A.I. Disclaimer: Work for this assignment was completed with the aid of artificial intelligence tools and comprehensive documentation of the names of, input provided to, and output obtained from, these tools is included as part of my assignment submission.

Most of the research was assisted by AI as that was my personal means of researching and iterating tests for this project. I used my personal knowledge from my experiences and this NLP class to guide my AI prompts and strategies to test on the neural network.

My favorite part about using AI for this project was visualizing my results quite easily, the method described later in this document.

A graph with blue lines

Description automatically generated

Before I was able to run the code, I had library install issues. Gensim does not work with the latest version of Numpy so I used Claude to figure out how to reinstall the correct version.

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

Then, the neural network was running on my CPU so I had to fix that by installing the CUDA library, which I was easily guided by Claude.

A screenshot of a computer

Description automatically generated

The next issue occurred when the labels were getting the wrong data type. The given code caused issues, but Claude identified a solution of coercing them to long types, which fixed the code on the second attempt.

A screenshot of a computer error

Description automatically generatedA screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

Once the code was working in its most basic form, I wanted to visualize it, so I asked Claude to see if there were any easy to use libraries, but I figured it would take more time than necessary for this assignment, and I ended up skipping these.

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated  
A screen shot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

I wanted to try something unique to attempt to get a higher accuracy. I personally figured that TF-IDF would be helpful for this. I didn’t specify TF-IDF initially because I wanted to see if Claude would either assume I meant TF-IDF or if it may have a different idea, so I treated the first few conversations here as brainstorming.   
  
A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generated

Here, Claude started to use frequencies so it was picking up lots of very common words, so I needed to adjust it by explicitly using TF-IDF.  
A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated  
  
A screenshot of a computer program

Description automatically generated

Then with the TF-IDF, I needed to adjust the model’s weights accordingly to the TF-IDF results:

A screenshot of a computer program

Description automatically generatedA screen shot of a computer program

Description automatically generated  
  
A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

Since proj4.ipynb uses a MulticlassRNN class, I made sure to update it to use a rtype=”tfidf” input.

A screenshot of a computer program

Description automatically generatedA computer screen shot of text

Description automatically generated  
  
A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

The alpha variable adjusted how much of the TF-IDF results would modify the final model weights.

Next, I prompted Claude to brainstorm for planning for more ideas to improve accuracy.

A screenshot of a computer

Description automatically generatedA black screen with white text

Description automatically generated

I was able to add a ReLU layer, which improved my accuracy by about 5-10% just by itself! I continued working with Claude to see if there was anything else I could add, which led to me adding Attention. There were other bits of code here that I didn’t add since they were a bit more advanced and abstract that what I was comfortable with submitting without understanding.

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

I then moved to a new chat with Claude to get a fresh start. I wanted to visualize my results in a dynamic graph, so that I can record the settings of each run, then just paste the text document into the Claude Artifact, and it worked very well!

A screenshot of a computer

Description automatically generated

A screenshot of a black and white screen

Description automatically generated A screenshot of a computer error message

Description automatically generated   
I added a text box to dynamically add and edit the results for each run, but Claude’s first iteration had an error.   
A screenshot of a phone

Description automatically generated

After fixing the error, I was able to just continuously use this interactive graph to visualize the progress of my experiments. The most important change was switching from RNN to GRU. The one time the score was 0% was when I set the learning rate to 5e-8.

The one time I got the score to 100% was when I cheesed the experiment by using 99% training and 0.5% testing, which means there was only a very small test size and it was overfitting.   
GRU 128 hidden dims, 2 n\_layers, 0.2 dropout\_rate, True bidirectional, 4 batch\_size, 99% trainSplitPercent, True ReLU: 100%

LSTM 128 hidden dims, 1 n\_layers, 0.2 dropout\_rate, True bidirectional, 4 batch\_size, 80% trainSplitPercent, 50% validationSplitPercent, True ReLU, average padding, 5e-8 rnn\_lr: 0.00%

A screenshot of a computer

Description automatically generated

I just wanted to get a list of more brainstormed ideas on improving the model, which Claude ended up writing the LSTM implementation. This just used the library version, which I was unaware of, and I was originally planning on trying to write it myself.  
  
A screenshot of a computer

Description automatically generatedA black screen with white text

Description automatically generated  
A screen shot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA computer screen shot of a program

Description automatically generatedA black screen with white text

Description automatically generated

I also used Claude directly in my IDE, Cursor.   
I was just trying some wild ideas like seeing if Huffman Encoding could potentially be used on the data for organizing it and encoding it so it would be more compressed. But alas, it was immediately shut down by Claude (lol).  
  
A computer screen shot of a program

Description automatically generated

A screenshot of a black and white text

Description automatically generated

I figured softmax was being used in the code but I couldn’t find it. Turns out, it is being used implicitly in the nn.CrossEntropyLoss function.

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

Similarly, forward() is defined but it is not explicitly called. Claude clarified that it is being implicitly called in the net() and rnn\_model() functions. A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

Earlier, I had received the Attention code snippets, but it wasn’t easily merge-able with my current code, so I tried to get the Cursor version of Claude to merge it, but it was unable to since I am working in a Jupyter Notebook. I was still able to work with it and do some error correction.

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA computer screen shot of a program code

Description automatically generatedA computer screen shot of a program code

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generatedA screenshot of a computer program

Description automatically generatedA screen shot of a computer

Description automatically generatedA screen shot of a computer program

Description automatically generatedA screen shot of a computer program

Description automatically generatedA screenshot of a computer error

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generatedA computer screen shot of a program code

Description automatically generatedA computer screen shot of a program code

Description automatically generatedA screen shot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generated

A black screen with white text

Description automatically generated

A screenshot of a computer error

Description automatically generated

A computer screen shot of a program code

Description automatically generatedA screen shot of a computer program

Description automatically generatedA screenshot of a computer program

Description automatically generated

Also when the LSTM mode was implemented, it encountered bugs, so I had Claude work on it to get it fixed quickly.

A screenshot of a computer error

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screen shot of a computer program

Description automatically generated

A screenshot of a computer program

Description automatically generatedA screenshot of a computer

Description automatically generated

Finally, the last part that I used AI for was modifying the padding. Instead of using max, I wanted to use average and pad up to the average or truncate down to the average. I just had Claude do it since I knew it was a relatively simple adjustment that it should be able to get on its first try, which it did.

A screenshot of a computer program

Description automatically generatedA screen shot of a computer program

Description automatically generatedA screenshot of a black screen

Description automatically generated