# Twitter Analysis to Predict the Satisfaction of Telecom Company Customers

Latifah Almuqren
Department of Computer Science,
Warwick University,
Coventry, UK.
L.Almuqren@warwick.ac.uk

Alexandra I. Cristea
Department of Computer Science,
Warwick University,
Coventry, UK
A.I.Cristea@warwick.ac.uk

## ABSTRACT
This research is aimed at mining Arabic tweets to measure *customer satisfaction* toward Telecom companies in Saudi Arabia, and to predict the ratio of *customer churn*. This report starts with a review of previous research in using Twitter to measure user satisfaction and subjectivity analysis for Arabic. Then, it provides our approach and future plan.

## Keywords
Semantic Sentiment Analysis (SSA), Arabic, Twitter, Sentiment, Customer Churn, Customer Satisfaction.

## 1. RESEARCH AIM
The global competition facing companies in the labour market, drive companies to strive for enhancing customer satisfaction, as much research correlates customer satisfaction with customer loyalty [5]. Traditionally, customer satisfaction has been measured through customer interviews and questionnaires, but these cannot measure it in real time [10]. Therefore, new research is needed to measure customer satisfaction based on real time methods.

Semantic Sentiment Analysis (SSA) in Arabic micro-blogs is still an inadequately researched area, and Arabic lexicons are few and limited [15]. Additionally, the Arabic language is quite challenging for Natural Processing Language (NLP) [11], due to the variety of forms in Arabic language, such as Modern Standard Arabic (MSA) and the informal Arabic [4]. Developing NLP tools for Arabic text requires an understanding of the unique Arabic internal structure [4].

However, current subjectivity and sentiment analysis tools are designed mainly for the English language and there is a severe lack of tools for Arabic. Accordingly, this study examines microblogging site mining techniques for the purpose of capturing user satisfaction towards Telecom companies in Saudi Arabia, and how we can use that data to provide recommendations to these companies. In addition, this study intends to introduce a notion of customer interaction for Saudi Telecommunication Companies, based on a prediction model of the 'lost customer' phenomena (*customer churn*). Moreover, this study intends to contributes towards Arabic Sentiment Analysis (ASA), by building an Arabic dialect lexicon.

## 2. RELATED WORK
This research plans to use Semantic Sentiment Analysis (SSA) of Arabic tweets to measure customer satisfaction.

Collines et al. [9] measured public transport rider satisfaction towards transit system services using the riders' tweets on Twitter. This research helped the transit system to improve the service quality and safety monitoring, by adding more personnel. They analysed the tweets of riders along the Chicago Transit Authority (CTA) rapid transit system, using a Sentiment Strength Detection Algorithm (SentiStrength), to detect rider sentiments in real time.

Other research has reviewed and classified the state of the art, based on the different methods used for Arabic subjectivity and sentiment analysis which are: supervised learning using machine learning methods [3,14], unsupervised learning using sentiment lexicons [1,16] and a hybrid approach, which combines the two techniques [2,16].

Arabic language is rich, but comparing it with similar languages, there is a lack of a large corpora [12]. Several attempts have been made to accomplish Arabic corporas, such as [13, 20]. Unfortunately, some studies have indicated that there are some shortcomings in the existing corpora, such as the availability of the corpora, the strict procedure for a permission of reusing aggregation data and most of the existing corpora are not free or just for subscribers [20]. There are some attempts to build a wide polarity Arabic lexicon of MSA [3]. In contrast, there are few attempts to build an Arabic dialect lexicon, especially a Saudi dialect lexicon [8].

## 3. APPROACH AND METHOD
To accomplish the aim of the study, we use thus a *hybrid approach* [2,16]. Based on our literature review, the following research steps have already started (below, the level of current accomplishment is also given):

1. **Selecting the Saudi Telecom Companies**: STC, Zain, and Mobily their services, and their official Twitter accounts.
2. **Extracting patterns:** So far, Node XL [18] was used to retrieved a maximum of 2000 Arabic tweets using the following hash tags: #STC, #Mobily, #Zain, @STCcare, @STC,@Mobily1100,@Mobily,@Zainksa, and @ZainHelpSA. These are based on the names of the companies in step 1 above. The result showed the most frequent 8 services provided by Saudi Telecommunication companies that are mentioned in the customers tweets: Internet, speed, coverage radio, services after sell, call centre, explorer, systems, and fibre communication.
3. **Data Collection:** Build a corpus of Arabic SSA messages via the Twitter semantic (search) API [1,10,17], using a Python script searching for real time tweets that mention Telecom companies using the hashtags (as defined in step 2) to monitor the latest sentiments of Telecom customers continuously, for six months, starting from January 2016 (the process has already started).
4. **Designing an exploratory survey**: This refers to surveying Telecom Companies' customers which published in Twitter for public consumption. The aims are: to find out if the

Telecom companies users used the official Twitter accounts of the Saudi Telecom companies for communication; to define the user satisfaction metrics through the user's perspective, and to collect some behaviours that can be correlated with customer churn. Survey elements were constructed based on the measures specified by the Saudi Communications and Information Technology Commission and from other related researches [7].

