International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. 11, No.1 (Apr-2022)
https://dx.doi.org/10.12785/ijcds/1101105

AT-ODTSA: a Dataset of Arabic Tweets for Open Domain Targeted Sentiment Analysis

Shaaban Sahmoud[1], Shadi Abudalfa[2*] and Wisam Elmasry[3]

[1]*Department of Computer Engineering, Fatih Sultan Mehmet Vakif University, Istanbul, Turkey*
[2]*Information Technology Department, University College of Applied Sciences, Gaza, Palestine*
[3]*Department of Computer Engineering, Istanbul Kultur University, Istanbul, Turkey*

Received 31 Jul. 2021, Revised 31 Dec. 2021, Accepted 9 Mar. 20122, Published 15 Apr. 2022

Abstract: In the field of sentiment analysis, most of research has conducted experiments on datasets collected from Twitter for manipulating a specific language. Little number of datasets has been collected for detecting sentiments expressed in Arabic tweets. Moreover, very limited number of such datasets is suitable for conducting recent research directions such as target dependent sentiment analysis and open-domain targeted sentiment analysis. Thereby, there is a dire need for reliable datasets that are specifically acquired for open-domain targeted sentiment analysis with Arabic language. Therefore, in this paper, we introduce AT-ODTSA, a dataset of Arabic Tweets for Open-Domain Targeted Sentiment Analysis, which includes Arabic tweets along with labels that specify targets (topics) and sentiments (opinions) expressed in the collected tweets. To the best of our knowledge, our work presents the first dataset that manually annotated for applying Arabic open-domain targeted sentiment analysis. We also present a detailed statistical analysis of the dataset. The AT-ODTSA dataset is suitable for train numerous machine learning models such as a deep learning-based model.

Keywords: , Arabic Tweets, Open-Domain Targeted Sentiment Analysis, Sentiment Analysis, Target Dependent

1. Introduction

The wealth of sophisticated sensors motivated researchers to develop techniques that recognize human activities for assisting numerous humans [1]. Human activity recognition is an important task in implementing numerous smart technologies such as smart homes [2]. Such task received a high interest these days with availability of Internet of Things (IoT) technology [3] [4].

The social media has become a major part in our life through including social relations with other people who share similar personal activities or career interests. Popularity of social media sites assists in generating a massive data for different topics. The availability of tremendous public opinions opens the door to researchers and scholars to mine people's polarity of opinions toward almost every topic of interest. This introduces what is popularly known as sentiment analysis. Sentiment analysis which is also known as opinion mining is one of the major tasks under umbrella of natural language processing (NLP) [5] [6].

The main goal of sentiment analysis is classifying polarity of opinions [7] [8][9]. State-of-the-art systems for sentiment analysis deal mainly with three levels of annotation granularity towards the input: document, sentence, aspect, or phrase (word) [10]. Our research work concentrates on short sentences that are usually used in writing tweets. This kind of sentences is called micro-blog in social media.

Most sentiment analysis tools available are based on a targeted independent strategy. Accordingly, these tools may fail in assigning or detecting the correct sentiment polarity in the case of micro-blog sentences where sentences may contain more than one target/topic. Recent research direction is based on classify the micro-blog toward a specific target. This research direction is referred to as "target-dependent sentiment analysis" [11][12]. In more complex scenarios, the system is responsible for detecting the targets from micro-blog sentences in the first step, then in the second step, the sentiment polarities are identified based on detected targets. This scenario entitled as "open-domain targeted sentiment analysis" [13].

Based on our knowledge, applying open-domain targeted sentiment analysis with Arabic micro-blogs is very limited. One of the main challenges that delay the progress in this direction is the severe lack of convenient Arabic datasets. In this work, we fill this gap by presenting a new

*E-mail address: ssahmoud@fsm.edu.tr, *Coresponding author: sabudalfa@ucas.edu.ps, w.elmasry@iku.edu.tr*

dataset for Arabic open-domain targeted sentiment analysis (AT-ODTSA).

The AT-ODTSA dataset is collected from Twitter to gather large number of Arabic tweets. The dataset includes tweets with formal and slang Arabic language to mimic the real environment of Arabic social media. Thus, our dataset provides many challenges and open the door for scholars to deals with numerous research problems in the field of Arabic sentiment analysis. The AT-ODTSA dataset is available free on the Internet for researchers and scholars [14].

Furthermore, the AT-ODTSA dataset includes labels for identifying the target and opinion expressed in each tweet. Thereby, our dataset has also contributed to apply entity recognition to real Arabic tweets. All tweets included in the dataset were manually annotated by three experienced annotators, whereas all of them are Arabic native speakers. We provided with this work a detailed statistical analysis for the dataset along with some validations to the annotation process.

The main originality of this paper can be summarized as follows.

- Unlike other Arabic language datasets that focus on a limited region, our dataset includes tweets from many Arabic countries who speak in different language dialects.
- To simulate real scenarios, our dataset includes a high variety of topics from many domains.
- The annotation process is applied manually and carefully to overcome the issues of other Arabic datasets where the annotation was carried out automatically which generates many false annotated tweets.
- To add more difficulty, our dataset contains texts written in both formal and dialect Arabic languages.

