COVID-19's Impact on the Telecommunications Companies

Latifah Almuqren[1, 2] and Alexandra I. Cristea[2][0000-0002-1454-8822]

[1] Princess Nourah bint Abdul Rahman University, Riyadh, Saudi Arabia

[2] Durham University, Durham, UK
Laalmuqren@pnu.edu.sa

Abstract. Now the world is witnessing most significant challenges due the Covid-19 crisis. Beyond health effects, it has social and *economic effects*. With the enormous amount of data available and the widespread use of social web globally, research can and should use it to provide solutions. Customer satisfaction is known to affect *customer churn* (customers leaving companies), which is a problem affecting many companies in competitive and volatile markets – like the current one. One easily available open source of customer opinions are *tweets* – more relevant now in the online world. Whilst Natural Language Processing (NLP) on tweets is not new, few studies target *customer satisfaction*, and NLP body of research on *Arabic tweets* is modest; we are not aware of any other study on this during a global pandemic. Our research thus aims to propose *a new model based on Twitter mining to measure customer satisfaction during Covid-19,* as well as *compare customer satisfaction before and during the crisis*. This is a use case for the largest Telecom companies in Saudi Arabia, and we involve the popular method of Sentiment Analysis (SA) for the task. We additionally propose a *new Saudi lexicon* and apply it to *monitor real-time customer satisfaction on Twitter* using three different transfer network models on Arabic sentiment analysis. Also, this research evaluates using these models on Arabic Sentiment Analysis as the *first study comparing between three different transfer network models on Arabic text*.

Keywords: Sentiment Analysis, Customer Satisfaction, Arabic.

1 Introduction

With the growing use of social media sites worldwide over the last ten years, SA has recently become a prominent and useful technique for capturing public opinion in many different disciplines. Sentiment analysis or 'opinion mining' refers to the computational processing of opinions, feelings and attitudes towards a particular event or issue [21]. SA has a vital function for real-life applications and decision-making process in various domains [3].

In the context of the global Covid-19 pandemic and the quarantine, there were several impacts on different sector, one of the critical sectors is Telecommunication (telecom) [6]. As an inevitable consequence of Covid-19 a lot of people used the telecom service for working from their home or for entertainment. Which constituted a burden on the telecom companies to increased network flexibility and this affected the telecom investments [6]. Several reports issued to addressing the impact of Covid-19 on telecom companies [6], [19]. On other hand, there is a lack of studies addressed the customer satisfaction toward telecom company services during the Covid-19 and the quarantine. Therefore, our study aims to find out what impact Covid-19 had on the customer satisfaction toward telecom company services and the implications for potential decisions. While, the sentiment have a crucial role in many decisions, including business-related ones. The detection of sentiment polarity, however, is a challenging task, due to the sentiment resources limitations in different languages. It remains a largely unexplored research area for the Arabic language [3]. This is due to the complexity of Arabic [3]. It has many forms, which are Classical Arabic (CA), as in the book of Islam's Holy Quran, Modern Standard Arabic (MSA) used in newspapers, education and formal speaking, Dialectical Arabic (DA) which is informal everyday spoken language, found in chat rooms, and social media platforms. Additionally, DA forms differs from one Arab country to another. Thus, unfortunately, resources used for one Arabic country cannot be applied to another. Hence, DA analysis, targeted here, is complicated, requiring a native speaker.

This paper attempts to alleviate this matter by focusing on Arabic Sentiment Analysis and provide solutions to one of the challenges that faces Arabic SA by creating a *Saudi dialectical data set* extracted from Twitter. In addition, we evaluate applying the transformer network a Robustly Optimized BERT Pretraining (RoBERTa) [16] on the Arabic data set and compare it to AraBERT [5] and hULMonA [10] as two transfer models designed for Arabic language. We chose RoBERTa because it was shown to outperform Bidirectional Encoder Representations (BERT) in sentiment classification tasks [16], although BERT is the best in many NLP tasks [16]. To the best of our knowledge, this is the first attempt of using RoBERTa for Arabic sentiment analysis. Moreover, this paper predicts customer satisfaction before and during Covid-19, in order understand how the pandemic and its traceable effects has influenced customers as well as telecoms companies. The main contributions of this paper have measured the impact of a pandemic (Covid-19) and quarantine on customer satisfaction toward telecom company services, compare customer satisfaction before and during the crisis and compare for the first-time different transfer network models RoBERTa, AraBERT, and hULMonA on Arabic Sentiment Analysis.

