ReDASPersuasion at ArAIEval Shared Task: Multilingual and Monolingual Models For Arabic Persuasion Detection

Fatima Zahra Qachfar
fqachfar@uh.edu
University of Houston
Houston, TX, USA

Rakesh M. Verma
rmverma2@central.uh.edu
University of Houston
Houston, TX, USA

Abstract

To enhance persuasion detection, we investigate the use of multilingual systems on Arabic data by conducting a total of 22 experiments using baselines, multilingual, and monolingual language transformers. Our aim is to provide a comprehensive evaluation of the various systems employed throughout this task, with the ultimate goal of comparing their performance and identifying the most effective approach. Our empirical analysis shows that *ReDASPersuasion* system performs best when combined with multilingual "XLM-RoBERTa" and monolingual pre-trained transformers on Arabic dialects like "CAMeLBERT-DA SA" depending on the NLP classification task.

1 Introduction

In recent years, the detection of persuasion techniques in text has gained a significant attention in research. Persuasion techniques can be used for either positive or negative ends. On one hand, persuasion can be used to convince people to support noble causes, promote social justice, and bring about positive change. However, these same techniques can also be exploited by individuals with ill intentions to manipulate and deceive others for personal gain or to perpetuate harmful beliefs and behaviors. Moreover, persuasion techniques can be employed with malicious intent including : 1. phishing scams, 2. propagandistic content (Barrón-Cedeño et al., 2019), 3. fallacy argumentation (Habernal et al., 2017, 2018), and 4. coercive extortion tactics.

With the increasing use of Arabic language in various forms of media, including social media, and news articles, it has become crucial to develop effective methods for identifying persuasive strategies in Arabic text. This task is challenging due to the complexities of the Arabic language, which includes various dialects, nuances, and cultural references (Glenn et al., 1977) that can affect the interpretation of persuasive elements. Researchers have employed various approaches, such as rhetoric methods (Koch, 1983), and deep learning models (Brahem et al., 2022), to automatically detect propaganda and persuasion in Arabic text. These techniques aim to identify specific linguistic features, such as sentiment analysis, and lexical semantics commonly used in persuasion.

Through this task, we have realized the importance of taking into consideration the different Arabic nuances and dialects in Multi-label and binary classification tasks. We also observe that Arabic writing styles vary immensely depending on the type of the data (paragraphs vs tweets) where *paragraphs* mainly use Modern Standard Arabic (MSA) or Classic Arabic (CA) while *tweets* contain a diversity of Arabic dialects using code-switching with other foreign languages and emojis.

We begin by providing an overview of the ArAiEval Shared-Task in Section 2. Next, we present a detailed description of the various systems utilized in our empirical study. We then delve into the preprocessing methods employed in Section 4, before presenting the results in Section 5. An error analysis is provided in Section 6, followed by a discussion section offering insights and perspectives on the task at hand. Finally, we summarize our findings and outline potential avenues for future research in Section 8.

2 Dataset and Tasks

Hasanain et al. (2023) organized the ArAIEval 2023 Shared-Task which includes two tasks in the Arabic language. The first task introduces *persuasion technique detection* while the second task introduces *disinformation detection* (Mubarak et al., 2023). Previously, Alam et al. (2022) described 20 propaganda techniques in the WANLP 2022 Shared Task adopting the same techniques as (Da San Martino et al., 2019) to Arabic news articles.

In this paper, we solely focus on the first task to investigate existing systems and enhance its per-

Proceedings of the The First Arabic Natural Language Processing Conference (ArabicNLP 2023), pages 549–557
December 7, 2023 ©2023 Association for Computational Linguistics

formance. For the persuasion technique detection task, the organizers offered two subtasks : i) *Task 1A*, and ii) *Task 1B*. We will describe these two subtasks in the following sections.

2.1 Task 1A : Binary Classification

This task involves classifying instances as either "true" or "false", where "true" indicates the presence of persuasion techniques in a given text, and "false" implies their absence. We report the class distribution in each subset (training, dev, and test) in Table 4 in the supplementary material.

2.2 Task 1B: Multi-Label Classification

The task involves assigning one or more labels from a predefined set, representing 23 types of persuasion techniques used in propaganda. Similarly, Piskorski et al. (2023) provided shared-task on a multilingual setting for multi-label classification. They have mapped these 23 techniques to six major categories (1. *Justification*, and 2. *Simplification*, and 3. *Distraction*, and 4. *Call*, and 5. *Manipulative wording*, and 6. *Attack on reputation*.). This is a multi-label classification problem where multiple propaganda techniques might be present in the same example. Samples with no persuasion technique are labeled with "no technique". We also report all the persuasion techniques in Table 5 in Appendix A.1.1.

