2020 IEEE/ACM 13th International Conference on Utility and Cloud Computing (UCC)

Capturing Public Concerns About Coronavirus Using Arabic Tweets: An NLP-Driven Approach

Mohammed Bahja
School of Computer Science
University of Birmingham
United Kingdom
m.bahja@bham.ac.uk

Rawad Hammad
Department of Computer Science and Digital Technologies
University of East London
United Kingdom
r.hammad@uel.ac.uk

Mohammed Amin Kuhail
College of Technological Innovation
Zayed University
United Arab Emirates
mohammad.kuhail@zu.ac.ae

Abstract—**This In order to analyze the people reactions and opinions about Coronavirus (COVID-19), there is a need for computational framework, which leverages machine learning (ML) and natural language processing (NLP) techniques to identify COVID tweets and further categorize these in to disease specific feelings to address societal concerns related to Safety, Worriedness, and Irony of COVID. This is an ongoing study, and the purpose of this paper is to demonstrate the initial results of determining the relevancy of the tweets and what Arabic speaking people were tweeting about the three disease related feelings/emotions about COVID: Safety, Worry, and Irony. A combination of ML and NLP techniques are used for determining what Arabic speaking people are tweeting about COVID. A two-stage classifier system was built to find relevant tweets about COVID, and then the tweets were categorized into three categories. Results indicated that the number of tweets by males and females were similar. The classification performance was high for relevancy (F=0.85), categorization (F=0.79). Our study has demonstrated how categories of discussion on Twitter about an epidemic can be discovered so that officials can understand specific societal concerns related to the emotions and feelings related to the epidemic.**

Keywords-Natural Language Processing; Arabic Text Mining; Coronavirus; COVID-19.

I. INTRODUCTION

According to the Lancet COVID-19 tracking dashboard [1], more than 11 million confirmed COVID-19 cases worldwide. The vast majority of these cases appeared initially in Mainland China, and some Asian countries including Thailand, Philippines and Korea, and soon spread across the USA and Europe. The rapid spread of this virus is implausible, as of 20th July 2020, 14348858 confirmed cases including 603,691 deahs according to World Health Organization (WHO) [2]. This situation manifested itself in news and social media around the globe. A massive change in the communities' behavior, especially in affected areas, has been noticed. Furthermore, a huge impact is expected on a range of sectors including economy, tourism, stock markets to mention but a few. This paper observes the Arabic community interaction with COVID-19 through the lenses of social media, particularly Twitter.

This in progress research attempts to present an initial result of the following questions: R1. Dataset Distribution Analysis: What proportion of male and female users tweeted about COVID, what were the polarities of the tweets by male and female users, and what were the proportions of tweets that discussed topics related to the different characteristics—Safety, Worry, Irony?

R2. Classification Performance Analysis: What was the agreement among annotators' labels that were used as the ground truth in this study, what was the classification performance to detect the tweets relevant to COVID, and how well were the classifiers able to distinguish between tweets on the different disease characteristics?

The rest of this paper is organized as follows. Section Two presents brief literature review about COVID-19 and related technologies that can help, i.e., Natural Language Processing (NLP). Section Three presents the research method used to carry out this research. Section Four presents the initial results from this study. Section five presents a discussion, followed by limitations (Section 6) and conclusion (Section 7).

II. LITERATURE REVIEW

This section of the paper presents brief facts about COVID-19 and its history and alludes to Natural Language processing as a technology used to analyze related tweets in Arabic language.

A. COVID-19 in brief

The outbreak of the novel coronavirus, i.e., COVID-19, in China has spread to many other countries. It is one of the types of coronavirus, which are common across the world. Due to the recency of this outbreak, there is a limited research published around it. Delineating the cause of this virus from a biological perspective is beyond the purpose of this paper. Nevertheless, we are trying to look into the impact of this virus on a specific area through the lenses of tweets published by internet users. The virus breakout started in December 2019 not very far from 30 January 2020 when the WHO issued a global health emergency call based on the continuously evolving cases reported the distribution of confirmed cases as shown in Figure 1 below. The global death toll is also jumping rapidly according to internationally-recognized report, i.e., Johns Hopkins University confirmed the death of 602,507 as of 20th July 2020 [3].