2 Related Research

a. Customer Satisfaction

Customer satisfaction (CS) is defined as the outcome of using a service, resulting from the comparisons that the buyer makes with other similar providers in terms of the rewards and costs of the service [11]. Social media includes a variety of platforms that allow people to share and exchange information, producing an abundance of data. Literature [13], [23] showed though that mining social media data is important for both marketers and customers, as —1) it provides a wealth of information about customers for the company [13], 2) it enables the extraction of information hidden in data that can benefit businesses by exploring their customers' views about particular products or services, 3) it helps to develop a recommendation system for maintaining existing customers or gaining new ones, and 4) it is also useful for building confidence among customers and stakeholders[23]. However, only a few studies measure customer satisfaction, via social media mining and all these studies mined English tweets [18] . Mostafa [18] mined random tweets on Twitter to find the consumer's sentiments towards certain brands through sentiment analysis. The findings proved that there is a positive consumer sentiment towards famous brands. Both studies used the sentiment analysis to measure the satisfaction. In addition, both studies measured the customer satisfaction in business field using English tweets.

b. Using transformer language model with Sentiment Analysis

Recently, pre-trained language models have achieved good results with different NLP target tasks, due to ability of these models to learn with few parameters [10]. Previous approaches depended on features [12]. The Universal Language Model Fine-tuning (ULMFiT) pre-trained language model [12], composed of three ASGD Weight-Dropped Long Short-Term Model (AWD-LSTM) and an LSTM layer [17], is perfectly accurate on different NLP tasks. The newest language model is BERT [9]. It used a Transformer network [28]. BERT outperformed the other pre-trained language models, due to its ability to manipulate context from both directions. Another pre-trained language model is RoBERTa [16], which is an enhanced version of the BERT model [9]. It trained on larger data, with longer series, and is the best model on the General Language Understanding Evaluation (GLUE) benchmark [24], Stanford Question Answering Dataset (SQuAD) [20] and ReAding Comprehension from Examinations (RACE) [14].

c. Using transformers with Arabic Sentiment Analysis (ASA)

The use of transfer language models is still new for ASA studies. there is a lack of studies has been published so far [2], [10], [5]. Al-Twairesh and Al-Negheimish [2] used the BERT model [9] on an Arabic tweets dataset. Their generic and sentiment-specific word-embedding model outperformed the BERT model. They explained that this was because the BERT model was trained on Wikipedia, which is written in MSA, whereas dialects are used on Twitter. in Addition, there is a first Arabic universal Language model hULMonA [10], it based on ULMFiT. The results showed that hULMonA achieved state-of-art in Arabic sentiment analysis. The most recent Arabic universal Language model is AraBert [5].

3 Data Processing

a. Data Sets

To build the data sets, we used Python to fetch Arabic tweets based on certain search keys. The hashtags used in the search were the ones that indicated different Saudi telecom companies: STC, Mobily and Zain. Then we grabbed the top hashtags mentioning these telecom companies, which were: #STC, #Mobily, #Zain, #الاتصالات_السعوديه, #موبايلي and #زين_السعودية.

The aim was to *monitor the telecom customers' sentiments continuously*. The raw data set comprised 3.5 million Arabic tweets. Then, we chose the sample of Saudi tweets randomly from the data set to construct our first corpus, AraCust1, Table1.

Our AraCust1 corpus is comprised of 20,000 Saudi tweets collected randomly from the data set, representing the **customer satisfaction situation before Covid-19**. The second corpus, which we called AraCust2, has been collected during Covid-19 from second of March until second of May, the period from registering the first case of Covid-19 in Saudi Arabia, Table 1. This corpus represents the **customer satisfaction situation during Covid-19**. Thus, we filtered tweets based on user location and time-zone, to identify Saudi tweets.