The remainder of this paper is structured as follows: Section 2 briefly reviews some related studies to target-independent sentiment analysis datasets. Section 3 explains the methodology used for collecting data. Section 4 describes statistics related to our presented dataset. Section 5 discusses the validation process. Finally, Section 6 concludes the paper and reveals some suggestions for future work.

2. Literature Review

In this section, we present some related works recently published in the literature. Based on our knowledge, many datasets can be found for target-independent and target-dependent sentiment analysis [15] [16] [17] [18]. Unfortunately, most of the available datasets are for English language and there is small number of datasets for other languages such as Arabic [19] [20] [21]. To show novelty of our work, we classified previous related work into two categories: Target independent Arabic sentiment analysis and Target-dependent Arabic sentiment analysis.

In the literature, the available datasets for Arabic language mostly focus on target-independent sentiment analysis [22] [23]. For instance, the authors in [24] addressed two approaches for the Arabic language sentiment analysis which are corpus-based and lexicon-based. The author started by gathering an annotated dataset and then a set of steps were applied for building the lexicon. The conducted experiments showed some improvements in the accuracy of the proposed system. In another research work [25], the authors presented a new corpus collected from Twitter for Arabic sentiment analysis. This dataset has 36K tweets labelled into negative and positive only. However, the authors applied automatic methods for the annotation task using distant supervision and self-training mechanisms. In addition, they released another 8K tweets dataset which it is manually annotated for more accurate classification tasks.

In [26], the authors presented a new big dataset with detailed description for Arabic Sentiment Analysis. The dataset was gathered from Twitter, and the authors was launched a competition to encourage the research in this domain for Arabic language. The dataset is larger than previous datasets and according to authors it was annotated using high-quality techniques. In [27], a new Arabic tweets corpus of 40K tweets from several topics was introduced. moreover, the authors employed three different deep learning models, namely, Convolutional Neural Network (CNN), Long-Short Term Memory (LSTM), Recurrent Convolution Neural Network (RCNN) to be trained using the proposed dataset and solve the Arabic sentiment analysis task.

A Multi-Dialect Arabic Sentiment Twitter Dataset (MD-ArSenTD) was proposed to analyze tweets collected from Egypt and the United Arab Emirates (UAE) using different deep learning models [28]. In [29], a massive amount of tweets (6M) for Arabic sentiment analysis were collected and labeled using emojis sentiment lexicons. The authors validated the introduced dataset using several standard classifiers.

On the other hand, very limited works have been achieved in target-dependent Arabic sentiment analysis domain. In the sake of brevity, we only present some related works that are close to our research based on our perspectives. There is a dataset [30] that collected for Arabic stance detection and can be used for Arabic target-dependent sentiment analysis. However, the annotation process is automatically applied. Similarly, Baly et al. collected a dataset [31] for target-based sentiment analysis. However, there work is used only with Levant's dialect Arabic language.

Based on the aforementioned literature review, somebody might argue that the previous works in sentiment Arabic analysis domain have the following downsides [33][34] [35]: (i) there are a lot of studies concerned

TABLE I. Comparison between popular Arabic datasets

Dataset	Size	Classes	Annotation Approach	Purpose
ASTAD [25]	36000	2	Using Emojis	Target-independent
Ara-SenTi-Tweet [22]	17573	4	Manual	Target-independent
Gold Standard [23]	8868	4	Manual	Target-independent
ASAD [26]	8000	3	Manual	Target-independent
OCA [24]	500	2	Manual	Target-independent
ASA [32]	2000	2	Manual	Target-independent
Our AT-ODTSA	3000	3	Manual	Target-dependent

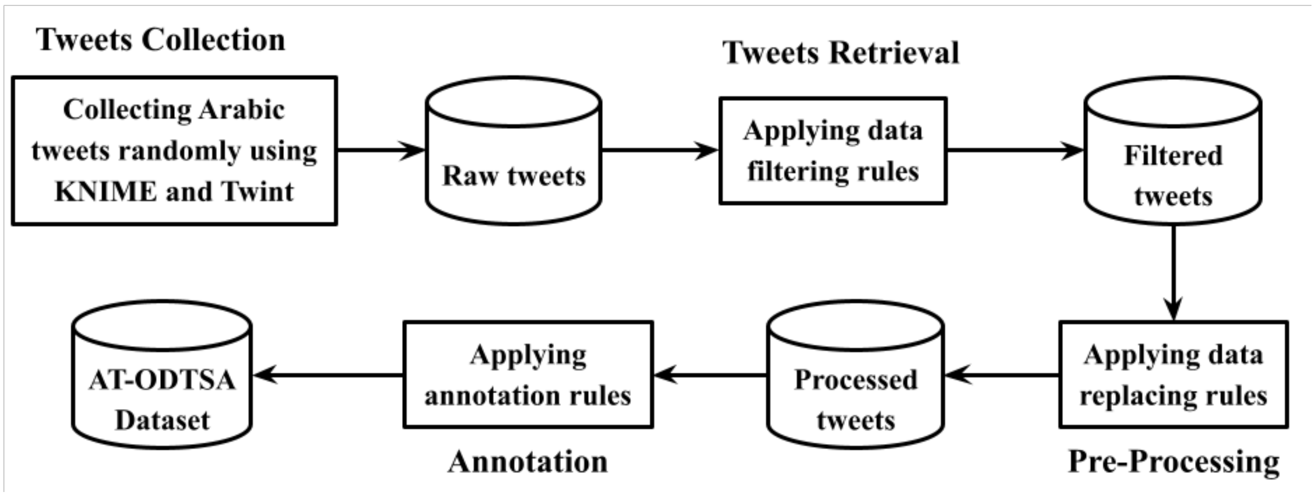


Figure 1. Methodology of creating the AT-ODTSA dataset.

