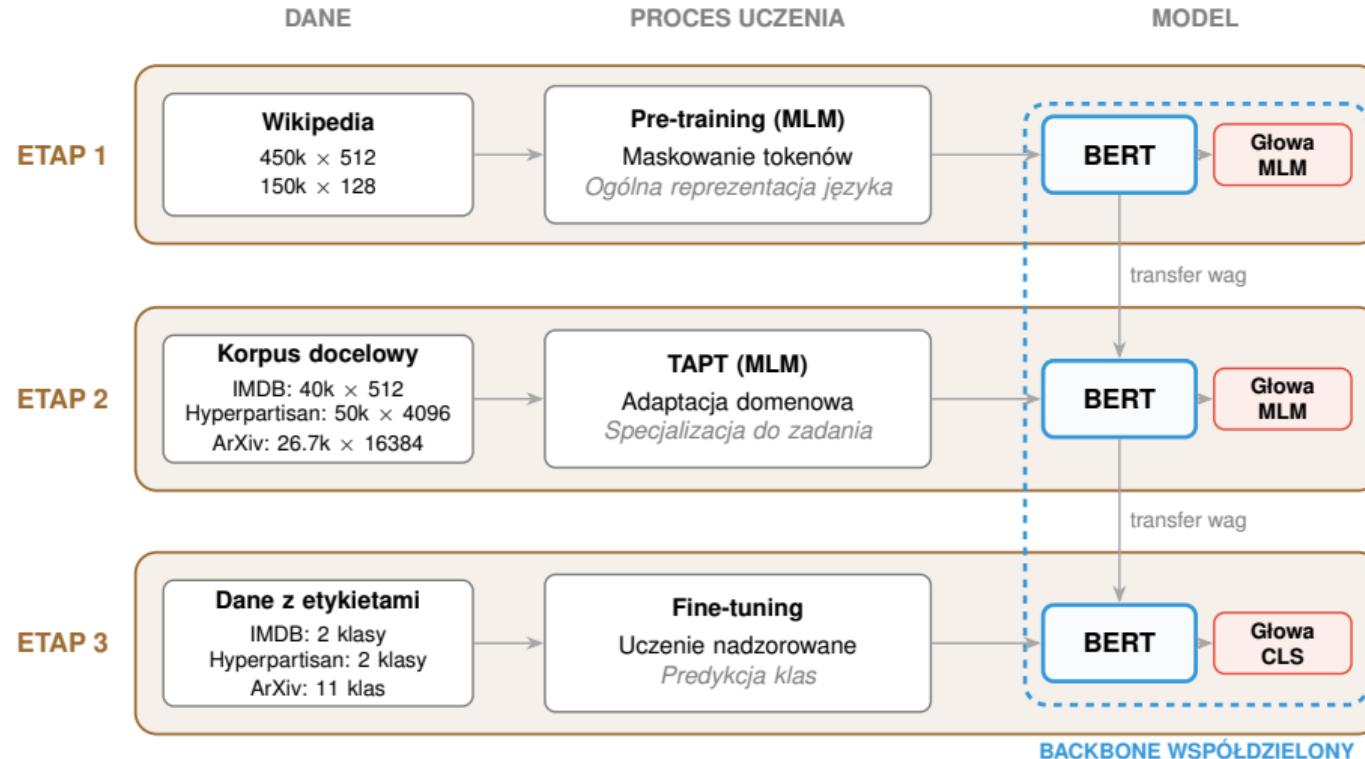
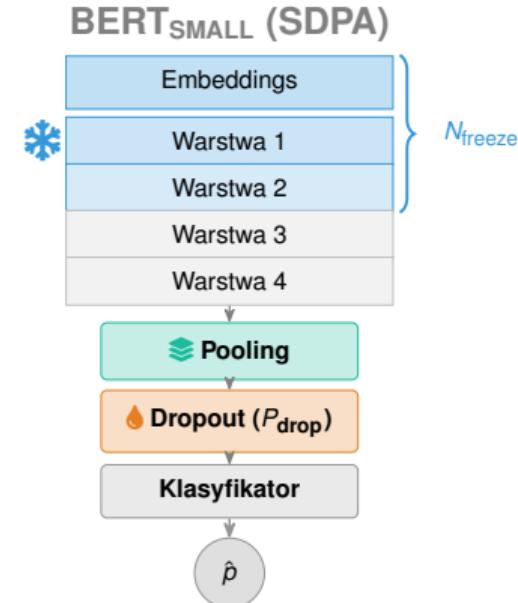
