Comp1804 Coursework – Dataset links

Applied Machine Learning

Here you can find the links to the datasets for the coursework for Applied Machine Learning.

Instructions

1. Click on the link for your chosen sub-task. Clicking on the link will lead you to a Google Drive folder.
2. Once in the folder, you will see files whose names follow this structure: "XXX_surnames***.csv" and/or "XXX_surnames***.zip".
3. Look for the **.csv file** where the letters after "surnames" include the **first letter of your surname.**
   a. For example, my surname is Garasto so I'm looking for the letter G. If I saw a .csv file whose name contains "surnamesG", I would download it. Or, if I saw a .csv file whose name contains "surnamesEFGH", I would download it.
   b. If your surname starts with the letter A and you saw a .csv file whose name contains "surnamesA", you would download it. Or, if you saw a .csv file whose name contains "surnamesABCD", you would download it.
4. Check if there is also a **.zip file** where **the letters after "surnames" include the first letter of your surname**. If so, download that one too.
   a. For example, my surname is Garasto so I'm looking for the letter G. If I saw a .zip file whose name contains "surnamesG", I would download it. Or, if I saw a .zip file whose name contains "surnamesEFGH", I would download it.
   b. If your surname starts with the letter A and you saw a .zip file whose name contains "surnamesA", you would download it. Or, if you saw a .zip file whose name contains "surnamesABCD", you would download it.

Note that:

- The folders with "Peer review" data and "Amazon review" data **only have** .csv files.
- The folders with "Skin lesion" data and "Adverts" data **have both** .zip and .csv files.
- You DO NOT need to tell me which sub-task you choose. Just make sure to specify the sub-task on your submission on Moodle, including in the file names, as per instructions.

Links

- Sub-task 1: "Peer reviews" dataset link
- Sub-task 2: "Skin lesions" dataset link
- Sub-task 3: "Adverts" dataset link
- Sub-task 4: "Amazon reviews" dataset link

[3] P. Tschandl, C. Rosendahl, and H. Kittler, H, "The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions". *Scientific data*, vol. 5, no. 1, pp.1-9., 2018.