World Academy of Science, Engineering and Technology
International Journal of Electrical and Computer Engineering
Vol:16, No:04, 2022

Topic Modelling Using Latent Dirichlet Allocation and Latent Semantic Indexing on SA Telco Twitter Data

Authors : Phumelele Kubheka, Pius Owolawi, Gbolahan Aiyetoro

Abstract : Twitter is one of the most popular social media platforms where users can share their opinions on different subjects. As of 2010, The Twitter platform generates more than 12 Terabytes of data daily, ~ 4.3 petabytes in a single year. For this reason, Twitter is a great source for big mining data. Many industries such as Telecommunication companies can leverage the availability of Twitter data to better understand their markets and make an appropriate business decision. This study performs topic modeling on Twitter data using Latent Dirichlet Allocation (LDA). The obtained results are benchmarked with another topic modeling technique, Latent Semantic Indexing (LSI). The study aims to retrieve topics on a Twitter dataset containing user tweets on South African Telcos. Results from this study show that LSI is much faster than LDA. However, LDA yields better results with higher topic coherence by 8% for the best-performing model represented in Table 1. A higher topic coherence score indicates better performance of the model.

Keywords : big data, latent Dirichlet allocation, latent semantic indexing, telco, topic modeling, twitter
Conference Title : ICHCE 2022 : International Conference on Human and Computer Engineering
Conference Location : Cape Town, South Africa
Conference Dates : April 14-15, 2022

Open Science Index, Electrical and Computer Engineering Vol:16, No:04, 2022 publications.waset.org/abstracts/147818.pdf