Table 1. Companies and the Total Number of Unique Tweets for AraCust1and AraCust2

Company	Twitter Handle	# of Unique Tweets AraCust1	AraCust2
STC	@STC_KSA, @STCcare	7,590	7000
Mobily	@Mobily, @Mobily1100	6,460	4004
Zain	@ZainKSA, @Zain-HelpSA	5,950	6000
Total		**20,000**	**17,004**

To clean the data set, we eliminate the non-Arabic tweets. Also, re-tweets were discarded. We filter all the features that unnecessary and will decrease the classifiers accuracy e.g., links, user mention, punctuation marks, and stop words. Then, applying pre-processing producers on the data set (tokenisation and normalisation). Normalization such as removing Kishida and uniting the same letters with different shapes. The cleaning and pre-processing were done using Toolkit (NLTK) library in Python. Examples before & after pre-processing (AraCust1) are shown in Table 2.

Table 2. Subset of the AraCust1 corpus before pre-processing

Tweet in Arabic before pre-processing	Label	Company	Tweet in English	After pre-processing
@So2019So @STCcare غيري الشركه	Negative	STC	Change the Company	غيري شركه
@alrakoo @mmshibani @GOclub @Mobily اشكرك😊	Positive	Mobily	Thank you	اشكرك

b. Annotation for corpora

We need to add a new aspect to the data set, which is the sentiment label. We used the binary labels (Positive, Negative) for both corpuses. Each label expresses the strongest emotion for each tweet, following previous recommendations [3]. This means that a given tweet might have other emotions associated, which we discard for the time being. The annotation process was done manually for the AraCust1 corpus by three annotators that have the experience with the annotation process. Every annotator needs to assign one label per tweet for the whole data set.

To identify the difficulty of the annotation task, the inter-annotator agreement (IAA) was used, to measure the trustworthiness of the annotation scheme. If there were more than two annotators, then Fleiss' Kappa [8] is recommended to measure the reliability of the annotation process. Therefore, we applied it, obtaining a value of 0.50, based on the level of acceptance, which is considered moderate [15]. Kappa, *k*, can be defined as:

$$k = \frac{P - Pe}{1 - Pe} \tag{1}$$

The factor *1-Pe* computes the degree of agreement that is attainable above chance, and *P-Pe* represents the degree of agreement achieved above chance. If the annotators are in complete agreement, then $k = 1$. If there is no agreement, then $k < 0$. The AraCust2 corpus annotation done using AraSTw lexicon that created and evaluated by the authors.

4 Model Construction

a. Evaluation Metrics

To evaluate the performances of the models, we used four metrics suitable for binary classification [7] : micro averages of (Precision (Pr), Recall (Rc), and F1), Accuracy (Ac). The micro average is suitable for binary-classes, especially if the classes are imbalanced, because the micro-average totals the contribution of all classes to the average metric calculation [22]. It aggregates the precision and recall of the classes.

b. RoBERTa Model Construction

We used RoBERTa model, BERT based, using the parameters seed=42 to build on the random weight, the precision floating for GPU, the batch size =16 and 64 for the maximum sequence length. After visualizing the number of tokens in most tweets, the maximum tokens per tweet is 30 tokens. Therefore, we padded all tokens up to this size. After that, the model converts the word to an integer. The model using discriminative fine-tuning and gradual unfreezing. The model freezes all the groups but the two last ones. It used the minimum numerical gradient: 1.00E-03 and minimum loss divided by 10: 3.02E-04. To implement the sentiment analysis task, it used discriminative fine-tuning and gradual unfreezing. That means to predict the next token, based on the present series of tokens in the sentiment corpus, with various learning rates, from 1e-02 to le-06. After that, the model unfreezes the output layer, after each epoch, layer by layer. It freezes all the groups except for the last two layers.

