A deep sentiment analysis of Tunisian dialect comments on multi-domain posts in different social media platforms

Emna Fsih, Rahma Boujelbane and **Lamia Hadrich Belguith**
ANLP-RG, MIRACL Lab. University of Sfax, Tunisia
{emnafsih, rahma.boujelbane}@gmail.com
lamia.belguith@fsegs.usf.tn

Abstract

With the emergence of social media, the Tunisian Dialect (TD), as the other Arabic dialects, started having a wide representation in the written form. It has switched from a purely oral language to a written form without normalizing or utilizing any orthographic convention or standard. Therefore, it is necessary to investigate these opinions and analyze them in order to extract useful knowledge. For this we propose in this paper an approach to create a richly annotated corpus. Then, by exploiting this corpus, we compare the effectiveness of machine learning and transfer learning models to build fine grained sentiment analysis models. The BERT model, Fine tuned with TD data, achieved the best result.

1 Introduction

Opinion Mining (OM) or Sentiment Analysis (SA) can be defined as the task of detecting, extracting and classifying opinions on some issues/facts (Vinodhini and Chandrasekaran, 2012). Indeed, the proliferation of social media has allowed collecting data for low-resourced languages. Facebook, Instagram and YouTube have become very popular tools for sharing videos and communication. Nowadays, these social networks are intensively used by different kinds of companies to develop their activities and this is through analysing the opinions of internet users. This analysis is constrained by the availability of adequate tools and resources. The dilemma is accentuated when it comes to dealing with a poorly endowed language.

In this work, we process comments posted on content related to news and companies in Tunisia. These comments are mainly written in Tunisian dialect but they could integrate other languages namely Arabic dialects, standard Arabic or other foreign languages and sometimes applying a code switching between two or more languages, which generates a huge orthographic heterogeneity. Thus a word belonging to a such corpus may be an emoji, a word written in Latin script or a word written in Arabic script. Each form allows designating more than one language. Indeed, a word in Latin script can be a word in French (or also in English). Table 1 illustrates an example of comment forms. In this case, the automatic identification of the words written as TD is not trivial. The lexicon in Arabic script can also be either written a standard Arabic word, dialect word or a TD word. We wish through this work to propose a deep analysis for different types of comments present in different social media platforms. For this we propose a method which consists of building a fine-grained annotated corpus and to train different machine-learning and deep-learning models with the resulted corpus. These resources could be later useful and beneficial for other NLP tasks.

The remainder of this paper is organized as follows. Section 2 discusses some related works. Section 3 describes the proposed approach for TD opinion analysis. Section 4 details the Tunisian Dialect sentiment analysis resources. Section 5 presents sentiment analysis models for TD comments. Finally, section 6 concludes and points to possible directions for future work.

Forms	**Comment**
Latin script	t3jbny //I like you//
Arabic script	ممتازة //Excellent//
Emoji	♥ ✌
Latin script in French	trés bonne //very good//
Latin script in dialect	mo7tarem //respectable//

Table 1: Comment forms.

2 Related work

Studies on sentiment analysis of Arabic dialect in social networks have rapidly grown in recent years. Several researchers have attacked the treatment of this field on several aspects and levels.(Alahmary et al., 2019) for example, in order to develop sentiment and emotion annotation Twitter corpus for Saudi dialect, have created SDCT a corpus in Saudi dialect by annotating 32,063 tweets into two classes containing 17707 positive tweets and 14356 negative tweets. Using these resources, the authors have compared the effectiveness of the LSTM, Bi-LSTM models and the SVM model. The experimental results have showed that, in their context, the use of Bi-LSTM (F_score 94%) is more efficient than the use of LSTM and SVM.

