# Cross-Multilingual, Cross-Lingual and Monolingual Transfer Learning For Arabic Dialect Sentiment Classification

**Naaima Boudad** ( naaima.boudad@gmail.com )
ENSIAS, Mohammed V University in Rabat
**Rdouan Faizi**
ENSIAS, Mohammed V University in Rabat
**Rachid Oulad Haj Thami**
ENSIAS, Mohammed V University in Rabat

**Research Article**

**Keywords:** Natural language processing, Transfer learning, Pre-trained Language Models, and Sentiment Classification

**Posted Date:** July 19th, 2023

**DOI:** https://doi.org/10.21203/rs.3.rs-3167222/v1



**Additional Declarations:** No competing interests reported.

**Version of Record:** A version of this preprint was published at Social Network Analysis and Mining on December 6th, 2023. See the published version at https://doi.org/10.1007/s13278-023-01159-9.

# Abstract

Transfer learning have recently proven to be very powerful in diverse Natural language processing (NLP) tasks such as Machine translation, Sentiment Analysis, Question/Answering. In this work, we investigate the use of transfer learning (TL) in Dialectal Arabic sentiment classification. Our main objective is to enhance the performance of Sentiment classification and overcome the low resource issue of Arabic dialect. To this end, we use Bidirectional Encoder Representation from Transformers (BERT) to transfer contextual knowledge learned from language modelling task to sentiment classification. We particularly use the multilingual models mBert and XLM-Roberta, the specific Arabic models ARABERT, MARBERT, QARIB, CAMEL and the specific Moroccan dialect Darijabert. After carrying out downstream fine-tuning experiments using different Moroccan SA datasets, we found that using TL significantly increase the performance of sentiment classification in Moroccan Arabic. Nevertheless, though specific Arabic models have proven to perform much better than multilingual and dialectal models, our experiments have demonstrated that multilingual models can be more effective in texts characterized by an extensive use of code-switching.


# 1 Introduction

The advent and the growth of social media platforms has led to an exceptional growth in user-generated data. Given the massive amounts of data that is increasingly being introduced inline sentiment analysis has become a very popular research topic amongst researchers. However, there some challenging problems that seem to hinder researchers from working on languages such as Arabic. The latter language is actually made up of multiple dialects that differ, in various ways, from each other and from the official form of Arabic called Modern Standard Arabic (MSA). Moreover, Arabic social media users mostly use their mother tongue rather than MSA to comment or express their opinions. Dealing with this diversity of informal writing presents, thus, a critical challenge of Arabic sentiment analysis.

Moroccan Arabic is one of Arabic dialects that differ significantly from MSA. The growth of using Moroccan dialect on social media and the lack of dedicated resources require more focus on how to take advantage of existing systems to improve processing and understanding of this Arabic variety.

The latest significant advances in NLP-related research has resulted in the emergence of a number of Pre-trained language models (PLM) which have proven to be very successful in diverse NLP tasks such as Machine translation, Sentiment Analysis and Question/Answering. In fact, since the introduction of transformers in 2017, several variants of such models have been proposed. These include include BERT (Devlin et al., 2018), RoBERTa (Liu et al., 2019), and BART (Lewis et al., 2019) among others. The success of transformer-based pre-trained language models has attracted the attention of Arabic PLN community and important efforts have been deployed to produce Arabic Transformer Models such as ArabicBERT (Safaya et al., 2020), AraBERT (Antoun et al., 2020), MARBERT (Abdul-Mageed et al., 2020), GigaBERT (Lan et al., 2020), CAMELbert (Inoue et al., 2021).

We propose in this work using transfer learning to take advantage of existing language models to improve the processing and understanding of the Moroccan Arabic variety. Our objective is to examine to what extent the multilingual, monolingual or even dialectal transfer learning can perform in Moroccan Sentiment Analysis. In this respect, an attempt will be made the following questions:

1. Are cross-multilingual transfer learning enough to achieve good performance in Moroccan Sentiment Analysis?
2. Can specific Arabic models outperform cross-multilingual models?
3. Is transfer learning from a small Moroccan language model more appropriate than transfer learning from a large Arabic multi-dialects language model?