c. AraBERT Model Construction

It is BERT based model; it trained on different Arabic datasets. It used the BERT basic configuration [9]. Except, it added a special pretraining prior the experiment specific to Arabic Language. It used "ال" "Al" before the word, it is a prefix without meaning, by using a Fast and Accurate Arabic Segmenter (Farasa) [1] to segment the word.

d. HULMonA Model Construction

HULMonA is the first Universal Language Model (ULM) for Arabic language. It is based on ULMFiT [12]. It pretrained on Large Arabic corpus and fine-tuned to many tasks. It consists of three stages: 1. training AWD-LSTM model [17] on Arabic Wikipedia corpus, 2. finetuning the model on a destination corpus, 3. and for text classification, they included a classification layer on the model.

5 Experiment Results, Discussion and Evaluation

a. Model Results

Comparing the results of using RoBERTa, AraBERT, and hULMona models using the micro average of different metrics, Table 3 shows that the results favour to the AraBERT model with 94.0% accuracy.

Table 3. Comparing between RoBERTa, AraBERT, and hULMonA Models

Model	Accuracy	F1	Recall	Precision
RoBERTa model	92.1	92.2	92.0	91.1
AraBERT model	94.0	92.6	92.1	93.0
hULMonA model	90.8	79.8	89.0	84.0

b. Discussion

To get the reasons behind the obtained results, we analysed the three models' architecture. AraBERT outperformed the two others models due to: 1. It trained on different Arabic data sets Modern Standard Arabic data sets and dialectical data sets. 2. It applied a special pretraining specific to Arabic Language. 3. It used Farasa [1] a pre-processing tool directed to the Arabic language; It outperformed the state-of-art MADAMIRA [2].

Although, hULMonA trained on different Arabic data sets, it performed worse than other models, because it based on ULMFiT [12] which is one directional model. Additionally, it lacks to appropriate pre-processing for Arabic text.
Regarding RoBERTa which a BERT based model, it transfers each Arabic letter to Latin character and every Latin character face one integer number. That is the reason for decreasing the performance of the model.

c. Predicting Customer Satisfaction

We used AraBERT model which achieved the higher result to predict the customer satisfaction for the AraCust1 corpora based on the predefined companies STC, Mobily and Zain. First, we calculated the customer satisfaction as follows:

cust_sat = total_ratings /(2*num_customers)

where:

num_customers = len(ratings)

total_ratings = sum(ratings) (the summation of all ratings) rating: binary rating.

Then we divided the corpus based on the company. We calculated the average accuracy of predicted customer satisfaction using the model. Our result showed that the predicted customer satisfaction (using the model) for the three companies STC, Mobily and Zain before Covid-19, 33.27%, 29.88% and 333.24% were below 50%.

d. Evaluate using the Sentiment Analysis Approach

This study has used a sentiment analysis to design a new real-time model to measure customer satisfaction. To evaluate the sentiment analysis approach is effective to measure customer satisfaction, it has used a questionnaire to the same customers that we mined their tweets to evaluate the approach by comparing the predicted customer satisfaction (using the AraBERT model) with actual customer satisfaction (using the survey). It can be seen in Table 4 that the predicted and actual customer satisfaction rates are approximate. That results denote to efficiency of using sentiment analysis approach to achieve the goal of predicting the customer satisfaction of telecom companies based on the Twitter analysis.
These results should encourage the decision-makers to consider using Twitter analyses for measuring customer satisfaction and to include it as a new method for evaluating their marketing strategies.

Table 4: Rate of predicted customer satisfaction vs actual customer satisfaction

Company	Predicted Customer Satisfaction	Actual Customer Satisfaction
STC	33.27%	20.1%
Mobily	29.88%	22.89%
Zain	33.24%,	22.91%

7 Predicting Customer Satisfaction during Covid-19

After proving the efficiency of using AraBERT model on the Arabic dialect text- specially- Saudi tweet corpus in term of applying sentiment analysis approach to predict the customer satisfaction; we used lexicon-approach to annotate the AraCust2. We used the generated AraSTw lexicon to annotate AraCust2 using binary sentiment (positive, negative). Secondly, we inspected AraCust2 manually, to check the tweets' sentiment; the positive tweets were more than the negative tweets.
After that, we applied AraBERT to predict the customer satisfaction as it performed best with the AraCust1 corpus. The accuracy was 90% and F_{avg} =88.8, as in Table 5.