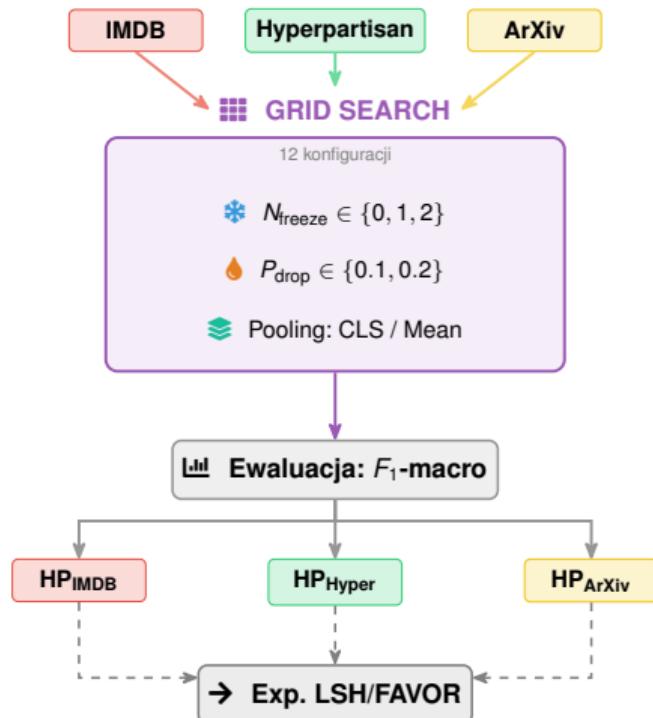