Also, (Mohammed and Kora, 2019) have applied three Deep Learning models (DL): CNN, LSTM and RCNN in a network that combines some of the characteristics of both CNN and LSTM neural networks in the sense that CNN is used as a strong feature extractor and LSTM layer applies the recurrent neural network architecture on those extract features on 40k Arabic tweets for Arabic sentiment analysis. The best accuracy achieved was 88% by LSTM model. In addition (Al-Smadi et al., 2019) proposed a method for Arabic sentiment analysis. Authors have used LSTM implementation. Experiments have been done on Arabic hotels' reviews dataset as (19,226) for training and (4802) for testing and recorded an F1-Score of 82.6%. Similarly, (Dahou et al., 2019) used CNN algorithm to perform Arabic sentiment analysis, experimental results were evaluated on five different Arabic sentiment datasets. The best accuracy achieved was 93.28%. Recently, research has gone beyond deep learning approaches and explored the advances offered by transfer learning using transformer architectures based on the encoder-decoder pattern. (Moudjari et al., 2020) evaluates deep learning models (LSTM, CNN and BERT), to classify if an Algerian tweets (9000 tweets) as either positive, negative or neutral. The best results in term of accuracy were obtained with the CNN and LSTM models with 76% and 75%, respectively. On the other hand, BERT gave the worst results in term of accuracy with 68%.

In (Abdul-Mageed et al., 2020) implementing ALBERT and MARBERT models: i) ARBERT is a large scale pre-training masked language model focused on Modern Standard Arabic (MSA). ii) MARBERT is a large scale pre-training masked language model focused on both Dialectal Arabic (DA) and MSA. Both models implemented for multiple text classification tasks: (1) sentiment analysis (SA), (2) social meaning (SM), (3) topic classification (TC), (4) dialect identification (DI), etc in Arabic. The sentiment analysis model achieved the F-score of 71.50% when applied with MARBERT model. Similarly, (Abuzayed and Al-Khalifa, 2021) proposed sentiment detection for Arabic dialect language by augmenting data proposed by the shared task in (Abu Farha et al., 2021) to analyse the sentiment of tweets. By using the MARABERT model, they obtained an F1-score of 86%. (Abdel-Salam, 2021) have also fine-tuned MARBERT for sentiment classification tweets in Arabic dialect: MSA, Egyptian, Maghrebi dialect, etc and MARBERT model achieved an accuracy of 69.57%.

Sentiment analysis for Tunisian dialect: To develop sentiment analysis model for Tunisian Dialect (TD), many efforts have been made to develop resources such as annotated corpus, lexicon and models. (Mdhaffar et al., 2017) have created the Tunisian Sentiment Analysis Corpus (TSAC) from Facebook official pages of Tunisian radios and TV channels. TSAC contains 17k comments written only with Arabic letters and extracted from Facebook. This corpus has been annotated by a native speaker with 8215 positive and 8845 negative comments. An MLP classifier was then applied to build a sentiment analysis model that achieved an F1-score of about 78%. The TunBERT (Messaoudi et al., 2022) was trained on a TSAC dataset (Mdhaffar et al., 2017) including 7452 comments on Tunisian dialect in Social media and Tunisian Election Corpus (TEC) (Sayadi et al., 2016) 3042 tweets obtained from twitter about Tunisian elections in 2014. It is composed of MSA and Tunisian content. It achieved great results on the TSAC corpus with an accuracy of 96.98% compared to 81.2% on the TEC corpus.

3 Proposed approach

Given the orthographic heterogeneity of the Tunisian dialect in social networks we propose in this work, a process to build deeper Sentiment Analysis (SA) models. For this we exploited, at first, the resources publicly available in the state of the art and we endowed them with more annotations. Using these resources we tried to learn different sentiment analysis models with the aim of

Corpus	TSAC	TSAC+	TSAC+	
	Facebook		YouTube	Instagram
Number of comments	17060	16000	49000	7000
Number of words	113196	45536	126288	11962
Number of vocabulary items	42123	40809	97460	9940

Table 2: Statistics about the corpus.

