**Task Description**

**Description**

Websites to be scrapped 10 -12.

Pages per website 20-30 estimated.

Types of website: News websites. Two categories of websites will be targeted. Fake content and Real news websites of south Asia.

Authorization: Not sure about all but signing would not be required, since news portals have no sign in restriction.

Data to be scrapped:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sr#** | **Title** | **News description** | **URL** | **Label (TRUE or Fake)** | **Category ( Politics, Sports, Showbiz, world, tech, Health)** |
|  |  |  |  |  |  |

1. News scrapped from fake news websites would be labelled as Fake.
2. News from genuine news websites would be labelled as True in label column.
3. Image or video is NOT to be extracted. Only textual data is required.
4. Need approx. 10K plus records(tuples).

Real news Websites:

1. <https://www.geo.tv/>
2. <https://www.dawn.com/>
3. <https://www.thenews.com.pk/>
4. <https://arynews.tv/>
5. <https://www.samaa.tv/>
6. <https://factly.in/>
7. <https://www.thehindu.com/>
8. <https://indianexpress.com/>
9. [https://timesofindia.indiatimes.com](https://timesofindia.indiatimes.com/)
10. <https://pakobserver.net/>

Fake news Websites:

To be provided soon.