SSN-NLP at CheckThat! 2024: Assessing the Check-Worthiness of Tweets and Debate Excerpts Using Traditional Machine Learning and Transformer Models

Sanjai Balajee Kannan Giridharan[†], Sanjjit Sounderrajan[†], B Bharathi[†] and Nilu R. Salim[†]

Sri Sivasubramaniya Nadar College of Engineering, Chennai, Tamil Nadu, India

Abstract

The rapid spread of misinformation on social media necessitates efficient methods for determining whether claims in tweets or transcriptions warrant fact-checking. Traditional approaches rely on professional fact-checkers or human annotators, which are labor-intensive and time-consuming. This paper presents automated methods using machine learning and natural language processing to streamline check-worthiness estimation. We leveraged various techniques, including transformer models, to capture contextual nuances and improve prediction accuracy. Our work focused solely on the English language dataset, and our methods ranked 13th on the leaderboard. Our findings demonstrate the effectiveness of these automated methods, highlighting their potential to significantly enhance the efficiency of fact-checking systems and promote information integrity in digital communication.

Keywords: Check-Worthiness Estimation, Fact-Checking Automation, Natural Language Processing (NLP), PoS Tagging, Machine Learning Classifiers, Transformer Models

1. Introduction

In today's digital age, the rapid dissemination of information through social media platforms and online forums has led to an increase in the spread of misinformation and fake news. Addressing this challenge requires effective methods for identifying claims that warrant further investigation and fact-checking. Traditionally, the task of check-worthiness estimation has relied on professional fact-checkers or human annotators who assess the verifiability and potential harm of claims [1]. However, these manual processes are labor-intensive and time-consuming, highlighting the need for automated solutions.

Our research aims to automate the task of check-worthiness estimation by leveraging machine learning and natural language processing techniques. We utilize a multi-genre dataset comprising tweets and transcriptions to evaluate the effectiveness of different models across various linguistic and cultural contexts [2]. By employing advanced algorithms and transformer models, we aim to enhance the accuracy and efficiency of check-worthiness estimation.

In this paper, we provide a comprehensive overview of the check-worthiness estimation task, emphasizing its significance in combating misinformation and promoting information integrity. We discuss the methodologies employed in existing approaches, including traditional machine learning algorithms and transformer-based models, and propose avenues for future research and model development. Our contributions seek to advance automated fact-checking systems that can effectively identify and flag potentially misleading or false claims in textual content, fostering a more informed and trustworthy information ecosystem [3].

CLEF 2024: Conference and Labs of the Evaluation Forum, September 09–12, 2024, Grenoble, France

*Corresponding author.

[†]These authors contributed equally.

sanjai2110173@ssn.edu.in (S. B. K. Giridharan); sanjjit2110378@ssn.edu.in (S. Sounderrajan); bharathib@ssn.edu.in (B. Bharathi); nilurs@ssn.edu.in (N. R. Salim)

https://www.ssn.edu.in/staff-members/dr-b-bharathi/ (B. Bharathi); https://www.ssn.edu.in/staff-members/nilu-r-salim/ (N. R. Salim)

0000-0003-3078-5470 (S. B. K. Giridharan); 0009-0008-0247-0475 (S. Sounderrajan); 0000-0001-7279-5357 (B. Bharathi); 0000-0001-6619-7027 (N. R. Salim)

2. Related Work

The task of check-worthiness estimation has gained considerable attention, particularly through the CLEF CheckThat! lab series. These efforts address the challenges of identifying claims in social media and other textual sources that warrant fact-checking. Nakov et al. [1] investigate the identification of check-worthy claims amidst the COVID-19 infodemic and the detection of fake news. Their research provides an in-depth examination of techniques for spotting misleading information on social media platforms.

The CLEF-2022 CheckThat! lab introduced additional tasks and datasets aimed at enhancing the automatic identification and verification of claims [4]. This included Task 1, which focused on identifying relevant claims in tweets [2].

Advancements have been made by incorporating machine learning and deep learning models. For example, Support Vector Machines (SVM) [5] and Random Forest [6] have been utilized effectively in classification tasks. Passive-Aggressive Classifiers have shown promise in fake news detection [7].

Transformer-based models have significantly advanced check-worthiness estimation. BERT [8], RoBERTa [9], XLM [10], and DeBERTa [11] have demonstrated high effectiveness in understanding complex linguistic patterns. Additionally, ensemble learning techniques have been identified as promising for improving model performance by combining the strengths of various algorithms [12].

These works collectively highlight the importance of leveraging machine learning and natural language processing techniques to automate the detection of claims that warrant fact-checking, thus contributing to the broader effort of combating misinformation and enhancing the reliability of information disseminated through social media.

