World Academy of Science, Engineering and Technology
International Journal of Computer and Information Engineering
Vol:15, No:02, 2021

# Short Text Classification for Saudi Tweets

**Authors :** Asma A. Alsufyani, Maram A. Alharthi, Maha J. Althobaiti, Manal S. Alharthi, Huda Rizq

**Abstract :** Twitter is one of the most popular microblogging sites that allows users to publish short text messages called 'tweets'. Increasing the number of accounts to follow (followings) increases the number of tweets that will be displayed from different topics in an unclassified manner in the timeline of the user. Therefore, it can be a vital solution for many Twitter users to have their tweets in a timeline classified into general categories to save the user's time and to provide easy and quick access to tweets based on topics. In this paper, we developed a classifier for timeline tweets trained on a dataset consisting of 3600 tweets in total, which were collected from Saudi Twitter and annotated manually. We experimented with the well-known Bag-of-Words approach to text classification, and we used support vector machines (SVM) in the training process. The trained classifier performed well on a test dataset, with an average F1-measure equal to 92.3%. The classifier has been integrated into an application, which practically proved the classifier's ability to classify timeline tweets of the user.

**Keywords :** corpus creation, feature extraction, machine learning, short text classification, social media, support vector machine, Twitter

**Conference Title :** ICCLSP 2021 : International Conference on Computational Linguistics and Semantic Prosody
**Conference Location :** Jeddah, Saudi Arabia
**Conference Dates :** February 18-19, 2021

Open Science Index, Computer and Information Engineering Vol:15, No:02, 2021 publications.waset.org/abstracts/130952.pdf