with target-independent Arabic sentiment analysis, whereas the target-dependent Arabic sentiment analysis field contains a lot of gaps that should be filled [36]. In other words, there is no adequate dataset that can be used for Arabic target-dependent sentiment analysis. This encouraged us to collect a new dataset for dealing with this research direction. Our dataset mimics an English dataset for target-dependent sentiment analysis that is compiled by Dong et al. [37]. (ii) All available target-dependent Arabic sentiment analysis datasets are annotated automatically, whereas our presented dataset is different and more accurate since we used manual annotation through three experts. (iii) Our presented dataset can be used for open-domain rather than for specific-topic propose as found in related works. Accordingly, we can declare that our presented dataset is unique and adds a novel contribution to the research community. TABLE I shows a comparison between some of available Arabic sentiment analysis datasets and our proposed dataset in terms of size, number of classes, annotation approach, and purpose.

3. Methodology

Our main goal in this paper is to gather a dataset of tweets in Arabic language from all Arabic countries and regions. Therefore, this dataset will include most of Arabic language dialects. In this section, we describe in detail the steps of gathering, annotating and properties of the created dataset. Figure 1 depicts the methodology that we followed to create the AT-ODTSA dataset.

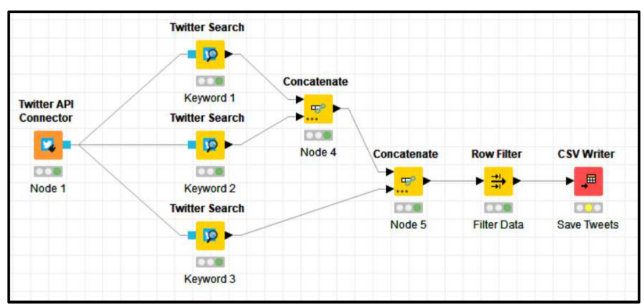


Figure 2. Example of KNIME components to retrieve and save a set of tweets.

A. Data Collection

We used two different data gathering tools to retrieve the tweets of our dataset which are a tool called KNIME and the Twint module. KNIME is a great data retrieving and data analysis open-source tool that includes many useful features. It is widely used in research domain for machine learning, data mining and visualization tasks. One of the best features of KNIME is its powerful graphical user interface which can be used as a drag and drop style to add and connect various nodes for downloading, processing and mining data [38]. Figure 2 display the used KNIME components to retrieve the tweets using three targets or keywords. Twint is an advanced Twitter scraping tool written in Python that allows for scraping Tweets from Twitter profiles without using Twitter's API. It is an easy-to-use and open-source library with many benefits [39].

We did not add any restriction for the Geo-locations of tweets or twitter users during the data retrieval to assess the randomness of the gathered tweets. The retrieval process started in 1-12-2020 and continued for three months but we did not restrict the date of tweets into a certain time period. We gathered more than 10,000 tweets that are randomly distributed across many Arabic countries. Based on the used methodology, we selected only 3000 tweets to be included in our dataset.

B. Rules of Data Collection

We aim to collect 3000 tweets in our dataset. The collected tweets target various topics and targets that are commonly discussed in Arab areas such as companies, organizations, press channels, famous actors, movies, products and Political figures. We gathered the dataset by searching for each target independently and collect a large number of tweets, then we filtered these tweets using the following rules to ensure the integrity and better quality for included tweets:

- Exclude any tweets with URLs only or target words only or hashtags only or emoticons only.
- The tweet should be only in Arabic language (one or two Non-Arabic words may be accepted as an exception if it can be excluded without affect the tweet meaning).
- No limits for the size of the tweet.
- No limits for posted date or time of the tweet.
- Tweets were selected without region limitations using any formal Arabic or slang Arabic language.
- Exclude any tweets containing inappropriate content such as sexual insinuations and insults.
- Some tweets include emoji symbols.

In addition, we applied some rules to organize the targets in the tweets of our dataset as follow:

- The target should be at most 3 words (three words targets are very limited).
- The target should be explicitly and exactly exists in the tweet.
- For each target, we should have at least five tweets.
- The target should be in Arabic language only.
- The tweet may have more targets other than our considered target.
- If the target contains more than one word, the words should be connected without any separation by other words.

We kept the tweets that satisfy all previous conditions and exclude the remaining tweets. Note that most of previous conditions are applied to ensure that the tweets have minimum properties to be treatable by selected classifiers or other machine learning algorithms. Moreover, some conditions were carried out to simplify the tracking of targets and the sentiment analysis algorithms, where we are planning to present another version of this dataset for more complex scenarios. Finally, we executed a filtering and replacing step to hide any existing twitter user IDs or mentions for privacy constrains.