The rest of this paper is structured as follows. In Section 2 we discuss related work, In Section 3 we describe pre-trained language models used in our study. Evaluation datasets are detailed in Section 4. In Section 5, we present our experiments and discuss results in section 6. Finally, we conclude in Section 7.

## 2 Related work

In the last two decades, a large number of research studies have been conducted on SA in Arabic. In most of these works, researchers have applied linear machine learning models together with traditional features such as n-gram. Some traditional supervised algorithms such as NB and SVM have frequently been used (Boudad et al., 2017). With the advance of deep learning and word embedding models and their introduction in many NLP fields, many Arabic Sentiment Analysis studies have explored neural models such as convolution models, recurrent models, and attention mechanisms to learn text representation,.

More recently, pre-trained language models have proven to be successful in text representation and transfer learning. In fact, several pre-trained language models such as Bidirectional Encoder Representation from Transformers (BERT) (Devlin et al., 2018), ELECTRA (Clark et al., 2020), GPTs (Brown et al., 2020; Radford et al., 2019), RoBERTa (Liu et al., 2019), have been able to achieve state-of-the-art results on different NLP tasks including Sentiment Analysis. In addition, parallel works extended these systems to multilingual settings such as mBert and Xlm-Roberta that are trained on multiple languages. Other models dedicated to specific languages other than English have been introduced. These include CamemBERT (Martin et al., 2019) for French and Bertje for Dutch (de Vries et al., 2019).

Arabic NLP community has also contributed to this rising work of dedicated language models and several Arabic specific models was proposed in the literature. In this respect, Antoun et al. (Antoun et al., 2020) introduced 15 Arabert variants, based on the same architecture of three original models Google's BERT (Devlin et al., 2018), OpenAI GPT2-base (Radford et al., 2019), and Electra (Clark et al., 2020). Their models were trained on a large corpus of 77 GB MSA Arabic text. Evaluation on Sentiment Analysis tasks showed that their pre-trained models outperform CNN and BiLSTM-CRF approaches on four different Arabic datasets HARD, ASTD-Balanced, ArsenTD-Lev and AJGT.

Abdul-Mageed et al. (Abdul-Mageed et al., 2020) proposed two Arabic models ARBERT and MARBERT. Following BERT pre-training setup, the first model was pre-trained on 61GB of MSA text, while the second was trained on 128GB of multi-dialect tweets. Evaluation of the two models on SA shows that MARBERT is more powerful than ARBERT, and it outperforms AraBERT in three dataset HARD, ASTD-Balanced and AJGT.

In 2021, Inoue et al. (Inoue et al., 2021) built three Arabic BERT-base models called CAMeLBERT-MSA, CAMeLBERT-DA, CAMeLBERT-CA that were pre-trained on MSA, dialectal, and classical Arabic text data respectively. In addition, the authors proposed CAMeLBERTMIX, which was pre-trained on a mixture of the three text genres. They evaluated their models on different NLP tasks including Arabic SA.

Ghaddar et al (Ghaddar et al., 2021) presented JABER and SABER, Junior and Senior Arabic BERT, respectively. These were trained on 115GB of MSA texts. Evaluation of the two models on a new collection of benchmark datasets for Arabic Language Understanding (ALUE) shows that JABER and SABER outperform all previous models.

There is no doubt that the introduction of pre-trained language models has improved the state of the art of Arabic sentiment Analysis. However, focus has mostly been on MSA and a little interest has been devoted to improve the performance of NLP tasks on Arabic dialects. In this context, Messaoudi et al. (Messaoudi et al., 2021) developed a Tunisian language model based on Bert architecture and pre-trained on a collection of 60 KB Tunisian dialect data. TunRoBERTa is another Tunisian model introduced by Antit et al. (Antit et al., 2022) that was trained using RoBERTa architecture on Tunisian texts and evaluated on two Tunisian SA datasets. Abdaoui et al. (Abdaoui et al., 2021) presented

the first Algerian language model: DziriBERT trained on 150 MB Algerian tweets using the same architecture of BERT base. Compared to existing models, DziriBERT achieves the best performance on two Algerian SA downstream datasets.