3 Systems

We will describe thoroughly the different systems we used during this task for an end of comparing their performance and finding the best system.

3.1 Baseline Algorithms

For the baseline models, we implement a pipeline object that extracts the TF-IDF features and vectorizes textual content using unigram and bigram count vectorizer. We choose four traditional baseline algorithms (LR (Wright, 1995), RF (Breiman, 2001), XGB (Chen et al., 2015), and SVM(Cortes and Vapnik, 1995)).

We have defined a search space of hyperparameters using distributed hyperparameter optimization package "HyperOpt" [1] (Bergstra et al., 2013) with 5 trials 2 cross-validation splits. The baseline hyperparameter tuning include parameters like regularization strength (C), number of maximum iteration (*max_iter*), and the number of estimators (*n_estimators*). The best estimator is used to predict the testing set. We include the best hyperparameters in Appendix A.3

[1] https://github.com/hyperopt/hyperopt

3.2 ReDASPersuasion System

Qachfar and Verma (2023) present a multilingual system for persuasion detection on a total of five languages (En, Fr, Ru, It, Po, Ge). This system leverages the power of multilingual transformers "XLM-RoBERTa" and language agnostic features to perform persuasion detection across multiple languages.

The initial structure of the *ReDASPersuasion* system is composed of three main components:

- A multilingual transformer model that can process input in various languages.
- A feature engineering module that extracts language-agnostic features suitable for cross-lingual classification of persuasion.
- A multi-label classification module that combines the transformer output with persuasion features using a dropout layer, a dense linear layer, and a sigmoid activation function to produce multiple classification labels.

For task 1A, We modify this system to perform a binary classification task by using the sigmoid activation function with one output node while in task 1B, we used the sigmoid activation function with one node per persuasion class (23 techniques). We also change the criterion loss function from "*BCEWithLogitsLoss*" for Multilabel classification in Task 1B to "*CrossEntropy*" loss for binary classification in Task 1A.

To prevent the model from predicting a combination of "no technique" and other techniques, we treat samples with the "no technique" label as having no label at all.

4 Preprocessing

As illustrated in Figure 1, ArAiEval's first task persuasion dataset contains two data types:

1) ***Paragraph*:** a passage from news articles written in Modern Standard Arabic (MSA) which does not include code-switching or any specific keywords unlike tweets.

2) ***Tweet*:** a social media message written in diverse Arabic dialects mixed from different regions containing code-switching, specific Twitter keywords and emojis.

Arabic Tweet		RT @USER My message to Y'all!! 😂#كورونا #صباح_الخير LINK
Preprocessing Steps	*KT*	[إعادة_التغريد] [مستخدم] #صباح_الخير #كورونا😂 My message to Y'all!! [موقع_الكتروني]
	CS	[إعادة_التغريد] [مستخدم] رسالتي لكم جميعا!!😂 #صباح_الخير #كورونا [موقع_الكتروني]
	EC	[إعادة_التغريد] [مستخدم] رسالتي لكم جميعا!! :وجه_بدموع_فرح: #صباح_الخير #كورونا [موقع_الكتروني]

***KT** :Keyword Translation.* ***CS** : Code Switching.* ***EC** : Emoji Conversion.*

Table 1: Preprocessing Techniques Applied to Arabic Text

We describe three preprocessing techniques we applied to the tweets to translate code-switched text to Arabic. An example of these techniques are shown in Table 1.

4.1 Keyword Translation (KT)

In the "tweet" data type, we have certain keywords like retweet (RT), username (@USER), and website (LINK). We replace these terms with Arabic words using regular expressions, maintaining the proper right-to-left alignment of Arabic words.

4.2 Code Switching (CS)

Arabic tweeters may use code-switching to express themselves more effectively, or to communicate with a diverse audience. For example, a user may start a tweet in Arabic, switch to English in the middle, and then finish it off in French. For each tweet, we automatically detect code-switching fragments using "*Lingua*" [2] Python package, and we translate it to Arabic using Google's translation API.

4.3 Emoji Conversion (EC)

In tweets, emojis are typically used to convey emotions or ideas. Mubarak et al. (2022) showed the importance of emojis in the detection of Arabic offensive language and hateful speech.