C. Pre-processing

Two versions of our dataset were created. The first version includes all of the gathered tweets in its raw form without deleting or changing any word, symbol or special characters. In the second version which will be publicly published and used in the classification tasks, we have done a second pre-processing round to ensure the integrity of tweets and to exclude the irrelevant information. The following steps were conducted in this stage:

- Replace all links by the term "URL"
- Replace all mentioned user names by "@User"
- Replace all hashtags by "#hashtag" if it is not the target itself.
- If the hashtag refers to the target, the hash symbol # and the underscore symbol _ were removed to match the target words.

Note that the deleting of user mentions and hashtags is significant for the privacy restrictions since our dataset will be published in public. On the other hand, removing the URLs is important for misleading the machine learning algorithms during the training process since it may include meaningless words that can also affect the training process.

D. Annotation Process

The annotation process was performed through three expert annotators (researchers and PhD students) in data mining and machine learning. We believe that asking experts to annotate the tweets is much better than asking normal persons who may misunderstand and increase the false negative or false positive classifications of the data. Indeed, we checked and analyzed a set of previous datasets in target dependent and multi-target dependent and we found that using general and inexpert annotators can significantly increase the annotation errors and reduce reliability of datasets. In our case, we give annotators the list of tweets as three columns, tweet text, target, and annotation. The annotators were asked to fill the third column by 0 to indicate neutral, 1 to indicate positively, and -1 to indicate negatively. Annotations were aggregated based on majority voting as normally done in similar datasets. During the tweets collecting process, we tried to balance the tweet sets across our annotations classes to create a balanced dataset.

TABLE II presents distributions of the three annotation classes in our dataset and statistics for the selected target. As shown in TABLE II the dataset is imbalanced with respect to sentiment classes but the percentages of classes are close to each other. We did not rearrange these percentages to make the dataset more realistic towards included topics (Targets).

TABLE II. DISTRIBUTIONS OF THE ANNOTATION CLASSES AND THE WORDS NUMBER IN OUR DATASET

Sentiment		Number of Words in the Target	
Positive	37.3%	One	30.9%
Negative	28.0%	Two	66.3%
Neutral	34.7%	Three	2.7%

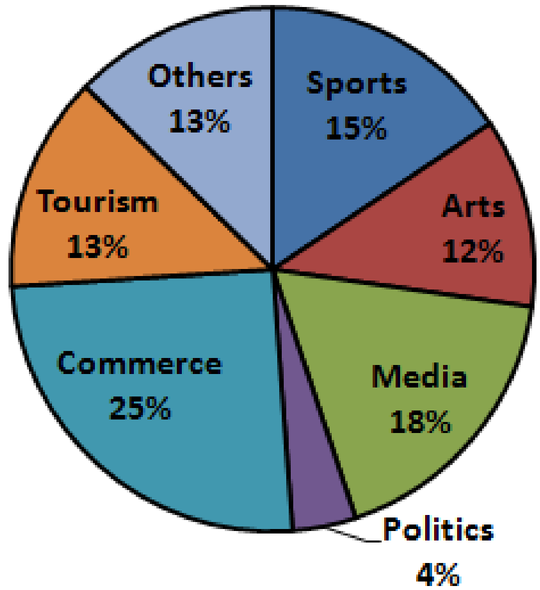
4. Dataset Statistics

In this section, we present general statistics and facts related to our presented dataset. Firstly, an overview of the dataset is introduced followed by results of our manual analysis.

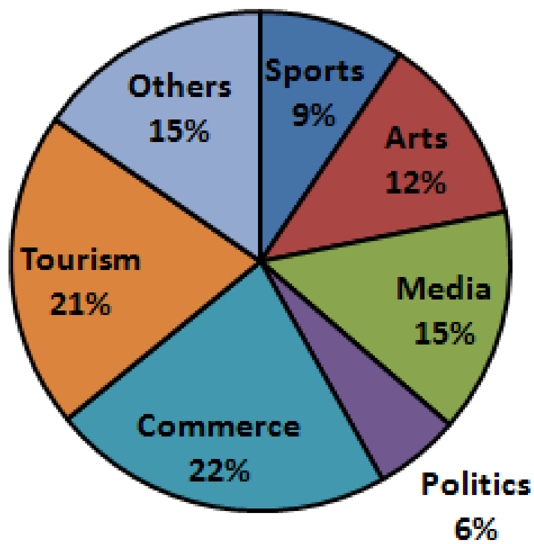
A. Dataset Overview

The AT-ODTSA dataset consists of a total of 3000 samples of tweets written in Arabic language. The reason behind this relatively small dataset size is that the authors released the collected tweets in the AT-ODTSA dataset as the first version and they intend to extend their work to add thousands of new tweets to the AT-ODTSA dataset in its next version. In addition to that, these tweets are collected through a systematic process by three independent experts. The tweet collection process lasted over three months and the chronological range of the collected tweets varies from January, 2011 to March, 2021. The tweets are collected to cover most of the Arabic countries in the Middle East region. Moreover, the AT-ODTSA dataset is dedicated for Arabic sentiment analysis purposes, and to accomplish this goal three features are included in the dataset, namely, Tweet, Target, and Class. The Tweet feature contains the text of the tweet in Arabic language, Target feature is the topic of the tweet, and the Class feature is deemed to be negative "-1", neutral "0", or positive "1".

