



Magyarország első egyeteme*
1367

Textmining a társadalomtudományi kutatásban – egy keretrendszer tapasztalatai

Hornyák Miklós*

tanársegéd

Pécsi Tudományegyetem

Közgazdaságtudományi Kar

Kvantitatív Menedzsment Intézet

Bencze Péter

hallgató

Pécsi Tudományegyetem

Természettudományi Kar

Matematikai és Informatikai Intézet

Fraudológusok Társasága

*Big Data Szakmai Konferencia
Pécsi Tudományegyetem
Szentágothai János Kutatóközpont
Big Data Kutatócsoport
Pécs, 2017.11.30.*



PÉCSIKÖZGÁZ
ahonnan a karrier indul!

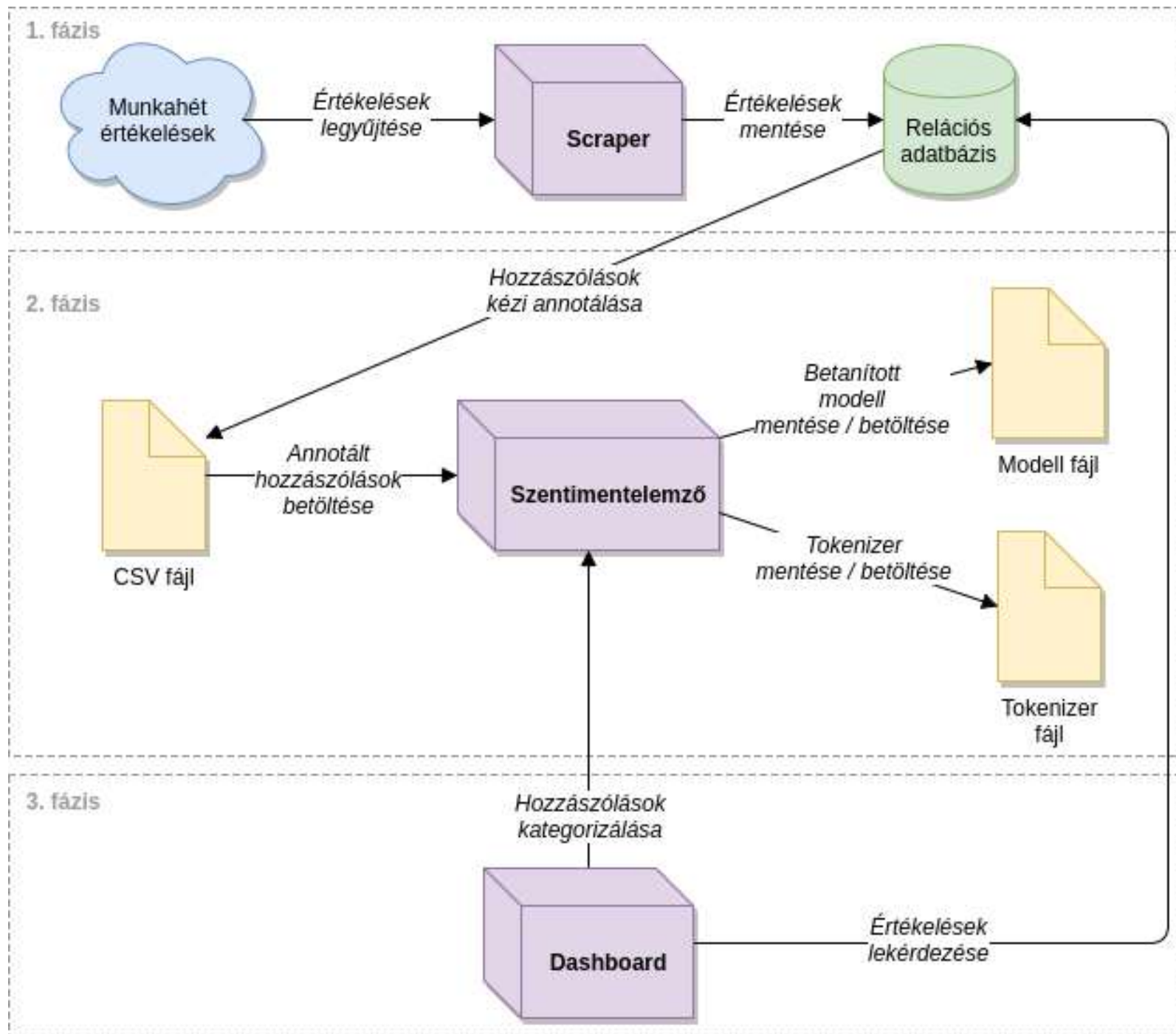
- Társadalomtudomány és big data
- A keretrendszer és a folyamat
- Adatgyűjtő modul
- Adatelemző modul
- Adatvizualizációs modul