Pełny Proces Treningu

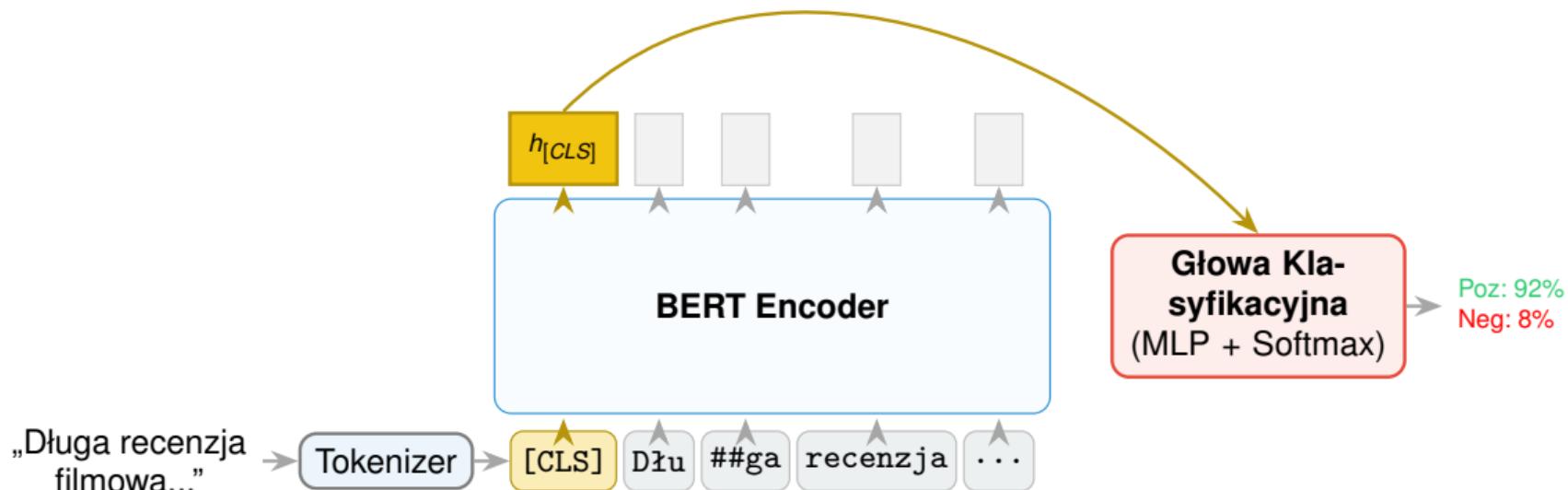


Optymalizacja hiperparametrów Finetuningu

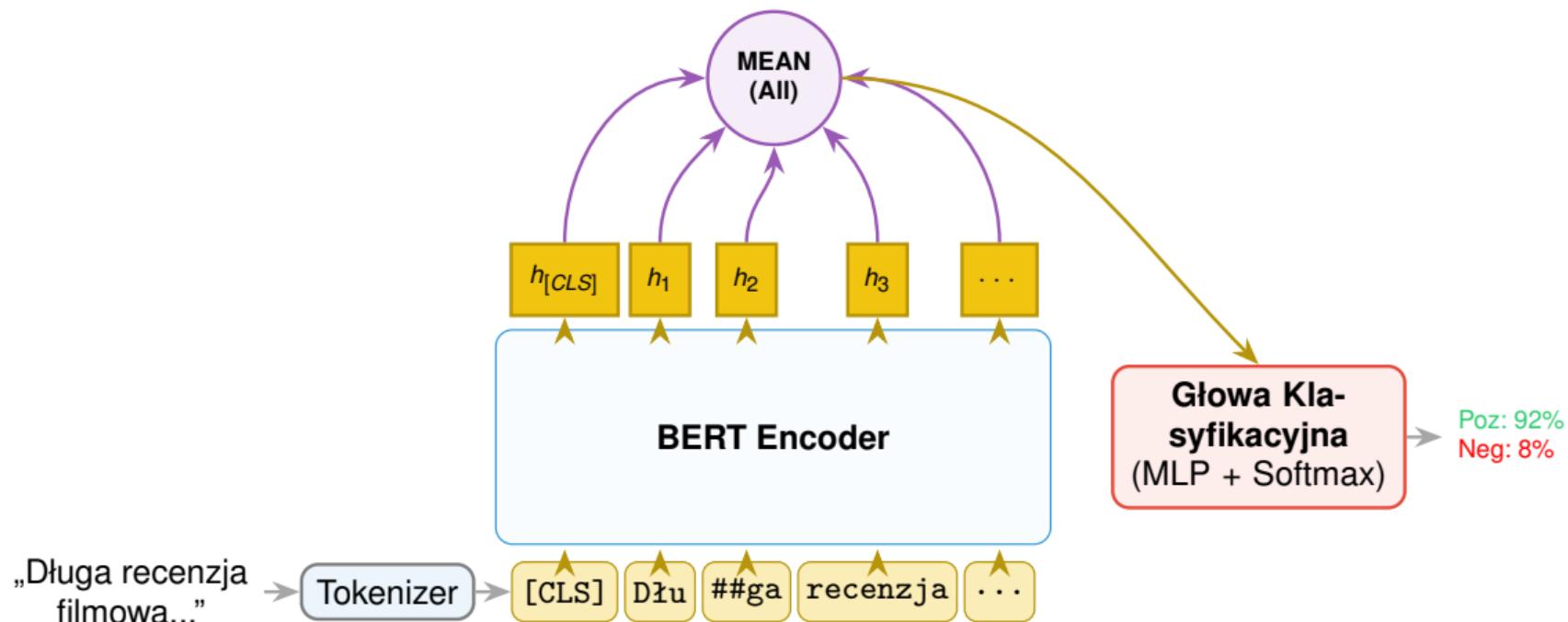


Cel: Wybranie optymalnych hiperparametrów finetuningu Metryka: $F_1\text{-macro}$

