LayerNorm weights. A linear scheduler with a 10% warmup phase and gradient clipping (max norm = 1.0) guarantees training stability.

5. Result Analysis

Table 2 shows the comparative results of different models for subjectivity detection in News Articles in English language using macro-averaged precision (Pr), recall (Re), and F1-score (F1). Among ML models, Naive Bayes (NB) achieved the highest weighted F1-score of 0.64. Besides Random Forest (RF) achieved 0.60 and SVM achieved 0.59. However, their macro F1-scores range is 0.49–0.55, which is indicates that model faces challenges for the minority SUBJ class. Among the DL models, CNN+LSTM demonstrated the best performance with a weighted F1 score of 0.64%. The CNN and LSTM models achieve weighted F1-scores of 0.61 and 0.62.

Transformer-based models performed better than ML DL models. BERT-base-uncased achieved a weighted F1-score of 0.74, while RoBERTa-base improved to 0.81. Our proposed EnsembleNet architecture which is built on RoBERTa-large. It achieves the best performance with a weighted F1-score of 0.82 and a macro F1-score of 0.77. These results highlight that our proposed EnsembleNet model's ability to capture nuanced subjectivity through feature fusion. In traditional ML models like Naive Bayes (F1: 0.64), SVM (F1: 0.59), and Random Forest (F1: 0.60) are struggled with lower F1-scores because they used on TF-IDF, which missed contextual depth and falters with imbalanced SUBJ data. In dL models like CNN (F1: 0.61), LSTM (F1: 0.62), and CNN-LSTM (F1: 0.643) do slightly better but are held

Table 2
Performance comparison across different model architectures and strategies on the test set, where Pr, Re, and F1 denote macro-averaged(M) and weighted(W) precision, recall, and F1-score respectively.

Model	Pr (W)	Re (W)	F1 (W)	Pr (M)	Re (M)	F1 (M)
ML Models						
Naive Bayes	0.63	0.65	0.64	0.55	0.54	0.55
SVM	0.61	0.58	0.59	0.52	0.52	0.51
Random Forest	0.59	0.61	0.60	0.49	0.49	0.49
DL Models						
CNN	0.61	0.62	0.61	0.52	0.51	0.51
LSTM	0.62	0.61	0.62	0.54	0.54	0.54
CNN + LSTM	0.64	0.66	0.64	0.55	0.55	0.55
Transformer Models						
BERT-base-uncased	0.74	0.74	0.74	0.68	0.68	0.68
RoBERTa-base	0.81	0.81	0.81	0.77	0.76	0.77
RoBERTa-large	**0.82**	**0.83**	**0.82**	**0.80**	**0.76**	**0.77**

back by static embeddings and lack of robust imbalance handling. Our proposed model EnsembleNet with RoBERTa-Large achieves a higher F1-score (0.82) due to contextual embeddings, multi-sample dropout, multi-head ensemble, and focal loss, which address overfitting and class imbalance. Three dropout layers (p=0.3) create diverse features, reducing overfitting and improving generalization for the minority SUBJ class. Three classification heads and a parallel path average predictions, stabilizing results and boosting SUBJ and OBJ accuracy. Using focal loss with $\alpha = 1$ and $\gamma = 2$, it targets tough SUBJ cases, improving recall and balancing performance.

6. Error Analysis

The confusion matrix for RoBERTa-large, enhanced with the EnsembleNet architecture, provides valuable insights into its misclassification patterns on the test set. Out of 215 true OBJ instances, it accurately predicted 199, but unfortunately, 16 were misclassified as SUBJ. When it comes to the 85 true SUBJ instances, only 50 were correctly identified, leaving 35 incorrectly labeled as OBJ. This highlights a concerning false negative rate for SUBJ at 41.2%, which points to the challenges faced by the minority class due to its lower support and the complexity of its linguistic features.

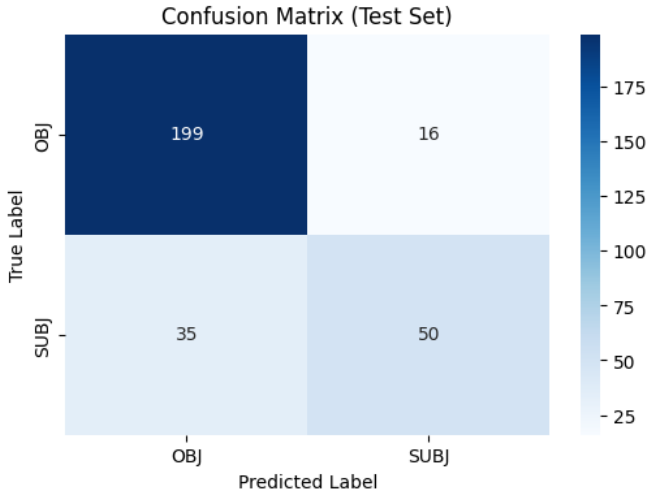

Figure 3: Confusion matrix demonstrating the proposed model's classification performance.