- Big data adatforrások
 - Kérdőívezés -> közösségi érzékelők
 - Irodalomkutatás támogatása -> openaccess
- Technológiák
 - Szövegbányászat
 - Duo-mining
- Eredmények vizualizálása
 - Intuitív grafika -> szófelhők
- Tudományos életben validált eszközök
 - Python vs. R
 - Closed vs. opensource
- Nyelvi probléma
 - Magyar nyelv
 - LIWC szótár



Keretrendszer és folyamat



- 1. Adatgyűjtés
- 2. Elemzés
- 3. Vizualizáció

Scraping modul

- Nincs közvetlen hozzáférésünk az adatbázishoz, az adatokat kézzel leszedni nagyon időigényes lenne
 - Megoldás: webscraper + XPath
- Az oldal a lapozáshoz JavaScript-et használ, ezért ezért a scraper-nek rendelkeznie kell JavaScript értelmezési és végrehajtási képességekkel
 - Megoldás: Selenium – felhasználói interakciók imitálása

1. *Adatgyűjtés*
2. *Elemzés*
3. *Vizualizáció*

Keras – neural network - supervised learning

- Egy olyan modellt állítunk elő, amely képes viszonylag nagy pontossággal eldönteni egy hozzászólásról, hogy az pozitív vagy negatív hangvételű (szentimentelemzés)
- Tanulási fázisában olyan hozzászólásokra van szükség, amelyekről már tudjuk, hogy pozitív vagy negatív kategóriába tartoznak (tanuló adatbázis)
- Sequential modell: lineárisan egymásra épülő rétegek
 - 300 neuron
 - pozitív komment 1-hez, negatív esetén 0-hoz lesz közelebb

1. Adatgyűjtés
2. Elemzés
3. Vizualizáció

Dash – analitikus webes alkalmazások

Hogyan oszlanak meg a munkahelyek az értékelési szempontok szerint?