In order to cover a wide range of topics in the AT-ODTSA dataset, the Target feature contains 234 different topics. These topics can be merged into seven major categories as illustrated in TABLE III. From TABLE III, it can be noticed clearly that the AT-ODTSA dataset is rich of a wide range of topics belonging to various interests and domains. Furthermore, all categories have a representative amount of tweets and topics inside. Figure 3 depicts the distribution of tweets and topics for all categories.



(a)



(b)

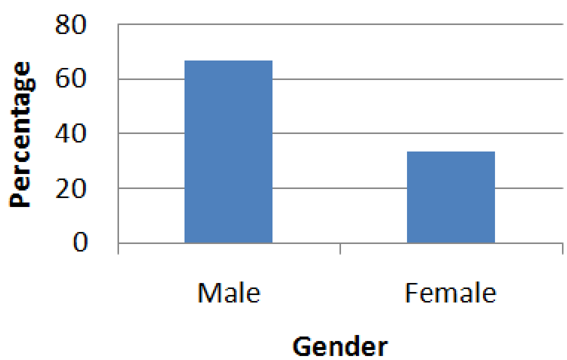
Figure 3. Distribution of: (a) Tweets, and (b) Topics.

B. Manual Analysis

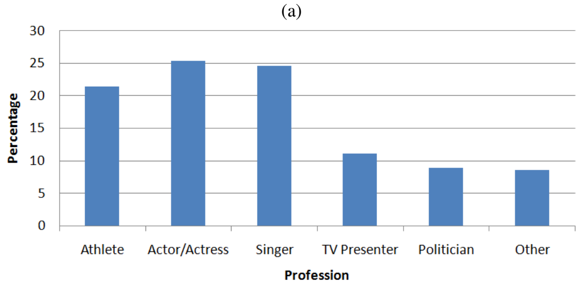
The aim of the manual analysis is to gain deep insights and understand the characteristics of the new dataset. The analysis focused only on the following issues: diversity, the language being used, and the sentiment expression. To emphasize the diversity of the AT-ODTSA dataset, a subset of tweets, which is related to the celebrities, are analyzed thoroughly. Figure 4 shows the distribution of these tweets according to celebrities' gender, profession, and nationality. Therefore, the AT-ODTSA dataset is not only diverse in topics but also it has a great deal of variety in its low level.

TABLE III. STRUCTURE OF THE AT-ODTSA DATASET.

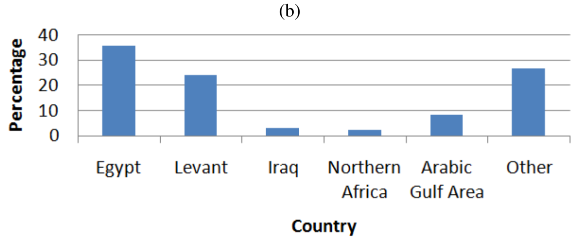
Categories	Number of Topics	Number of Tweets	Size of Tweets (%)	Subcategories
Sports	22	468	15.6	Soccer, Basketball players, Tennis players,Formula 1 drivers, Technical directors, Matches, Cups, Sport clubs.
Arts	29	350	11.67	Actors, Actresses, Singers, Artists.
Media	34	528	17.6	TV Presenters, Social media influencers, Songs,TV channels, TV series, Cinema movies, TV show, Theatre.
Politics	13	118	3.93	Politicians, News, Public figures.
Commerece	52	755	25.17	Companies, Brands, Products, Restaurants, Banks, Shopping malls.
Toursim	48	400	13.33	Historical and famous places, Islands, Cities, Resorts, Towers, Streets, Parks.
Others	36	381	12.7	Trending events, Animals, Institutes, Hospitals,Foods, Drinks, Flowers, Opinions.
Total	**234**	**3000**	**100**	



(a)



(b)



(c)

Figure 4. Distribution of tweets for the celebrities according to: (a) gender, (b) profession, and (c) country.

TABLE IV presents information about the language being used as well as the sentiment expression in the collected tweets. Regarding the language being used in tweets, there are two languages, i.e., Dialectal Arabic (DA) and Formal Arabic (FA). Obliviously, most of the tweets are written in the DA which indicates that users tend to use the DA rather than the FA when they discuss daily matters and topics. Finally, most tweets are expressed explicitly which indicates that users prefer to express their ideas and opinions explicitly to interact with others easily. However, a considerable amount of tweets is expressed implicitly which is also a common case in the Arabic language.

TABLE IV. PERCENTAGES OF THE USED LANGUAGES AND SENTIMENT EXPRESSION.

Used language		Sentiment expression	
DA	83.57	Explicit	75.31
FA	16.43	Implicit	24.69

5. Validation Process

To validate the annotation process, we used two validation rounds. The first round was conducted after collecting 67% of the dataset. After reporting results of this round, the annotators added more rules to unify the strategy in annotating the tweets. Additionally, the annotators fixed all labels that override the rules. On the other hand, the second round of validation was conducted after collecting the whole dataset. Each validation round included 300 tweets which are selected randomly from the dataset. Equally important, distinct tweets are used in the first and second rounds. TABLE V shows the results of both rounds. Figure 5 shows the mechanism of the validation process applied in this study.

