Project Proposal 266 Amazon Review Sentiment Analysis
Section: Mike, Wednesday 6:30 PT
Team Members: Tejasvi Kalakota, Mohammad Hafezi, Patrick Yim

- **What do you plan to do?** We plan to run a sentiment analysis on a dataset of Amazon reviews to see what correlates to positive and negative reviews from 0 to 5 stars. We are trying to end up with the highest prediction of accuracy for classification of these reviews by running various algorithms on the dataset.
- **Why is it important, and why is it challenging?** This project is important because although sentiment analysis has been a topic of interest for many years, advances in the AI/ML space have allowed for more accurate predictions. As direct consumers of Amazon on a regular basis, we decided that our reliance on reviews is a standard direction for when purchasing items, which led to the interest in analyzing and predicting sentiment based on a dataset of such reviews.
- **What dataset(s) will you use?** We will be pulling the Amazon review dataset from this kaggle source: https://www.kaggle.com/datasets/bittlingmayer/amazonreviews/data
- **What algorithms might you use? Are good implementations available, or will you need to write your own? (Don't worry if you can't answer this well at this stage of the course.)** Review sentiment analysis is something that has been commonly visited throughout the years, with quite a few examples out there in the field currently. However, we are planning on taking 3 different approaches that utilize the course material that we have learned, from CNNs to RNNs, BERT and Word2Vec. It will also be very interesting to see how the accuracy changes based on the modernization of AI/ML and advances within the space.
- **References to *at least* four papers related to your proposal:**
  - https://www.sciencedirect.com/science/article/pii/S2949719124000074: This paper references the advances within the sentiment analysis space and will serve as contextual information
  - https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3886135: This paper, although somewhat dated from 2021, serves as a reference for potential similar algorithms and we will be using the various cross-entropy and accuracy scores to see general trends
  - https://www.mdpi.com/1999-4893/17/2/59: This paper serves as a recent advancement of sentiment analysis using BERT
  - https://www.kaggle.com/code/anshulrai/cudnnlstm-implementation-93-7-accuracy: This link will serve as a direct implementation of cudnnlstm and the underlying accuracy score will be a top score to try and improve upon

- **Hugging Face Transformers -** emotion detection
- Dataset: https://www.kaggle.com/datasets/bittlingmayer/amazonreviews/data
- Each of us find a research paper around hugging face/emotion detection use case and then come back
- Check with TAs and instructor Mike if our use case/algorithms are on the right path