Identifying Mobile Owner based on Authorship Attribution using WhatsApp Conversation

Badr Mohammd Almezaini
Department of Information Systems
Taibah University
Madina al Munawwara, Saudi Arabia
bmalmoziny@gmail.com

Muhammad Asif Khan
Department of Information Systems
Taiabah University
Madina al Munawwara, Saudi Arabia
asifkhan2k@yahoo.com

Summary— Social media is increasingly becoming a part of our daily life for communicating each other. There are various tools and applications for communication and therefore, identity theft is a common issue among users of such application. A new style of identity theft occurs when cybercriminals break into WhatsApp account, pretend as real friends and demand money or blackmail emotionally. In order to prevent from such issues, data mining can be used for text classification (TC) in analysis authorship attribution (AA) to recognize original sender of the message. Arabic is one of the most spoken languages around the world with different variants. In this research, we built a machine learning model for mining and analyzing the Arabic messages to identify the author of the messages in Saudi dialect. Many points would be addressed regarding authorship attribution mining and analysis: collect Arabic messages in the Saudi dialect, filtration of the messages' tokens. The classification would use a cross-validation technique and different machine-learning algorithms (Naïve Baye, Support Vector Machine). Results of average accuracy for Naïve Baye and Support Vector Machine have been presented and suggestions for future work have been presented.

Key words: whatsapp, naïve bayes, support vector machine, authorship attribution, text classification

1. Introduction

Social media has become an essential element in our daily lives which has changed lifestyle. platforms have recently become a primary element of our daily lives. These days people depend on social media for almost every need, such as daily news, weather, significant events to entertainment, communicating with friends and family, feedback on goods/services and locations, workplace management, and following up with the newest in fashion. There are various social media platforms such as Facebook, Twitter, WhatsApp, and YouTube, and these platforms may vary in the type of content provided. However, all of them are driven by user-generated content.

WhatsApp application is one of the primary forms of social media since 2009, which allows users to instant messaging in the form of text, images, audio, or video, in addition to sharing the location and ability to create groups consisting of several members [1]. WhatsApp application is the most common chatting application than other chatting applications, and it has become the most common way of contacting each other [2]. The massive popularity and rapid growth of social media platforms attracted criminals and cyber-criminals to commit cybercrime. Internet with its services has provided modern and inventive methods for fraudsters to perform illegal actions [3][4]. Many studies have discussed solving cybercrime from a data science perspective by natural language processing (NLP) using a Text Mining (TM) approach for Evaluating Event Credibility on Twitter or Spam Filtering in the email. Data mining has many applications which can protect or contribute to improving a product or service. For example, sentiment analysis reveals what the text carries in terms of feelings, whether positive, negative or neutral. Towards the subject of the text, it is widely used in marketing, customer service etc. and TM can also contribute to the medical field by analysing patients' opinions about a drug [5]. Text mining is used in a wide range of areas and can prevent or mitigate the harm caused by information crimes on social media.

Cybercriminals are using internet to commit crimes such as identity theft, unauthorized access to information, defacing websites, illegal financial transaction etc. The latest technique is to break into someone's WhatsApp account and request money or personal information from friends of the owner of the WhatsApp account. The problem is to recognize whether the message is sent by actual friend of the WhatsApp's account holder or a hacker is behind the message. This has triggered us to investigate a solution of such problem with a perspective of data science specifically in text classification area.

One can suspect on WhatsApp message when the message is in different dialect or writing style as it used to be in regular conversation with friends. Data science could help to overcome such problem. There are few researches related to authorship attribution in Arabic language, but there is no specific research which has solved identify theft in Arabic language of WhatsApp application using authorship attribution. Therefore, there is a need to develop model which can evaluate WhatsApp messages and recognize whether the messages sent by the actual author.

Manuscript received July 5, 2021
Manuscript revised July 20, 2021
https://doi.org/10.22937/IJCSNS.2021.21.7.36

2. Literature Review

A K-Nearest Neighbour (KNN) classifier combined with stylometry features for classifying short historical Arabic texts written by different authors were trained on limited data against two text documents per author, where the average text length was about 550 words per document [6]. In Stylometric Features, they used Character N-grams and Character Count of Alphabets(bi-gram, character trigram, and character tetra-gram) and Rare Words Frequency (They are the words that are repeated in a text at a low frequency) every single word that appeared once or twice in each document per author was considered a rare word feature.