In this work, we hope to help bridge the gap between the success of MSA SA and Moroccan SA. We investigate the efficiency of transfer learning in Moroccan sentiment analysis by using different variant of BERT Based language Models.

# 3 Language Models

BERT's model architecture is a multi-layer bidirectional transformer encoder based on the original implementation described in (Vaswani et al., 2017). In this section, we describe BERT based models selected to perform different types of transfer learning. All these models have the same configuration of Bert Base, which is implemented with 12 layers, 768 hidden size and 12 self-attention heads.

Cross-multilingual transfer learning

To perform cross-multilingual transfer learning, we consider the two following multilingual models:

- **m-BERT** (Pires et al., 2019): the version multilingual of Bert trained on the Wikipedia pages of 104 languages with a shared wordPiece vocabulary.
- **Xlm-Roberta** (Conneau et al., 2019): a multilingual version of the RoBERTa model. It is pre-trained on 2.5TB of data across 100 languages data from Common Crawl including 28GiB Arabic Data. It is trained on Masked language modeling (MLM) task using only monolingual data with wordPieces tokenization. In our experiments, we use the base version that has 12 encoder blocks, 768 hidden dimensions, 12 attention heads and 270M parameters. XLM-R achieves state-of-the-arts results on multiple cross lingual benchmarks.

Cross-lingual transfer learning

To transfer learning from Arabic language modeling to Moroocan Arabic dialect sentiment Analysis, that following Arabic models have been opted for:

- **Arabertv02-twitter** (Antoun et al., 2020): It is an Arabic monolingual model based on Bert Architecture and trained on 77GiB of Arabic news collected from different media in different Arab regions. Arabert trained with masked language modeling (MLM) and next sentence prediction (NSP) objectives. It uses the BERTbase configuration with 512 maximum sequence length, and a total of ∼136M parameters.
- **Marbert** (Abdul-Mageed et al., 2020): an Arabic multi dialects model based on Bert Architecture and trained on 128GiB of Arabic tweets. It is trained with masked language modeling (MLM) objective using BERTbase configuration with a total of ∼163M parameters and 218 maximum sequence.
- **Qarib** (Abdelali et al., 2021): An Arabic BERT model trained on a collection of ~ 420 Million tweets and ~ 180 Million sentences from different available Arabic corpus. The pre-training was performed on predicting random missing words with a total of ∼135M parameters.
- **Camel-DA** (Inoue et al., 2021): An Arabic BERT model trained on a 54GB text from a range of dialectal corpora. It is trained with the whole word masking with a total of ∼163M parameters. The maximum sequence length was limited to 128 tokens for 90% of the steps and 512 for the remaining 10%.

Monolingual transfer learning

In an aim to evaluate the impact of both the size and language similarity spectra, we consider **DarijaBert** a Moroccan Arabic dialect BERT model trained on a small size corpus. It is developed and made available by AIOX Labs, an AI Moroccan company. It is trained on 691MB of Moroccan dialects text from web stories, tweets and YouTube comments, using masked language modeling (MLM) task with Bert base configuration and a total of ∼147M parameters.

The table below presents some statistics of the aforementioned models.

Table 1
*Statistics of the different pre-trained language models*

| Model | type | Arabic Vocabulary size | Num Tokens | Data Size | Num of parameters |
|---|---|---|---|---|---|
| **XLM-RoBERTa-base** | Multilingual | 14K | 2.9B | 2.5TB | 270M |
| **mBert-base** | Multilingual | 5K | 0.15B | 42GB | 167M |
| **AraBERTv0.2-twitter** | Arabic | 64K | 2.5B | 77GB | 136M |
| **Qarib** | Arabic | 64K | 14.0B | - | 135M |
| **MarBERTv2** | Arabic | 100K | 15.6B | 128GB | 163M |
| **Camel-DA** | Arabic | 30K | 5.8B | 54GB | 109M |
| **Darijabert** | Moroccan Dialect | 80K | 2.5B | 691MB | 147M |

# 4 Moroccan SA Datasets

In this section, we describe the four Moroccan datasets selected to conduct sentiment classification fine-tuning:

- **MSAC**: The Moroccan Sentiment Analysis Corpus Arabic created by Oussous et al. (Oussous et al., 2020) was collected from 2,000 Moroccan tweets and labelled manually into two classes, namely negative and positive.
- **ElecMorocco**: Moroccan Arabic dataset collected by Elouardighi et al. (Elouardighi et al., 2017) from Facebook comments written in Modern Standard Arabic or in Moroccan Dialectal Arabic about the Morocco's Legislative Elections of 2016. Comments are labelled into 3673 positive and 6581 negative classes.
- **MSDA**: an open data set of social data content in several Arabic dialects collected by Boujou et al. (Boujou et al., 2021). The data is collected from the Twitter social network and consists of + 50K tweets in five Arabic dialects: Moroccan, Algerian, Tunisian, Egyptian and Lebanese. We consider this dataset as the most challenging since tweets are from multiple dialects, and some tweets are written in a mixture of Arabic with other Latin languages such as English and French.
- **MAC**: Moroccan Arabic Corpus is built by Garouani et al. (Garouani and Kharroubi, 2021) and consists of 18,087 tweets written in Standard Arabic and Moroccan dialect. Tweets are manually labeled into positive, negative, mixed and neutral classes.

Since there is no official split for these datasets, we divided the datasets following the standard split: 80% training, 10% evaluation, and 10% test. We release the dataset splits and the code for easier comparison and reproducibility of our results. Table 2 shows the number of samples by training, evaluation and testing sets.

Table 2
split of training, evaluation and testing sets

| | classes | total | train | Eval | test |
|---|---|---|---|---|---|
| **MSAC (2classes)** | pos | 1000 | 801 | 107 | 92 |
| | neg | 1000 | 799 | 93 | 108 |
| **ElecMorocco (2classes)** | pos | 3673 | 2930 | 369 | 374 |
| | neg | 6581 | 5273 | 656 | 652 |
| **MSDA (3 classes)** | pos | 6792 | 5384 | 688 | 720 |
| | neg | 15385 | 12309 | 1542 | 1534 |
| | neu | 30033 | 24075 | 2991 | 2967 |
| **MAC (4classes)** | pos | 9897 | 7881 | 988 | 1028 |
| | neg | 4039 | 3272 | 374 | 393 |
| | neu | 3508 | 2811 | 370 | 327 |
| | mixed | 643 | 505 | 77 | 61 |

# 5 Experiments

To investigate the performance of transferring contextual embedding from multilingual, Arabic language and Moroccan dialect modelling to Moroccan SA downstream task, we perform fine-tuning for each model using training and validation datasets and we ran classification prediction on the test sets.

In fact, while the corpus used for training DarijaBert is a dialect-specific corpus, it is also smaller and may not provide enough contexts for sentiment classification fine-tuning. On the other hand, the corpus used for training Arabic and multilingual models is considerably larger but may not have enough coverage of the Moroccan dialect, which could lead to biases in the output embedding.

For the implementation of different TL experiments, we utilized the publically available Pytorch Transformer library1. As for the environment of implementation, we used Google Colaboratory tool with 4 NVIDIA Tesla V100 GPU 16G RAM. The used models are publically available on HuggingFace library.

As cleaning was done for most of the datasets, we performed only some light pre-processing such as removing diacritics, tatweel and repetition of more than two characters.

For text classification tasks, BERT considers the final hidden vector C $\in \mathbb{R}^{H}$ of the special token [CLS] as the representation of the whole sequence. A standard classification loss is computed using a simple Softmax classifier (Devlin et al., 2018), i.e., $Log\left(softmax\left(C^{\perp}W\right)\right)$, where $W$ is the classification layer weight $W \in \mathbb{R}^{K \times H}$, and K is the number of classes. During fine-tuning, all the weights from BERT as well as the weight matrix $W$ are trained jointly by maximizing the log-probability of the correct class.

Following the fine-tuning procedure of BERT (Devlin et al., 2018), we set the sequence length and batch size to 128 and 32, respectively. We fine-tuned each model for 2,3,4,5 epochs on training data and we measured the results on the development set. The best fine-tuning learning rate was selected among 5e-5, 4e-5, 3e-5, and 2e-5 on the development dataset.