Instead of removing all emojis from tweets like (Bennessir et al., 2022), we choose to convert them to Arabic descriptive text since emojis might hold meaning in the context of a short deceptive tweet representing positive or negative sentiment. For this we add Arabic language support to the "*emoji*" [3] Python package using normalized representations from the latest release of Unicode Common Locale Data Repository (CLDR) [4] to avoid broken Unicode. We create a dictionary of Arabic emoji representations based on the *emojiterra* website.[5]

[2] https://github.com/pemistahl/lingua

[3] https://github.com/carpedm20/emoji/

[4] https://github.com/unicode-org/cldr/raw/release-43/common/annotations/ar.xml

[5] https://emojiterra.com/copypaste/ar/

5 Experimental Results

We ran all classification experiments on a high performing cluster machine with an Intel® Xeon® Gold 6252 (3.70GHz) processor with 24 cores and 48 threads running Linux Red Hat Enterprise Server 8.6 with Nvidia® Volta V100 GPUs.

For task 1A, the initial structure of the *ReDASPersuasion* system with "XLM-RoBERTa" (Conneau et al., 2020) achieves the best F1-Micro score of 0.7336 on the test set.

For task 1B, the *ReDASPersuasion* system with "CAMeLBERT-DA SA" (Inoue et al., 2021) fine-tuned on sentiment analysis for Dialect Arabic (DA) achieves the best performance on the testing set with a F1-Micro score of 0.5584.

According to Table 3, the combination of *ReDASPersuasion* and "XLM-RoBERTa" yields the highest F1-score macro weighted strategy, with a value of 0.1449, for task 1B.

Our investigation reveals that during the development process, the *ReDASPersuasion* system powered by "XLM-RoBERTa" shows the most promise, with the *ReDASPersuasion* system using monolingual "CAMeLBERT-DA SA" coming in a close second. However, when it comes to the testing phase, one method excels in the first task, while the other method excels in the second task, as evidenced by their respective F1-Micro scores.

In Table 3 Task 1A, logistic regression, majority class baseline, and *ReDASPersuasion* system with "DistilBERT" all achieve an F1-score of 0.6581 which means these models fail to accurately predict the test set, as they simply assign the majority class "true" to all samples.

Due to the lack of visibility during the testing phase evaluation, we accidentally submitted wrong prediction results from the *ReDASPersuasion* system using "DistilBERT" instead of the intended top-performing *ReDASPersuasion* system using "XLM-RoBERTa". We have also encountered technical difficulties on the ArAIEval competition's hosting platform, *CodaLab*, mostly stemming from their HTTP backend server.

	Models	Task 1A Evaluation		Task 1B Evaluation	
		F1-score (Micro)	F1-score (Macro)	F1-score (Micro)	F1-score (Macro)
Baselines	LR (Wright, 1995)	0.7799	0.4382	0.4701	0.0393
	RF (Breiman, 2001)	0.7761	0.4687	0.4647	0.0582
	XGB (Chen et al., 2015)	0.7452	0.5971	0.4417	0.0951
	Linear SVM (Cortes and Vapnik, 1995)	0.7954	0.5076	0.5178	0.0699
	Random-Guess Baseline	0.5405	0.4774	0.0938	0.0573
	Majority-Class Baseline	0.7799	0.4382	0.4595	0.0337
ReDASPersuasion with Multilingual Transformers	mBERT (Devlin et al., 2019)	0.8263	0.6639	0.5922	0.1453
	DistilBERT (Sanh et al., 2020)	0.7992	0.6306	0.5658	0.1295
	XLM RoBERTa (Conneau et al., 2020)	**0.8764**	**0.8017**	**0.6454**	**0.1884**
ReDASPersuasion with Monolingual Transformers	AraBERT (Antoun et al.)	0.8340	0.7597	0.6064	0.1792
	MarBERT (Abdul-Mageed et al., 2021)	0.8224	0.7059	0.6249	0.1194
	CAMeLBERT-DA SA (Inoue et al., 2021)	0.7954	0.7001	0.6048	0.1594
	CAMeLBERT-MIX SA (Inoue et al., 2021)	0.8340	0.7030	0.6215	0.1813