In a study [7] there were 53,205 tweets collected for 20 tweeters. The researcher avoids using stemming because many tweets in the dataset are written in Dialectal Arabic and they used Naïve Bayes for classifying the tweets and We use the Bag-Of-Words (BOW) approach and compute the weights using the Term Frequency-Inverse Document Frequency (TFIDF). 80% of the dataset will be used as a training set for the classifier, and the remaining 20% is used as a test set. For the trained model, the accuracy is 88.5%. These high values for both measures reveal that the trained model is robust. The test set has to be evaluated to determine the classification process's accuracy and how the model can correctly identify the authors. The results show that the accuracy is 61%.

A decision tree algorithm for classifying over two self-collected datasets was practiced by [8]. The first dataset consists of Arabic texts from varied areas obtained from the Arabian scientific encyclopaedia, includes 373 documents under 8 categories. Whereas, the second dataset is a set of prophetic traditions (Hadiths) obtained from the Prophetic encyclopaedia. It involves 453 documents under 14 categories. Feature selection used Term Frequency (TF), Document Frequency (DF), and the combined frequency (TF/DF). The accuracy is 93% for the scientific dataset and 91% for the literary dataset (hadith).

In a study [9] researchers worked on sentiment analysis positive or negative on Twitter data in Egyptian dialect. They used Support Vector Machine (SVM) and Naïve Bayes (NB) as classifier algorithms with a combination of unigram and bigram as feature selections. The accuracy would be 65% when using NB and 72% when using SVM.

In a study conducted by [10], the dataset included 21929 poems of 114 poets. The total number of words was 12311402, with 1673465 words in the training phase and 89456 words. feature selection is lexical features, character features, structural features, poetry features, syntactic features, semantic features, and specific words features. they applied three classification techniques and testing on the features, the maximum accuracy acquired in utilizing NB, SVM, Linear Discriminant Analysis (LDA) is 99, 12% of accurate attribution by applying LDA.

In a study difficulty of Arabic comparative opinion sentences was addressed by [11] in which three tasks were completed. First, identified a comparative from the non-comparative statement and gained an f-measure of 63.73%, based on the linguistic classification. in the second phase, the author used three machine-learning algorithms (NB, KNN, and SVM) a result of about 86.63%. Finally, when used a combined approach of linguistic and machine learning and the result of an f-measure of about 88.87%.

Researchers [12] investigated Hadiths' unique linguistic features. The dataset consists of 3150 Hadiths from the Sahih Al-Bukhari book. The evaluation outcome shows that neural networks classify the Hadith with 94 % accuracy because neural networks can work with complex (high-dimensional) input data.

The performance of a particular dataset (Scopus) with five categories was evaluated by [13]. The categories include Medicine (all), Mathematics(all), Finance, Agricultural & Biological Sciences (all), and Engineering (all). In this research, KNN results gave better performance than the Naïve Bayes classifier. Further, boosting and bagging improved the accuracy using Naïve Bayes Classifier, whereas the performance of KNN remained unchanged and better than the Naïve Bayes classifier. The best performance for Bayesian Boost (Naïve Bayes) 76% and Bayesian Boost (KNN) 80%.

Naıve Bayes classifiers were utilized on training and test data set of WhatsApp group include four users [14]. Bigrams and 4-grams could not give entirely correct results. Classifiers trained on 6-grams and 7-grams succeed in recognizing the correct authors in all data size categories. Therefore, it can be said that – with the proper selection of n-gram size and enough data – character based classifiers work very well in determining authorship in Swiss German WhatsApp messages. Although the classifier has difficulty on shorter n-grams, bi- to 4-grams, it has to be remarked when the full size and half volume training and test sets are utilized, the performance for these size n-grams is still well above chance level.

In another study researchers [15] utilized data taken from WhatsApp that includes personal and group texts message from four different users. The dataset includes 76,000 words of messages in a combination of two languages — Hindi and English. Naïve Bayes Classifier and Support Vector Machine and Conditional Tree and Random Forest are applied. Word n-gram and Character n-gram are then extracted from the dataset. Feature selection executed using weightage methods, i.e., term frequency (TF), Term Frequency-Inverse Document Frequency (TF-IDF), and Binary Weight. SVM provides an accuracy of 94.862% in the case of Word unigram and an accuracy of 95.079% in the case of Character 3-gram. Similarly, Naïve Bayes Classifier computed the maximum accuracy of 94.455% and 93.259%.

