

Natural Language Processing Application in Construction Domain: An Integrative Review and Algorithms Comparison

Abdolmajid Erfani, S.M.ASCE¹; and Qingbin Cui, A.M.ASCE²

¹Ph.D. student, Dept. of Civil and Environmental Engineering, Univ. of Maryland, College Park, MD 20742. Email: erfani@umd.edu

²Professor, Dept. of Civil and Environmental Engineering, Univ. of Maryland, College Park, MD 20742. Email: cui@umd.edu

ABSTRACT

As one of the main sources of information in construction projects, analyzing and exploiting text data emerged as a fast-growing body of literature in the construction domain recently. Various applications appear in a wide range of domains, specifically exploration of different Natural Language Processing (NLP) techniques. Usage ranges from construction contracts, design requirements, risk registers, change orders, claims and litigation documents, and safety reports. In this paper, the authors present a systematic review of the current NLP body of knowledge in the construction research domain. Various machine learning and deep learning-based NLP techniques, as well as their applications in construction research, are documented. Further, the authors introduce potential knowledge gaps and future research directions. In particular, this paper compares the performance of these NLP techniques through a risk classification problem. Analysis reports that Bidirectional Encoder Representations from Transformers (BERT) model outperforms other NLP models and achieves 80% of accuracy in the risk classification task.