Table 2: Evaluation Results on the Development Set using ArAiEval Scorer

	Models	Task 1A Evaluation		Task 1B Evaluation	
		F1-score (Micro)	F1-score (Macro)	F1-score (Micro)	F1-score (Macro)
Baselines	LR (Wright, 1995)	0.6581	0.3969	0.3629	0.0302
	RF (Breiman, 2001)	0.6600	0.4190	0.3585	0.0378
	XGB (Chen et al., 2015)	0.6640	0.5732	0.3275	0.0688
	Linear SVM (Cortes and Vapnik, 1995)	0.6600	0.4031	0.3760	0.0412
	Random-Guess Baseline	0.4771	0.4598	0.0868	0.0584
	Majority-Class Baseline	0.6581	0.3969	0.3599	0.0279
ReDASPersuasion with Multilingual Transformers	mBERT (Devlin et al., 2019)	0.6899	0.5656	0.4923	0.1083
	DistilBERT (Sanh et al., 2020)	0.6581	0.3969	0.4523	0.0568
	XLM RoBERTa (Conneau et al., 2020)	**0.7336**	0.6684	0.5555	**0.1449**
ReDASPersuasion with Monolingual Transformers	AraBERT (Antoun et al.)	0.7117	0.6967	0.5154	0.1344
	MarBERT (Abdul-Mageed et al., 2021)	0.7197	0.6826	0.5549	0.0988
	CAMeLBERT-DA SA (Inoue et al., 2021)	0.7217	**0.7007**	**0.5584**	0.1313
	CAMeLBERT-MIX SA (Inoue et al., 2021)	0.7217	0.6712	0.5565	0.1372

Table 3: Evaluation Results on the Testing Set using ArAiEval Scorer

6 Error Analysis

Another limitation we faced in this task is the imbalanced nature of the data and the small number of examples in certain persuasion techniques. For example, "*Appeal to Popularity*" persuasive technique occurs only twice in the training set and once in both the dev and test set as described in Table 5. Thus, the system was unable to accurately predict any of the labels for that particular class resulted in a zero F1 score, which had a negative impact on the overall performance in multi-label classification.

To provide a more detailed examination of the prediction errors, we present the confusion matrices for the top-performing models on both tasks in Figure 3, and Figure 4 in the supplementary material. As shown in Figure 4, "Name_Calling-Labeling" and "Loaded_Language" had the highest accuracy rates, whereas all other persuasive technique were inaccurately predicted. This can be attributed to the limited quantity of training data available for these categories.

7 Discussion

As shown in Figure 2, our Arabic dialect identification process reveals that the ArAiEval dataset encompasses a diverse array of dialects, with the most prominent languages being Saudi Arabian, Egyptian, and Palestinian dialects.

Different dialects have different vocabularies, and certain words or phrases might be interpreted differently in another dialect or deemed offensive. A concrete example would be the word "العافية" in Egyptian dialect means "health" while the same word means "fire" in Moroccan dialect. These differences in Arabic dialects can significantly impact persuasion strategies. An interesting take would be to consider the unique features of each dialect.

8 Conclusion

We described our systems for the two subtasks of the ArAiEval 2023 shared task on persuasion detection in Arabic to detect a total of 23 persuasion techniques for multi-label classification. We experiment with different combinations of multilingual and monolingual transformers. We have proven that the *ReDASPersuasion* model can benefit from both the multilingual "XLM-RoBERTa" transformer and the monolingual Dialect Arabic "CAMelBERT-DA SA" model depending on the NLP task. This task was an opportunity to evaluate the *ReDASPersuasion* model in depth and to conduct an error analysis to enhance our persuasion detection model for future works.

Limitations

Each dialect has its own unique features, such as vocabulary, grammar, and pronunciation, which can impact the way messages are conveyed and received by audiences. Therefore, one of the shortcomings of the *ReDASPersuasion* system is to detect persuasive words in the various Arabic dialects in classification.

Because of time constraints, we were unable to apply training data augmentation; however, we have translated the SemEval 2023 shared task (Piskorski et al., 2023) into Arabic text, which we will add to the imbalanced training dataset in future work to further analyze the behavior of the *ReDASPersuasion* system.

Ethics Statement

The dataset used in this paper, provided by the organizers (Hasanain et al., 2023), already adheres to data privacy regulations by eliminating all usernames and links from the tweets.

To ensure reproducible and ethical research, we provide models' hyperparameters used to achieve our experiments. Our work complies with the ACL Ethics Policy.

Acknowledgements

We would like to thank ArAiEval 2023 shared task organizers for their efforts in preparing Arabic datasets for propaganda detection including persuasive and disinformative cues.

We also thank the reviewers for their constructive feedback. This research was partly supported by NSF grants 1433817, 1950297, and 2210198, ARO grants W911NF-20-1-0254 and W911NF-23-1-0191, and ONR award N00014-19-S-F009. Verma is the founder of Everest Cyber Security and Analytics, Inc.

References

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2021. ARBERT & MARBERT: Deep bidirectional transformers for Arabic. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7088–7105, Online. Association for Computational Linguistics.

Firoj Alam, Hamdy Mubarak, Wajdi Zaghouani, Giovanni Da San Martino, and Preslav Nakov. 2022. Overview of the WANLP 2022 shared task on propaganda detection in Arabic. In *Proceedings of the Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 108–118, Abu Dhabi, United Arab Emirates (Hybrid). Association for Computational Linguistics.