978-0-7381-2394-3/20/$31.00 ©2020 IEEE
DOI 10.1109/UCC48980.2020.00049

Figure 1: Distribution of COVID-19 cases as of 2nd June 2020

Initial investigations express that COVID is not so different from similar viruses and infectious diseases, however its impact is growing exponentially from many perspectives. On one hand, from health perspectives, studies are not mature enough to judge its full effect. For instance, a recent clinical study [4] aimed to evaluate the clinical characteristics of COVID-19 in pregnancy and the intrauterine vertical transmission potential of COVID-19 infection. Nine livebirths were recorded in the study and no neonatal asphyxia was observed in new-born babies. However, these finding remain initial due the insufficient sample, i.e., only nine pregnant.

Figure 2: Coronavirus impact on economy: USA perspective [5]

On the other hand, COVID impact on other perspectives such as economy, tourism, businesses are immensely rising. World Economic Forum [5] identified the economic effects of COVID-19 around the world. Figure 2 reflects how severe this time in comparison with other financial crises: USA as an example. As of 14th May, 2020, the US Treasury Department has said it plans to borrow nearly $3 trillion in the second quarter of 2020 to mitigate the risk of COVID-19. This also affecting the whole world according to the report published by Deloitte Insights and Wall Street Journal report on 11th May 2020 indicated that four different challenging scenarios are expected as a result of COVID-19 global impact [6].

While the studies have identified major impact of COVID-19 on socio-economic aspects, the focus on mental health of the people due to vast changes in lifestyles resulting out of lockdowns, physical distancing, and other containment strategies; and the resulting economic breakdown could increase the risk of mental health problems [7]. Therefore, it is very much essential that the public mental health need to be assessed at regular intervals in order to analyze the impact of pandemic. Innovative technologies such as ML, AI, and NLP etc. can be used to analyze the people's feelings and opinions using their interactive data on online platforms.

B. NLP and Public Opinions

Public opinions can be identified on various platforms, especially social media platforms such as Facebook and Twitter. Public opinions carry a sense of psychological expressions, which reflect the emotions, attitudes of the public towards an event or topic. NLP in this context can be very effective for analyzing sentiments by mining hat mines the data for emotional cues based on predefined keywords and ascertains the polarity of the public view on the topic [8-10]. Accordingly, opinion mining and sentimental analysis using NLP and various classification algorithms have shown greater accuracies in classification of opinions [11]. Thus, recent studies have identified NLP to be effective in the analysis and classification of public opinions, deriving meaningful outcomes with high levels of accuracy.

C. NLP and Epidemiology

Electronic Health Records, diagnosis or symptom descriptions and patient information sheets contain free text, which may be challenging for the analysts to analyze the information. Classification of the information from such large datasets of free text can be simplified by the use of NLP, which can provide accurate and quality information from the free text in relation to the various contexts, limiting the drawbacks of manual and conventional diagnostic coding. NLP has been gaining popularity in the healthcare, considering its efficiency in analyzing free text (which is increasing in formats of case sheets, health records, monitoring reports, diagnosis text etc.), and providing highly accurate information [12]. NLP is used in the study of the epidemiology of various diseases or disorders, such as Allergic Drug Reactions [11], psychiatric disorders [13], Statin Side Effects, angina pectoris and cancer. Thus, recent studies have shown that NLP is being increasingly adopted in studying the epidemiology of various disorders, with classification of the unstructured and analyzing it to make meaningful outcomes, which can help healthcare practitioners in taking quality based efficient decisions.

NLP is also being integrated with other technologies such as AI and ML for managing covid-19. Recently an automatic framework using deep learning model called as Covid-19 detection neural network (COVNet) to extract visual details from chest scans for the detection of Covid-19 [18]. Predictive models were identified to be increasingly adopted in various studies to aid decision-making in various aspects such as diagnosis, tracking, research etc. However, the results of these models include high risk of bias [14]. Therefore, there is a need to develop more rigorous models and validate the existing ones. However, using predictive models for analyzing public opinions revealed promising results. For example, a recent study [15] has achieved 91% and 74% accuracy using Naive Bayes and logistic regression classification respectively for shorter online tweets. However, both methods reflected poor accuracy for longer tweets. A similar study [16] that has analysed 1.9 million Tweets related to coronavirus collected from January 23 to March 7, 2020 through sentiment analysis identified that fear related to the unknown nature and information about Covid-19 was dominant in all topics. Similarly, another recent study [17], trained deep models to classify each tweet into different emotions including anger, anticipation, disgust, fear, joy, sadness, surprise and trust using BERT classifier and analyzed that emotions change among different regions and cultures. NLP techniques in this context of analyzing tweets in different languages can be a reliable and effective solution. Thus, there is a wider scope for integrating NLP and other techniques such as ML, AI, and deep learning for developing more effective predictive models which can accurately predict the mental health of the public.