	LEX1	LEX2	LEX3	LEX4
Simple words	227	312	3213	4041
Phrasal	331	643	7385	7673
Foreign words	11	12	19	45
Foreign phrasal	7	26	34	66
Emoticons	0	0	0	61
Positive opinion indicator	283	419	3638	4581
Negative opinion indicator	293	589	6960	7775
Total	576	1035	10651	12356

Table 3: Statistics of TD sentiment analysis lexicon.

providing comments with more analysis tags. In addition, to treat Arabic dialects comments that may be presented to comment publications intended for Tunisian people, we tested and compared the performance of the TD model and the model learned on a corpus augmented with such content. We detail in what follows the different stages of this process.

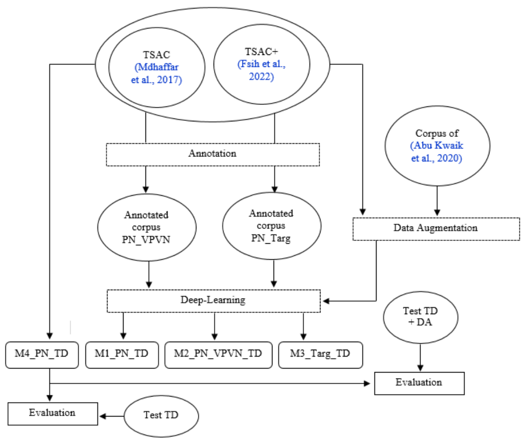

Figure 1: The proposed process to develop fine grained SA models

4 Tunisian Dialect Sentiment Analysis Resources

TSAC: We exploited the TSAC (Tunisian Sentiment Analysis Corpus) proposed by (Mdhaffar et al., 2017). This corpus contains 17k comments written with Arabic letters. It has been annotated into 8212 positive comments and 8854 negative comments of opinions expressed on Facebook and taken from Tunisian TV and radio pages during the period from January 2015 to June 2016.

TSAC+: It is a corpus developed by (Fsih et al., 2022). It contains 65K comments scraped from Tunisian TV channels during a period spanning over January 2016 through March 2019. In order to broaden the spectrum of sentiment analysis, we gathered by 7k comments containing opinions on content presented by influence's from different fields on Instagram. TSAC+ has been annotated into positive (42k) and negative (30k) comments. It has been annotated using TD lexicon. More statistics on the corpus are shown in table 2.

Corpus of Arabic dialect: Arabic Tweets Sentiment Analysis Dataset (ATSAD) presented in (Abu Kwaik et al., 2020) collected from Twitter consists of 42k tweets which are classified as positive and negative, which is available at `https://github.com/motazsaad/arabic-sentiment-analysis`. Table 4 shows the statistics of the corpus.

TD Sentiment analysis lexicon: These resources have also been developped by (Fsih et al., 2022). They contain 12356 entries classified into two main categories: simple words and phrases associated with their inflected forms (**LEX2**) and also with the corresponding ambiguous forms (**LEX3**). The following example shows the different forms corre-

Number of tweets	42693
Number of words	439518
Number of vocabulary items	64629
Number of emoji's	455
Number of positive tweets	18106
Number of negative tweets	24588

Table 4: The statistics of the Arabic dialect tweets.

sponding to the lexical entry "ناجح" //successful// (See Figure 2). Based on this process, we enriched the lexicon with new entries taken from TSAC+ (**LEX4**). Table 3 shows details about the TD sentiment analysis lexicon.

```
<mot num="1">
<FormN Gender="MASC_SG">ناجح</FormN>
<FormF num="1-1" Gender="FEM_SG">ناجحة</FormF>
<FormF num="1-2" Gender="MASC_PL">ناجحين</FormF>
<FormF num="1-3" Gender="FEM_PL">ناجحات</FormF>
<FormA num="1">najeh</FormA>
<FormA num="2">naje7</FormA>
<FormA num="3">nejeh</FormA>
<FormA num="4">neje7</FormA>
<FormA num="5">najeh</FormA>
<FormA num="6">naje7</FormA>
</mot>
```