Wissam Antoun, Fady Baly, and Hazem Hajj. Arabert: Transformer-based model for arabic language understanding. In *LREC 2020 Workshop Language Resources and Evaluation Conference 11–16 May 2020*, page 9.

Alberto Barrón-Cedeño, Israa Jaradat, Giovanni Da San Martino, and Preslav Nakov. 2019. Proppy: Organizing the news based on their propagandistic content. *Information Processing & Management*, 56(5):1849–1864.

Mohamed Aziz Bennessir, Malek Rhouma, Hatem Haddad, and Chayma Fourati. 2022. icompass at arabic hate speech 2022: Detect hate speech using qrnn and transformers. In *Proceedinsg of the 5th Workshop on Open-Source Arabic Corpora and Processing Tools with Shared Tasks on Qur'an QA and Fine-Grained Hate Speech Detection*, pages 176–180.

James Bergstra, Daniel Yamins, and David Cox. 2013. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *International conference on machine learning*, pages 115–123. PMLR.

Houda Bouamor, Nizar Habash, Mohammad Salameh, Wajdi Zaghouani, Owen Rambow, Dana Abdulrahim, Ossama Obeid, Salam Khalifa, Fadhl Eryani, Alexander Erdmann, et al. 2018. The madar arabic dialect corpus and lexicon. In *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018).*

Bechir Brahem, Hatem Haddad, et al. 2022. icompass at wanlp 2022 shared task: Arbert and marbert for multilabel propaganda classification of arabic tweets. In *Proceedings of the The Seventh Arabic Natural Language Processing Workshop (WANLP)*, pages 511–514.

Leo Breiman. 2001. Random forests. *Machine learning*, 45:5–32.

Tianqi Chen, Tong He, Michael Benesty, Vadim Khotilovich, Yuan Tang, Hyunsu Cho, Kailong Chen, Rory Mitchell, Ignacio Cano, Tianyi Zhou, et al. 2015. Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale.

Corinna Cortes and Vladimir Vapnik. 1995. Support-vector networks. *Machine learning*, 20:273–297.

Giovanni Da San Martino, Seunghak Yu, Alberto Barrón-Cedeño, Rostislav Petrov, and Preslav Nakov. 2019. Fine-grained analysis of propaganda in news article. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 5636–5646, Hong Kong, China. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. Bert: Pre-training of deep bidirectional transformers for language understanding.

E.S. Glenn, D. Witmeyer, and K.A. Stevenson. 1977. Cultural styles of persuasion. *International Journal of Intercultural Relations*, 1(3):52–66.

Ivan Habernal, Raffael Hannemann, Christian Pollak, Christopher Klamm, Patrick Pauli, and Iryna Gurevych. 2017. Argotario: Computational argumentation meets serious games. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 7–12, Copenhagen, Denmark. Association for Computational Linguistics.

Ivan Habernal, Patrick Pauli, and Iryna Gurevych. 2018. Adapting serious game for fallacious argumentation to german: Pitfalls, insights, and best practices. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, page in press, Miyazaki, Japan. European Language Resources Association (ELRA).

Maram Hasanain, Firoj Alam, Hamdy Mubarak, Samir Abdaljalil, Wajdi Zaghouani, Preslav Nakov, Giovanni Da San Martino, and Abed Alhakim Freihat. 2023. ArAIEval Shared Task: Persuasion Techniques and Disinformation Detection in Arabic Text. In *Proceedings of the First Arabic Natural Language Processing Conference (ArabicNLP 2023)*, Singapore. Association for Computational Linguistics.

Go Inoue, Bashar Alhafni, Nurpeiis Baimukan, Houda Bouamor, and Nizar Habash. 2021. The interplay of variant, size, and task type in Arabic pre-trained language models. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, Kyiv, Ukraine (Online). Association for Computational Linguistics.

Barbara Johnstone Koch. 1983. Presentation as proof: The language of arabic rhetoric. *Anthropological linguistics*, pages 47–60.

Hamdy Mubarak, Samir Abdaljalil, Azza Nassar, and Firoj Alam. 2023. Detecting and reasoning of deleted tweets before they are posted.

Hamdy Mubarak, Sabit Hassan, and Shammur Absar Chowdhury. 2022. Emojis as anchors to detect arabic offensive language and hate speech.

Jakub Piskorski, Nicolas Stefanovitch, Giovanni Da San Martino, and Preslav Nakov. 2023. Semeval-2023 task 3: Detecting the category, the framing, and the persuasion techniques in online news in a multilingual setup. In *Proceedings of the the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2343–2361.