D. *NLP for Arabic Language*

The focus on processing Arabic texts and speech has been one of the primary areas of research among Arabic NLP community. Due to its complexity and intrinsic variations, research is not only being carried out on Modern Standard Arabic but also on various Arabic dialects spoken in other regions such as North Africa, the Middle East, Pakistan etc. Research focusing on NLP and Arabic, mostly include machine translation, corpus compilation, parsing of Arabic dialects, and Arabic dialect identification [18,19] developed system to differentiate Modern Standard Arabic and Egyptian Arabic using a linear SVM classifier and achieved 85.5% accuracy. In a similar study by Tillmann et al. [20] achieved 89.1% accuracy. Similarly, Malmasi et al. [18] analyzed performance of different methods to discriminate Modern Standard Arabic and five other Arabic dialects including Egyptian, Jordanian, Palestinian, Syrian, and Tunisian using meta-classifier and achieved 74% accuracy. Lexical, morphological, and syntactic features were identified to be important factors for identifying and discriminating various dialects [21]. Increasing research in relation to Arabic NLP has led to various developments such as Ara Vec, a set of Arabic word embedding models [22], improved generalization of Arabic text classifiers [23], use of metaheuristic algorithms, and integration of NLP with Deep Learning and Machine Learning. Though NLP in Arabic text processing has been achieving good results in various studies using different procedures and methods, there are few challenges associated with Arabic, which are discussed in the following section.

E. *Challenges in NLP for Arabic Language*

Arabic language can be considered more complex than English, as it does not possess vowels; rather, diacritics are placed above or below letters. This characteristic of Arabic involves both structural and lexical ambiguity, as various diacritics may lead to various meanings [23]. Other complexities such as the use of dots (used for differentiating letters) in Arabic text, design of letters, vocabulary (leading to different meanings), absence of capital letters etc. makes it more challenging for developing text translators and classifiers [24]. To conclude, Arabic is a Semitic language, which contrasts from Indo-European lingos phonetically, morphologically, syntactically and semantically [25].

III. Research Methods

A combination of NLP and ML techniques were used in this study for determining what information about COVID related information such as safety, worry, and irony, which were being discussed by the people on Twitter. A 2-stage classifier system was built for finding relevant tweets on COVID and then categorizing these into three disease categories: safety, worry, and irony.

A. *Dataset Distribution*

Data Collection: Tweets were collected from 16 February, 2020 and 10th July, 2020, resulting in 782,391 tweets using Twitris 2.0. COVID has been one of the trending topics on twitter during January and February 2020, as the disease has been progressive in the Mainland China, with increasing number of deaths and rapidly spreading across the world. Streaming application program interface (API) from Twitris system [26] was used to collect the tweets in Arabic. Initially we used nine keywords in Arabic to identify the tweets related to COVID, which included the following:
#كورونا, #فيروس_كورونا, #فيروس_كورونا_المستجدّ,
#كورونا_فيروس, #فيروس_كورونا_الصين, #فيروس_كورونا_الجديد,
#فيروس كورونا, #كوفيد_19

Initial search used the term 'COVID' in Arabic corresponds to '*كورونا*', which resulted in capturing large number of tweets. Later using the semantic concept, two terms were utilized, 'COVID' and 'COVID Virus' corresponds to "*فيروس_كورونا*' in Arabic, which has significantly improved the quality of tweets.

Pre-Processing: After the initial search, 782,391 were collected. All retweeted tweets, duplicated tweets were deleted; all hashtags are cleaned; Arabic stopping words were deleted; link and search keys are cleaned as well. The number of tweets after refining, deleting, and cleaning included 365,498 tweets. We found lots of duplicated tweets and this is mainly due to the fact that people are campaigning on Twitter to raise awareness about the issue so that they can take precautions. They usually prepare a tweet list which they share among themselves and start massive tweeting in order to create TREND.

