Newsbot Midterm Reflection
By: Codie Munos
ITAI 2373

This assignment was definitely a game changer for me personally. I ended up doing it solo and there were definitely challenges along the way. I would get stuck on one module for a few hours at a time trying to troubleshoot. It led to me finding out cool shortcuts and different ways to approach the goal at hand. One being that working with API keys, you can use the secret tab on collab to pull it rather than doing a manual upload every time.

On another note, the most challenging module for me personally was module 5. The syntax parsing and semantic analysis itself was okay, but my visuals were coming out funky for the first 20 runs. I seeked assistant from AI to work through this one if I am being completely honest. I had spent a day on Module 5 and needed to move on quickly as I was falling behind already.

The integration of the different NLP tasks was tricky trying to keep the pipeline in tack without it losing sustainability. I had issues with my model after integration, but had to make minor fixes in previous code cells to initialize certain functions. Thankfully, this process did not take as long as prior ones.

Despite the struggles, I do feel as if this can help out professionals in the field regarding research and statistics. This model does not solely need to be based on BBC News, it can in fact be tweaked to analyze other datasets so long as the code cells are changed to match correct titles in the dataset. Although, this program may not be the most ethical in regards to it needing more training on categorizing articles with limited text offered. Depending on the sample size and text size, this bot will work great. I will say that it is not 100% perfect to replace a human, but definitely can guide a human in the right direction and filter a fair amount of null articles out.

Next, I would like to venture out of my comfort zone and attempt to work more with multilingual NLP. This is different from cross-lingual which is something new to my ears. Cross-lingual models can transfer the knowledge from language to language, but multilingual models can actually handle multiple languages on their own. I would like to learn how tokenization across languages works in models like mBERT and XLM-R. That will be experimental due to challenges such as pitfalls in tokenization, bias, cultural ignorance, and the evaluation across the languages.

Due to me completing the project solo, I took on the full load of the lab which only made me more dedicated. I would love to work with a team in the future as I like to be a team player, but I do not know how I will work through team problems whenever people get unprofessional. I personally feel like that will be my biggest struggle in this field, is to be an equal and not a leader. I have been managing teams and if I personally see the immaturity, I am quick to call it out and request for everyone to come together for the purpose of the tasks. In college it is different, we are all equals and if one person does not want to work together, it makes it hard for everyone else.

For the future employers, I would explain to them the purpose of this model, the challenges encountered, the solution to the problem, and display the working result of the model. I would answer questions regarding the model that they would have along with accept any feedback they give. This field is constantly moving forward and changing, therefore I must be okay with constructive criticism. No one ever got better by staying in one place, we must progress as a unit for the better future of our field.

Overall, this project allowed me to test my knowledge along with gaining some pride in my work. This is a useful lab for any beginners interested in model training.