NN at CheckThat! 2023: Subjectivity in News Articles Classification with Transformer Based Models

Krishno Dey[1], Prerona Tarannum[1], Md. Arid Hasan[1] and Sheak Rashed Haider Noori[1]

[1]Daffodil International University, Daffodil Smart City, Birulia 1216

Abstract

The CheckThat! Lab is a challenging lab designed to address the issue of disinformation. We participated in CheckThat! Lab Task 2, which is focused on classification of subjectivity in news articles. This shared task included datasets in six different languages, as well as a multilingual dataset created by combining all six languages. We followed standard preprocessing steps for Arabic, Dutch, English, German, Italian, Turkish, and multilingual text data. We employed a transformer-based pretrained model, specifically XLM-RoBERTa large, for our official submission to the CLEF Task 2. Our results were impressive, as we achieved the 1st, 1st, 2nd, 5th, 2nd, 2nd, and 3rd positions on the leaderboard for the multilingual, Arabic, Dutch, English, German, Italian, and Turkish text data, respectively. Furthermore, we also applied BERT and BERT multilingual (BERT-m) models to assess the subjectivity of the text data. Our study revealed that XLM-RoBERTa large outperformed BERT and BERT-m in all performance measures for this particular dataset provided in the shared task.

Keywords

Subjectivity-checking, News-articles, Transformer Models, BERT, BERT-m, XLM-RoBERTa large

1. Introduction

Subjectivity in Natural language refers to features of language used to express judgments, opinions, and conjectures [1]. The development of systems that can automatically distinguish between subjective and objective texts has gained popularity in recent years. This is a challenging task, as subjective text often expresses the personal opinions or beliefs of the author, while objective text presents facts or information in a neutral manner [2, 3]. One of the challenges in dealing with subjective text is that it frequently reflects the personal opinions or beliefs of the author, making it difficult to maintain a completely objective standpoint. One way to approach this task is to use a binary classification approach, in which systems are trained to identify whether a text sequence (a sentence or a paragraph) is subjective or objective. This task has been previously explored in a number of research papers, with systems achieving promising results on a variety of datasets [4, 5, 6]. However, existing systems have a number of limitations. First, they are often trained on datasets that are biased toward one type of text, such as news articles or social media posts. This can lead to systems that are not able to generalize to

CLEF 2023: Conference and Labs of the Evaluation Forum, September 18–21, 2023, Thessaloniki, Greece
krishno.cse@diu.edu.bd (K. Dey); prerona15-14134@diu.edu.bd (P. Tarannum); arid.cse0325.c@diu.edu.bd (Md. A. Hasan); drnoori@daffodilvarsity.edu.bd (S. R. H. Noori)
0000-0002-3194-5483 (K. Dey); 0000-0002-3292-1870 (P. Tarannum); 0000-0001-7916-614X (Md. A. Hasan); 0000-0001-6937-6039 (S. R. H. Noori)

other types of text. Second, existing systems often rely on hand-crafted features, which can be time-consuming and expensive to create. Third, existing systems are often not able to handle the nuances of human language, such as sarcasm and irony [7]. Fourth they usually can not deal with different languages [8, 9].

We used state-of-the-art approaches for the subjective-objective text classification task in this paper. In our work, we used fine-tuned XLM-RoBERTa large [10], BERT [11] and BERT multilingual(BERT-m) which are attention models that has been proven effective for a number of natural language processing (NLP) tasks. Our work is able to learn the underlying features of text without the need for hand-crafted features. Additionally, our work is able to handle multiple human languages. We perform our study on a dataset of news articles in six languages: Arabic, Dutch, German, Italian, German, and Turkish. The dataset also contained a multi-lingual subset which was created in a combination of all six languages[12]. The dataset was offered by CLEF2023 Task 2, and it contains a thousands news articles, each of which has been labeled as subjective or objective [13]. Our study uses the transformer-based XLM-RoBERTa large pretrained model to achieve excellentresults in all six languages and multilingual data. On the Arabic dataset, our work achieved the joint 1^{st} position in the leaderboard with a macro F1 score of 0.79. On the Dutch dataset, our work achieved the 2^{nd} position in the leaderboard with a macro F1 score of 0.76. On the English dataset, our work achieved the 5^{th} position in the leaderboard with a macro F1 score of 0.73. On the German dataset, our work achieved the joint 2nd position in the leaderboard with a macro F1 score of 0.74. On the Italian dataset, our work achieved the 2^{nd} position in the leaderboard with a macro F1 score of 0.71. On the Turkish dataset, our work achieved the 3^{rd} position in the leaderboard with a macro F1 score of 0.81. Lastly, On the multi-lingual dataset, our work achieved the 1^{st} position in the leaderboard with a macro F1 score of 0.82. Originally we used XLM-RoBERTa large for shared task 2 offered in CLEF2023 and run all the experiments with train set and dev set. Once the test set was published, we subsequently ran all of the experiments again using test set. While re-running all the experiments we used XLM-RoBERTa large as well as the BERT and BERT-m models to compare their performance on datasets of the shared task 2. These findings imply that our method is an effective way of classifying texts into those that are subjective and those that are objective. In addition to the above, our work has a number of other advantages. Our system is able to learn the underlying features of text without the need for hand-crafted features. This makes our system more scalable and generalizable to new types of text.