Fatima Zahra Qachfar and Rakesh Verma. 2023. ReDASPersuasion at SemEval-2023 task 3: Persuasion detection using multilingual transformers and language agnostic features. In *Proceedings of the 17th International Workshop on Semantic Evaluation (SemEval-2023)*, pages 2124–2132, Toronto, Canada. Association for Computational Linguistics.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter.

Raymond E Wright. 1995. Logistic regression.

A Appendix

A.1 Data Distribution

We will describe the class and type distribution of the different subsets provided in Task 1 for persuasion detection.

A.1.1 Class Distribution

We observe that the binary class distribution is imbalanced of an approximately 1 to 4 True to False ratio.

Binary Class Distribution			
	Train	Dev	Test
True	1919	202	331
False	509	57	172

Table 4: Binary Labels in Arabic Task 1A

A.1.2 Data Type Distribution

As illustrated in Figure 1, most samples are categorized as paragraph data type, accounting for over 65% of the total samples in each subset. This introduces new challenges to the classification task where the structure of tweets and article paragraph news differ substantially.

Total Number of Persuasion Techniques			
	Train	Dev	Test
Appeal to Authority	48	5	14
Appeal to Fear Prejudice	108	12	15
Appeal to Hypocrisy	56	7	17
Appeal to Popularity	2	1	1
Appeal to Time	10	2	2
Appeal to Values	37	4	29
Causal Oversimplification	128	15	12
Consequential Oversimplification	33	3	3
Conversation Killer	28	3	7
Doubt	143	16	21
Exaggeration Minimisation	292	33	40
False Dilemma No choice	32	3	6
Flag Waving	63	7	25
Guilt by Association	13	1	1
Loaded Language	1574	176	253
Name Calling Labeling	692	77	133
Obfuscation Vagueness Confusion	240	28	25
Questioning Reputation	383	43	89
Red Herring	8	1	3
Repetition	25	3	6
Slogans	70	8	25
Straw Man	6	1	2
Whataboutism	9	1	2

Table 5: Persuasion Techniques in Arabic Task 1B

A.2 Dialect Language Identification

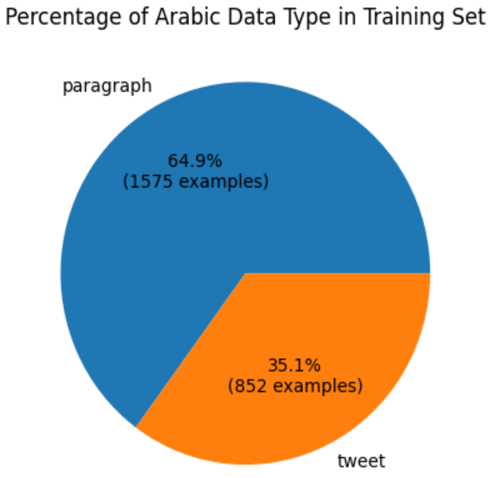
For Arabic dialect language detection, we used the bert-base-arabic model provided by CAMel (Computational Approaches to Modeling Language) Laboratory on the HuggingFace Hub [6] trained on MADAR (Bouamor et al., 2018) Twitter dataset which contains Arabic dialect tweets originating from 25 regions. In Figure 2, we observe that the top five Arabic dialects originate from Saudi Arabia, Egypt, Kuwait, Palestine, and Jordan throughout the training, dev, and test sets.

These different Arabic dialects from different regions have distinct grammatical structures, vocabularies, and idiomatic expressions that can be challenging to reconcile within one classification model. In this manner, we fine-tune the CAMELBERT-MIX SA model (Inoue et al., 2021) on our tasks which shows significant performance in predicting persuasive writing in Arabic text.

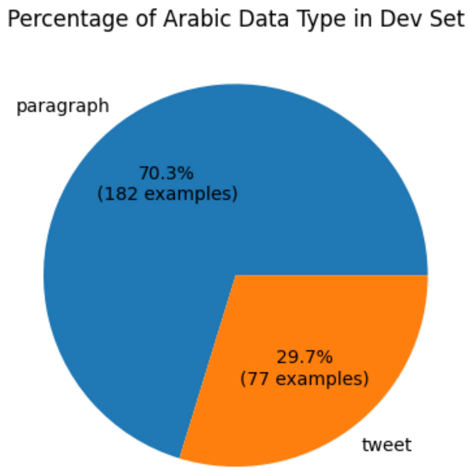
A.3 Model Hyperparameters

In this section, we report in Tables 6 , and 7 all the hyperparameters used in optimization for transformer and baseline models respectively.

