

Leaving no one behind? Embeddings-based topic modelling and LLMs to explore international cooperation projects implemented in Greece

Alessandro Meneghini¹, Arjuna Tuzzi²

¹Università di Padova – alessandro.meneghini.1@phd.unipd.it

²Università di Padova – arjuna.tuzzi@unipd.it

Abstract

The international debate on sustainable development puts the importance on pursuing a development that can reduce the inequalities within and between countries. In the context of the European Union (EU), the goal of reduction of inequalities is pursued through the Cohesion Policy, a set of funding and initiatives aiming at reducing the economic, social, and territorial disparities that still exist within the EU. While research in political science and economics used approaches based on policies and financial analysis, this research proposes to analyze a large corpus of project texts to identify the extent to which these initiatives are in line with the key areas assumed for the development of capabilities by the Human Development Index for inclusive and equitable development. For this study, the project summaries implemented in Greece (2014-2020) have been retrieved from the EU Platform Kohesio. A word embedding transformation and a topic modelling procedure using the BERTopic library have been used to extract topics related to the main capabilities (health, education, and income generation) using the cosine similarity measure. We employ Mistral-7b for identifying the beneficiaries of the projects to obtain a descriptive picture of the themes and strategic orientations of these projects. The results describe the country's developmental perspectives mainly in terms of economic development, including local capacity-building initiatives, but with marginal inclusion of vulnerable populations.

Keywords: word embeddings, project analysis, information retrieval, Mistral-7b, BERTopic.