

Semantic Search for NIC Codes

Prof. Pritesh Patil, Akash Varde, Vijay Hissal, Yash Inamdar

Department of Information Technology

AISSMS Institute of Information Technology, Pune, Maharashtra, India

akashvarde22@gmail.com, vijayhissal3487@gmail.com, yashinamdar3349@mail.com

Abstract: *Correct industry classification from textual business descriptions is important for regulatory enforcement, taxation, and policy decisions. Manual classification is time-consuming and leads to errors because of differences in terminologies and contexts. This work introduces a semantic search method utilizing BERT-based Natural Language Processing (NLP) methods for improving the precision of mapping business descriptions to the correct National Industrial Classification (NIC) code. The suggested system tokenizes input descriptions, performs embedding extraction through a fine-tuned BERT model, and uses cosine similarity to establish the most appropriate NIC code. The method is compared with common keyword-based approaches and proves to exhibit outstanding improvement regarding classification accuracy and relevance. An interactive interface is further constructed, enabling users to enter business descriptions and obtain the most appropriate NIC code. Experimental results verify that the system gives a stable and scalable solution for automatic industry classification*

Keywords: Semantic Search, BERT-based NLP Model, Text Classification, Industry Classification, Machine Learning, Cosine Similarity