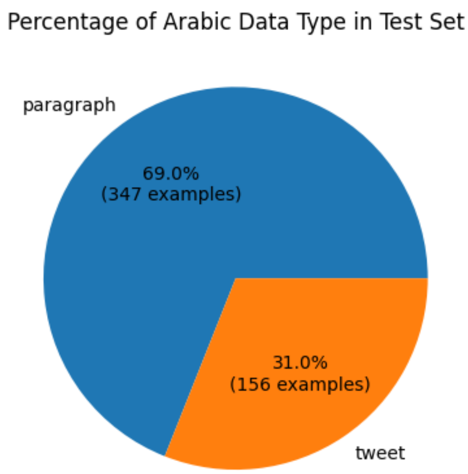
[6] https://huggingface.co/CAMeL-Lab/bert-base-arabic-camelbert-msa-did-madar-twitter5



(a) Training Set

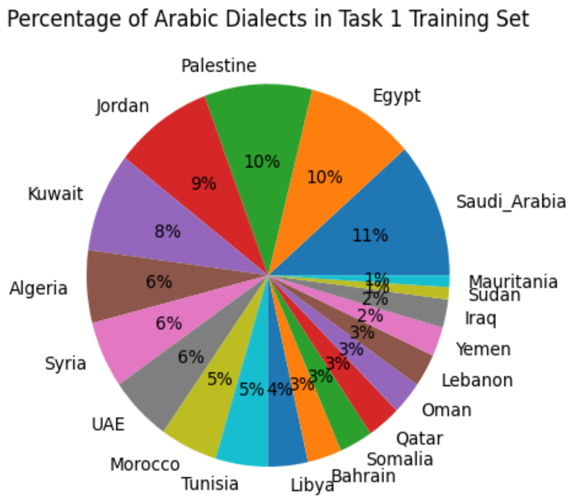


(b) Development Set

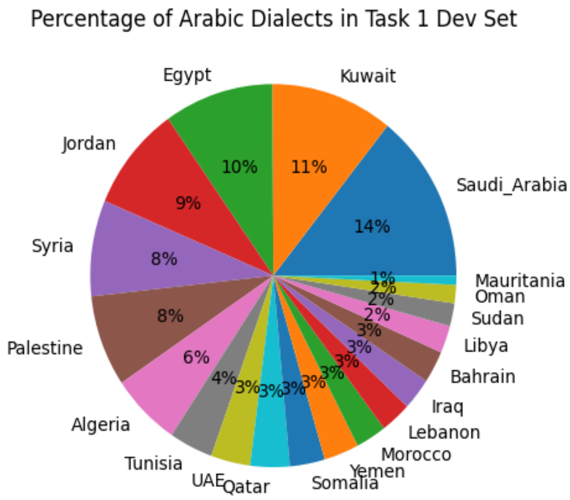


(c) Testing Set

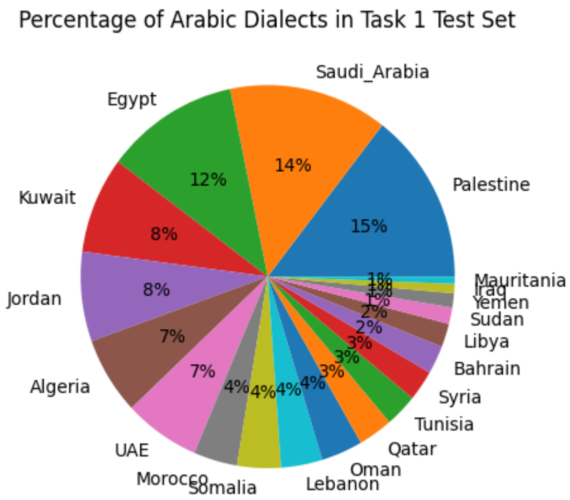
Figure 1: Text Type Distribution in Task 1



(a) Training Set



(b) Development Set



(c) Testing Set

Figure 2: Arabic Dialect Identification in Task 1

Hyperparameters	Range Or Value
Batch Size	8
Random Seed	42
Learning Rate	2e-05
Number of Epochs	10
Max Length	512
Total Steps	600
Optimizer	AdamW

Table 6: Hyperparameters for System Implementation

Fine-tuned SVM

Hyperparameter	Value
C	1
max_iter	1000

Fine-tuned RF

Hyperparameter	Value
criterion	gini
n_estimators	200

Fine-tuned LR

Hyperparameter	Value
C	100
penalty	L2

Fine-tuned XGB

Hyperparameter	Value
max_depth	6
n_estimators	200

Table 7: Hyperparameters for Baseline Models

A.4 Best Model Performance

After conducting a total of 22 experiments across two subtasks using 11 models. We will focus on the two best performing models in each subtask on the testing subsets:

i. Best performing *ReDASPersuasion* with Multilingual Transformers: "XLM RoBERTa" on Task 1A, and

ii. Best performing *ReDASPersuasion* with Monolingual Transformers: "CAMeLBERT-DA SA" on Task 1B.