The rest of the paper is organized as follows: In the Related Works section, we list all comparable works, in the Experimental Methodology section, we list the methodology we employed, in the Experimental Analysis section, we offer a thorough analysis of our work, and in the Conclusion and Future Work section, we summarize our findings, as well as outline potential future prospects.

2. Related Works

Due to the urgent necessity to counteract misinformation and deception, academics from all around the world have paid close attention to subjectivity checking and fact checking. Numerous studies have been performed to create efficient methods and procedures for correctly recognizing subjective claims and confirming the veracity of factual information. However, the dynamic nature of online information and the advent of fresh platforms and channels for communication mean that there are constantly new problems to be solved. As a result, there is a constant need for new research and invention to improve the subjectivity checking and fact checking systems' precision, effectiveness, and scalability. Researchers can help to the creation of solid and trustworthy techniques to guarantee the integrity of information in an increasingly digital and interconnected world by continuing to study this area. Moreover the study conducted by many researchers have discussed the challenges of subjectivity and sentiment analysis, two important tasks in natural language processing [14, 15]. They identified several challenges that need to be addressed, for instance, lack of materials for languages other than English, resource reliance on social and cultural factors, and the need to consider the dialog structure and topic modeling in social media. They concludes that there are still many challenges to be overcome, but it will become increasingly important for the business community to find solutions involving these tasks..

The recognition of both subjective genres, like editorials, and objective genres, like business or news, has been examined in earlier works on genre classification [16]. Since subjective and objective sentences can coexist in most documents, it has been discovered that classifying subjectivity at the sentence level rather than the document level is more useful. For example, even though it's commonly accepted that newspaper stories are objective, 44% of the sentences in a news collection were found to be subjective. Sentence-level subjectivity classification proves valuable in extracting intricate subjectivity details, such as expressions of opinions, identification of opinion holders, and establishing relationships between opinions.

In the domain of opinion piece recognition, Wiebe et al. presented a study that focuses on utilizing subjective language cues to enhance the performance of opinion piece recognition[17]. The authors discovered the significance of various factors, such as unique words, collocations, adjectival and verbal clues, in effectively identifying subjective language. Additionally, they determined that the density of subjective clues in the surrounding context plays a crucial role in determining the subjectivity of a word. By employing the k-nearest-neighbor classification algorithm and implementing leave-one-out cross-validation, they attained an impressive classification accuracy of 94% on a substantial test dataset. This achievement demonstrated a significant 28% decrease in error compared to the baseline. Riloff et al. looked into a number of ways to advance the field of subjectivity analysis [18]. First, they demonstrated how a significant amount of labeled training data for upcoming learning algorithms may be produced via high-precision subjective categorization. Second, they demonstrated that linguistically richer subjective expressions can be learned using an extraction pattern learning technique than can be learned using single words or set phrases. Third, they added these newly identified extraction patterns to their initial high-precision subjective classifier. This bootstrapping procedure produced significantly higher recall with barely any precision loss.Wang et al. proposed a semi-supervised learning technique known as self-training for classifying sentence subjectivity [19]. In their work, they incorporated decision tree classifiers and adapted a selection metric called VDM. Experimental results on the MPQA corpus demonstrated that self-training with NBTree and VDM yielded better performance compared to other combinations. For the subjective classification of sentences, it produced outcomes similar to supervised learning models. Additionally, Hajj et al. proposed an affective computing framework for classifying sports articles as either objective or subjective [20]. The framework comprises several steps, including article extraction, POS tagging, feature reduction using a modified cortical algorithm (CA*), and classification utilizing multiple classifiers like SVM, LMSVM, and CA*. By applying feature reduction techniques, the framework achieved improved accuracy by eliminating around 40% of the features. Notably, among all the tested classifiers, CA* exhibited the highest accuracy, reaching 85.6%.