Labelling Process and Data Annotation: We have tweets to the four themes including: safety, worry, irony, and news, as explained below:

- Safety: tweets about safety and whether people feel safe or not. Also, if any tweets relevant to health instruction or general instruction about protection from the viruses. (1 'Yes', 0 'No')

- Worry: tweets about any general concerns and worrying (1 'Yes', 0 'No')
- Irony: tweets that contains Humour and/or Sarcasm (1 'Yes', 0 'No').
- News: tweets from new agencies in Classic Arabic language (1 'Yes', 0 'No'). We included this theme to clean any News agencies tweets as we mainly interested in public opinions and concerns. We have found that there are large number of tweets and duplicated tweets come from News agencies.

Around 150k tweets were identified under three themes including safety, worry and irony.

B. Classification Performance

Unclassified Arabic tweets in all categories were initially cleaned, and tweets were extracted from xls file to text file. Models were then built using TfidfVectorizer - WordBagging and train Test split for the corpus. Supervised classification techniques including LinearSVC (SVC), Multinomial Naive Bayes (MNB), DecisionTreeClassifier (DTC) were implemented on COVID dataset for: (1) classifying whether a tweet was relevant or nonrelevant, and (2) if relevant, further categorizing the tweets into the disease characteristics. Supervised techniques rely on labelled data, in this case tweets that are manually labelled as relevant to COVID virus, as well as the category it belongs to: COVID issues, COVID safety, COVID worry, and COVID irony. They "learn" the nature of the tweets in the different groups and subgroups. The class of News was used to classify if the tweets were relevant or irrelevant as we do not consider news in this study, as it only reports the incidents but do not show public opinions.

IV. Results

A. Dataset Distribution (Addressing RQ1)

Overall, 34.5% (126,002/365498) of tweets contained a retweet and 66.5% (242967/65498) contained a URL. Figure 3 shows the number of tweets classified in each category according to the best classification model multinomial Naïve Bayes.

Tweets by gender were found by using twitter usernames using genderize API [27]. According to genderize, 35.5% (129897/365498) of the tweets were posted by males, 30.3% (111023/365498) were posted by females, and 34.1% (124578/365498) were by unknown gender. Class imbalance in the categories was identified (Figure 3). As there is no treatment (vaccine) for COVID, and due its nature of progress and spreading across the world, most of the tweets identified were related to worry and irony, reflecting negative opinions of the people.

B. Classification Performance

In the first stage of the categorization process for the ground truth tweets, tweets were first classified as being relevant or not relevant to COVID. Tweets that were relevant were then categorized as being about Safety, Worry, and Irony. To train the classifiers and evaluate their performance, 151,437 tweets were manually labelled.

The distribution of relevant themes in these three categories is shown in figure 4. Observing figures 3 and 4, it can be analyzed that the distribution of the labelled gold standard dataset was similar to the distribution of the large data corpus, except for a larger portion of tweets related to Safety.

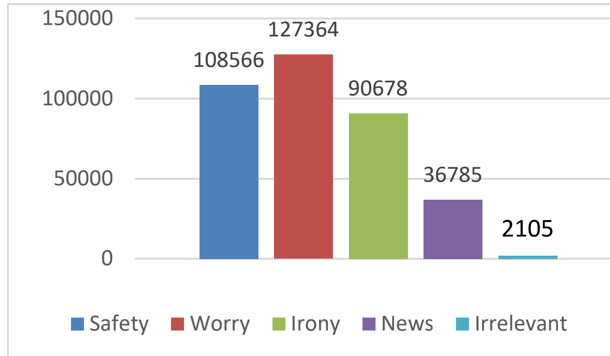

Figure 3. COVID-19 tweets classified.

Interrater Reliability: Fleiss kappa values for relevant or not was .76. Fleiss kappa values for Safety, Worry, and Irony were .94, .93, and .88, respectively. This indicates substantial to almost perfect agreement among the raters [28]. Given substantial interrater reliability, a model needed to be built based on the gold standard dataset.