We carefully examine the confusion matrix plots to gain insights into the performance of our models. By doing so, we can determine which classes posed more difficulty for the classifiers.

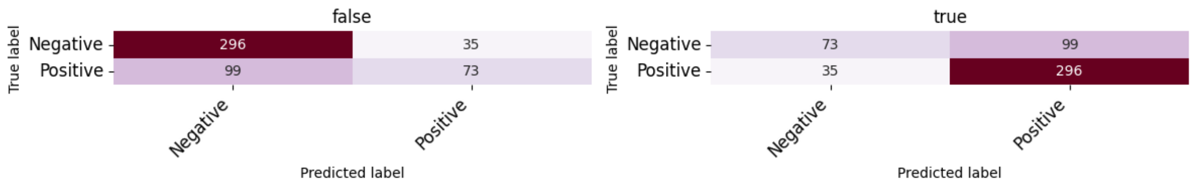


Figure 3: Confusion Matrix of *ReDASPersuasion* employing **XLM RoBERTa** for Binary Persuasion Classification on *Task 1A Testing Set.*

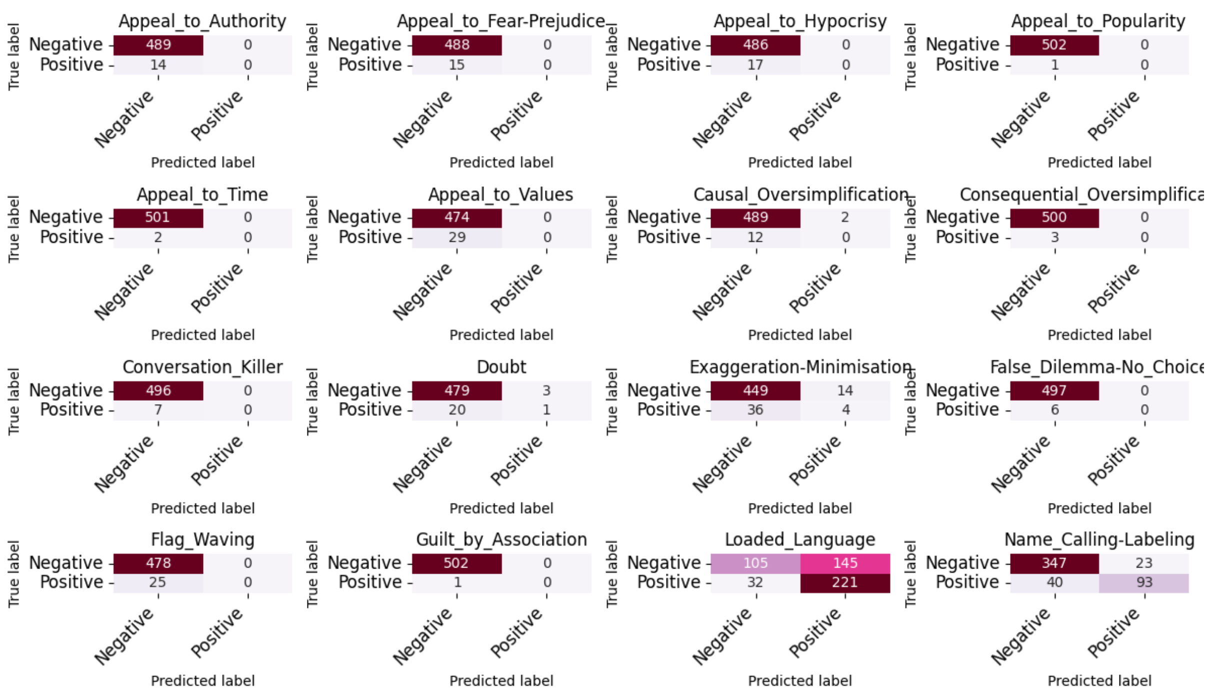


Figure 4: Confusion Matrix of *ReDASPersuasion* employing **CAMeLBERT-DA SA** for Multi-Label Persuasion Classification on *Task 1B Testing Set.*