Figure 2: Example of lexicon forms

4.1 Fine grained sentiment analysis annotation

Polarity Opinion annotation: The idea of this paper is to provide a deep analysis to the TD comments, for this and using the TD sentiment analysis lexicon we have annotated the corpus of TSAC+ using the following logic : we calculate for each comment the number of occurrences of positive and negative words. If the number of occurrences of positive words is greater than the negative words, the comment will be annotated as very positive and vice versa as very negative. For example, the comment "magnifique bravo stevy aymen" //wonderful bravo stevy aymen// contains two positive opinion indicators ("magnifique" //wonderful//, bravo) and zero negative indicators, the number of occurrences of positive words is greater than negative, the comment is then annotated as very positive. The table below illustrates the opinion annotation statistics with degree of polarity.

	TSAC+
Number of positive comments	20170
Number of negative comments	18115
Number of comments very positive	20794
Number of comments very negative	13630

Table 5: Statistics of the annotation with polarity.

Target annotation: We also sought to annotate the target of the opinion, for this, we extracted,at first, different target indicators like "حلقة "//Episode//, "برنامج" //Program//, "قناة" //Channel// from TSAC+ corpus. Five target categories "episode, program, person, subject and channel" were distinguished. Each category has its corresponding indicators. For instance, the category person has two types of indicators: 305 of them are named entities and 901 are adjectives used to designate a person such as "منشط" //A television presenter//, "ضيف" //A guest//, "مذيع" //Broadcaster//. Comments having no indicators were classified as neutral. Table 6: reveals the number of comments in each target category.

Categories	Number of comments
Episode	755
Program	1349
Subject	183
Channel	614
Person	36570
Neutral	33250

Table 6: Statistics of target comments.

Comments	Polarity	Target
ya m3alem ya sami bravo //Great sami bravo//	Positive	Person
حلقة عالية //Great episode//	Positive	Episode
ملا تفاهة //What is this nonsense//	Negative	Natural
برنامج فاشل //Failed program//	Negative	Program

Table 7: Examples of TD comments.

5 Sentiment Analysis Models for TD Comments

5.1 Standard sentiment analysis model (M1_PN_TD):

Most of the works cited in the state of the art have focused on the construction of the usual sentiment analysis models, i.e they only detect Positive (P) and Negative (N) polarities. For this, we first used 59K comments of the corpus cited in the previous section to build a model able to detect Positive and Negative TD comment (M1_PN_TD). For this, we

M1_PN_TD			
Polarity	**Precision**	**Recall**	**F-measure**
Positive	72%	91%	81%
Negative	81%	53%	64%
Accuracy		75%	
M4_PN_AD			
Polarity	**Precision**	**Recall**	**F-measure**
Positive	83%	92%	87%
Negative	87%	75%	81%
Accuracy		85%	

Table 8: Evaluation of BERT model for Arabic dialect tweets classification.

M1_PN_TD Machine-learning model				
Classifiers	**Polarity**	**Precision**	**Recall**	**F-measure**
SMO	Positive	84%	96%	89%
	Negative	88%	61%	72%
NB	Positive	82%	98%	89%
	Negative	93%	54%	69%
DT	Positive	81%	94%	87%
	Negative	81%	55%	66%
Accuracy		SMO=84.66%	NB=83.85%	DT=81.13%
M1_PN_TD Deep-learning model				
Polarity		**Precision**	**Recall**	**F-measure**
Positive		84%	96%	90%
Negative		89%	63%	74%
Accuracy			86%	

Table 9: Evaluation of Machine-learning and Deep-learning models for Positive-Negative classification.

examined at first the effectiveness of three different text mining algorithms including SMO, Naive Bayes (NB) and Decision Tree (DT) model. Moreover, given the efficiency of transformer architecture on different NLP tasks, we fine tuned a BERT (Bidirectional Encoder Representations from Transformers) model learned on dialectal data (Fsih et al., 2022). The different models were based on a test corpus containing 13000 sentences. Results in table 9 show that the BERT model gives the best result with an accuracy of 86%.

