

# EconBERTa - teaching computer to understand language of economics.

Piotr Antoniak

3<sup>rd</sup> December, 2021

## Abstract

Working with text data, also called Natural Language Processing, has a quite long history in the field of economics. However, latest advances in dealing with this unstructured type of data are used rarely. This work introduces EconBERTa and SEconBERTa - transformers trained on economics texts from RoBERTa checkpoint. I show that adapting them on a large data set, yield models that are quite good at “understanding” economics texts. EconBERTa can be further trained on downstream tasks such as classification while SEconBERTa excels at embeddings extraction and computing sentence/document similarity. Models and the code are available at [\[link\]](#).

## 1 Introduction

## 2 Data

### 2.1 Textual Data

## 3 Methodology

## References

- Devlin, Jacob et al. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*. arXiv: [1810.04805 \[cs.CL\]](#).
- Kingma, Diederik P. and Jimmy Ba (2014). *Adam: A Method for Stochastic Optimization*. arXiv: [1412.6980 \[cs.LG\]](#).

# Appendices

Appendix A:

- aaa