EnsembleNet is making waves with its innovative techniques that really boost performance. The multi-sample dropout helps to cut down on overfitting by creating a variety of feature representations. Meanwhile, the multi-head ensemble works to stabilize predictions across different classifiers, which likely helps to reduce those 16 false positives for OBJ. The focal loss, set with alpha and gamma, does a great job of focusing on those misclassified SUBJ instances, which boosts the recall to 0.59 when compared to simpler models. Still, the ongoing issue of 35 false negatives indicates that capturing complex subjective expressions is quite a challenge. This could be due to either limited training data or the need for better attention mechanisms. EnsembleNet shows strong performance in dealing with imbalanced data, but there's definitely way to improve on reducing those SUBJ false negatives.

7. Conclusion

The EnsembleNet with RoBERTA-large architecture, shows the impressive effectiveness in detecting subjectivity in news articles. This architecture achieve highest weighted F1-score of 0.82 and macro F1-score of 0.77, it clearly outperforms than all machine learning methods, like Naive Bayes, which only scores 0.64 of F1 score, as well as simpler deep learning models such as CNN+LSTM, also at 0.64. The model's multi-sample dropout improves feature diversity, the Multi-head ensemble ensures efficient predictions, and the focal loss effectively handle class imbalance and boost SUBJ recall to 0.59. The confusion matrix shows that the model increases true positive and true negative rates and decreases false positive and false negative rates than other transfomer, ML and DL models. These results highlight EnsembleNet with RoBERTa-Large ability at capturing subtle nuances in subjectivity, making it an innovative method for tackling imbalanced datasets in English text analysis.

8. Limitations

Our proposed architecture EnsembleNet with RoBERTa-Large has its strengths, it does come with some limitations. For instance, the model has a tendency to produce false negatives, misclassifying 35 SUBJ instances as OBJ, which results in a 41.2% error rate for the minority class. This is largely due to its lower support of just 85 instances and the complex linguistic features involved. The reliance on the pre-trained RoBERTa-large model may limit its ability to adapt to the individual subtleties of many domains without further fine-tuning. Besides, the efficiency of focus loss alpha and gamma could be limited by the size of the dataset, implying that using larger or diverse training data could help to reduce errors. Future study should improved attention processes or other regularization technique to overcome these difficulties.

9. Declaration on Generative AI

During the preparation of this work, We made limited use of AI-assisted tools such as ChatGPT and Grammarly. These tools were used only for minor tasks, including checking grammar, correcting spelling mistakes, and rephrasing some sentences to improve readability. All scientific contributions, experimental design, analysis, and conclusions presented in this paper were fully conceived, written, and verified by us. We carefully reviewed and edited all AI-assisted suggestions and take full responsibility for the final content of the manuscript.

References

[1] A. I. Paran, M. S. Hossain, S. H. Shohan, J. Hossain, S. Ahsan, M. M. Hoque, Semanticcuetsync at checkthat! 2024: finding subjectivity in news article using llama, Faggioli et al.[22] (2024).