5.2 Fine grained sentiment analysis models

Deep sentiment analysis model (M2_PN_VPVN_TD): We used the corpus annotated with 4 sentiments tags: Very Positive (VP), Very Negative (VN), positive and negative tags to study a sentiment analysis model. The same algorithms as before have been adopted. The best result was obtained using the BERT algorithm with an accuracy of 77% (Table 10).

Target of opinion detection model (M3_Targ_TD): On the corpus annotated with Target (Targ) tags, we also learned, using the same algorithms as the other experiments, a machine and deep-learning models for the detection of the target of the opinion. The resulted models have been projected on the same test corpus used in the experiments presented in this paper. The obtained results are given in table 11. The BERT model achieves the higher results with an accuracy of 96%.

multi-Arabic Dialect Deep sentiment analysis model (M4_PN_TD_AD): We wish through this experiment, studying the effect of the models designed for the TD on a mixed content including TD comments and dialectal comments. For this we introduced at first to the test corpus 21k comments (8k comments belonging to the different Arabic dialects and 13k comments in TD) on which we projected the model (M1_PN_TD). Then we introduced to the training data, dialectal comments con-

M2_PN_VPVN_TD Machine-learning model				
Classifiers	**Polarity**	**Precision**	**Recall**	**F-measure**
SMO	Positive	58%	79%	67%
	Negative	64%	35%	46%
	Very-Positive	73%	61%	66%
	Very-Negative	54%	66%	60%
NB	Positive	59%	36%	45%
	Negative	61%	51%	56%
	Very-Positive	50%	83%	63%
	Very-Negative	70%	42%	52%
DT	Positive	54%	80%	64%
	Negative	53%	35%	42%
	Very-Positive	72%	51%	60%
	Very-Negative	54%	48%	51%
Accuracy		SMO=62.56%	NB=55.33%	DT=57.85%
M2_PN_VPVN_TD Deep-learning model				
Polarity		**Precision**	**Recall**	**F-measure**
Positive		74%	86%	80%
Negative		71%	45%	55%
Very-positive		86%	89%	87%
Very-negative		67%	71%	69%
Accuracy			77%	

Table 10: Evaluation of Machine-learning and Deep-learning models for VeryPositive-VeryNegative classification.

taining 14k positive and 19k negative comments and then we compared the results. Table 8 shows the obtained results. We noticed that the augmentation performed with the dialect data improves the results. The model trained on this corpus achieves an accuracy of 85%.

6 Conclusion

In this paper, we presented an approach to build resources to analyse opinions written in Tunisian dialect. Indeed, we built a corpus containing 72k commentaries annotated in polarity and opinion target. Then, we developed four automatic opinion analysis models. The first allows to predict if the opinion is positive or negative. The second predicts whether the opinion is positive, very positive, negative or very negative. The third makes it possible to predict the target of the opinion. We expanded the spectrum of lexical coverage at the level of the fourth model. All these resources will be open for access once the paper is published. In future work we aim to test the effectiveness of other transformer models and to further extend the lexical coverage of the model to cover more ambiguities such as code switching or other languages.

References

Reem Abdel-Salam. 2021. WANLP 2021 shared-task: Towards irony and sentiment detection in Arabic tweets using multi-headed-LSTM-CNN-GRU and MaRBERT. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 306–311, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Muhammad Abdul-Mageed, AbdelRahim Elmadany, and El Moatez Billah Nagoudi. 2020. Arbert & marbert: deep bidirectional transformers for arabic. *arXiv preprint arXiv:2101.01785*.

Ibrahim Abu Farha, Wajdi Zaghouani, and Walid Magdy. 2021. Overview of the WANLP 2021 shared task on sarcasm and sentiment detection in Arabic. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 296–305, Kyiv, Ukraine (Virtual). Association for Computational Linguistics.

Kathrein Abu Kwaik, Stergios Chatzikyriakidis, Simon Dobnik, Motaz Saad, and Richard Johansson. 2020. An Arabic tweets sentiment analysis dataset (ATSAD) using distant supervision and self training. In *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*, pages 1–8, Marseille, France. European Language Resource Association.