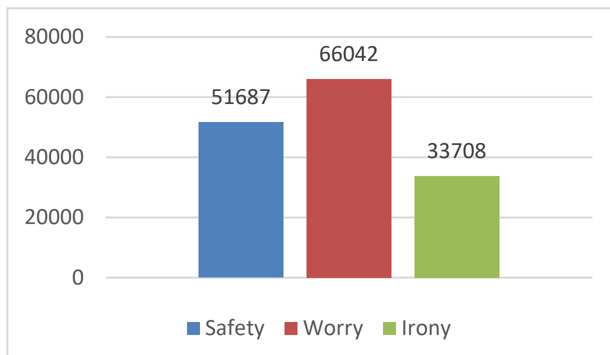

Figure 4. Number of tweets from the labelled dataset

Classification and Performance: The performance of different classifiers on the 151437 pre-processed twitter data to find the relevancy of the tweets towards COVID is presented in Table 1.

TABLE 1. CLASSIFIERS PERFORMANCES FOR DETECTING RELEVANT TWEETS

Classifier	TP	FP	Precision	Recall	F1 score	AUC
SVC	.821	.395	.811	.824	.83	.89
MNB	.880	.360	.881	.891	.85	.94
DTC	.834	.471	.832	.853	.79	.85

TP: True Positive; FP: False Positive; AUC: Area Under Curve

Weka toolbox [29] was used to extract unigram features. Considering the AUC values ranging from 0.85 to 0.94, it can be identified that classifiers performed well on the dataset. As identified from table 1, MNB outperformed other classifiers based on the F-measure (0.85) and AUC (0.94). As the length of the tweets (data sets that have a large variance in document length) in the dataset are large, MNB classifiers performed well [30]. The class imbalance was affecting the classifier performance. Although the AUC value was high (0.94), the classifier predicted a tweet was relevant more often than not relevant as 87.6% (8435/9630) of the tweets belonged to the relevant category.

The performance of different classifiers on the 8435 pre-processed twitter data to find the relevancy of the tweets towards COVID and to find categorical classification (Safety, Worry, Irony) presented in Table 2. Again, the classifiers performed well with AUC values ranging from 0.82 to 0.95. With this dataset, MNB outperforms other classifiers again.

TABLE 2. CLASSIFIERS PERFORMANCES FOR DETECTING THREE CATEGORIES WITHIN RELEVANT TWEETS

Classifier	TP	FP	Precision	Recall	F1 score	AUC
SVC	.682	.123	.705	.698	.695	.821
MNB	.785	.085	.797	.796	.795	.952
DTC	.633	.121	.763	.635	.730	.863

From the analysis of above results, it can be identified that two-stage classifier system was found to be having high precision and recall performance for categorizing the tweets into relevant and not relevant, and further classifying the relevant tweets into the three categories.

V. DISCUSSION

Similarities between males and females was identified with respect to number of tweets, with majority of tweets being negative. In addition, the two-stage classifier performed well at two levels: relevancy and categorization. Though there are a greater number of "safety" tweets identified, it was observed that the number of "worry" tweets were higher than expected. Some examples of tweets which are classified as "worry" include: " يجب وقف التخويف والفزع والتعامل مع اي ازمه باحترافيه وتوازن وتماسك ، ويجب التعامل مع السلطات بالاستجابه ، ويجب على السلطات التعامل بحزم ، وعلى وسائل الاعلام والمغردين نشر تعليمات وارشادات وتهدئه النفوس.". Words such as "تخويف", "فزع", "نفوس" could be the reason why some tweets were classified as worry.

Focusing on the classification analysis, it is interesting to identify that MNB classifier outperformed the other more popular classifiers in text analytics. Naive Bayes is one of the simplest classification models available to us, but it is nonetheless among the most effective for this dataset. This result is non intuitive but not surprising when we consider that using text for classification is relatively imprecise compared with other types of data.

VI. LIMITATIONS

Although we adopted trustworthy methods for data collection and analysis, there are few limitations which are identified in this study. Firstly, only Arabic language tweets are considered in this study, which certainly limits the strength of this study. Secondly, we used key words COVID, COVID VIRUS, COVID SAFETY etc. However, as COVID is trending on twitter in the recent past and it is non-treatable, we could not identify large number of tweets related to safety. Thirdly, only 65.9% (240920/365498) of the tweets were labelled by the gender API using the profile name, limiting the gender analysis of complete list of collected tweets.