Two evaluation measures for reporting the results of

TABLE V. RESULTS OF APPLYING THE FIRST AND SECOND ROUNDS OF VALIDATION.

First Round				Second Round			
Targets		Sentiments		Targets		Sentiments	
Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score	Accuracy	F1-Score
0.72	0.84	0.87	0.93	0.89	0.94	0.91	0.95

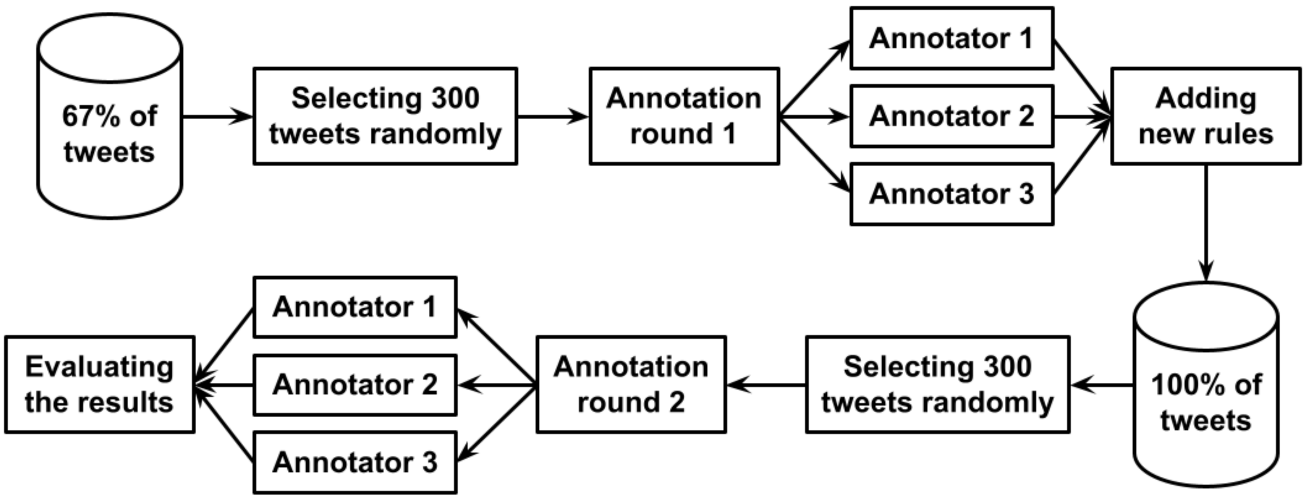


Figure 5. Mechanism of the manual validation process.

validation are utilized, namely, accuracy and F1-Score [40]. The accuracy is the ratio of all samples (labels) that are annotated with same value by all annotators (a complete agreement with all annotators). It is calculated using the following formula [41]:

$$Accuracy = \frac{Complete\,Agreement\,Samples}{All\,Samples} \quad (1)$$

The F1-Score which is also called F-score or F-measure is calculated using the precision and recall values of the test. The recall metric (also known as sensitivity or true positive rate) is the number of samples that are classified correctly as positives divided by the number of all positive samples. The precision metric is the number of samples that are classified correctly as positives divided by the number of all samples that are classified as positives. The best and worst values for the F1-Score metric are 1 and 0, respectively. It is calculated using Equation (2) [41].

$$F1\_Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2)$$

It is worth noting here that the F1-Score metric is basically used with binary classification (positive and negative classes) and there are different modifications [42]. In this study, the macro-average F1-Score formula is exploited since it is a straightforward method and it can be computed simply by averaging the recall and precision metrics on different sets.

Based on findings in TABLE V, it can be concluded that the second round provides fewer errors in comparison with the first round. This note shows that the accuracy of results is high and there is a harmony between the annotators.

6. Conclusions and Future Work

This paper offers an eye opener to researchers on research gaps in the specific area of Arabic sentiment analysis within the domain of text mining. The AT-ODTSA dataset is presented which contains Arabic tweets to address the problem of open-domain targeted sentiment analysis. The collected tweets were annotated with targets (topics) for positive, negative, and neutral classes. The AT-ODTSA dataset has also contributed to apply entity recognition through real Arabic tweets. Furthermore, a detailed statistical analysis of the dataset is reported in this paper and the annotations process is validated manually as well. The authors are currently working on using the AT-ODTSA dataset to propose a new model for Arabic open-domain targeted sentiment analysis. Future work will also consider working on developing actuate systems for Arabic open-domain targeted sentiment analysis.

References

[1] K. Rasch, "Smart assistants for smart homes," Ph.D. dissertation, KTH Royal Institute of Technology, 2013.

[2] M. Chan, E. Campo, D. Estève, and J.-Y. Fourniols, "Smart homes—current features and future perspectives," *Maturitas*, vol. 64, no. 2, pp. 90–97, 2009.

[3] I. Lee and K. Lee, "The internet of things (iot): Applications, investments, and challenges for enterprises," *Business Horizons*, vol. 58, no. 4, pp. 431–440, 2015.

[4] A. M. V. Ramadas, "Smart-homes activity pattern recognition: A comparative study," Ph.D. dissertation, UNIVERSIDADE DO PORTO, 2018.

