AI at Qure.ai

Let’s do a quick end to end project which represents the work of an AI guy/gal here at qure.ai. We will do

* Data collection
* Pre-processing
* Model training
* Model deployment

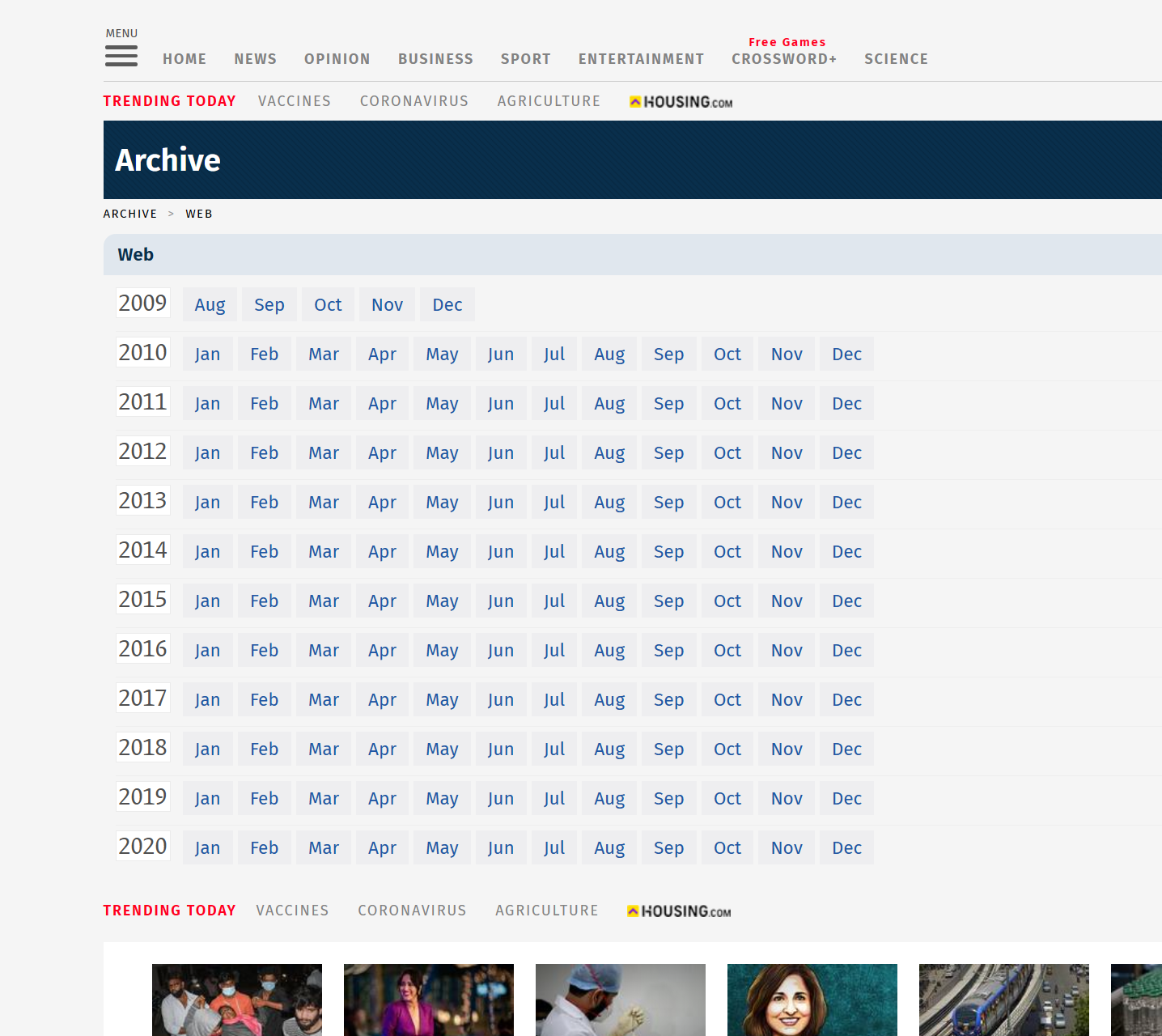
The problem we are going to solve is news text classification. You will be assessed on

