# Multiclass news classification with RuBert

Maxim Kishik

May 2024

**Abstract**

This document will provide you with guidelines how to fine-tune RuBert to identify class of news article.

## 1 Introduction

News articles encompass a wide range of topics and cater to diverse audiences. Identifying the class of a news article, such as politics, business, sports, or technology, is essential for efficient information retrieval, personalization, and content recommendation. Results and expierence from this work will be used in my diploma work about extraction information of popular events from news articles.

### 1.1 Team

**Maxim Kishik** prepared this document.

## 2 Related Work

In this section, you will describe in details the existing approaches to the problem you work on. For each approach, you need to provide a reference.

https://arxiv.org/abs/2107.06785 is a sample reference to the previous art. https://www.fruct.org/publications/volume-31/fruct31/files/Lag.pdf is a sample reference to the previous art in Russian.

## 3 Model Description

RuBERT (Russian, cased, 12-layer, 768-hidden, 12-heads, 180M parameters) was trained on the Russian part of Wikipedia and news data. We used this training data to build a vocabulary of Russian subtokens and took a multilingual version of BERT-base as an initialization for RuBERT. Additionally model got several linear layers and SoftMax layer at the end. All layers except added are frozen.

# 4 Dataset

An example dataset we will use is TopicNet/Lenta. Initial size of dataset is 263556 rows and contains only train part. I removed elements that appeared in dataset less than 6000 times, used label encoder for target variable and splitted dataset into train/val/test with presence of all classes in each group with batch size of 8.

| | Train | Valid | Test |
|---|---|---|---|
| Articles | 147588 | 63252 | 52710 |

Table 1: Statistics of the modified Lenta dataset.

# 5 Experiments

## 5.1 Metrics

Used metrics is F1 score with micro average.

## 5.2 Experiment Setup

Settings used during experiment:
Both experiments: AdamW optimizer and linear scheduler with warmup.
First experiment: learning-rate=1e-5, epochs=3.
One run had model with all parameters non-frozen, another model had all parameters frozen except classifier.
Unfrozen model F1 score: 0.8661623898889315
Frozen model F1 score: 0.5810225967062428
Second experiment: learning-rate=2e-5, epochs=5.
One run had model with all parameters non-frozen, another model had all parameters frozen except classifier.
Unfrozen model F1 score:
Frozen model F1 score:

# 6 Results

# 7 Conclusion

In this section, you need to describe all the work in short: what you have done and what has been achieved. E.g. you have collected a dataset, made a markup for it and developed a model showing the best results compared to other models.

## References