Subjectivity classification has traditionally been tackled using basic classification algorithms like linear classifiers, Support Vector Machines, and others. However, recent advancements have shifted the focus towards Transformer-based models. A BERT-based multitask learning (MTL) system that integrates sentiment and subjective detection was introduced by Satapathy et al. [21] The suggested system makes use of a Neural Tensor Network to improve task performance as well as a multilayer multitask LSTM for polarity and subjectivity detection. Results indicate a 2-4% improvement in subjectivity detection. The experiments also show that the MTL setup outperforms BERT-based individual task embeddings. In addition, they claimed that polarity and subjectivity detection are related tasks even when trained on different datasets. Similarly, Pant et al. conducted an extensive experiments on the Wiki Neutrality Corpus (WNC) and focuses on the detection of subjective bias using BERT-based models [22]. The suggested BERT-based ensembles perform significantly better than cutting-edge techniques (5.6 F1 score difference). The BERT-based ensemble, which consists of RoBERTa, ALBERT, DistillRoBERTa, and BERT, is the best-performing architecture, according to the study, which looks into sentence-level subjective bias detection on a huge Wikipedia corpus.

The use of transformer-based models has seen a significant increase across various domains, including fact checking, fake news detection, along with subjectivity classification [23, 24, 25]. The flexibility and agility of these models make them appropriate for a variety of Natural Language Processing tasks. For instance a methodology for identifying previously fact-checked assertions was described in the proposed system by Pritzkau et al. [26]. The task was calculating the degree to which a given claim and a group of fact-checked statements are comparable. The method relies on RoBERTa, a modified form of BERT, and specifically makes use of BERT. The results shown how dependent the system was on semantic similarity, and performance was affected by the comparison of semantic and lexical similarity. Huertas-Garca et al. suggested a novel method to identify false information and categorize news themes by combining Doc2Vec with several cutting-edge transformer-based models [27]. Doc2Vec and transformer-based models work together to produce performance that is superior to either strategy by itself. They oversampled the unbalanced data to enhance the performance of the classifiers and employed a sliding window method to get over the sequence length restrictions imposed by transformer-based models. They completed the CLEF 2021's shared classification problem with a macro-average F1-score of 67.65% and the misinformation detection job with a macro-average F1-score of 41.43%.

Over the past few years, the CLEF CheckThat! laboratories have featured shared tasks focused on automatically identifying and verifying assertions, subjectivity in political debates and tweets and these challenges have led to significant findings in the field of assertion identification and verification [28, 29, 30, 31, 32, 33]. Despite the advancements made in subjectivity detection and classification, there is still significant untapped potential in utilizing transformer-based models for the classification of subjectivity in news articles. The use of transformer models, such as BERT and RoBERTa, has revolutionized natural language processing tasks and has shown promising results in various domains.

Table 1
Data Split and Distribution

Class label	Train	Dev	Test	Total
Multi-lingual				
OBJ	4371	300	300	4971
SUBJ	2257	300	300	2857
Total	6628	600	600	7828
Arabic				
OBJ	905	227	363	1495
SUBJ	280	70	82	432
Total	1185	297	445	1927
Dutch				
OBJ	489	107	263	859
SUBJ	311	93	237	641
Total	800	200	500	1500
English				
OBJ	532	106	116	754
SUBJ	298	113	127	538
Total	830	219	243	1292
German				
OBJ	492	123	194	809
SUBJ	308	77	97	482
Total	800	200	291	1291
Italian				
OBJ	1231	167	323	1721
SUBJ	382	60	117	559
Total	1613	227	440	2280
Turkish				
OBJ	422	100	111	633
SUBJ	378	100	129	607
Total	800	200	240	1240