VII. CONCLUSION AND FUTURE WORK

Considering the findings in relation to gender constraints, it can be concluded that the proportion of tweets made by the female and male Arabic twitter users were similar in terms of number of tweets. Though the majority of tweets were about "worry", it was identified that "safety" tweets were more than expected, which might be due to the words such as "حماية", "تعليمات", "نظافة" in the collected text or tweets. There were not many tweets about irony, and we think this is due to the seriousness of the event of Coronavirus. The classification performance was high for relevancy (F=0.85), categorization (F=0.79). This is one of the first studies to report successful creation of an automated content classification tool to analyze COVID-related tweets, specifically in the area of epidemiology. Through analysis of Arabic public opinions, this model would help advance the field's technological and methodological capabilities to harness social media sources for disease surveillance research. This work will be further developed to detect fake news that might impact the validity of information published on social media [31]. Also, there is a plan to generalize this model to capture opinion outside pandemic domain, e.g., measuring the impact of using educational chatbots [32] on students' performance or the role of online learning to substitute school learning.

REFERENCES

[1] Acharya, R., & Porwal, A. (2020). A vulnerability index for the management of and response to the COVID-19 epidemic in India: an ecological study. The Lancet Global Health. doi: 10.1016/s2214-109x(20)30300-4.

[2] World Health Organization (2020). WHO Coronavirus Disease (COVID-19) Dashboard. Available at: https://covid19.who.int/ [Accessed: 20 July 2020]

[3] Coronavirus 2019-nCoV, CSSE. Coronavirus 2019-nCoV Global Cases by Johns Hopkins CSSE., Available: https://coronavirus.jhu.edu/map.html [Accessed: 20 July 2020]

[4] Chen, H., Guo, J., Wang, C., Luo, F., Yu, X., Zhang, W., Li, J., Zhao, D., Xu, D., Gong, Q., Liao, J., Yang, H., Hou, W., Zhang, Y. (2020). Clinical characteristics and intrauterine vertical transmission potential of COVID-19 infection in nine pregnant women: a retrospective review of medical records. The Lancet.

[5] World Economic Forum (2020) The economic effects of COVID-19 around the world. Available: https://www.weforum.org/agenda/2020/02/coronavirus-economic-effects-global-economy-trade-travel [Accessed: 1 June 2020]

[6] Deloitte Insights, (11 May 2020) The world remade by COVID-19 https://www2.deloitte.com/global/en/pages/about-

deloitte/articles/covid-19/covid-19-scenarios-and-impacts-for-business-and-society-world-remade.html?id=us:2el:3dp:wsjspon:awa:WSJCIO:2020:WSJFY20 [Accessed: 20 May 2020]

[7] Moreno, C., Wykes, T., Galderisi, S., Nordentoft, M., Crossley, N., & Jones, N. et al. (2020). How mental health care should change as a consequence of the COVID-19 pandemic. The Lancet Psychiatry. doi: 10.1016/s2215-0366(20)30307-2

[8] Wen H., Lin P., Hu Y. (2012) Analysis and Design of Internet Monitoring System on Public Opinion Based on Cloud Computing and NLP. In: Wang F.L., Lei J., Gong Z., Luo X. (eds) Web Information Systems and Mining. WISM 2012. Lecture Notes in Computer Science, vol 7529. Springer, Berlin, Heidelberg

[8] Bahja, M., & Razaak, M. (2018). Automated Analysis of Patient Experience Text Mining using a Design Science Research (DSR) Approach. The Eighth International Conference on Business Intelligence and Technology. ISBN: 978-1-61208-614-9

[9] Rajwadi, M., Glackin, C., Wall, J., Chollet, G., Cannings, N. (2019) Explaining Sentiment Classification. Proc. Interspeech 2019, 56-60, DOI: 10.21437/Interspeech.2019-2743.

[10] Aich, Animikh & Hegde, Chetana & Krishna, Akshay & .V, Akhilesh. (2019). Analysis of Customer Opinion Using Machine Learning and NLP Techniques.