In a study [16] used Arabic-language tweets to analyze opinions, and these tweets were classified within certain hashtags. For example, there was a hashtag on economic and sports topics, and more than one machine learning technique was used. Then a model was built to compare the results of the classifier and give the highest accuracy.

3. Method and Data Preparation

In order to analyse data from WhatsApp messages, Rapid Miner tools has been used which promotes action, including predictive analysis in their work processes with its user-friendly, well-healed library of data science and machine learning algorithms over its all-in-one programming surrounding like Rapid Miner Studio. It involves fundamental data mining such as data cleansing, filtering, and clustering. The tool is also suitable for weak scripts. Rapid Miner presents data mining and machine-learning methods, including data loading, transformation, pre-processing and visualization, predictive analytics, statistical modelling, evaluation, and deployment. It also supports multiple languages including Arabic language.

3.1 Arabic WhatsApp messages authorship attribution

We used Rapid Miner to perform the Arabic WhatsApp messages authorship attribution. It comprises of two main phases i.e. Training, and Testing phases. The two phases consist of four steps: collecting WhatsApp messages, WhatsApp messages pre-processing, classification using naive Bayes, and support vector machine with different feature selection and validation for check a model accuracy.

3.2 Collecting data of WhatsApp messages

Finding WhatsApp messages was a challenging task because the messages contain private information that no one desires to share. WhatsApp messages are not like tweets on Twitter, which everyone can read and share with others by retweet. However, we could not find suitable WhatsApp messages on the internet, as most of the messages are in English or other languages. Therefore, The WhatsApp groups on our phones were used for the research purpose after acquiring the group participants' permission. We avoided WhatsApp groups with many multimedia or broadcasts because we desired to get more text messages of many messages written by the user to train the machine to reach a higher accuracy in identifying the participants. It is hard to clean the message that includes broadcasts, so we had to remove it manually. Broadcast messages are not written by the sender and do not express his words, such as articles or poetry forwarding by WhatsApp users, and they are usually anonymous sources. However, WhatsApp recognizes some broadcast messages that the user forwarding, contains the phrase " forwarded, "making detect this kind of message sometimes is easy, but sometimes the user does not forward for broadcast rather copy it then paste.

3.3 Dataset description

As mentioned earlier, datasets are messages taken from one of the groups on WhatsApp that was not created for research purposes. The messages are in Arabic with different dialects of the Kingdom of Saudi Arabia. The group on WhatsApp includes eight Participants, four of whom used them for research purposes after obtaining their approval to use the research messages. We aimed at the active people in the group, based on the number of messages sent by individual participant (P). The dataset consists of 2,498 messages in which P1, P2, P3 and P4 sent 616, 590, 550 and 742 messages respectively. The messages were taken randomly from the beginning of the group creation from 2019 to 2021 for each participant.

The number of words in the messages sent by P1 is 2833, P2 is 2250, P3 is 2713 and number of words in the messages sent by P4 is 2638. This condition of words in align with the research (Maciej, 2014) which states the minimum number of words per text must be 2500 words to achieve good accuracy. However, newer research (Iqbal et al., 2020) shows that the minimum size of words needs to be at least 2000 words to achieve good accuracy.

3.4 Text pre-processing

This phase of text pre-processing is essential in determining the quality of feature extraction, and classification phase. This phase includes replace and map tokenizing, filtering stop words, finding Arabic stems, generating n-gram for words and characters, and filtering tokens by length. The messages from WhatsApp were exported and personal information were removed. The participant name was replaced by Pn where P denotes participant and n is for participant number. Figure 1 shows the WhatsApp messages dataset in textual form

Figure 1. Messages dataset after replacing real names

The test file was transformed into an excel sheet and different symbols such as a ‘ – ’ replaced by ‘ $ ’ sign. Likewise some sentences such as ‘ تم حذف الرسالة ’ which appear by WhatsApp and not written by a participant removed as well as date and time. Also there were some sentences in English and links needed to be removed, so we used map and replace operator. The regular expression in map is [a-zA-Z] to eliminate English words where replace with single space as it appears in figure10 , and [-!"#$%&'()*+,./:;<=>?@\[\\\]_`{|}~] to remove email and links.

Process Document from Data is used to generate word vectors from string attributes and remove any unwanted words. Following processes are carried out in this operator. This operator includes many parameters which strongly effects the accuracy of the model, for example, prune method, determine if frequent or infrequent words must be ignored for word list building and how the frequencies are defined .and vector creation to create word vector by using TF_IDE which is feature selection.