3. Experimental Methodology

3.1. Data

Text sequences in the following six languages make up the dataset offered for CLEF 2023 task 2 Subjectivity in News Articles: Arabic, Dutch, English, German, Italian, and Turkish. There is also a multilingual dataset made up of all six languages. This binary classification task's goal is to identify if a sentence or paragraph conveys the author's subjective position or offers an objective stance on the subject at hand. For researchers and programmers working on natural language processing tasks, this dataset offers a useful resource because it enables them to train and test their systems across a variety of languages. In order to ensure consistency and dependability across languages, Ruggeri et al. present a standardized framework for annotating subjective and objective text sequences [34]. The dataset details are presented in Table 1, providing a comprehensive description of the dataset.

Table 2
The official evaluation results and the overall ranking for Task 2: Subjectivity in News Articles

Language	Model	Macro F1	F1 (SUBJ class)	Rank
Multi-lingual	XLM-RoBERTa large	0.82	0.81	1st
Arabic	XLM-RoBERTa large	0.79	0.67	Joint 1st
Dutch	XLM-RoBERTa large	0.76	0.71	2nd
English	XLM-RoBERTa large	0.73	0.73	5th
German	XLM-RoBERTa large	0.74	0.67	Joint 2nd
Italian	XLM-RoBERTa large	0.71	0.58	2nd
Turkish	XLM-RoBERTa large	0.81	0.80	3rd

3.2. Data Pre-Processing and Cleaning

The datasets for CLEF 2023 problem 2 were compiled using information scraped from Twitter, where symbols, URLs, and invisible letters are present as noise. We used a number of pre-processing approaches to clear the data. We started by removing URLs and extraneous letters from the text. Then, we removed any common stopwords that didn't add anything to the meaning. Then, we eliminated usernames and hashtag signs to further clean up the dataset for analysis and modeling. Although the thorough pre-processing step did not particularly improve the classification results. Finally, to break up a sentence into words or tokens, we utilized the word tokenize method from nltk package.

3.3. Models

We employed transformer-based models like BERT [11] large uncased, BERT multilingual(BERT-m) and XLM-RoBERTa large [10] large to evaluate the subjectivity of the text data across all six languages and the multilingual dataset.These algorithms are renowned for their capacity to properly capture contextual data, allowing precise forecasts of the objectivity of news stories. Through exposure to labeled data and iterative weight adjustments, the models' parameters were optimized during the training process to reduce prediction errors. We aimed to improve the ability of these robust language models to distinguish between subjective and objective textual content in a wide range of languages, enabling more thorough and dependable analysis of news articles across various linguistic contexts. We fine-tuned these models so they could perform binary classification tasks. We used Learning rate of 2e-5 and the batch size was 16 for each of

Table 3
Comprehensive Breakdown of the Classification Results. Bold numbers indicate the highest F1 score achieved between XLM-RoBERTa large and BERT.

Class label	Model	Accuracy	Precision	Recall	F1 Score
		Multi-lingual			
OBJ	XLM-RoBERTa large	0.820	0.80	0.86	**0.83**
SUBJ			0.85	0.78	0.81
OBJ	BERT-m	0.751	0.70	0.87	0.78
SUBJ			0.83	0.64	0.72
		Arabic			
OBJ	XLM-RoBERTa large	0.856	0.95	0.87	**0.91**
SUBJ			0.58	0.78	0.67
OBJ	BERT	0.782	0.82	0.93	0.87
SUBJ			0.27	0.11	0.16
		Dutch			
OBJ	XLM-RoBERTa large	0.764	0.72	0.90	**0.80**
SUBJ			0.85	0.61	0.71
OBJ	BERT	0.61	0.59	0.81	0.69
SUBJ			0.65	0.38	0.48
		English			
OBJ	XLM-RoBERTa large	0.728	0.70	0.77	0.73
SUBJ			0.77	0.69	0.73
OBJ	BERT	0.716	0.64	0.91	**0.75**
SUBJ			0.87	0.54	0.66
		German			
OBJ	XLM-RoBERTa large	0.763	0.85	0.79	**0.82**
SUBJ			0.63	0.71	0.67
OBJ	BERT	0.659	0.72	0.81	0.765
SUBJ			0.49	0.36	0.41
		Italian			
OBJ	XLM-RoBERTa large	0.766	0.85	0.82	**0.84**
SUBJ			0.55	0.62	0.58
OBJ	BERT	0.707	0.74	0.91	0.82
SUBJ			0.36	0.14	0.20
		Turkish			
OBJ	XLM-RoBERTa large	0.812	0.74	0.93	**0.82**
SUBJ			0.92	0.71	0.80
OBJ	BERT	0.683	0.62	0.81	0.70
SUBJ			0.78	0.57	0.66

