

Klasyfikacja danych tekstowych z użyciem sentence transformers + wielowarstwowej sieci neuronalowej

1 Cel Projektu

Celem projektu jest zastosowanie klasyfikacji tekstu przy użyciu sentence-transformerów z własną siecią neuronalną feed forward jako head. Projekt powinien obejmować etapy związane z:

- Pobranie i przygotowaniem danych (huggingface datasets)
- Wytrenowaniem i walidacją modelu (pytorch)
- Prezentacją wyników i wniosków

2 Temat Projektu

Projekt będzie polegał na wyborze etykietowanego zestawu danych tekstowych i zastosowaniu na nim klasyfikacji za pomocą transformacji tekstu do wektorów przy użyciu sentence-transformerów a następnie wytrenowaniu klasyfikatora.

W analizie powinno zostać uwzględnione:

- Opis wybranego zestawu danych
- Wstępna analiza danych
- Preprocessing danych (jeśli jest taka potrzeba)
- Opis metody klasyfikacji i motywacja do jej wyboru
- Trening i walidacji modelu (pytorch)
- Ewaluacja modelu z użyciem odpowiednich metryk (torchmetrics)

- Analiza wyników i wnioski

Dane: Użytkownicy mogą użyć dowolnego zestawu danych tekstowych, które wybiorą.

3 Opis Technologii

Projekt polega na napisaniu funkcjonalności w języku Python z wykorzystaniem w dowolnym stopniu biblioteki pytorch.

4 Zasady zaliczenia projektu

Należy udostępnić repozytorium na github prowadzącemu oraz poinformować go o tym na adres email: kamil@czarnogorski.pl z jego nazwą oraz swoim imieniem i nazwiskiem.

5 Materiały

Wszystkie pomoce naukowe oraz dokumentacje zostały udostępnione podczas zajęć w repozytorium na github.com /Google Drive do którego studenci otrzymali dostęp na pierwszych zajęciach.

6 Termin oddania projektu

Dwa tygodnie przed ostatnim zjazdem.

7 Kontakt

kamil@czarnogorski.pl