Natural Language Processing with Disaster Tweets CS523

Aaron Daniels

Patrick Charles Kay

Reed Sarasua Tucker

The initial setup for the notebook was challenging. I couldn't get this to run in Kaggle notebook or in VS code on my desktop. I tried following the error messages and various versions of the keras, tensorflow, and keras\_nlp packages, but it wasn't helpful. I ended up installing WSL on my Windows 11 machine. This allowed me to run the notebook locally with VS Code.

The initial score for the NLPnotebook was .57. We found the best configuration to be changing the optimizer to Adafactor, the preset to "distil\_bert\_base\_multi", and BATCH\_SIZE = 60. We also removed the hyperlinks from the text. Our final score was .816 which is a vast improvement from .57.

If we had more time, we would have liked to explore other ways of cleaning the text. We also removed rows with null values, but it didn't help. The batch size was set to 60, but we could have tried this with other values. Overall, I think we were successful at improving the accuracy metric as it increased by 24 percent.