the subtasks. We conducted training using the XLM-RoBERTa large for 3 epochs each for Arabic, Dutch, English, German, Italian, and Turkish languages. For the multi-lingual dataset, we ran for 5 epochs to capture specific language nuances. Additionally, we utilized the BERT large uncased model for Arabic, Dutch, English, German, Italian, and Turkish, running it for 3 epochs, while for the multi-lingual dataset dataset, we extended training to 8 epochs where we employed the BERT multilingual model. The model achieved good convergence and generalization on the test data after only three epochs, except for the multilingual dataset. XLM-RoBERTa large had a total of 559,892,482 trainable parameters, while BERT large uncased had 335,143,938 parameters. Additionally, BERT-m consisted of a total of 177,854,978 parameters.

This comprehensive training approach aimed to optimize the models' subjectivity detection capabilities across the various languages in the dataset.

4. Experimental Analysis

The performance of our work for Task 2: Subjectivity in News Articles is displayed in Table 2. The table presents the macro F1 scores and F1 scores, specifically for the SUBJ class. These results showcase how effectively our strategy performed on the official test set. The table also includes the overall ranking, highlighting our position compared to other competitors. The outcomes validate the resilience and accuracy of our method, demonstrating its ability to assess subjectivity across multiple languages.

Table 3 presents a comprehensive breakdown of the classification results for each language. After obtaining the test labels, we re-ran all the experiments and updated the results accordingly. By utilizing XLM-RoBERTa large, we achieved F1 scores of 0.83, 0.91, 0.80, 0.82, 0.84, and 0.82 for the Multilingual, Arabic, Dutch, German, Italian, and Turkish datasets, respectively, for the OBJ class. These scores surpass those achieved by the BERT and BERT-m models for the same datasets. However, for the English dataset, the BERT models attained a higher F1 score of 0.75 for the OBJ class, while achieved a score of 0.73.

The table 3 also reveals that XLM-RoBERTa large, BERT and BERT-m perform better in terms of F1 score for the OBJ class compared to the SUBJ class in all languages. Moreover, XLM-RoBERTa large outperforms BERT and BERT-m in terms of accuracy for all languages. Although the BERT model produced a better F1 score than for the English dataset, still surpasses BERT in terms of accuracy.

The table 3 provides clear evidence that XLM-RoBERTa large outperforms BERT and BERT-m for Arabic, Dutch, English, German, Italian, Turkish, and Multilingual text datasets. Furthermore, XLM-RoBERTa large demonstrates better performance in handling long-range dependencies compared to BERT due to its self-attention mechanism.

5. Conclusion and Future Work

In this research, we used transformer-based models, fine-tuned XLM-RoBERTa large, BERT and BERT-m, to analyze the subjectivity of news articles. We conducted a comparative analysis of these models to evaluate their performance. Our investigation used a dataset of news articles from the CLEF2023 shared task, which included six different languages and a multilingual dataset.

Our findings show that XLM-RoBERTa large outperforms BERT and BERT-m in effectively detecting subjective news articles. Additionally, we observed that XLM-RoBERTa large exhibited superior performance across multiple languages, highlighting its versatility and effectiveness in subjective news classification tasks. In order to construct more reliable models, we would like to perform deeper research with a larger dataset in the future. Our objective is to leverage a diverse range of machine learning and deep learning algorithms to create models that effectively classify the subjectivity of text. By doing so, we aim to make significant advancements in the field of subjectivity classification and contribute to existing research efforts.

References

[1] A. Banfield, Unspeakable sentences (routledge revivals): narration and representation in the language of fiction, Routledge, 2014.