[11] Bahja, M., & Lycett, M. (2016). Identifying patient experience from online resources via sentiment analysis and topic modelling. Proceedings Of The 3Rd IEEE/ACM International Conference On Big Data Computing, Applications And Technologies - BDCAT '16. doi: 10.1145/3006299.3006335

[12] Qiuyue, Z. (2018). Natural Language Processing in Perinatal Epidemiology. [online] Available at: https://dash.harvard.edu/handle/1/37945648 [Accessed: 29 Feb 2020]

[13] Alexander, T. (2018). Natural Language Processing to Study Epidemiology of Statin Side Effects. [online] Available at: https://grantome.com/grant/NIH/RC1-LM010460-01 [Accessed: 29 Feb 2020]

[14] Wynants, L., Van Calster, B., Collins, G., Riley, R., Heinze, G., & Schuit, E. et al. (2020). Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ, m1328. doi: 10.1136/bmj.m1328

[15] Samuel, J., Ali, G., Rahman, M., Esawi, E., & Samuel, Y. (2020). COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification. Information, 11(6), 314. doi: 10.3390/info11060314

[16] Xue, J., Chen, J., Chen, C. Zheng, C., Li, S., Zhu, T. (2020). Public discourse and sentiment during the COVID-19 pandemic: using Latent Dirichlet Allocation for topic modeling on Twitter. Cornell University. https://arxiv.org/abs/2005.08817

[17] Li, I., Li., Y., Li, T., Suzumura et al. (2020). What are We Depressed about When We Talk about COVID19: Mental Health Analysis on Tweets Using Natural Language Processing. Cornell University. https://arxiv.org/abs/2004.10899

[18] Malmasi, S., Refaee, E., & Dras, M. (2015). Arabic Dialect Identification using a Parallel Multidialectal Corpus. In Proceedings of PACLING

[19] Malmasi, S., & Zampieri, M. (2016). Arabic Dialect Identification in Speech Transcripts. Proceedings of the Third Workshop on NLP for Similar Languages, Varieties and Dialects, pages 106–113, Osaka, Japan.

[20] Tillmann, C., Mansour, S., & Al-Onaizan, Y. (2014). Improved Sentence-Level Arabic Dialect Classification. In Proceedings of the VarDial Workshop.

[21] Darwish, K., Sajjad, H. & Mubarak, H. (2014). Verifiably Effective Arabic Dialect Identification. In Proceedings of EMNLP.

[22] Soliman, A., Eissa, K., & El-Beltagy, S. (2017). AraVec: A set of Arabic Word Embedding Models for use in Arabic NLP. Procedia Computer Science, 117, 256-265. doi: 10.1016/j.procs.2017.10.117

[23] Khaddaj, A., Hajj, H., & El-Hajj, W. (2019). Improved Generalization of Arabic Text Classifiers. Proceedings of the Fourth Arabic Natural Language Processing Workshop, pages 167–174 Florence, Italy.

[24] Babych, B., & Hartley, A. (2003). Improving machine translation quality with automatic named entity recognition, in Proc. EACL-EAMT, 2003, pp. 1–8.

[25] Shaalan, K., Siddiqui, S., Alkhatib, M., & Abdel Monem, A. (2018). Challenges in Arabic Natural Language Processing. Computational Linguistics, Speech And Image Processing For Arabic Language, 59-83. doi: 10.1142/9789813229396_0003

[26] Jadhav A, Purohit H, Kapanipathi P, Ananthram P, Ranabahu A, Nguyen V, Mendes P, Smith A, Cooney M, Sheth A. (2010). Twitris 2.0: semantically empowered system for understanding perceptions from social data. Semantic Web Application Challenge at ISWC; 2010; Shanghai, China.

[27] Genderize. [2017-06-06]. Genderize API https://genderize.io/ webcite.

[28] Landis J., & Koch G. (1977). The measurement of observer agreement for categorical data. Biometrics. 1977;33(1):174

[29] Frank, E., Hall, M., Witten, I. (2016). Data mining: Practical machine learning tools and techniques. Burlington: Morgan Kaufmann. The WEKA workbench

[30] McCallum, A., & Nigam, K. (1998) A comparison of event models for naive bayes test classification. AAAI-98 workshop on learning for text categorization; 1998; Madison. pp. 41–8.

[31] Safieddine, F., & Hammad, R. (2020). Fake News: Origins and Political Impact. In Handbook of Research on Recent Developments in Internet Activism and Political Participation (pp. 103-121). IGI Global.

[32] Bahja, M., Hammad, R., & Hassouna, M. (2019, September). Talk2Learn: A Framework for Chatbot Learning. In European Conference on Technology Enhanced Learning (pp. 582-586). Springer, Cham.