Proces Klasyfikacji (BERT Standard)



Proces Klasyfikacji (Mean Pooling - Wszystkie Tokeny)



Wyniki

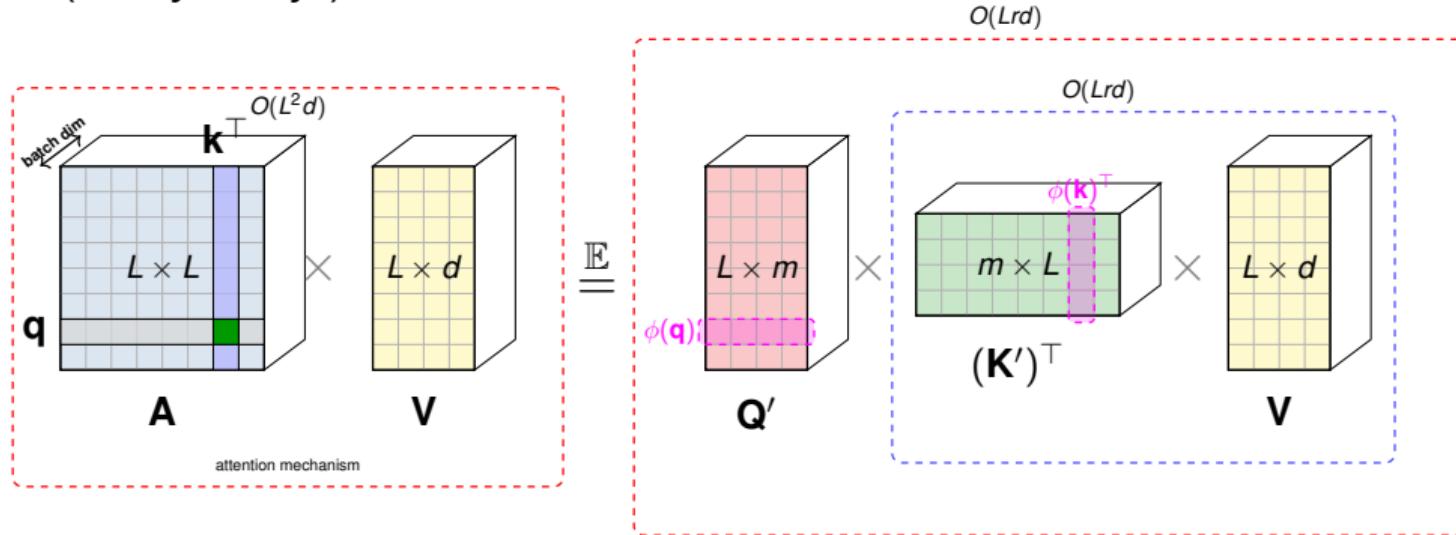
Tabela: Wyniki F1-macro na zbiorze testowym

Model	IMDB	Hyperpartisan	Arxiv
<i>TF-IDF+LR</i>	—	—	—
<i>CLS pooling</i>			
$N_{freeze} = 0, P_{drop} = 0.1$	93.22	—	—
$N_{freeze} = 0, P_{drop} = 0.2$	93.22	—	—
$N_{freeze} = 1, P_{drop} = 0.1$	92.95	—	—
$N_{freeze} = 1, P_{drop} = 0.2$	92.67	—	—
$N_{freeze} = 2, P_{drop} = 0.1$	92.97	—	—
$N_{freeze} = 2, P_{drop} = 0.2$	92.93	—	—
<i>Mean pooling</i>			
$N_{freeze} = 0, P_{drop} = 0.1$	92.90	—	—
$N_{freeze} = 0, P_{drop} = 0.2$	92.93	—	—
$N_{freeze} = 1, P_{drop} = 0.1$	92.99	—	—
$N_{freeze} = 1, P_{drop} = 0.2$	92.87	—	—
$N_{freeze} = 2, P_{drop} = 0.1$	92.83	—	—
$N_{freeze} = 2, P_{drop} = 0.2$	92.85	—	—

