

Sentiment Analizáló

Mesterséges Intelligencia Házifeladat Specifikáció

Gosztolya Máté

FZRNZW

A szentimentum analízálás egy olyan folyamat, amely során érzelmeket azonosítunk szövegek alapján, például meghatározzuk, hogy egy adott szöveg pozitív, negatív vagy semleges érzelmi tartalmú-e. A projekt célja egy PyTorch alapú mesterséges intelligencia modell létrehozása, amely képes érzelmeket azonosítani szövegekben, például Twitter vagy Reddit bejegyzésekben.

A projektemben egy kaggle datasetet használok, ami 37 ezer Reddit, és 163 ezer Twitter bejegyzést tartalmaz, minden bejegyzéshez meg van határozva, hogy annak érzelmi töltete negatív (-1), semleges (0), vagy pozitív (1). A dataset linkje: <https://www.kaggle.com/datasets/cosmos98/twitter-and-reddit-sentimental-analysis-dataset?resource=download>

Végeredményül egy olyan Mesterséges Intelligencia modellt szeretnék készíteni, ami képes mondatokról meghatározni, hogy azok a három kategóriából melyikbe tartoznak.

A projektet Pythonban készítem, terv szerint az alábbi könyvtárak felhasználásával:

- torch: PyTorch mély tanulási keretrendszer.
- torch.nn: PyTorch neurális hálózatok moduljai.
- torch.optim: PyTorch optimalizálási algoritmusok.
- pandas: Adatmanipulációhoz és -kezeléshez használt könyvtár.
- sklearn.model_selection: Scikit-learn könyvtár modulja a modell kiválasztásához.
- sklearn.feature_extraction.text: Scikit-learn könyvtár modulja a szöveges adatok feature extraction-jéhez.
- sklearn.metrics: Scikit-learn könyvtár metrikáinak modulja.
- numpy: Numerikus számításokhoz használt könyvtár.
- json: JSON formátumú adatok kezeléséhez használt könyvtár.

A projekthez a következő Github repositorit fogom használni:

<https://github.com/mategm88/SentimentAnalyser.git>