Challenges to be faced start from providing the server and appropriate computer to save the huge amount of data, learn the Python program, issues such as frequent retweets in the corpus, and ending with the bureaucracy involved in arranging for collaboration permissions with the Saudi Telecom Companies, to provide us historical data of their customers, to use in building the prediction model.

## 4. OUTCOMES

The uniqueness of this study relies in the endeavour of using Twitter mining to predict potential customer loss (churn) in Saudi Telecom Companies, which has not been attempted before. Another outcome will be that of building a comprehensive Saudi dialect lexicon.The final contribution of this study will be capitalised as recommendations to these companies, based on monitoring in real time their customers' satisfaction in Twitter.

## 5. FUTURE PLAN AND TIMELINE

The timeline for the research encompasses September 2015 to August 2022, for the whole duration of the part-time PhD study.

Further steps should be accomplished to achieve the aim of this study, as follows (see also Figure 1).

1. **Manual annotation:** Cooperate with Arabic native speakers to annotate the corpus with strongly positive, positive, neutral, negative and strongly negative sentiment.
2. **Pre-Processing Tweets**: Normalise and tokenise all tweets.
3. **Building an In-Domain Arabic Lexicon**: Use a corpus-based approach to build the lexicon, using all data from the same domain [1]. In corpus-based approaches, the words of the lexicon are formed from the corpus using a seed list of known sentiment words [1]. There are different approaches to find words of similar or opposite polarity of the word. One of the prominent works in this approach was [19]. They used the AltaVista search engine to find the sentiment of a given word, by calculating the association strength between that word and a set of positive words minus the association strength between the word and a set of negative words. The association strength is measured using Pointwise-Mutual Information (PMI) [19].
4. **SSA**: Detect customer satisfaction using SSA for each tweet, applying classifiers with proven high accuracy for Arabic text: Support Vector Machines and Naïve Bayes [3,10,14]. Then use the results and the customer historical data in building a prediction model.
5. **Applying performance metrics** to the classifiers.

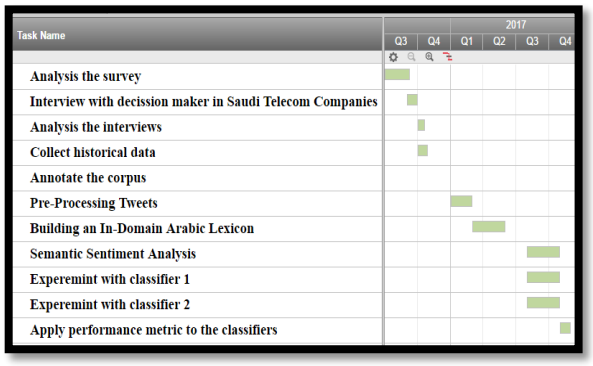


**Figure 1: Future plan time line**

## 6. REFERENCES


[1] Abdulla, N., Majdalawi, R., Mohammed, S., Al-Ayyoub, M. and Al-Kabi, M. 2014. Automatic lexicon construction for Arabic sentiment analysis. *In Proceedings of the Future Internet of Things and Cloud (FiCloud)*, (Barcelona,Spain, 2014). IEEE , 547-552. DOI= 10.1109/FiCloud.2014.95

[2] Abdul-Mageed,M., Kübler, S. and Diab, M. 2014. SAMAR: A system for subjectivity and sentiment analysis of Arabic *social media*. *Computer Speech and Language*. 28, 1 (January. 2014),20-37.