Tabela: Optymalne hiperparametry wybrane na podstawie F1-macro na zbiorze walidacyjnym

Zbiór danych	N_{freeze}	P_{drop}	Pooling
IMDB	2	0.2	Mean
Hyperpartisan	—	—	—
ArXiv	—	—	—

FAVOR+ (Modyfikacja)



$$\phi(x) = \frac{1}{\sqrt{m}} \exp\left(-\frac{1}{2}\|x\|^2\right) [\exp(\omega_1^\top x), \dots, \exp(\omega_m^\top x)]^\top$$

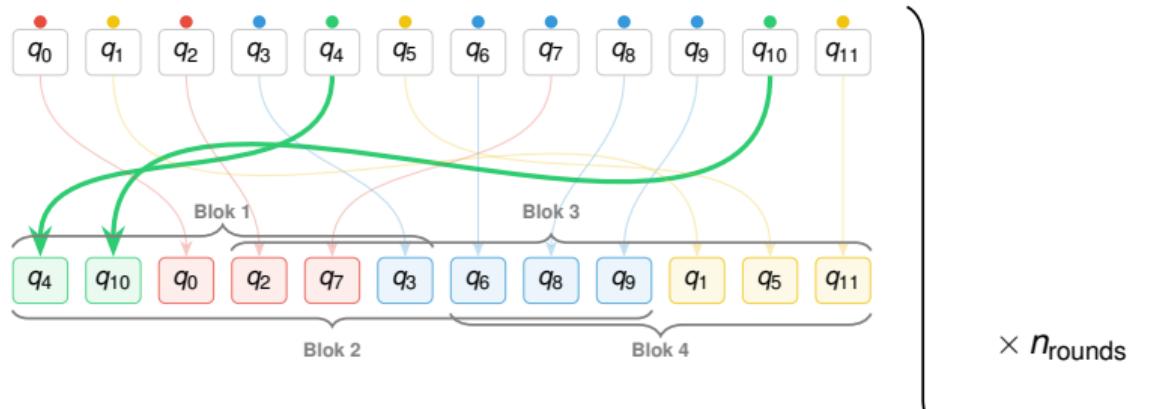
$$\phi^*(x) = \frac{1}{\sqrt{2m}} \exp\left(-\frac{1}{2}\|x\|^2\right) [\exp(\omega_1^\top x), \dots, \exp(\omega_m^\top x), \exp(-\omega_1^\top x), \dots, \exp(-\omega_m^\top x)]^\top$$

$$\begin{aligned} \mathbb{E} [\phi(q)^\top \phi(k)] &= \mathbb{E} [\phi^*(q)^\top \phi^*(k)] = \exp(q^\top k) \\ \text{Var}(\phi^*(q)^\top \phi^*(k)) &\leq \text{Var}(\phi(q)^\top \phi(k)) \end{aligned}$$

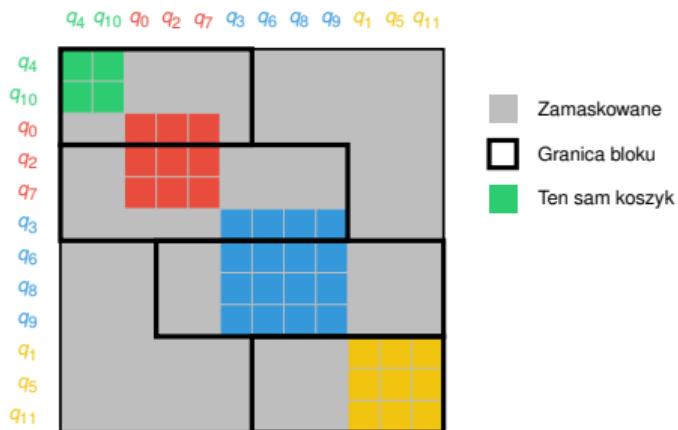
LSH Attention

1. HASHOWANIE

Przypisanie koszyka (kolor)