[2] M. Shokri, V. Sharma, E. Filatova, S. Jain, S. Levitan, Subjectivity detection in english news using large language models, in: Proceedings of the 14th Workshop on Computational Approaches to Subjectivity, Sentiment, & Social Media Analysis, 2024, pp. 215–226.
[3] R. Suwaileh, M. Hasanain, F. Hubail, W. Zaghouani, F. Alam, Thatiar: subjectivity detection in arabic news sentences, in: Proceedings of the International AAAI Conference on Web and Social Media, volume 19, 2025, pp. 2587–2602.
[4] 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.
[5] J. Struß, F. Ruggeri, A. Barrón-Cedeno, F. Alam, D. Dimitrov, A. Galassi, G. Pachov, I. Koychev, P. Nakov, M. Siegel, M. Wiegand, M. Hasanain, R. Suwaileh, W. Zaghouani, Overview of the clef-2024 checkthat! lab task 2 on subjectivity in news articles, CEUR Workshop Proceedings 3740 (2024) 287–298. Publisher Copyright: © 2024 Copyright for this paper by its authors.; 25th Working Notes of the Conference and Labs of the Evaluation Forum, CLEF 2024 ; Conference date: 09-09-2024 Through 12-09-2024.
[6] S. Tran, P. Rodrigues, B. Strauss, E. M. Williams, Accenture at checkthat!-2023: Impacts of back-translation on subjectivity detection., in: CLEF (Working Notes), 2023, pp. 507–517.
[7] L. Escouflaire, A. Descampe, C. Fairon, Unveiling subjectivity in press discourse: A statistical and qualitative study of manually annotated articles, Discours. Revue de linguistique, psycholinguistique et informatique. A journal of linguistics, psycholinguistics and computational linguistics (2024).
[8] E. Jakaza, M. Visser, 'subjectivity'in newspaper reports on 'controversial'and 'emotional'debates: An appraisal and controversy analysis, Language Matters 47 (2016) 3–21.
[9] K. Dey, P. Tarannum, M. A. Hasan, S. R. H. Noori, Nn at checkthat!-2023: Subjectivity in news articles classification with transformer based models., in: CLEF (Working Notes), 2023, pp. 318–328.
[10] M. R. Biswas, A. T. Abir, W. Zaghouani, Nullpointer at checkthat! 2024: identifying subjectivity from multilingual text sequence, arXiv preprint arXiv:2407.10252 (2024).
[11] S. Gruman, L. Kosseim, Clac at checkthat! 2024: a zero-shot model for check-worthiness and subjectivity classification, Faggioli et al.[22] (2024).
[12] 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, 2025.
[13] 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.
[14] 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: CEUR Workshop Proceedings, volume 3740, CEUR-WS, 2024, pp. 590–597.
[15] F. Haq, M. T. A. Shawon, M. A. Mia, G. S. Md. Mursalin, M. I. Khan, KCRL@DravidianLangTech 2025: Multi-pooling feature fusion with XLM-RoBERTa for Malayalam fake news detection and classification, in: B. R. Chakravarthi, R. Priyadharshini, A. K. Madasamy, S. Thavareesan, E. Sherly, S. Rajiakodi, B. Palani, M. Subramanian, S. Cn, D. Chinnappa (Eds.), Proceedings of the Fifth Workshop on Speech, Vision, and Language Technologies for Dravidian Languages, Association for Computational Linguistics, Acoma, The Albuquerque Convention Center, Albuquerque, New Mexico, 2025, pp. 624–629. URL: https://aclanthology.org/2025.dravidianlangtech-1.107/.
[16] K. Salas-Jimenez, I. Díaz, H. Gómez-Adorno, G. Bel-Enguix, G. Sierra, Jk_pcic_unam at checkthat! 2024: analysis of subjectivity in news sentences using transformers based models, Faggioli et al.[22] (2024).
[17] S. Zehra, K. Chandani, M. Khubaib, A. Aun Muhammed, F. Alvi, A. Samad, Checker hacker at checkthat! 2024: detecting check-worthy claims and analyzing subjectivity with transformers, Faggioli et al.[22] (2024).

[18] F. Antici, A. Galassi, F. Ruggeri, K. Korre, A. Muti, A. Bardi, A. Fedotova, A. Barrón-Cedeño, A corpus for sentence-level subjectivity detection on english news articles, arXiv preprint arXiv:2305.18034 (2023).
[19] F. Alam, J. M. Struß, T. Chakraborty, S. Dietze, S. Hafid, K. Korre, A. Muti, P. Nakov, F. Ruggeri, S. Schellhammer, V. Setty, M. Sundriyal, K. Todorov, V. V., The clef-2025 checkthat! lab: Subjectivity, fact-checking, claim normalization, and retrieval, in: C. Hauff, C. Macdonald, D. Jannach, G. Kazai, F. M. Nardini, F. Pinelli, F. Silvestri, N. Tonellotto (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2025, pp. 467–478.
[20] F. Alam, J. M. Struß, T. Chakraborty, S. Dietze, S. Hafid, K. Korre, A. Muti, P. Nakov, F. Ruggeri, S. Schellhammer, V. Setty, M. Sundriyal, K. Todorov, V. Venktesh, Overview of the CLEF-2025 CheckThat! Lab: Subjectivity, fact-checking, claim normalization, and retrieval, in: J. Carrillo-de Albornoz, J. Gonzalo, L. Plaza, A. García Seco de Herrera, J. Mothe, F. Piroi, P. Rosso, D. Spina, G. Faggioli, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction. Proceedings of the Sixteenth International Conference of the CLEF Association (CLEF 2025), 2025.
[21] A. Conneau, Unsupervised cross-lingual representation learning at scale, arXiv preprint arXiv:1911.02116 (2019).
[22] J. Devlin, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).
[23] T. Wolf, Huggingface's transformers: State-of-the-art natural language processing, arXiv preprint arXiv:1910.03771 (2019).