[3] Ahmed, S., Pasquier, M. and Qadah, G. 2013. Key issues in conducting sentiment analysis on Arabic social media text. *In 9th International Conference on Innovations in Information Technolog (IIT)*, (Abu Dhabi, UAE, March 17-19, 2013). IEEE, 72-77. DOI= 10.1109/Innovations.2013.6544396

[4] Albraheem, L. and Al-Khalifa, H. S. 2012. Exploring the problems of sentiment analysis in informal Arabic. *In Proceedings of iiWAS*, (Bali,Indonesia, December, 2012). IIWAS '12, ACM, New York, NY, 415-418. DOI= 10.1145/2428736.2428813

[5] Allola, A. 2011. Measuring the quality of services provided by Jawwal Co. from the viewpoint of customers in the provinces of Gaza Strip. Master thesis, University of Alazhar, Gaza, Palestine. Retrived January 19,2016 from www.alazhar.edu.ps/Library/aattachedFile.asp?id_no=004473

[6] Al-Twairesh,N., Al-Khalifa, H. and Al-Salman, A., Subjectivity and sentiment analysis of Arabic: Trends and challenges. 2014. *In Proceeding of the 11th International Conference on Computer Systems and Applications*,(Doha, Qatar, November 10-13, 2014). IEEE/ACS, 148-155. DOI= 10.1109/AICCSA.2014.7073192

[7] Athanassopoulos, A. D. and Iliakopoulos, A. 2003. Modeling customer satisfaction in telecommunications: Assessing the effects of multiple transaction points on the perceived overall performance of the provider. *Production and Operations Management*. 12, 2 (June. 2003) 224-245. DOI= 10.1111/j.1937-5956.2003.tb00502.x

[8] Azmi, A. and Alzanin, S. 2012. Aara'- A System For Mining The Polarity Of Saudi Public Opinion Through E-Newspaper Comments. *Journal of Information Science*. 40, 3 (March, 2012),398-410.

[9] Collins, C., Hasan, S. and Ukkusuri, S. V. 2013. A novel transit rider satisfaction metric: Rider sentiments measured from online social media data. *Journal of Public Transportation.* 16, 2 (2013), 21-45. DOI= http://dx.doi.org/10.5038/2375-0901.16.2.2.

[10] Duwairi, R.M. and Qarbaz,I. 2014. Arabic Sentiment Analysis Using Supervised Classification. *In Proceeding of the International Conference on Future Internet of Things and Cloud (FiCloud),* (Barcelona, Spain, November 27-29, 2014). IEEE Computer Society, 579-583. DOI= 10.1109/FiCloud.2014.100

[11] Farghaly A. and Shaalan, k. 2009. Arabic natural language processing: Challenges and solutions. *Transactions on Asian Language Information Processing*. 8, 4(December. 2009), 1-22.

[12] Korayem, M., Crandall, D. and Abdul-Mageed, M. 2012. Subjectivity and sentiment analysis of arabic: A survey. *In Advanced Machine Learning Technologies and Applications.* (2012). Springer Berlin Heidelberg, 128-139.

[13] Maamouri, M., Bies, A., Buckwalter, T. and Mekki, W. 2004. The penn arabic treebank: Building a large-scale annotated arabic corpus. *In the Proceeding of the NEMLAR conference on Arabic language resources and tools* (September, 2004). 466-467

[14] Mountassir, A., Benbrahim, H. and Berrada, I., 2012. In Bramer, M., and Petridis, M, A cross-study of Sentiment Classification on Arabic corpora. *Research and Development in Intelligent Systems XXIX*, Springer London ,2012, 259-272.

[15] Mourad, A. and Darwish, K. 2013. Subjectivity and Sentiment Analysis of Modern Standard Arabic and Arabic Microblogs. *In Proceedings of the 4th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, ( Atlanta, Georgia, June, 2013). Association for Computational Linguistics, 55–64.

[16] Ohana, B. and Tierney, B. 2009. Sentiment classification of reviews using SentiWordNet. *In Proceedinf of the 9th. IT & T Conference*, (Dublin, Ireland, October 22-23,2009). Dublin Institute of Technology, 13.

[17] Pak, A. and Paroubek, P. 2010. Twitter as a Corpus for Sentiment Analysis and Opinion Mining. *In Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10), European Language Resources Association (ELRA),* (Valletta, Malta, 2010), 1320-1326.

[18] Smith, M.A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C., Capone, T., Perer, A. and Gleave, E. 2009. Analyzing (social media) networks with NodeXL. *In Proceedings of the fourth international conference on Communities and technologies* ( New York, USA, June, 2009).ACM, 225-264.

[19] Turney, P.D. 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. *In Proceedings of the 40th annual meeting on association for computational linguistics* (Stroudsburg, PA, USA, July, 2002). Association for Computational Linguistics, 417-424.

[20] Zaghouani, W. 2014. Critical survey of the freely available Arabic corpora. *In Proceedings of the Workshop on Free/Open-Source Arabic Corpora and Corpora Processing Tools, LREC* ( May, 2014). 1-8