Data Sentiment Analysis in Omni Channel models

**Retrieving Data**

The customer feedback data was retrieved from the Amazon website by the means of a web scraper. The web scraper used the library *requests\_html*. The scraper parsed the data and generated json files that were later converted into text files.

**Data**

Sustainable clothing products under the tag “Climate Pledge Friendly” were selected and “Most recent” customer reviews were collected. Data consisted of the *Title, Star rating* and *Body Text.*

**Analysis**

The analysis was done by writing a script in *python* programming language (version 3.10.8). The following libraries were used: *pandas, seaborn, matplotlib.pyplot, nltk, transformers.*

First the *Title, rating* and *body* were changed into a data frame consisting of rows and columns where each row consisted of one complete feedback, with columns being “title, rating and body” (Chart 1). Then the *nltk* library was used to apply *Sentiment Analyser* function on every entry in every row. The function analysed the words in the entries and gave a “positive”, “negative”, “neutral” and “compound” score to every entry. A particular score indicated the intensity of the that specific emotion in the data row based on the rating given and the words used in the title and body. For example, a higher positive score indicated that the emotion behind the data row was more positive. Furthermore, the compound score is the sum of all three scores. Higher compound score indicates the emotion behind the data row is more positive. All the score columns were added to the feedback data frame. (Chart 2)

Then the *seaborn* and *matplotlib* libraries were used to create graph 1 which shows the rating group (on x-axis) along with their average compound scores (on y-axis) (Graph 2). The higher star rating has higher average compound scores. The same libraries were used to create graphs 3, 4 and 5 which show the rating group (on x-axis) along with their positive, neutral, and negative ratings respectively (on y-axis).

Then the *transformers* library was used to apply emotion sentiment analysis the customer data. The emotion function was applied to the *body* column which gave out a specific emotion for every feedback row based on the words used in the body of the customer feedback. The function also provided with the emotion intensity rating which showed how strong the provided emotion was. Scores closer to 1 indicated stronger feelings of that particular emotion. The emotion and the scores were added to the feedback data frame (Chart 3). In the end a countplot was created to show how many feedbacks of different emotions are there (graph 6). This graph can be used to estimate the average customer emotion towards the product.