Fine-tuning BERT on a small dataset can be unstable (Devlin et al., 2018). Moreover, randomness in machine learning models can happen because of randomly initialized weights and biases, dropout regularization, and optimization techniques (Dodge et al., 2020). The accuracy of the model can be influenced by the random seed value choice. Therefore, following recommendations put forward in the literature (Ameri et al., 2021), we ran several random restarts on the development set using the same hyper-parameters and changing only the random seeds to select a statistically reliable accuracy.

[1] https://github.com/huggingface/transformers

# 6 Results and Analysis

After the fine-tuning step, the best checkpoint model and tokenizer were saved. We run the sentiment prediction on test sets and report F1-macro averaging and accuracy in Table 3.

As a baseline, we consider the Arabic SA system developed in (Boudad et al., 2019) and based on a combination of CNN and ski-gram word2vec. We display also previous work results on each dataset. It is worthwhile to note that we cannot use previous works as baseline since they used cross-validation method.

A look at the results above reveals that pre-trained Bert-based models outperform the baseline results on different datasets which proves that using deep transfer learning is more powerful that using classical deep learning regardless of the size and the similarity language of the training corpus. The average results show that Bert-Based transfer learning outperforms CNN by a margin between + 3.83 and + 10.45 Macro-F1. More specifically, we showed that the gain becomes more important when dealing with multiple classes and unbalancing datasets. For instance, in MAC dataset that has four classes considerably unbalanced, the CNN model gets the worst score 59.35 Macro-F1 caused by its low recall on mixed and neutral classes. This means that CNN cannot make generalizations in less populated classes and that class imbalance biases the output of CNN towards the most populated class. However, the Arabic language model “Qarib” is able to yield 82.05 Macro-F1 (over the performance of CNN by 22.7 points) on this dataset, which proves the effectiveness of using Bert-based TL when dealing with challenging datasets.

Comparing the results of monolingual TL against cross-multilingual TL, we can see that DarijaBert even trained on only a small scale of data (691MB), outperforms in most of the cases the multilingual models. It is worth noting that mBert is almost systematically below all other Bert-based models. One possible explanation is that mBert is trained on only WIKIPEDIA texts. On the other hand, using the multilingual model XLM-R trained on 2.5TB texts leads overall to slightly lower results (-1.98 Macro-F1) than the results obtained with monolingual TL. This is interesting as it shows that language similarity matters more than the size of corpora.

Evaluating the efficiency of transferring Arabic contextual knowledge to Moroccan dialect SA demonstrates that using Arabic Bert-based models further enhances performance and allows the system to reach the best scores. The four Arabic specific models: Arabert, Marbert, Qarib and Camel-DA are approximatively equivalent. On average, the Qarib model performs the best, followed by the Arabert Model. Interestingly, both models outperform Marbert model pre-trained on 128Gib much larger than the size of training AraBert (77GB). A possible explanation is that Qarib and Arabert corpora are collected from different categories of Arabic text and include both standard and dialectal Arabic, while Marbert is pre-trained only on Arabic dialectal tweets. This confirms that having a mixed language variety in pre-training data is more useful. In addition, we observe a significant gap between results obtained using Monolingual Transfer and the best-obtained performance. This decrease on performance shows that language similarity solely is not enough to build good in-domain representations.

Table 3
F1-macro averaging and accuracy obtained using different models on Moroccan SA (*results using cross-validation method)

