

Hesaplamalı Anlambilim Dersi 2.Ödevi

Konusu: Sınıflandırma

Ne üzerinde?

<https://huggingface.co/datasets/turkish-nlp-suite/TrGLUE> koleksiyonunun cola (2 sınıflı) ve sst2 (2 sınıflı) veri kümeleri üzerinde

Veri kümelerinin eğitim ve test ayrımları mevcuttur.

sst2 için eğitim kümesini 60K→5K, test kümesini 9K→1K düşürerek kullanınız.

cola için eğitim kümesini 8K→5K, test kümesini değiştirmeden kullanınız.

A) Parametre sayısı büyük eşit 7 milyar olan en az 3 llm in (cosmos dpo, gemma2 9b it vb.) zero shot, 3 shot ve 5 shot performansları ölçümü (%20 test kümesi üzerinde)

B) Berturk, e5, jina vb. embedding lerle (en az 5) elde edilen temsillerin RF, SVM ve LogReg ile eğitim kümesiyle eğitimden sonra test kümesinde performans ölçümü

Bonus: llm lerin farklı prompt lar için başarıları inceleme

Sonuçlarınızı sayısal grafiklerle / tablolarla sununuz.

Derste sunmak üzere bulgularınızı anlatan bir sunum hazırlayınız. Yorumlarınız önemli.

Sınıfta sunum yapmayanlar bu ödevden en fazla %50 alabilirler.

Bu ödevde en çok 2 kişilik gruplar halinde çalışabilirsiniz. Grup halinde çalışanların 2 sinin de sunumda bulunmaları gerekmektedir.

Örnek LLM listesi:

https://huggingface.co/spaces/malhajar/OpenLLMTurkishLeaderboard_v0.2

Örnek embedding listesi:

<https://huggingface.co/spaces/mteb/leaderboard>

Teslim edilecekler: Kod, sunum

Teslim şekli: online.yildiz.edu.tr

Son teslim tarihi: 14 Nisan 2025 09:30