3. Experiment Setup

3.1. Dataset Description

In this study, we utilized a dataset encompassing four languages: English, Spanish, Arabic, and Dutch, as released by the CLEF CheckThat! organizers. The dataset comprises sentence IDs, text snippets extracted from tweets, debates, or speech transcriptions, and a class label indicating whether a claim is check-worthy (Yes) or not (No). Table 1 presents the distribution of the dataset across the four languages.

Table 1
Dataset Distribution for English, Spanish, Dutch, and Arabic

Language	Label	Train + Dev	Dev-Test
English	YES	5651	108
	NO	17,882	210
Spanish	YES	3826	509
	NO	21,124	4491
Arabic	YES	2656	377
	NO	5910	128
Dutch	YES	509	318
	NO	1202	577

However, our research focuses exclusively on the **English** subset of the dataset.

3.2. Dataset Preprocessing

In this section, we describe the data preprocessing steps undertaken to prepare the dataset for training and evaluation. The preprocessing pipeline consists of several key stages, including text cleaning, tokenization, stopword removal, punctuation removal, URL removal, spelling correction, part-of-speech (POS) tagging, dependency parsing, and feature extraction.

3.3. Feature Extraction Methods

The feature extraction process begins with the linguistic analysis of text data using natural language processing (NLP) techniques. Initially, the input sentences are subjected to part-of-speech (POS) tagging and dependency parsing. POS tagging assigns grammatical categories to each word, distinguishing between parts of speech such as nouns, verbs, adjectives, etc., while dependency parsing reveals the syntactic relationships between words, delineating the structure of the sentence through dependencies like subject-verb or modifier-modified relationships.

Figure 1 shows the dependency relations

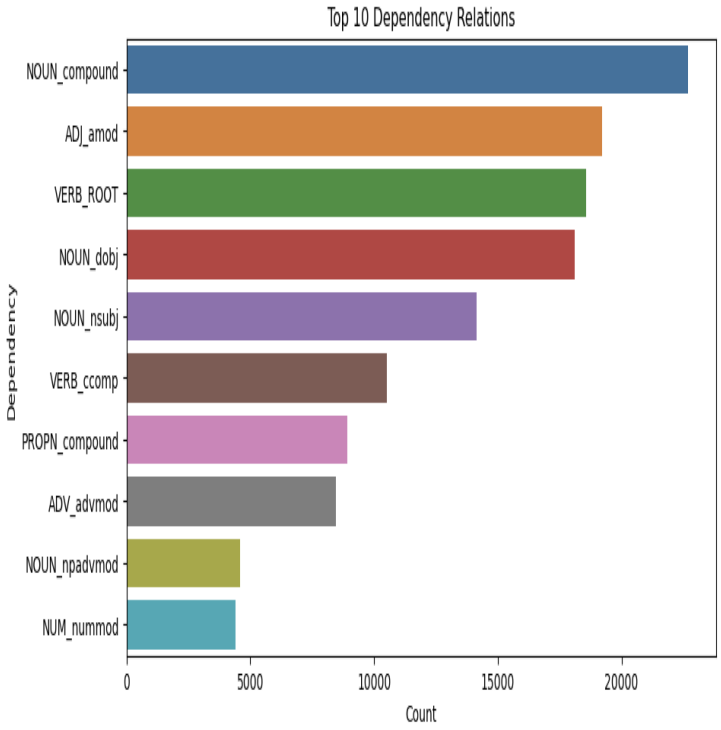

Figure 1: Distribution of the Top 10 Dependency Relations

Subsequently, these syntactic analyses are leveraged to extract relevant features capturing the linguistic structure of the text. Feature extraction involves converting the sequences of POS tags and dependency labels into meaningful representations. This process encompasses aggregating POS tags into vectors, capturing the distribution of different grammatical categories in the text, and encoding dependency relationships into feature vectors, emphasizing crucial syntactic dependencies.

Finally, the feature vectors are combined with sentence embeddings generated using a pre-trained transformer model such as Sentence-BERT. These embeddings capture the semantic content of the text at the sentence level, enabling the extraction of high-level semantic features. This combination allows for a richer and more nuanced understanding of the textual data. The combined feature representation undergoes data scaling to normalize the feature values and dimensionality reduction using principal component analysis (PCA) to reduce computational complexity and potentially enhance model performance.The resulting reduced-dimensional feature vectors serve as input to machine learning models.

Figure 2 illustrates the feature extraction pipeline.