Tokenize is the task of turning raw text files into a well-defined sequence of linguistically meaningful units. The messages are split into a stream of words by removing all punctuation marks, brackets, hyphens, numbers, and symbols. Although Removing stop words is another common step in text preprocessing but in our case could not help in enhancing the model because every participant uses a different dialect which helps to detect the participants. However, the stop words are the most frequent appearance, and unimportant words useless in information retrieval and text mining. For Arabic, stop words include pronouns, prepositions, adverbs, days of the week, and months of the year.

An n-gram is the sequence of items from a sequence of text. The items can be syllables, letters, words. The n-grams typically are obtained from a text . An n-gram of size 1 is referred to as a "unigram"; size 2 is a "bigram"; size 3 is a "trigram" where n indicates to number, and gram indicates words or characters. Set Role operator is used to select the label which is “participants” attribute will act as the target attribute for the learning operators. Once we applied all the operators in the process document from data, for all the messages in the dataset, a document by term matrix is generated. Figure 2 shows example set from the “process document from data” operator.

Row No.	user	اغر	امين	ابو	ابيض	اتصل	اتفق	اجل	احد
1014	P4	0	0	0	0	0	0	0	0
1015	P4	0	0	0	0	0	0	0	0
1016	P4	0	0	1	0	0	0	0	0
1017	P4	0	0	0	0.689	0	0	0	0
1018	P4	0	0	0	0.689	0	0	0	0
1019	P4	0	0	0	0.689	0	0	0	0
1020	P4	0	0	0	0.689	0	0	0	0
1021	P4	0	0	0	0.472	0	0	0	0
1022	P4	0	0	0	0.535	0	0	0	0
1023	P4	0	0	0	0.445	0	0	0	0
1024	P4	0	0	0	0.474	0	0	0	0
1025	P4	0	0	0	0	0	0	0	0
1026	P4	0	0	0	0	0.837	0	0	0
1027	P2	0	0	0	0	0	0	0	0
1028	P4	0	0	0	0	0	0.521	0	0
1029	P4	0	0	0	0	0	0.746	0	0
1030	P2	0	0	0	0	0	0	0	0

Figure 2. Document by term matrix

The output of the "process document form data" operator is given to the set role operator. Figure 3 is the word list that has been generated from the "Process data from Document" operator. Here we can see a word list containing all the different words in the document and the number of their occurrence next to it in the "Total Occurrences" column.

Word	Attribut...	Total O...	Docum...	P3	P2	P1	P4
آخر	آخر	2	2	0	1	0	1
آمين	آمين	3	3	0	0	0	3
أبو	أبو	2	2	0	0	0	2
أبيض	أبيض	6	6	0	0	0	6
أتقى	أتقى	3	3	0	0	0	3
أجل	أجل	2	2	0	0	0	2
أحد	أحد	2	2	0	0	0	2
أختياره	أختياره	2	2	0	0	0	2
أرجو	أرجو	6	6	0	0	0	6
أرسلنا	أرسلنا	4	4	0	4	0	0

Figure 3. Wordlist in all dataset

4. Results and Discussion

In order to evaluate the model a 10-fold cross-validation method was used. In the 10-fold cross-validation method, we divided the data set is into ten parts or (folds). One fold was chosen as the test set, while the other nine folds were used for training. Ten iterations would run for the cross-validation. We used cross-validation to avoid Overfitting, it refers to a problem that occurs if a classifier fits the training dataset very well and doesn't generalize well to the test dataset [17]. We computed the confusion matrix entries True Positives, True Negatives, False Positives, and False Negatives in each iteration. Table 1 depicts the best classification score achieved in using NB 59.34% ,this accuracy is achieved by applying NB on Character Bi-gram with TF-IDF where prune method is absolute, prune method above 999 and below is 2

Table 1: Accuracy of NB and SVM classifiers

Feature	Accuracy of attribution using (NB) with pruning	Accuracy of attribution using (SVM) with pruning
Character Uni-gram	58.58%	59.74%
Character Bi-gram	59.34%	57.94%
Character Tri-ram	55.69%	58.70%
Character Tetra-gram	56.53%	59.02%
Word uni-gram	57.98%	58.02%
Word Bi-gram	57.94%	56.89%
Word Tri-gram	58.14%	57.29%
Word Tetra-gram	57.86%	56.70%
Average of the all used features	57.75%	58.03%

