
Классификация товаров по ОКПД2 кодам

A Preprint

Фирсов Сергей
Кафедра интеллектуальных систем
МФТИ
firsov.sa@phystech.edu

Всеволод Михайлович Старожилец
Кафедра интеллектуальных систем
Фореक्सис
vsevolod.starozhilets@antirutina.net

Abstract

Исследование направлено на решение задачи классификации товаров по ОКПД 2 кодам с использованием кратких текстовых описаний. Основная цель - повышение точности и сокращение ресурсозатратности классификации, анализируя влияние глубины классификатора ОКПД 2. Для достижения этих целей предлагается метод построения текстовых эмбедингов с использованием нейросетевых технологий, таких как spaCy. Задача усложняется необходимостью предварительной обработки данных для трансформации исходных описаний в стандартизированные короткие тексты, адаптированные для анализа. Используются данные государственных закупок 2022 года, охватывающие около 40% открытых источников, что обеспечивает достаточный объем и разнообразие информации для анализа. Также часть этих данных будет критерием оценки работы программы. Новизна заключается в применении методов машинного обучения к индустриальной задаче, что обещает улучшение в процессах логистики, учёте и анализе в сфере закупок.

Keywords ОКПД 2 code · text analysis · task of classification

1 Введение

Список литературы