| | | MSAC (2 classes) Moroccan AD | | ELEC (2 classes) MSA + Moroccan AD | | MSDA (3 classes) Multi AD | | MAC (4 classes) MSA + Moroccan AD | | Macro Average | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Ma-F1 | Acc | Ma-F1 | Acc | Ma-F1 | Acc | Ma-F1 | Acc | Ma-F1 | Acc |
| Previous works | **CNN** (Oussous et al., 2020) | - | 95.50* | - | - | - | - | - | - | | |
| | **SVM** (Elouardighi et al., 2017) | - | - | - | 78.00 | - | - | - | - | | |
| | **SGD Classifier** (Boujou et al., 2021) | - | - | - | - | 74.00 | 77.00 | | | | |
| | **LSTM** (Garouani et al., 2021) | - | - | - | - | - | - | - | 89.60* | | |
| Baseline | **CNN + Word2Vec** | 89.00 | 89.00 | 80.75 | 81.80 | 77.96 | 79.21 | 59.35 | 67.68 | 76.76 | 79.42 |
| Transfer Learning | **mBert** | 90.70 | 90.70 | 85.03 | 86.18 | 82.34 | 85.85 | 64.29 | 80.83 | 80.59 | 85.89 |
| | **XLM-Roberta** | 92.00 | 92.00 | 89.96 | 90.56 | 81.38 | 85.35 | 66.10 | 82.04 | 82.36 | 87.48 |
| | **Arabertv02-twitter** | **95.90** | **95.90** | 89.82 | 90.99 | 81.70 | 85.58 | 79.08 | 88.67 | 86.62 | 90.28 |
| | **Marbertv2** | 95.40 | 95.40 | **90.15** | **91.18** | 81.42 | 85.16 | 77.16 | 87.84 | 86.03 | 89.89 |
| | **Qarib** | 95.60 | 95.60 | 89.39 | 90.60 | 81.80 | 85.62 | **82.05** | **89.89** | **87.21** | **90.42** |
| | **Camel-DA** | 95.00 | 95.00 | 85.82 | 86.99 | **82.14** | **85.96** | 77.11 | 86.01 | 85.01 | 88.49 |
| | **DarijaBert** | 94.97 | 94.97 | 83.80 | 86.11 | 81.32 | 85.41 | 77.29 | 86.35 | 84.34 | 88.21 |

# 7 Conclusion

In this work, we explored the use of Bidirectional Encoder Representation from Transformers (BERT) models in a sentiment classification of Moroccan dialect texts. We found that using deep transfer learning is more powerful that using classical deep learning regardless of the size and the similarity language of the training corpus. Comparing the performances of using multilingual, Arabic language and Moroccan dialect models, shows that Arabic languages trained on mixed data of dialect and MSA outperforms multilingual and specific dialect models. In future works, we intend to develop a new Moroccan language model trained on a large corpus, and explore more downstream NLP on Moroccan dialect. Furthermore, we will attempt to handle Arabizi scripts and French-Arabic code switching that is likely to help improve the performance of Moroccan dialect NLP tools.