Abeer Abuzayed and Hend Al-Khalifa. 2021. Sarcasm and sentiment detection in arabic tweets using bert-

M3_Targ_TD Machine-learning model				
Classifiers	**Target**	**Precisison**	**Recall**	**F-measure**
SMO	Person	87%	91%	89%
	Program	98%	83%	90%
	Episode	99%	87%	93%
	Subject	78%	47%	58%
	Channel	94%	90%	92%
	Neutral	93%	92%	92%
NB	Person	53%	98%	68%
	Program	97%	91%	94%
	Episode	95%	81%	88%
	Subject	98%	47%	38%
	Channel	94%	89%	92%
	Neutral	91%	48%	63%
DT	Person	91%	83%	87%
	Program	97%	85%	91%
	Episode	98%	84%	91%
	Subject	80%	40%	53%
	Channel	90%	89%	89%
	Neutral	89%	95%	92%
Accuracy		SMO=90.39%	NB=64.40%	DT=90%
M3_Targ_TD Deep-learning model				
Target		**Precision**	**Recall**	**F-measure**
Person		94%	96%	95%
Program		98%	94%	96%
Episode		97%	92%	95%
Subject		80%	80%	80%
Channel		99%	97%	98%
Neutral		97%	96%	97%
Accuracy			96%	

Table 11: Evaluation of Machine-learning and Deep-learning models for target classification.

based models and data augmentation. In *Proceedings of the sixth Arabic natural language processing workshop*, pages 312–317.

Mohammad Al-Smadi, Bashar Talafha, Mahmoud Al-Ayyoub, and Yaser Jararweh. 2019. Using long short-term memory deep neural networks for aspect-based sentiment analysis of arabic reviews. *International Journal of Machine Learning and Cybernetics*, 10(8):2163–2175.

Rahma M Alahmary, Hmood Z Al-Dossari, and Ahmed Z Emam. 2019. Sentiment analysis of saudi dialect using deep learning techniques. In *2019 International Conference on Electronics, Information, and Communication (ICEIC)*, pages 1–6. IEEE.

Abdelghani Dahou, Mohamed Abd Elaziz, Junwei Zhou, and Shengwu Xiong. 2019. Arabic sentiment classification using convolutional neural network and differential evolution algorithm. *Computational intelligence and neuroscience*, 2019.

Emna Fsih, Rahma Boujelbane, and Lamia Hadrich Belguith. 2022. Resources building for sentiment analysis on content disseminated by tunisian media in social networks. *Manuscript submitted for publication.*

Salima Mdhaffar, Fethi Bougares, Yannick Esteve, and Lamia Hadrich-Belguith. 2017. Sentiment analysis of tunisian dialects: Linguistic ressources and experiments. In *Third Arabic Natural Language Processing Workshop (WANLP)*, pages 55–61.

Abir Messaoudi, Ahmed Cheikhrouhou, Hatem Haddad, Nourchene Ferchichi, Moez BenHajhmida, Abir Korched, Malek Naski, Faten Ghriss, and Amine Kerkeni. 2022. *TunBERT: Pretrained Contextualized Text Representation for Tunisian Dialect*, pages 278–290. arXiv.

Ammar Mohammed and Rania Kora. 2019. Deep learning approaches for arabic sentiment analysis. *Social Network Analysis and Mining*, 9(1):1–12.

Leila Moudjari, Karima Akli-Astouati, and Farah Benamara. 2020. An Algerian corpus and an annotation platform for opinion and emotion analysis. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1202–1210, Marseille, France. European Language Resources Association.

Karim Sayadi, Marcus Liwicki, Rolf Ingold, and Marc Bui. 2016. Tunisian dialect and modern standard arabic dataset for sentiment analysis: Tunisian election context. In *Second International Conference on Arabic Computational Linguistics, ACLING*, pages 35–53.

G Vinodhini and RM Chandrasekaran. 2012. Sentiment analysis and opinion mining: a survey. *International Journal*, 2(6):282–292.