

E. Dritsas, G. Vonitsanos, I.E. Livieris, A. Kanavos, A. Ilias, C. Makris and A. Tsakalidis. [An efficient preprocessing tool for supervised sentiment analysis on Twitter data](#). In Advances in Information and Communication Technology, Springer, 2019 (accepted).



**Abstract** - Twitter Sentiment Classification is undergoing great appeal from the research community; also, user posts and opinions are producing very interesting conclusions and information. In the context of this paper, a pre-processing tool was developed in Python language. This tool processes text and natural language data intending to remove wrong values and noise. The main reason for developing such a tool is to achieve sentiment analysis in an optimum and efficient way. The most remarkable characteristic is considered the use of emojis and emoticons in the sentiment analysis field. Moreover, supervised machine learning techniques were utilized for the analysis of users' posts. Through our experiments, the performance of the involved classifiers, namely Naive Bayes and SVM, under specific parameters such as the size of the training data, the employed methods for feature selection (unigrams, bigrams and trigrams) are evaluated. Finally, the performance was assessed based on independent datasets through the application of  $k$ -fold cross validation.