

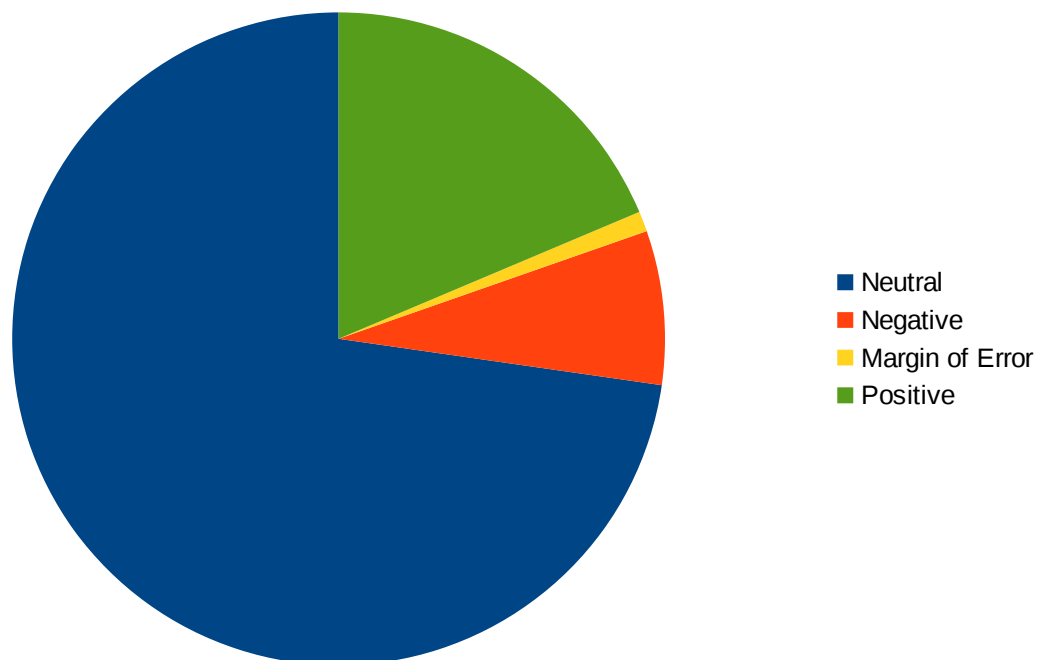
Zachary Shaffer
Python Sentiment Analysis
Work Overseen By: Janyl Jumadinova
Course: Artificial Intelligence
Allegheny College

The purpose of this implementation of a sentiment data analysis tool, in the context of Amazon reviews in the video game category, is to gain an overall sense of the products in the video game department of Amazon, and the mood of the user responses to said products.

I used the NLTK Python package for sentiment analysis. Not only did it seem to be the easiest to use, but when I found it's sentiment intensity measurement tool (Vader), it struck me as a brilliant plan to use NLTK rather than any other tool. Using python with text files and this tool was actually really easy, and very pleasant. The algorithm is a lexicon-based learning algorithm that uses data collected from social media websites, in conjunction with a lexicon of words and their respective values of positivity/negativity and intensity, to determine the positive/negative nature of given input and the intensity of that nature. I also found it really interesting that it has sections of the algorithm dedicated to handling negation words (don't, ain't, isn't, etc.), as well as utilizing punctuation and excessive user-capitalization to adjust the intensity.

The results of my data experiments (the terminal output, actually) are as follows:

Average compound value (intensity): 0.352263
Average positive percent value: 0.185148
Average neutral percent value: 0.722211
Average negative percent value: 0.075660



The sum of all of the averages I got from the data is 0.983019, which leaves a margin of error of 0.017 due to rounding. The intensity of the sentiments is independent of these values, and so it was not included in this chart. However, an average intensity of 0.352263 indicates that sentiments, both

positive and negative, are not very strong. While it is true that the compound value can come out as anything between -1 and 1, I accounted for this in my program to convert all negative intensity scores to positive, so the compound (intensity) value is not a true average, but rather an average of the strengths of the sentiments from the data collected. The overwhelmingly neutral nature of the reviews, as well as the low intensity score, leads me to believe that the majority of the reviews for this product section of Amazon are generally very technical and professionally written.

The sentiment analysis of this algorithm is related to machine learning in the sense that the algorithm learns from data collected from social media, so the sentiment analysis is very in tune with the sentiments and feelings of the “average joe,” which is another reason as to why I liked this algorithm.