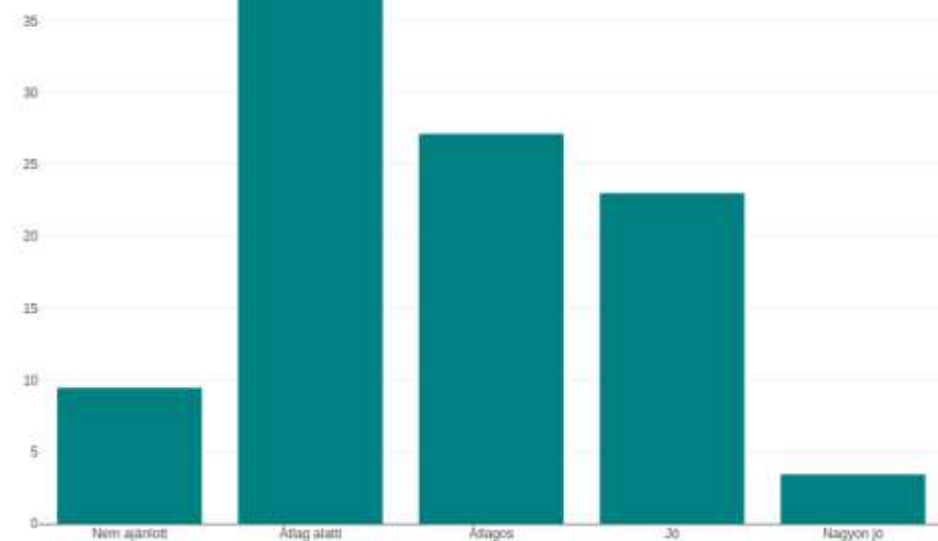
Az alábbi grafikonról leolvasható, hogyan oszlanak meg a vélemények a munkahelyekről, az értékelési szempontok szerint. Választhat, hogy az értéktengelyen a különböző értékelések száma, vagy ezek aránya jelenjen meg. A véleményezők a munkahelyeket egy-egy szempont szerint egy 1-től (nem ajánlott) 5-ig (nagyon jó) terjedő skálán pontozhatják.

Értéktengelyen megjelenő értékek:

Értékelések aránya

Értékelési szempont:

Fizetés



Hogyan alakulnak a munkahely értékelések a vállalkozás mérete szerinti csoportosításban?

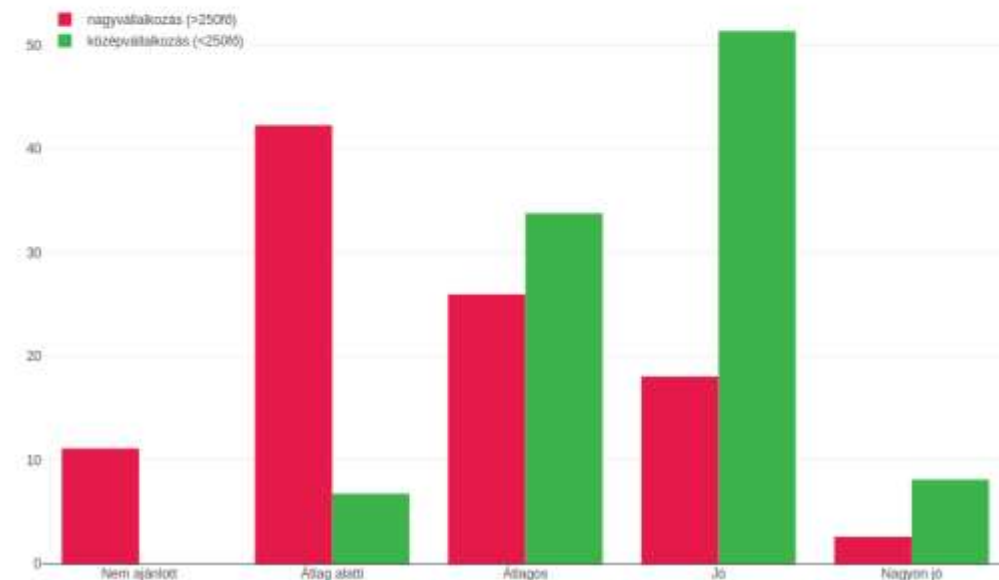
Az alábbi grafikonról leolvasható, hogyan alakulnak a munkahely értékelések a vállalkozás mérete szerinti csoportosításban. Választhat, hogy az értéktengelyen a különböző értékelések száma, vagy ezek aránya jelenjen meg. A véleményezők a munkahelyeket egy-egy szempont szerint egy 1-től (nem ajánlott) 5-ig (nagyon jó) terjedő skálán pontozhatják.

Értéktengelyen megjelenő értékek:

Értékelések aránya

Értékelési szempont:

Fizetés



Köszönjük a figyelmet!

Hornyák Miklós
hornyakm@ktk.pte.hu

Bencze Péter
benczepeter95@gmail.com