[5] B. Pang, L. Lee *et al.*, "Foundations and trends® in information retrieval," *Foundations and Trends® in Information Retrieval*, vol. 2, no. 1-2, pp. 1–135, 2008.

[6] A. Shoufan and S. Alameri, "Natural language processing for dialectical arabic: A survey," in *Proceedings of the second workshop on Arabic natural language processing*, 2015, pp. 36–48.

[7] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up? sentiment classification using machine learning techniques," *arXiv preprint cs/0205070*, 2002.

[8] A. Ghallab, A. Mohsen, and Y. Ali, "Arabic sentiment analysis: A systematic literature review," *Applied Computational Intelligence and Soft Computing*, vol. 2020, 2020.

[9] I. Gupta and N. Joshi, "Real-time twitter corpus labelling using automatic clustering approach," *International Journal of Computing and Digital Systems*, vol. 10, pp. 519–532, 2021.

[10] R. Dehkharghani, B. Yanikoglu, Y. Saygin, and K. Oflazer, "Sentiment analysis in turkish at different granularity levels," *Natural Language Engineering*, vol. 23, no. 4, pp. 535–559, 2017.

[11] S. I. Abudalfa and M. A. Ahmed, "Deep learning for target-dependent sentiment classification in social media," in *Smart Cities Symposium 2018*. IET, 2018, pp. 1–5.

[12] S. Abudalfa, "Comparative study on efficiency of using supervised learning techniques for target-dependent sentiment polarity classification in social media," *International Journal of Computing and Digital Systems*, vol. 7, no. 03, pp. 155–160, 2018.

[13] S. Abudalfa and M. A. Ahmed, "Open domain targeted sentiment classification using semi-supervised dynamic generation of feature attributes." *J. Univers. Comput. Sci.*, vol. 24, no. 11, pp. 1582–1603, 2018.

[14] AT-ODTSA Dataset, "GitHub," 2021. [Online]. Available: https://github.com/sabudalfa/AT-ODTSA

[15] M. Orbach, O. Toledo-Ronen, A. Spector, R. Aharonov, Y. Katz, and N. Slonim, "Yaso: A targeted sentiment analysis evaluation dataset for open-domain reviews," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, 2021, pp. 9154–9173.

[16] M. Hu, Y. Peng, Z. Huang, D. Li, and Y. Lv, "Open-domain targeted sentiment analysis via span-based extraction and classification," *arXiv preprint arXiv:1906.03820*, 2019.

[17] M. Mitchell, J. Aguilar, T. Wilson, and B. Van Durme, "Open domain targeted sentiment," in *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing*, 2013, pp. 1643–1654.

[18] O. Kotevska and A. Lbath, "Sentiment analysis of social sensors for local services improvement," *International Journal of Computing and Digital Systems*, vol. 6, no. 04, pp. 165–174, 2017.

[19] S. Abudalfa and M. Ahmed, "Survey on target dependent sentiment analysis of micro-blogs in social media," in *2017 9th IEEE-GCC Conference and Exhibition (GCCCE)*. IEEE, 2017, pp. 1–1.

[20] I. A. Farha and W. Magdy, "A comparative study of effective approaches for arabic sentiment analysis," *Information Processing & Management*, vol. 58, no. 2, p. 102438, 2021.

[21] S. O. Alhumoud and A. A. Al Wazrah, "Arabic sentiment analysis using recurrent neural networks: a review," *Artificial Intelligence Review*, pp. 1–42, 2021.

[22] N. Al-Twairesh, H. Al-Khalifa, A. Al-Salman, and Y. Al-Ohali, "Arasenti-tweet: A corpus for arabic sentiment analysis of saudi tweets," *Procedia Computer Science*, vol. 117, pp. 63–72, 2017.

[23] E. Refaee and V. Rieser, "An arabic twitter corpus for subjectivity and sentiment analysis." in *LREC*, 2014, pp. 2268–2273.

[24] N. A. Abdulla, N. A. Ahmed, M. A. Shehab, and M. Al-Ayyoub, "Arabic sentiment analysis: Lexicon-based and corpus-based," in *2013 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)*. IEEE, 2013, pp. 1–6.

[25] K. A. Kwaik, S. Chatzikyriakidis, S. Dobnik, M. Saad, and R. Johansson, "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*, 2020, pp. 1–8.

[26] B. Alharbi, H. Alamro, M. Alshehri, Z. Khayyat, M. Kalkatawi, I. I. Jaber, and X. Zhang, "Asad: A twitter-based benchmark arabic sentiment analysis dataset," *arXiv preprint arXiv:2011.00578*, 2020.

[27] A. Mohammed and R. Kora, "Deep learning approaches for arabic sentiment analysis," *Social Network Analysis and Mining*, vol. 9, no. 1, pp. 1–12, 2019.

[28] R. Baly, G. El-Khoury, R. Moukalled, R. Aoun, H. Hajj, K. B. Shaban, and W. El-Hajj, "Comparative evaluation of sentiment analysis methods across arabic dialects," *Procedia Computer Science*, vol. 117, pp. 266–273, 2017.