This is the best score for all features that have been employed in this experiment. The lowest classification result is 55.69% obtained by applying NB on Character Tri-gram. The best classification score obtained in using SVM 59.74% This number is achieved by applying SVM on Character Uni-gram with TF-IDF where prune method is absolute, prune above method 999 and below is 2. This is the best score for all features that have been employed in this experiment. The lowest classification result is 56.70% obtained by applying SVM on Word Tetra. To overcome the issue of low accuracy of results, we duplicated the data set that we had by 100.the low accuracy may be attributed to the lack of data that the machine needs to learn and then It has the ability in addition to the fact that the texts are short since they have not enough contextual information, which poses a great challenge for classification [18].

It is important to mention that original data was not enough for the classifier to learn on it to have the ability to classify the participants, short texts, and the small size of data, which may contribute significantly to the lack of high accuracy. there was difficulty collecting more data, especially as we mentioned previously data needs to be cleaned, so we had to multiply our dataset by 100% and Table 2 depicts the results.

Table 2: Accuracy of classifiers by duplicated data

Feature	Accuracy of attribution using (NB) with pruning	Accuracy of attribution using (SVM) with pruning
Character Uni-gram	90.80%	82.96%
Character Bi-gram	91.78%	77.34%

Feature	Accuracy of attribution using (NB) with pruning	Accuracy of attribution using (SVM) with pruning
Character Tri-ram	90.92%	82.31%
Character Tetra-gram	90.62%	84.02%
Word uni-gram	91.44%	82.31%
Word Bi-gram	92.64%	85.60%
Word Tri-gram	92.52%	85.14%
Word Tetra-gram	92.54%	84.68%
Average of the all used features	91.65%	81.8%

The best classification score obtained using NB 92.64%. This accuracy is achieved by applying NB on Word Bi-gram features with TF-IDF where prune method is absolute, prune above is 999 and prune low absolute is 2 to avoid words that appear less than 2. Therefore, the model ignores rare words that are not usually used by users. While the best classification score gained in using SVM 85.60%. This number is achieved by applying SVM on Char Word Bi-gram feature with TF-IDF where prune method is absolute, prune above method 999 and below is 2. The worst classification result is 77.34% which was acquired by applying SVM on Character Bi-gram The overall average accuracy 91.65% was obtained by applying NB for all features; whereas 81.8% was achieved by applying SVM. Therefore, NB obtained higher accuracy than SVM that cause NB is better than SVM with few training cases or small dataset.

Table 3. Average of precision and recall for all classifiers

Class	P1%	P2%	P3%	P4%
Average of recall with SVM classifier for all features	82.72	84.95	78.0	91.0
Average of recall with NB classifier for all features	78.32	92.71	87.89	93.6
Average of precision with SVM classifier for all features	82.4	79.23	86.71	83.2
Average of precision with NB classifier for all features	93.94	85.85	95.17	92.6

As the Table III depicts p1 highest precision was obtained with NB with Word Bi-gram, it means correct around 96.95% of the time and best recall obtained with SVM with word Bi-gram and the exact value repeat in Character tri-gram with SVM 92.49%, indicate a measure of how accurately our model is able to identify the relevant data. Best precision for P2 achieved Character Tetra-gram in NB, it is correct around 91.94% of the time and best recall obtained with word Bi-gram in, which is 95.59%. Best precision for P3 obtained with NB in word Tetra-gram, it is correct around 92.49% of the time and best recall obtained with Character tri-gram, which is 92.49%. The best precision with P4 obtained with word Bi-gram in NB is correct around 96.95% of the time .Best recall for P4 got with Character Tetra-gram 95.27%.

It is necessary to mention that the accuracy of the classifier increases with less number of participants, the highest accuracy obtained was about 85.81% between the third participant and the fourth participant. It is also important to accept that due to constraint of time we could not experiment with all the scenarios. As we mentioned, we need several scenarios in case we want the model to distinguish between two participants, for example, if we choose a Character tri-gram feature, we need to run six scenarios for each feature.

5. Conclusion and Future Work

The feature word-gram in both classifiers achieved better results than char-gram , where Word Bi-gram feature achieved higher accuracy in both classifier This may be that the participants' instant messages contain words of different dialects and thus these different and distinct words specific to each user or sentence include two words or more makes the accuracy increase thus the type of text play critical role .However At some times, accuracy may be misleading and cannot be relied upon, especially if the data are not balanced, so it is also important to rely on different types of metrics, not just accuracy.