Table 5. Evaluation results on AraCust2 using AraBERT model.

Label	Precision	Recall	F1-score	F avg	Accuracy
0	90.0	82.0	85.5	88.8	90.0
1	89.0	95.0	92.0		

8 Comparing between customer satisfaction percentage before and during Covid-19 using the AraBERT model

Because of the pandemic and its effects such as the quarantine, it enforced people to work from home and to connect with the outsider world depending on the technology. As a result, the telecom industry became an essential sector. Which constituted a burden on the telecom companies to offer a reliable and fast data and voice services. The Covid-19 pandemic made people to use the telecom services more than before. This resulted to increase the benefits of the Saudi telecom companies to 7% in the first quarter of 2020 [4]. As we noticed there are several reports that studied the impact of Covid-19 on the telecom sector from investment side [6], [19]. This is not the case in our study; it focused on the customers side toward the telecom services during Covid-19 by measuring the customer satisfaction toward the telecom companies in Saudi Arabia. Our result showed that comparing the predicted customer satisfaction (using the AraBERT model) before and during Covid-19 (Table 6), the customer satisfaction percentage for the three companies STC, Mobily and Zain before Covid-19, 33.27%, 29.88% and 33.24 %, were below 50%. However, the customer satisfaction during Covid-19 and quarantine was raised to 58.99%, 50.05% and 62.50% for STC, Mobily and Zain, respectively, Fig. 1.

Table 6. Percentage of customer's satisfaction before and during Covid-19 in Saudi Arabia.

Company	Customer's Satisfaction Before Covid-19	Customer's Satisfaction During Covid-19
STC	33.27%	58.99%
Mobily	29.88%	50.05%
Zain	33.24%,	62.50%

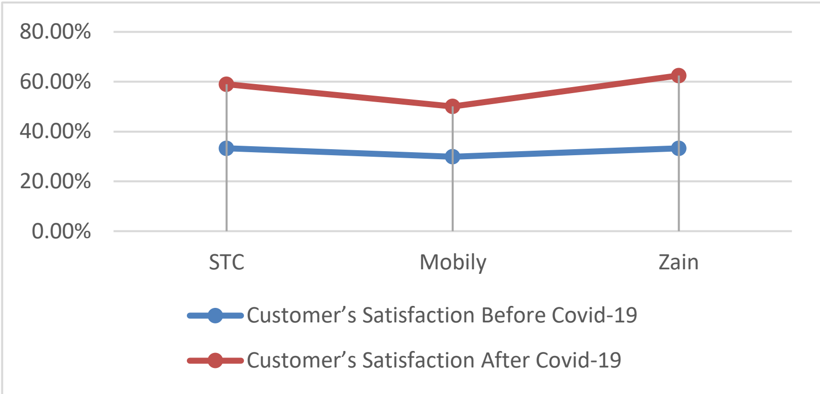

Fig. 1. Comparing between Customer satisfaction before and during Covid-19

9 Conclusion

This study *proposes a new model based on Twitter mining to measure customer satisfaction during Covid-19.* To find out *if Covid-19 has an impact on the telecom companies*. Interestingly, we found *that customer satisfaction toward telecom companies enhanced during Covid-19*, because of home quarantine and growing demand for Internet and voice services. Additionally, we *used the RoBERTa model for the first time for Arabic sentiment analysis* and compared it to AraBERT and hULMonA models for Arabic Sentiment Analysis. Datasets were collected before and during Covid-19. Surprisingly, results showed that AraBERT outperformed RoBERTa, and hULMonA based on several measures. A reason behind the obtained results may be due to the RoBERTa didn't train on Arabic dataset and hULMonA depending on forward layer rather than bidirectional layers. Still, initial results for hULMonA are promising, and further experimentations are planned to use.