[2] K. Vis, J. Sanders, W. Spooren, Diachronic changes in subjectivity and stance–a corpus linguistic study of dutch news texts, Discourse, Context & Media 1 (2012) 95–102.

[3] G. Pounds, Attitude and subjectivity in italian and british hard-news reporting: The construction of a culture-specific 'reporter'voice, Discourse studies 12 (2010) 106–137.

[4] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679.

[5] W. Medhat, A. Hassan, H. Korashy, Sentiment analysis algorithms and applications: A survey, Ain Shams engineering journal 5 (2014) 1093–1113.

[6] J. Serrano-Guerrero, J. A. Olivas, F. P. Romero, E. Herrera-Viedma, Sentiment analysis: A review and comparative analysis of web services, Information Sciences 311 (2015) 18–38.

[7] A. Rajadesingan, R. Zafarani, H. Liu, Sarcasm detection on twitter: A behavioral modeling approach, in: Proceedings of the eighth ACM international conference on web search and data mining, 2015, pp. 97–106.

[8] K. Dashtipour, S. Poria, A. Hussain, E. Cambria, A. Y. Hawalah, A. Gelbukh, Q. Zhou, Multilingual sentiment analysis: state of the art and independent comparison of techniques, Cognitive computation 8 (2016) 757–771.

[9] K. Denecke, Using sentiwordnet for multilingual sentiment analysis, in: 2008 IEEE 24th international conference on data engineering workshop, IEEE, 2008, pp. 507–512.

[10] 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, CoRR abs/1911.02116 (2019). URL: http://arxiv.org/abs/1911.02116. arXiv:1911.02116.

[11] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[12] A. Galassi, F. Ruggeri, A. B.-C. no, F. Alam, T. Caselli, M. Kutlu, J. M. Struss, F. Antici, M. Hasanain, J. Köhler, K. Korre, F. Leistra, A. Muti, M. Siegel, M. D. Turkmen, M. Wiegand, W. Zaghouani, Overview of the CLEF-2023 CheckThat! lab task 2 on subjectivity in news articles, in: Working Notes of CLEF 2023–Conference and Labs of the Evaluation Forum, CLEF '2023, Thessaloniki, Greece, 2023.

[13] A. Barrón-Cedeño, F. Alam, T. Caselli, G. Da San Martino, T. Elsayed, A. Galassi, F. Haouari,

F. Ruggeri, J. M. Struß, R. N. Nandi, G. S. Cheema, D. Azizov, P. Nakov, The clef-2023 checkthat! lab: Checkworthiness, subjectivity, political bias, factuality, and authority, in: J. Kamps, L. Goeuriot, F. Crestani, M. Maistro, H. Joho, B. Davis, C. Gurrin, U. Kruschwitz, A. Caputo (Eds.), Advances in Information Retrieval, Springer Nature Switzerland, Cham, 2023, pp. 506–517.

[14] A. Montoyo, P. Martínez-Barco, A. Balahur, Subjectivity and sentiment analysis: An overview of the current state of the area and envisaged developments, Decision Support Systems 53 (2012) 675–679. URL: https://www.sciencedirect.com/science/article/pii/S0167923612001339. doi:https://doi.org/10.1016/j.dss.2012.05.022, 1) Computational Approaches to Subjectivity and Sentiment Analysis 2) Service Science in Information Systems Research : Special Issue on PACIS 2010.

[15] A. Balahur, R. Mihalcea, A. Montoyo, Computational approaches to subjectivity and sentiment analysis: Present and envisaged methods and applications, 2014.

[16] J. Wiebe, T. Wilson, M. Bell, Identifying collocations for recognizing opinions, in: Proceedings of the ACL-01 Workshop on Collocation: Computational Extraction, Analysis, and Exploitation, Citeseer, 2001, pp. 24–31.

[17] J. Wiebe, T. Wilson, R. Bruce, M. Bell, M. Martin, Learning subjective language, Computational linguistics 30 (2004) 277–308.

[18] 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.

[19] B. Wang, B. Spencer, C. X. Ling, H. Zhang, Semi-supervised self-training for sentence subjectivity classification, in: Advances in Artificial Intelligence: 21st Conference of the Canadian Society for Computational Studies of Intelligence, Canadian AI 2008 Windsor, Canada, May 28-30, 2008 Proceedings 21, Springer, 2008, pp. 344–355.

