

Offenzív szövegek automatikus felismerése

Schwan Krisztián
III. évf. gazdaságinformatikus BSc

Témavezető: Dr. Farkas Richárd
egyetemi docens

SZTE TTIK Számítógépes Algoritmusok és Mesterséges Intelligencia Tanszék

Az utóbbi évtizedben a közösségi média szerepe jelentősen megnövekedett, amelynek hatása globálisan érzékelhető. Napjainkban rengeteg adat generálódik minden egyes másodpercben, amit egyenként, emberi erővel lehetetlen lenne nyomon követni. Ahhoz, hogy ebből a terjedelmű szövegből érdemleges információt tudjunk kinyerni, automatikus megoldásokat, természetesnyelv-feldolgozást (Natural Language Processing, NLP) alkalmazunk.

Dolgozatom alapját az OffensEval versenykiírás képezi, ahol a közösségi médiában, pontosabban a Twitteren előforduló offenzív tartalmú tweetek azonosítása és kategorizálása a fő célkitűzés. Adott egy több, mint tízezer tweetet tartalmazó adathalmaz, amik kategorizálva vannak három szempont alapján: offenzív-e az adott tweet, célzott-e a gyűlöletbeszéd és kikre irányul a támadás. Ez alapján kell meghatározni csaknem ezer tweetről, melyik kategóriákba tartoznak.

Dolgozatomban több dokumentum osztályozási módszert is megvizsgálok a versenyfeladat által biztosított adathalmaz tanító- és mintakészletén, kezdve a lineáris osztályozóktól (n-gram felhasználással), a mélyebb, GloVe szóbeágyazási modellel segített konvolúciós neurális hálós osztályozókig, továbbá az ELMo dinamikus szóbeágyazás alkalmazásával. Mindezt Python nyelven valósítottam meg, többek között a Scikit-learn gépi tanuló, illetve a Keras neurális háló könyvtár segítségével.

A legjobban teljesítő megoldásom a 2019-es verseny hivatalos résztvevőivel versenyképes eredményeket ér el.