Gépitanulás projekt terv – 2020

# 1, Csapat név: *BiZoDaJa*

* Tag1: *Biacsi Zoltán*
* Tag2: *Daniel János Róbert*

# 2, Választott feladat

* Feladat leírása
  + Érzelemdetektálás Twitteren. Programunk tetszőleges Twitter bejegyzésről képes lesz becslést adni arra, hogy a bejegyzés pozitív, vagy negatív véleményt fejt ki.
  + hol érhető el?
    - <https://github.com/EagleBeagle/BiZoDaJaProjekt>
  + feladat típusa: osztályozás
  + motiváció
    - A közösségi média egyre fontosabbá válik üzleti és marketing szempontból, az emberek bejegyzései alapján pontos képet kaphatunk egy termék fogadásával kapcsolatban. Ezen információk kinyerését nagyban megkönnyíti az érzelemdetektálás, amely emiatt egy fontos és számunkra érdekes probléma.
    - Platformként a Twittert választottuk, mivel ez az egyik legnépszerűbb közösségi oldal, és kiváló eszközöket biztosít az adatok kinyerésére. Kisebb módosításokkal viszont akár más oldalakkal is működhet a programunk.
* Adatbázis leírása
  + Adatbázis elérése: <https://www.kaggle.com/kazanova/sentiment140/>
  + Formátuma: .csv
  + A predikálandó érték egész szám formájában adott (0 = negatív, 4 = pozitív)
* Nyelv/ gépi tanuló csomag
  + Python/scikit-learn, Keras

# 3, Baseline rendszer leírása

* Jellemzők: Tweetek szövege vektorizált formában, érzelem címke
* A train és test adatokat 80-20%-os arányban osztjuk fel. A kiértékelést konfúziós mátrix, osztályozási jelentés és pontosság alapján végezzük.
* Baseline: lineáris SVM osztályozó SGD optimalizációval
* Kiértékelési metrikák: Precision, Recall, F-score, Accuracy