# References


1. Abdaoui, A., Berrimi, M., Oussalah, M., Moussaoui, A., 2021. Dziribert: a pre-trained language model for the algerian dialect. ArXiv Prepr. ArXiv210912346.
2. Abdelali, A., Hassan, S., Mubarak, H., Darwish, K., Samih, Y., 2021. Pre-training bert on arabic tweets: Practical considerations. ArXiv Prepr. ArXiv210210684.
3. Abdul-Mageed, M., Elmadany, A., Nagoudi, E.M.B., 2020. ARBERT & MARBERT: deep bidirectional transformers for Arabic. ArXiv Prepr. ArXiv210101785.
4. Ameri, K., Hempel, M., Sharif, H., Lopez Jr, J., Perumalla, K., 2021. CyBERT: Cybersecurity Claim Classification by Fine-Tuning the BERT Language Model. J. Cybersecurity Priv. 1, 615–637.
5. Antit, C., Mechti, S., Faiz, R., 2022. TunRoBERTa: A Tunisian Robustly Optimized BERT Approach Model for Sentiment Analysis. Atlantis Press, pp. 227–231.
6. Antoun, W., Baly, F., Hajj, H., 2020. Arabert: Transformer-based model for arabic language understanding. ArXiv Prepr. ArXiv200300104.
7. Boudad, N., Ezzahid, S., Faizi, R., Thami, R.O.H., 2019. Exploring the Use of Word Embedding and Deep Learning in Arabic Sentiment Analysis. Presented at the International Conference on Advanced Intelligent Systems for Sustainable Development, Springer, pp. 243–253.
8. Boudad, N., Faizi, R., Thami, R.O.H., Chiheb, R., 2017. Sentiment Classification of Arabic Tweets: A Supervised Approach. J Mob. Multimed. 13, 233–243.
9. Boujou, E., Chataoui, H., Mekki, A.E., Benjelloun, S., Chairi, I., Berrada, I., 2021. An open access NLP dataset for Arabic dialects: Data collection, labeling, and model construction. ArXiv Prepr. ArXiv210211000.
10. Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J.D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., 2020. Language models are few-shot learners. Adv. Neural Inf. Process. Syst. 33, 1877–1901.
11. Clark, K., Luong, M.-T., Le, Q.V., Manning, C.D., 2020. Electra: Pre-training text encoders as discriminators rather than generators. ArXiv Prepr. ArXiv200310555.
12. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., Stoyanov, V., 2019. Unsupervised cross-lingual representation learning at scale. ArXiv Prepr. ArXiv191102116.
13. de Vries, W., van Cranenburgh, A., Bisazza, A., Caselli, T., van Noord, G., Nissim, M., 2019. Bertje: A dutch bert model. ArXiv Prepr. ArXiv191209582.
14. Devlin, J., Chang, M.-W., Lee, K., Toutanova, K., 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. ArXiv Prepr. ArXiv181004805.
15. Dodge, J., Ilharco, G., Schwartz, R., Farhadi, A., Hajishirzi, H., Smith, N., 2020. Fine-tuning pretrained language models: Weight initializations, data orders, and early stopping. ArXiv Prepr. ArXiv200206305.
16. Elouardighi, A., Maghfour, M., Hammia, H., 2017. Collecting and processing arabic facebook comments for sentiment analysis. Springer, pp. 262–274.
17. Garouani, M., Chrita, H., Kharroubi, J., 2021. Sentiment analysis of Moroccan tweets using text mining.
18. Garouani, M., Kharroubi, J., 2021. MAC: an open and free Moroccan Arabic Corpus for sentiment analysis. Presented at the The Proceedings of the International Conference on Smart City Applications, Springer, pp. 849–858.
19. Ghaddar, A., Wu, Y., Rashid, A., Bibi, K., Rezagholizadeh, M., Xing, C., Wang, Y., Xinyu, D., Wang, Z., Huai, B., 2021. JABER: Junior Arabic BERt. ArXiv Prepr. ArXiv211204329.
20. Inoue, G., Alhafni, B., Baimukan, N., Bouamor, H., Habash, N., 2021. The interplay of variant, size, and task type in Arabic pre-trained language models. ArXiv Prepr. ArXiv210306678.

21. Lan, W., Chen, Y., Xu, W., Ritter, A., 2020. An empirical study of pre-trained transformers for Arabic information extraction. ArXiv Prepr. ArXiv200414519.
22. Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., Stoyanov, V., Zettlemoyer, L., 2019. Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension. ArXiv Prepr. ArXiv191013461.
23. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., Stoyanov, V., 2019. Roberta: A robustly optimized bert pretraining approach. ArXiv Prepr. ArXiv190711692.
24. Martin, L., Muller, B., Suárez, P.J.O., Dupont, Y., Romary, L., de La Clergerie, É.V., Seddah, D., Sagot, B., 2019. CamemBERT: a tasty French language model. ArXiv Prepr. ArXiv191103894.
25. Messaoudi, A., Cheikhrouhou, A., Haddad, H., Ferchichi, N., BenHajhmida, M., Korched, A., Naski, M., Ghriss, F., Kerkeni, A., 2021. TunBERT: Pretrained Contextualized Text Representation for Tunisian Dialect. ArXiv Prepr. ArXiv211113138.
26. Oussous, A., Benjelloun, F.-Z., Lahcen, A.A., Belfkih, S., 2020. ASA: A framework for Arabic sentiment analysis. J. Inf. Sci. 46, 544–559.
27. Pires, T., Schlinger, E., Garrette, D., 2019. How multilingual is multilingual BERT? ArXiv Prepr. ArXiv190601502.
28. Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., 2019. Language models are unsupervised multitask learners. OpenAI Blog 1, 9.
29. Safaya, A., Abdullatif, M., Yuret, D., 2020. Kuisail at semeval-2020 task 12: Bert-cnn for offensive speech identification in social media. pp. 2054–2059.
30. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Adv. Neural Inf. Process. Syst. 30.