[20] N. Hajj, Y. Rizk, M. Awad, A subjectivity classification framework for sports articles using improved cortical algorithms, Neural Computing and Applications 31 (2019) 8069–8085.

[21] R. Satapathy, S. R. Pardeshi, E. Cambria, Polarity and subjectivity detection with multitask learning and bert embedding, Future Internet 14 (2022) 191.

[22] K. Pant, T. Dadu, R. Mamidi, Towards detection of subjective bias using contextualized word embeddings, in: Companion Proceedings of the Web Conference 2020, 2020, pp. 75–76.

[23] I. Baris Schlicht, A. Magnossão de Paula, P. Rosso, UPV at CheckThat! 2021: Mitigating cultural differences for identifying multilingual check-worthy claims, 2021.

[24] E. Williams, P. Rodrigues, S. Tran, Accenture at CheckThat! 2021: Interesting claim identification and ranking with contextually sensitive lexical training data augmentation, 2021.

[25] S. Kumari, NoFake at CheckThat! 2021: Fake news detection using BERT, 2021.

[26] A. Pritzkau, NLytics at CheckThat! 2021: Detecting previously fact-checked claims by measuring semantic similarity, in: CLEF (Working Notes), 2021, pp. 582–591.

[27] Álvaro Huertas-Garcıia, J. Huertas-Tato, A. Martín, D. Camacho, CIVIC-UPM at CheckThat! 2021: Integration of transformers in misinformation detection and topic classification, in: CLEF (Working Notes), 2021, pp. 520–530.

[28] P. Nakov, A. Barrón-Cedeño, T. Elsayed, R. Suwaileh, L. Màrquez, W. Zaghouani,

P. Atanasova, S. Kyuchukov, G. Da San Martino, Overview of the clef-2018 checkthat! lab on automatic identification and verification of political claims, in: P. Bellot, C. Trabelsi, J. Mothe, F. Murtagh, J. Y. Nie, L. Soulier, E. SanJuan, L. Cappellato, N. Ferro (Eds.), Experimental IR Meets Multilinguality, Multimodality, and Interaction, Springer International Publishing, Cham, 2018, pp. 372–387.

[29] T. Elsayed, P. Nakov, A. Barrón-Cedeño, M. Hasanain, R. Suwaileh, G. D. S. Martino, P. Atanasova, Overview of the clef-2019 checkthat!: Automatic identification and verification of claims, arXiv preprint arXiv:2109.15118 (2021).

[30] T. Elsayed, P. Nakov, A. Barrón-Cedeño, M. Hasanain, R. Suwaileh, G. Da San Martino, P. Atanasova, Checkthat! at clef 2019: Automatic identification and verification of claims, in: L. Azzopardi, B. Stein, N. Fuhr, P. Mayr, C. Hauff, D. Hiemstra (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2019, pp. 309–315.

[31] A. Barrón-Cedeno, T. Elsayed, P. Nakov, G. Da San Martino, M. Hasanain, R. Suwaileh, F. Haouari, Checkthat! at clef 2020: Enabling the automatic identification and verification of claims in social media, in: Advances in Information Retrieval: 42nd European Conference on IR Research, ECIR 2020, Lisbon, Portugal, April 14–17, 2020, Proceedings, Part II 42, Springer, 2020, pp. 499–507.

[32] S. Shaar, F. Haouari, W. Mansour, M. Hasanain, N. Babulkov, F. Alam, G. Da San Martino, T. Elsayed, P. Nakov, Overview of the clef-2021 checkthat! lab task 2 on detecting previously fact-checked claims in tweets and political debates., in: CLEF (Working Notes), 2021, pp. 393–405.

[33] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, H. Mubarak, A. Nikolov, Overview of the clef-2022 checkthat! lab task 1 on identifying relevant claims in tweets (2022).

[34] F. Ruggeri, F. Antici, A. Galassi, K. Korre, A. Muti, A. Barrón-Cedeño, On the definition of prescriptive annotation guidelines for language-agnostic subjectivity detection, in: Proceedings of Text2Story—Sixth Workshop on Narrative Extraction From Texts, held in conjunction with the 45th European Conference on Information Retrieval (ECIR 2023), volume 3370, CEUR-WS. org, 2023, pp. 103–111.