Classification of Amazon Reviews using NLP

Business Understanding

- What problem are you trying to solve, or what question are you trying to answer?
    - Increasing transparency and consumer trust in online shopping through detection of product reviews written by bots. This allows customers to know that the reviews they see are genuine and to be more willing to buy products instead of avoiding the risk of being misled.
- What industry/realm/domain does this apply to?
    - E-commerce
- What is the motivation behind your project? (Saying you needed to do a capstone project for flatiron is not an appropriate motivation)
    - Many online storefronts like Amazon have rules for fake reviews, but given the scope of the site and the number of reviews posted daily, it is an almost impossible task for humans to solve. In 2020, an estimated 42% of Amazon reviews were thought to be fake. This can disillusion consumers if they cannot discern whether the rating for a product is real. This creates a less positive shopping experience and can hurt the sales of companies with legitimate reviews.

Data Understanding

- What data will you collect?
    - A large dataset that contains genuine reviews left on Amazon and reviews created by bots.
- Is there a plan for how to get the data (API request, direct download, etc.)?
    - Direct download from OSF, a website with open-source research data. Currently, I plan on using a csv file that was used in a research paper concerning the topic.
- What are the features you'll be using in your model?
    - Some features of interest will be the length of the reviews, TF-IDF, and the sentiment of the review. This may change as I work on my model, however.

Data Preparation

- What kind of preprocessing steps do you foresee (encoding, matrix transformations, etc.)?
    - Removal of blank or low-quality reviews, tokenization, lemmatization, and removal of punctuation and stop words.
- What are some of the cleaning/pre-processing challenges for this data?
    - Trying to determine what reviews would need to be removed due to being low quality. If looking for something like grammatical errors, how do I know if it is a human-mistake or a product of a bot. Things like emoticons (example: :/) can indicate that a human wrote the review, but these would be removed in the cleaning. Some of these processes may also take a long time to run given how large the set will be.

Modeling

- What modeling techniques are most appropriate for your problem?
    - Natural Language Processing is the best technique for analyzing reviews.
- What is your target variable? (remember - we require that you answer/solve a supervised problem for the capstone, thus you will need a target)
    - A variable that indicates whether the review is human-generated or computer-generated
- Is this a regression or classification problem?
    - This problem is classification since the goal is to determine whether the text is real (human written) or fake (bot-written)

Evaluation

- What metrics will you use to determine success (MAE, RMSE, Accuracy, Precision etc.)?
    - Accuracy and Precision will be the main metrics of success for this model as the goal is to spot potentially harmful reviews. The more accurate and precise the model, the greater its benefits to the user or company will be.

Tools/Methodologies

- What modeling algorithms are you planning to use (i.e., decision trees, random forests, etc.)?
    - I plan to use the Naïve Bayes Classification algorithm as it has uses in other similar problems such as spam filtering.