

MESTERSÉGES INTELLIGENCIA

6. forduló



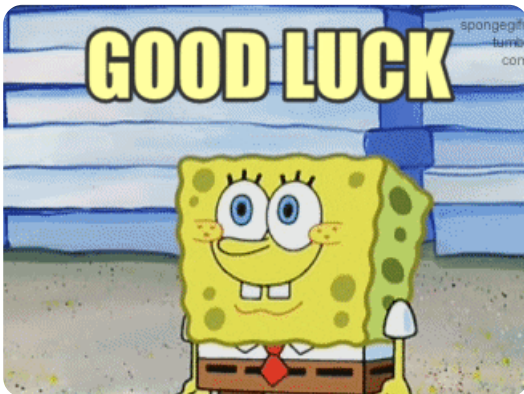
A kategória támogatója: Ulyssys Kft.

Ismertető a feladatlaphoz

Kezdj neki minél hamarabb, mert a feladatot a forduló záró időpontjáig lehet beküldeni, nem addig lehet elkezdni!

A feladatot javasoljuk NEM mobilon/tableten kitölteni!

Sok sikert!



Ebben a fordulóban 3 feladat lesz, egy könnyebb, egy közepes és egy komplexebb. Saját GPU-ra nem lesz szükség, helyette nyugodtan használj Colab-ot (<https://colab.research.google.com/>) !

Az utolsó feladathoz a kategória indítása után elérsz majd egy fire_imgaes.zip csatolmányt. Ha nem menne a notebookban a gdown letöltés, használd ezt a verziót!

Hogy ne fuss ki az időből, érdemes lehet előre ismerkedned:

a keras_ocr-rel: <https://github.com/faustomorales/keras-ocr>

Indítás utáni csatolmányok

1. feladat 2 pont

Melyik a kakuktktojás?

Válasz

- ☐ pandas
- ☐ seaborn
- ☐ python
- ☐ terrestrial
- ☐ anaconda

2. feladat 10 pont

Data Cleaning twitter

Van egy json fájlunk, ami 128837 twiterről leszedett bejegyzést tartalmaz. Minden itt tárolt bejegyzésnek megtalálható a szövege, hogy hányan likeolták és egy keyword ami a tweethez tartozik.

Sajnos mint a legtöbb social media felületen, itt is sok bot található, és az ő tweetjeik is bekerülhettek az adatok közé. Az egyik jellemzője a bot bejegyzéseknek, hogy éppen trending hashtaget használnak, ha van köze a bejegyzésükhöz, ha nincs.

Ahhoz hogy a bot bejegyzések egy részét kiszűrj, csak azokat a tweeteket tartsd meg, amelyek a keywordot tartalmazzák, de kizárólag NEM hashtag formátumban (#keyword). Tehát ha a kulcsszó bitcoin, akkor a "vegyél bitcoint" jó, az "olcsó #bitcoin" és a "bitcoin akció! #bitcoin" nem jó.

Az eredeti adat hány százaléka maradt meg?

adott: https://colab.research.google.com/drive/1xo3R_U3HMYaE8A1VtHW7F4gyNCy3KZqZ

Válaszok

- ☐ 82%
- ☐ 43%
- ☐ 51%
- ☐ 7%
- ☐ 15%
- ☐ 76%

3. feladat 15 pont

Elrejtett bizonyíték

Egy bűnügyi nyomozást vezetsz. Megalapozott gyanúd van egy vádlottal szemben, azonban a házában lefoglalt dolgok között nem találsz bizonyítékot.

Találtál viszont egy mappányi képet erdőtüzekről, amiket úgy látod gyakran nyitottak meg az elmúlt időszakban, de ránézésre nincs rajtuk semmi furcsa.

Próbálj ezen az adathalmazon elrejtett nyomokat keresni! (OCR (*optical character recognition*), keras_ocr) Mit találtál?

adott: https://colab.research.google.com/drive/1XVV4e4uE_OnmXVojz1Ww2nM-vm2kV-r1

Válasz

Megoldások beküldése