3.4. Basic ML Models

In our experiment, we utilized various machine learning models with hyperparameters optimized using GridSearchCV. The models and their respective hyperparameters are as follows:

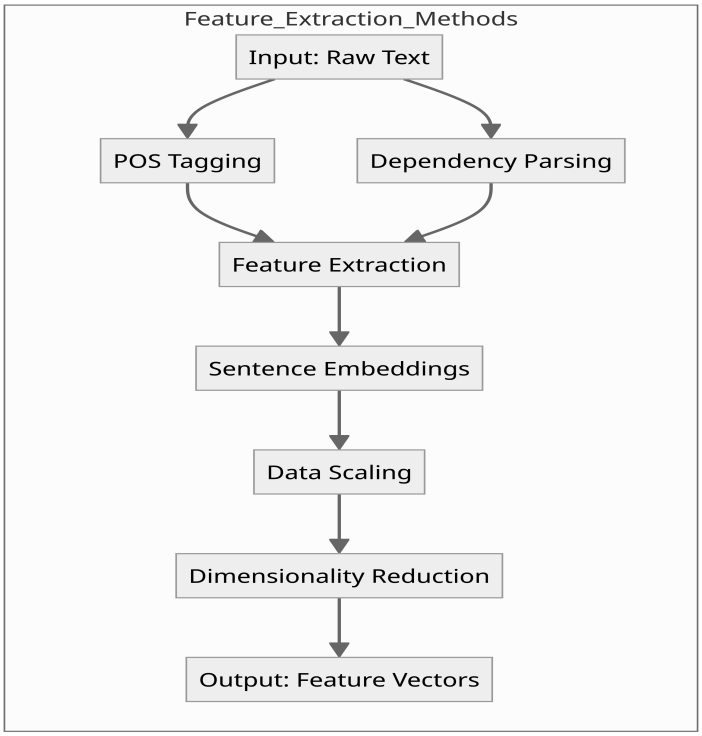

Figure 2: Feature Extraction Pipeline

- **Support Vector Machine (SVM)**: $C = 100$, $\gamma = 0.02$, *kernel*: rbf.
- **Random Forest Classifier**: *n_estimators*=300.
- **Logistic Regression**: $C = 0.1$, *solver*=liblinear.
- **XGBoost Classifier**: *learning_rate*=0.1, *max_depth*=6, *n_estimators*=1000.
- **CatBoost Classifier**: *depth*=5, *learning_rate*=0.05, *iterations*=1000.
- **K-Nearest Neighbors (KNN)**: *n_neighbors*=11, *metric*=euclidean.
- **Passive Aggressive Classifier**: $C = 0.01$.

The hyperparameters for each model were tuned using GridSearchCV to ensure optimal performance.

3.5. Transformer Models

In our experiment, we utilized several transformer models to evaluate their performance on our task. **BERT-base-uncased** is the original BERT model developed by Google in 2018. It is a pre-trained language model that uses a multi-layer bidirectional transformer encoder to generate contextualized representations of words in a sentence. **RoBERTa-base** is another variant of BERT developed by Facebook AI in 2019. It improves upon the original BERT model by using a different pre-training objective and a larger dataset. **XLM-RoBERTa-base** is a multilingual version of RoBERTa developed by researchers from the University of Montreal and Facebook AI in 2020. It is trained on a large corpus of text in multiple languages and can be fine-tuned for specific NLP tasks in any of these languages. **DeBERTa-v3-base** is a variant of BERT developed by researchers from Microsoft in 2021. It enhances BERT by using disentangled attention and ELECTRA-style pre-training, improving efficiency and performance on downstream tasks. Among these models, DeBERTa-v3-base demonstrated slightly better performance compared to the others.

The table shows the maximum sequence length, batch size, and learning rate used for each transformer model in our experiment (Table 2).

Model	Maximum Sequence Length	Batch Size	Learning Rate
BERT (bert-base-uncased)	128	32	2e-5
RoBERTa (roberta-base)	128	32	1e-5
XLM-RoBERTa (xlm-roberta-base)	128	32	3e-5
DeBERTa (deberta-v3-base)	128	32	2e-5

Table 2
Transformer Models and Hyperparameters

4. Results

For the challenge, the macro-average F1-score was employed as the official evaluation metric. This metric calculates the F1-score for each class individually and then computes the average of these scores.

The organizers provided three datasets for both languages: training, dev-test, and test. All models were trained using only the training dataset. Refer to the previous section for the hyperparameters used during training.

Table 3 summarizes the test F1 scores achieved by the models on the provided test dataset.

