

Text classification for marketing research using pre-trained general language models

David Dornekott
Universität Osnabrück

Cite as:

Dornekott David (2021), Text classification for marketing research using pre-trained general language models. *Proceedings of the European Marketing Academy*, 50th, (94327)

Paper from the 50th Annual EMAC Conference, Madrid, May 25-28, 2021



Text classification for marketing research using pre-trained general language models

Abstract

Marketing research increasingly involves the use of unstructured text data, such as user-generated content collected from social media platforms. A common task in rendering these data useful for analysis is classification of the text units, for example by topic or their expressed sentiment towards brands and products. Pre-trained general language models based on the transformer architecture have recently shown very promising results in these problem domains, but have so far not seen extensive use by marketing researchers. The goal of this short study is to evaluate these models on common marketing research data such as microblog postings and product reviews and compare them to other methods commonly encountered in the marketing literature. Overall, transformer models show impressive performance, warranting further exploration of their capabilities and use cases.

Keywords: *text classification; user-generated content; machine learning*

Track: Digital Marketing & Social Media