**Key Questions for the Project**

CDT Owen, CDT Persteins, CDT Hernon, and CDT Rennie

**1. How effective is the CNN at performing multi-label classification on Amazon reviews?**

The performance of our Neural Network was not very sufficient as when taking a look at our confusion matrices, our F1 score typically ranged from 38 to 45 percent roughly. This essentially means the model is almost guessing the reviews and cannot accurately predict scores. This is mostly seen with three- and four-star reviews being mixed up the most.

**2. What preprocessing steps improve CNN performance?**

Steps that help improve the performance of CNN is first removing the stop words, punctuation, and special characters from the data. Another preprocessing step is to try different labels and splits between test set and training set as results may vary based on these factors.

**3. Do different CNN kernel sizes and pooling methods enhance pattern recognition in**

**product categories?**

As we adjusted our kernel sizes in this model, the performance metrics decreased with an increasing factor of kernels. A kernel size of 3 had the highest f1 score of 0.44, which was visibly noticeable compared to the other factors (5 and 7). Kernel is known as the window that spans across the data to then shorten the matrix based on the highest numbers in the kernel size to pass into the next layer of the neural network. Enhancing pattern recognition in Amazon text reviews, our findings prove how it is important to keep a larger spanning size across the matrices.

**4. Which categories are challenging to classify, and why?**

The 3 and 4 score categories were challenging to classify for our model. This is because the score is extremely subjective as opposed to a 1 or a 5 score because the wording in the review is more emotional or easier to track using positive and negative language. Removing stop words and vectorizing normalized the data, but the language was very similar used in 3- and 4-star ratings making it hard for the model to accurately label the review text.

**5. How does the CNN compare to baseline models like Logistic Regression for multi-label**

**classification?**

Due to the fact that a CNN has a black-box machine learning component it is only helpful in comparing the output, making a logistic regression more interpretable. We did not account for a logistic regression for multi-label classification model, so it is difficult to compare direct metrics. However, a CNN is used for data that has a spatial component to it and has filters, kernels, and pooling; logistic regression uses tabular, vectorized data that relies on pre-processed features. The linear model consists of coefficients that can be helpful for multi-label but is not as effective as CNN when it comes to complex relationships or high-dimensional data.