[29] H. Abdellaoui and M. Zrigui, "Using tweets and emojis to build tead: an arabic dataset for sentiment analysis," *Computación y Sistemas*, vol. 22, no. 3, pp. 777–786, 2018.

[30] P. Sobhani, D. Inkpen, and X. Zhu, "A dataset for multi-target stance detection," in *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, 2017, pp. 551–557.

[31] R. Baly, A. Khaddaj, H. Hajj, W. El-Hajj, and K. B. Shaban, "Arsentd-lev: A multi-topic corpus for target-based sentiment analysis in arabic levantine tweets," *arXiv preprint arXiv:1906.01830*, 2019.

[32] M. Rushdi-Saleh, M. T. Martín-Valdivia, L. A. Ureña-López, and

J. M. Perea-Ortega, “Oca: Opinion corpus for arabic,” *Journal of the American Society for Information Science and Technology*, vol. 62, no. 10, pp. 2045–2054, 2011.

[33] A. M. Alayba, V. Palade, M. England, and R. Iqbal, “A combined cnn and lstm model for arabic sentiment analysis,” in *International cross-domain conference for machine learning and knowledge extraction*. Springer, 2018, pp. 179–191.

[34] M. Heikal, M. Torki, and N. El-Makky, “Sentiment analysis of arabic tweets using deep learning,” *Procedia Computer Science*, vol. 142, pp. 114–122, 2018.

[35] M. A. El-Affendi, K. Alrajhi, and A. Hussain, “A novel deep learning-based multilevel parallel attention neural (mpan) model for multidomain arabic sentiment analysis,” *IEEE Access*, vol. 9, pp. 7508–7518, 2021.

[36] M. Abdullah and M. Hadzikadic, “Sentiment analysis on arabic tweets: Challenges to dissecting the language,” in *International Conference on Social Computing and Social Media*. Springer, 2017, pp. 191–202.

[37] L. Dong, F. Wei, C. Tan, D. Tang, M. Zhou, and K. Xu, “Adaptive recursive neural network for target-dependent twitter sentiment classification,” in *Proceedings of the 52nd annual meeting of the association for computational linguistics (volume 2: Short papers)*, 2014, pp. 49–54.

[38] B. Jagla, B. Wiswedel, and J.-Y. Coppée, “Extending knime for next-generation sequencing data analysis,” *Bioinformatics*, vol. 27, no. 20, pp. 2907–2909, 2011.

[39] Twint, “Github,” 2019. [Online]. Available: https://github.com/twintproject/twint

[40] W. Elmasry, A. Akbulut, and A. H. Zaim, “Comparative evaluation of different classification techniques for masquerade attack detection,” *International Journal of Information and Computer Security*, vol. 13, no. 2, pp. 187–209, 2020.

[41] W. Elmasry, A. Akbulut, and A. H. Zaim, “A design of an integrated cloud-based intrusion detection system with third party cloud service,” *Open Computer Science*, vol. 11, no. 1, pp. 365–379, 2021.

[42] S. Puthiya Parambath, N. Usunier, and Y. Grandvalet, “Optimizing f-measures by cost-sensitive classification,” *Advances in Neural Information Processing Systems*, vol. 27, pp. 2123–2131, 2014.

Shaaban Sahmoud Shaaban Sahmoud is currently an assistant professor in computer engineering department at Fatih Sultan Mehmet Vakif University in Istanbul. He received the Bachelor and the Master degrees in Computer Engineering in 2006 and 2012 from the Islamic University of Gaza, Palestine. He finished his PhD in computer engineering from Marmara University, Turkey in 2019. His research interests include dynamic optimization, evolutionary algorithms, multi-objective optimization, computer vision, biometrics, and data mining.

Shadi Abudalfa Shadi Abudalfa received the BSc and MSc Degrees both in Computer Engineering from the Islamic University of Gaza (IUG), Palestine in 2003 and 2010 respectively. In 2013, he joined King Fahd University of Petroleum and Minerals (KFUPM) as a full time student (scholarship) to pursue the PhD degree. He completed his PhD in Computer Science and Engineering (CSE) at KFUPM in 2018. He is an Assistant Professor at the University College of Applied Sciences, Palestine. From July 2003 to August 2004, he worked as a Research Assistant at Projects and Research Lab in IUG. During same period, he worked as a Teaching Assistant at Faculty of Engineering in IUG. Abudalfa is a member of some scientific societies and serves as reviewer through various international conferences and journals. His current research interests include artificial intelligence, data mining, activity recognition, machine learning, and sentiment analysis.

Wisam Elmasry Wisam Elmasry received his B.Sc. and M.Sc. degrees in computer engineering from The Islamic University of Gaza (IUG), Gaza, Palestine in 2004 and 2010 respectively. He earned his Ph.D. (2019) in computer engineering at Istanbul Commerce University in Turkey. He currently holds the position of assistant professor in the Computer Engineering Department at Istanbul Kultur University. Prior to the Istanbul Kultur University, he was a Postdoctoral Researcher in the Electronics and Electrical Engineering Department at Istanbul Sabahattain Zaim University between 2020 and 2021. His areas of research interest include Image Steganography, Cryptography, Cyber Security, Deep Learning, Evolutionary Algorithms, and MANET Routing Protocols Security.