For future research, there is a way that can be examined to solve the problem of scam messages on WhatsApp, by classifying WhatsApp messages into scam messages or regular messages. we could use punctuation marks and emoji's in the classification of the messages and use more types of classifiers, in addition, full conversations will be added to each class to avoid the problem of short texts. however, there will be a challenge in collecting scam messages sent by WhatsApp, as there is no data source for this type of message on the Internet.

References

[1] P. Natalie, and T. Yue, T. (2020). "What about WhatsApp? A systematic review of WhatsApp and its role in civic and political engagement," First Monday, vol. 25, 2020, [https:// doi.org/ 10.5210 / fm. v25i12.10417]

[2] H. Tankovska, "Daily active users of WhatsApp status 2019," Statistica, 2021, [Retrieved on Jan 18, 2021]

[3] I. Shreen, and R. Tariq, "Identify theft and social media," International Journal of Computer Science and Network Security, vol. 18, pp. 43-55, 2018

[4] P. Poonam, K. Krishan, S. Bharanidharan, C. Kheng, "A theoretical review of social media usage by cybercriminals, "in Conference on Computer Communication and Informatics, 2017, India

[5] A. Abbe, C. Grouin, P. Zweigenbaum, B. Falissard, "Text mining applications in psychiatry: a systematic literature review," International Journal of Methods in Psychiatric Research, 25, pp. 86-100. 2016

[6] H. Fatma, M. Masnizah, A. Zahra, and A. Jowan, “Authorship attribution of short historical arabic texts using stylometric features and a KNN classifier withlLimited training data,” Journal of Computer Science, vol. 16, pp. 1334-1345, 2020

[7] A.-B. Jafar, T. Bashar, A.-A. Mahmoud, and B. Zaqaibeh, “Using big data analytics for authorship authentication of arabic tweets,” in 8th IEEE/ACM International Conference on Utility and Cluod Computing, 2015, Cyprus

[8] B. Mariam, E.Wael, T. Sara, and H. Amjad, “Sentiment classification techniques for arabic language: a survey,” in 7th International Conference on Information and Communication Systems, 2016, Jordan

[9] S. Amira, and R. Ahmed, “Sentence-Level Arabic Sentiment Analysis,” in International Symposium on Collaboration, Social Computing, New Media and Networks (SoMNet2012), Cairo

[10] A. -f. Ahmed, R. Mohammed, and M. Bellafkih,.”Machine learning for authorship attribution in arabic poetry, “ International Journal of Future Computer and Communication, vol. 6, pp. 42-46, 2017

[11] A. El-Halees, “Opinion mining from arabic comparative sentences, “ in International Arab Conference on Information Technology, 2012, Jordan

[12] A. Mohammad, I. Norisma, M. Rohana, J. Salinah, T. Dirk, and A. Gani, “Hadith data mining and classification: a comparative analysis, “ Artificial Inteliigence Review, vol. 46, pp. 113-128, 2016

[13] A. Muhammad, A. Tanvir, S. Fahad, and I. Muhammad, “Feature extraction based text classification using k-nearest neighbor algorithm,” Internationla Journal of Computer Science and Network Security, vol. 18, pp. 95-101, 2018

[14] R. Oppliger, “Automatic authorship attribution based on character n-grams in Swiss German,” in 13th Conference on Natural Language Processing (KONVENS), 2016, Germany

[15] S. Abhay, N. Ananya, and R. Reetika, “An investigation of supervised learning methods for authorship,” ACM Transactions on Asian and Low-Resource Language Information Processing, vol. 1, pp. 1-11, 2018

[16] A. Lama, S. Mostafa, and E. Fathy, “Arabic blogging Sentiment Analysis,” La Pensee, vol. 76, 2016

[17] W. Koehrsen, “Overfitting vs. Underfitting: A Complete Example,” [Retrieved from towardsdatascience: https://towardsdatascience.com/ overfitting-vs-underfitting-a-completeexample-d05dd7e19765]

[18] J. Chen, Y. Hu, J. Liu, Y. Xiao, and H. Jiang, “Deep Short Text Classification with Knowledge Powered Attention,” in Proceedings of the AAAI Conference on Artificial Intelligence. 33. 2019