FAKE NEWS DETECTOR

Nermine ALI

Description of the project

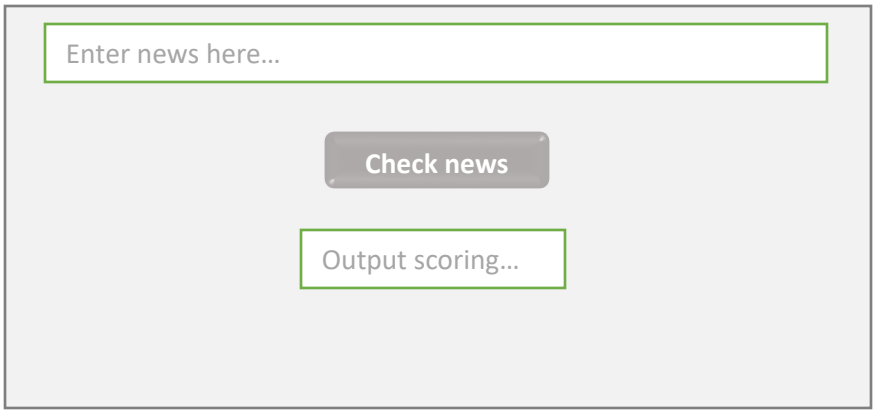
Fake news is one of the most trending media problems; it can have a very harmful effect since we cannot distinguish it from real news. This is where the fake news detector comes into play, to check the credibility of suspicious news before publishing it to the public. AI offers promising solutions for such tasks, especially in automating the fact checking process.
The goal of this project is to exploit the power of AI to fight the fake news problem. Python 3.x is the language of choice for this project.

Expected outcome*:*

- Web page: check button and textbox (see figure below)
- Input: news article (URL)
- Output: Score (0 to 100) for the credibility of the news

Evaluation of the source code:

- Module 1 (35%): Availability in trusted news sources
  Scoring according to the number of mentions in news sources
- Module 2 (65%) : AI Scoring



Hints to start

- To build the website I recommend using Django: https://docs.djangoproject.com/en/3.0/intro/
- Input: news article (URL), headlines…

Project Part 1: Algorithmic part

First, you need a database that contains trusted sources. I propose two options:

1. Get a list of *N* trusted sources containing (BBC, Reuters…): (name, facebook, twitter)
   ex.: Reuters, https://www.facebook.com/Reuters/, https://www.twitter.com/Reuters
   In each trusted source look if the news exists, using twitter API and Facebook API
   The idea is to match the input with content posted on the predefined sources official social media accounts.
   You can start with one source then generalize.
   Use Either FB or twitter. Both would be better.
   Based on the matches you can rate the credibility of the input.
2. Using NewsAPI: https://newsapi.org a gateway of all news

Based on the obtained result give a scoring if (0 to 100) if the news is true or fake.

Part 2: AI Scoring

- Step1: Search for a trained model (recommended) or train a model yourself.
  - Mainly, as an AI algorithm for this task, you will need **Natural Language Processing**
  - You might also look for transfer learning, it depends on what you can find on the internet
- Step2: Use this model to assess the article in question
- Step3: give a score (0 to 100) based on the output of your model

You can use **pytorch** as a framework: https://pytorch.org/ for AI. If you find better framework you can use it too, and mention this in your report.

Evaluation

1. 30% **Source code**: You should write a readable code and well documented. Add your code to your own repository. The repository must contain the sources, as well as a plain text file README.md that indicates the actual operational features and limitations.
2. 20% **Defense:** (10 min presentation, 5 min questions): You need to present your work in a formal way. You need to show:
   - the goal of the project
   - what did you do to achieve that goal (must include an analysis of the encountered difficulties and implemented solutions)
   - the results through a demonstration (most important part)

- no source code in the slides, only organizational charts are allowed if needed

3. 20% **Practice Analysis**: You will return a (4 to 5 pages) report, in one single PDF, before the defense. In this report you need to present your work in details:
    - implemented classes and used libraries
    - the functioning of your code
    - the encountered difficulties
    - the implemented solutions to overcome these obstacles

    The report must include a link to the repository must be in a **.zip file**
4. 30% **Continuous Progress Evaluation:** based on **git commits** in your own repository. You will need to send a preliminary version of the report halfway through the course.

PS: Plagiarism is prohibited and will be checked.

Useful links

- https://en.wikipedia.org/wiki/Natural_language_processing
- https://towardsdatascience.com/your-guide-to-natural-language-processing-nlp-48ea2511f6e1
- https://pytorch.org/tutorials/beginner/deep_learning_nlp_tutorial.html
- https://newsapi.org
- https://docs.djangoproject.com/en/3.0/intro/
- https://towardsdatascience.com/transfer-learning-in-nlp-fecc59f546e4

Contact

Please do not hesitate to contact:

- email: nermineali.lb@gmail.com
- Skype: Nermine Ali (skype meetings if needed outside lab hours preferably after 6 p.m)