Table 3
Test F1 Scores of Models on dev-test-set

Model	Test F1-score
Majority Baseline	0.000
Random Baseline	0.462
N-gram Baseline	0.599
SVM	0.597
Random Forest Classifier	0.530
Logistic Regression	0.610
XGBoost Classifier	0.606
CatBoost Classifier	0.614
KNN	0.540
Passive Aggressive Classifier	0.641
MLP Classifier	0.410
BERT-base-uncased	**0.851**
RoBERTa-base	0.843
XLM-RoBERTa-base	0.847
DeBERTa-v3-base	**0.876**

Overall, the transformer models outperformed the traditional machine learning algorithms. DeBERTa-v3-base achieved the highest F1-score on the test dataset.

5. Conclusion

In this work, we present our participation in CLEF 2024 CheckThat-Lab Task 1: Check-worthiness Estimation in Text. Our evaluation showed the superiority of transformer models over traditional machine learning algorithms, as measured by the macro-average F1-score. This highlights the importance of using advanced transformer-based approaches for natural language processing tasks. Future research could explore fine-tuning strategies and alternative architectures to improve performance.

Our team, SSN-NLP, ranked 13th out of 27 teams on the leaderboard with a macro F1-score of 0.706, using the BERT model. The leaderboard results are summarized in Table 4, showing that BERT-base-uncased performed better than other models.

Table 4
Leaderboard Score

Team Name	Model	F1-score
SSN-NLP	BERT-base-uncased	0.706

6. Perspectives for Future Work

- **Ensemble Methods**: Combining multiple models to leverage their individual strengths can potentially lead to higher accuracy and robustness. Future work could explore various ensemble techniques, such as stacking, boosting, or voting, to improve overall performance.
- **Larger and More Diverse Datasets**: The availability of larger and more diverse datasets can significantly impact the generalizability of the models. Future studies should aim to collect and utilize datasets that encompass a wider range of topics, languages, and cultural contexts to train more versatile and robust models.
- **Cross-Lingual and Cross-Domain Transfer Learning**: Exploring transfer learning techniques to adapt models trained on one language or domain to other languages or domains can broaden the applicability of check-worthiness estimation models.

References

[1] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, The clef-2022 checkthat! lab on fighting the covid-19 infodemic and fake news detection, in: M. Hagen, S. Verberne, C. Macdonald, C. Seifert, K. Balog, K. Nørvåg, V. Setty (Eds.), Advances in Information Retrieval, Springer International Publishing, Cham, 2022, pp. 416–428.

[2] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, H. Mubarak, A. Nikolov, Y. S. Kartal, J. Beltrán, Overview of the CLEF-2022 CheckThat! lab task 1 on identifying relevant claims in tweets, in: Working Notes of CLEF 2022—Conference and Labs of the Evaluation Forum, CLEF '2022, Bologna, Italy, 2022.

[3] P. Gencheva, P. Nakov, L. Màrquez, A. Barrón-Cedeño, T. Mihaylova, A context-aware approach for detecting worth-checking claims in political debates, in: Proceedings of the International Conference Recent Advances in Natural Language Processing, RANLP 2017, 2017, pp. 267–276.

[4] P. Nakov, A. Barrón-Cedeño, G. Da San Martino, F. Alam, J. M. Struß, T. Mandl, R. Míguez, T. Caselli, M. Kutlu, W. Zaghouani, C. Li, S. Shaar, G. K. Shahi, H. Mubarak, A. Nikolov, N. Babulkov, Y. S. Kartal, J. Beltrán, M. Wiegand, M. Siegel, J. Köhler, Overview of the CLEF-2022 CheckThat! lab on fighting the COVID-19 infodemic and fake news detection, in: Proceedings of the 13th International Conference of the CLEF Association: Information Access Evaluation meets Multilinguality, Multimodality, and Visualization, CLEF '2022, Bologna, Italy, 2022.

[5] F. Rossi, N. Villa, Support vector machine for functional data classification, Neurocomputing 69 (2006) 730–742.

[6] A. Liaw, M. Wiener, Classification and regression by randomforest, R news 2 (2002) 18–22.

[7] S. Gupta, P. Meel, Fake news detection using passive-aggressive classifier, in: Inventive Communication and Computational Technologies: Proceedings of ICICCT 2020, Springer Singapore, 2021, pp. 155–164.

[8] 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).

[9] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, V. Stoyanov, Roberta: A robustly optimized bert pretraining approach, arXiv preprint arXiv:1907.11692 (2019).

[10] G. Lample, A. Conneau, Cross-lingual language model pretraining, arXiv preprint arXiv:1911.02116 (2019).

[11] P. He, X. Liu, J. Gao, Deberta: Decoding-enhanced bert with disentangled attention, arXiv preprint arXiv:2111.09543 (2021).

[12] S. Salvador, R. Francisco, J. Carmen, H. Olga, Ensemble learning: Insights for machine learning ensemble methods, in: Proceedings of CEUR Workshop, 2023, pp. 251–264.