1. Code Quality and Approach – 50
2. Accuracy - 20
3. Architecture - 20
4. Presentation in the report (max 2 pages) – 10

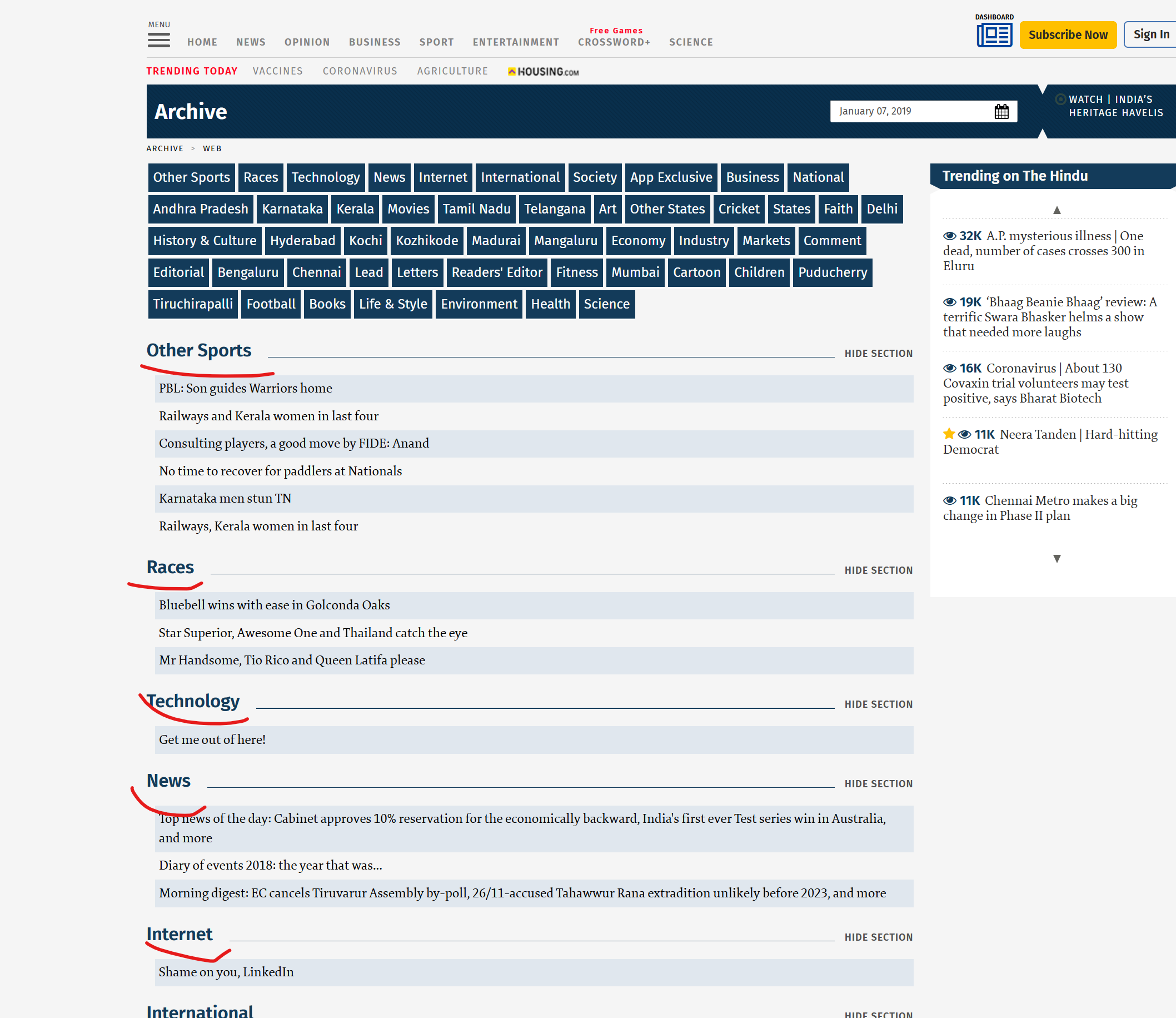
Please submit a report with your approach and relevant data. You are free to google and find as many relevant resources as possible. Just make sure to cite them in your report. This project is open-ended

# The Problem

Let us dive into the problem. We will have to first collect the data to do any machine learning. So, we will scrape news from *The Hindu*. Hindu maintains an archive of all articles here: <https://www.thehindu.com/archive/web/>



Here’s how example page for a day looks like:



You will get links of all the news articles for that day here. You can use them to scrape the text. You can use highlighted text as classes. You can also explore merging some of these into one class. You can also discard data from few classes if they do not make sense – like ‘News’ above. You need not scrape all the days – just scrape as much as you think fit. Make sure to explain your thought process in your report.

Train a reasonable model with this data – does not have to be cutting edge but should not be super inaccurate either. Your thought process around pre-processing, model training and improvement matters more than the actual model itself – be sure to present those in the report.

Once we train the model, we gotta deploy it. Create a web app with extremely simple UI where you can put in a new URL of the Hindu news article and it shows the predicted class.

Write a short (max 2-page) report and supply the code in a zip. Make sure you code is well documented too.