3. UWAGA BLOKOWA

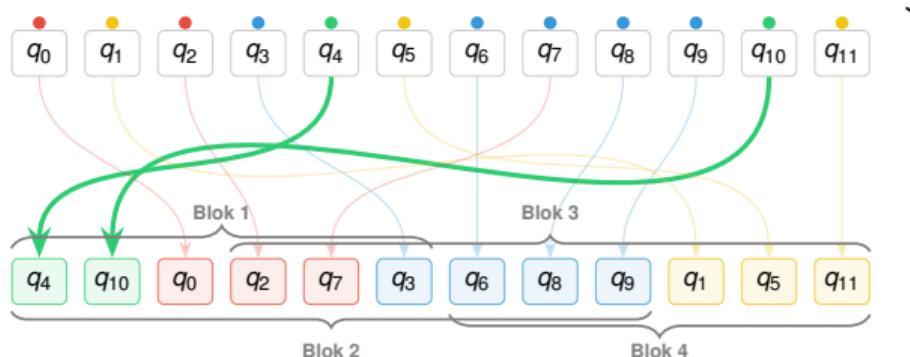


$$\frac{1}{n_{\text{rounds}}} \sum_{r=1}^{n_{\text{rounds}}}$$

LSH Attention (Modyfikacja)

1. HASHOWANIE

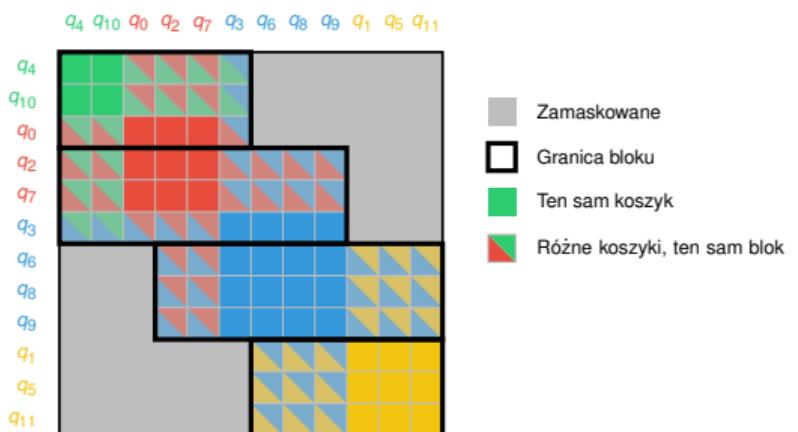
Przypisanie koszyka (kolor)