Acknowledgment

This research was funded by the Deanship of Scientific Research at Princess Nourah bint Abdulrahman University, through the Pioneer Research Funding Program (Grant No# PR-1441-3).

References

1. Abdelali, A., et al. *Farasa: A fast and furious segmenter for arabic*. in *Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: Demonstrations*. 2016.
2. Al-Twairesh, N. and H.J.I.A. Al-Negheimish, *Surface and Deep Features Ensemble for Sentiment Analysis of Arabic Tweets.* 2019.
3. Al-twairesh, N.S., *Sentiment Analysis of Twitter : A Study on the Saudi Community.* 2016.
4. Awalan. *Saudi Arabia: Corona pushes profits of telecom companies up*. 2020 [cited 2020 1/6/2020]; Available from: https://www.awalan.com/Article/5130/-السعودية-كورونا-يدفع-أرباح-شركات-الاتصالات-صعوداً.
5. Baly, F. and H. Hajj. *AraBERT: Transformer-based Model for Arabic Language Understanding*. in *Proceedings of the 4th Workshop on Open-Source Arabic Corpora and Processing Tools, with a Shared Task on Offensive Language Detection*. 2020.
6. Casey, M. and C. Wigginton, *Understanding the sector impact of Covid-19 Telecommunications*, 2020, Deloitte.
7. Chicco, D. and G.J.B.g. Jurman, *The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation.* 2020. **21**(1): p. 6.
8. Davies, M. and J.L.J.B. Fleiss, *Measuring agreement for multinomial data.* 1982: p. 1047-1051.
9. Devlin, J., et al., *Bert: Pre-training of deep bidirectional transformers for language understanding.* 2018.
10. ElJundi, O., et al. *hULMonA: The Universal Language Model in Arabic*. in *Proceedings of the Fourth Arabic Natural Language Processing Workshop*. 2019.
11. Gustafsson, A., M.D. Johnson, and I. Roos, *The effects of customer satisfaction, relationship commitment dimensions, and triggers on customer retention.* Journal of marketing, 2005. **69**: p. 210-218.
12. Howard, J. and S.J.a.p.a. Ruder, *Universal language model fine-tuning for text classification.* 2018.
13. Kennedy, H.J.J.o.B. and E. Media, *Perspectives on sentiment analysis.* 2012. **56**(4): p. 435-450.
14. Lai, G., et al., *Race: Large-scale reading comprehension dataset from examinations.* 2017.
15. Landis, J.R. and G.G. Koch, *Landis_Jr__Koch_Gg_1977_Kappa_and_Observer_Agreement.* Biometrics, 1977. **33**: p. 159-174.
16. Liu, Y., et al., *Roberta: A robustly optimized bert pretraining approach.* 2019.

17. Merity, S., N.S. Keskar, and R.J.a.p.a. Socher, *Regularizing and optimizing LSTM language models.* 2017.
18. Mostafa, M.M.J.E.S.w.A., *More than words: Social networks' text mining for consumer brand sentiments.* 2013. **40**(10): p. 4241-4251.
19. Natasha Veligura, et al., *COVID-19's Impact on the Global Telecommunications Industry* May,2020, International Finance Corporation (IFC).
20. Rajpurkar, P., et al., *Squad: 100,000+ questions for machine comprehension of text.* 2016.
21. Ravi, K. and V.J.K.-B.S. Ravi, *A survey on opinion mining and sentiment analysis: tasks, approaches and applications.* 2015. **89**: p. 14-46.
22. Sokolova, M., G.J.I.p. Lapalme, and management, *A systematic analysis of performance measures for classification tasks.* 2009. **45**(4): p. 427-437.
23. Tundjungsari, V.J.I.J.o.C.S. and E.E. Volume, *Business Intelligence with Social Media and Data Mining to Support Customer Satisfaction in Telecommunication Industry.* 2013. **1**.
24. Wang, A., et al., *Glue: A multi-task benchmark and analysis platform for natural language understanding.* 2018.