Problem Statement Worksheet (Hypothesis Formation)

How can Amazon use their product reviews to better understand their customers feedback. For example, categorizing reviews as good, bad, or neutral, will help organize customer feedback, so Amazon can focus on those problematic issues.

H

1 Context

Amazon is a well known company that sells many products, including instruments, online, and so has a lot of customer reviews. These reviews can be useful in helping understand what makes a product poor, but with so many reviews it can be hard to analyze them. Thus, this is where sentiment analysis comes in, where it can categorize these reviews as positive, negative, or neutral, making them organized for easier analysis.

2 Criteria for success

Achieving a high score in the evaluation metric used for the model that categorizes these reviews as positive, negative, or neutral, is crucial to success here. A good target would be about 80%, where the evaluation metric will likely be the F1 score.

3 Scope of solution space

The focus here will be on the written reviews for each product. These reviews are labeled with ratings, which can be converted into classes to then categorize the reviews in the appropriate way.

4 Constraints within solution space

- Data is heavily skewed to having most reviews being positive, which may make it hard for a model to predict negative reviews - perhaps the most important category to predict
- Some reviews have the instrument brand names, or other names, which can be random, unique, and uncommon with respect to the english language.

5 Stakeholders to provide key insight

The dataset comes from Kaggle, but it is a set of Amazon reviews, so the results should be presented to the Amazon stakeholders or product managers, specifically the ones that focus on the musical instrument products.

6 Key data sources

The entire dataset comes from Kaggle:
https://www.kaggle.com/datasets/eswarchandt/amazon-music-reviews?select=Musical_instruments_reviews.csv

H D E I P