2. SORTOWANIE

Grupowanie wg koszyków

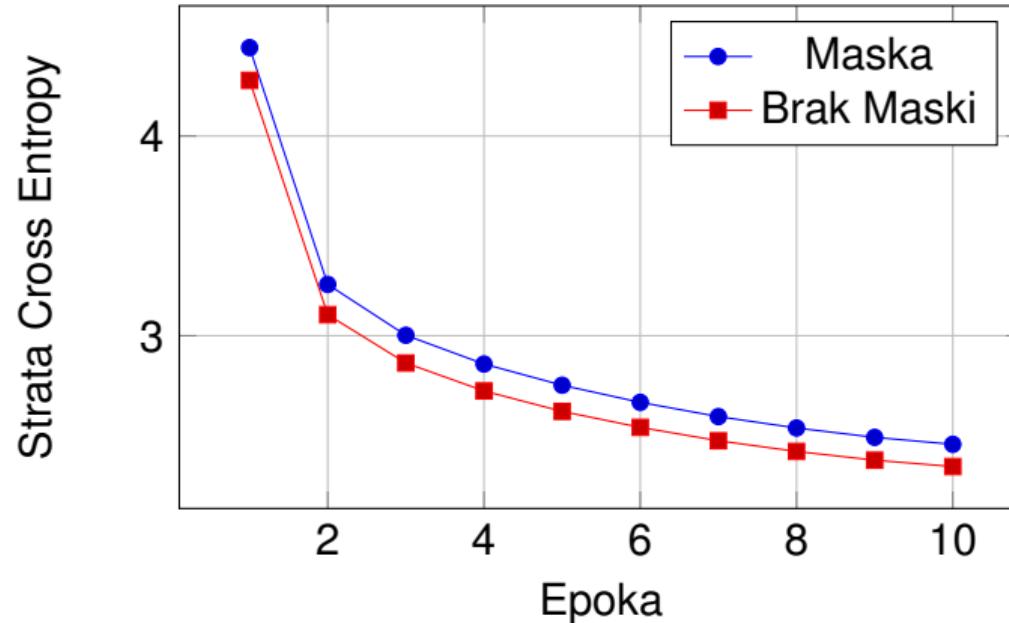
3. UWAGA BLOKOWA



$\times n_{\text{rounds}}$

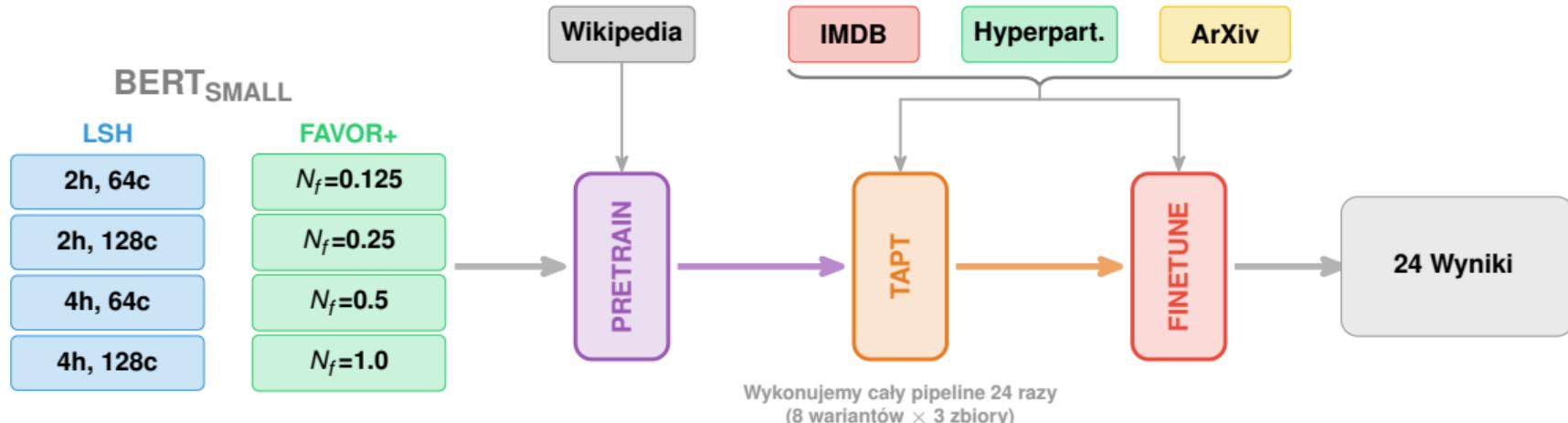
$$\frac{1}{n_{\text{rounds}}} \sum_{r=1}^{n_{\text{rounds}}}$$

LSH Bez Maski



Rysunek: Porównanie pretreningu LSH – na zbiorze Wikipedia w konfiguracji `num_hashes = 2, chunk_size = 64` – w zależności od maskowania wewnętrz koszyków. Wykres przedstawia przebieg straty cross-entropy w kolejnych epokach dla dwóch wariantów: `mask_withing_chunks=true (maska)` oraz `mask_withing_chunks=false (brak maski)`. Czas treningu i zużycie GPU różnią się w stopniu znikomym.

Warianty atencji LSH i FAVOR+



Cel: Porównanie mechanizmów LSH i FAVOR+ z różnymi parametrami

Metryki: F_1 -macro, czas, RAM

Wyniki