**Week 2 Assignment**

Prestha Khanal

Presidential Graduate School

TECH 405 Artificial Neural Network and Deep Learning

Professor Acharya

3rd October, 2024

**Introduction:**

This is followed by a brief description of the problem solved by the use of the particular dataset featured in the study. The dataset under consideration seems to be connected with reviews of products on Amazon which can give valuable information about customers’ satisfaction, performance of offered goods, and trends on the market. The first problem is that we need to classify these reviews to find common patterns that can be helpful in business with reference to different aspects of activity like product modification or advertisement approaches.

The problem can be framed as follows: That is how can we utilize customers’ feedback received on Amazon reviews to improve product lines and meet the customers’ needs and expectations? This entails knowledge of some aspects such as: ratings distribution, review content themes, and features positive ratings relationship.

**Finding**

The findings from the EDA are as follows:

1. Distribution of Ratings: The distribution of ratings likely shows how users rate products, whether it can give a broad idea of the general customer satisfaction or if those ratings are a mere distortion of actual perception of product quality.
2. Text Analysis of Reviews: Specific simple terms to customers in reviews are as follows: Personal findings of a customer are identified by common words in the reviews as negative or positive.
3. Sentiment Patterns: Popular ratios calculated on the scores or classifications got from the review text may indicate the proportion of positive and negative feedbacks and thus provide insight on the customers’ attitude.
4. Word Clouds: These word clouds make a linear review text more visually explicit and underline the most frequently used words which gives a reader a general idea of what is repeatedly mentioned.

**Approach**

The approach taken to tackle the challenges presented by the dataset involved several steps:

1. Data Cleaning: First pre-processing operations included as basic data cleaning operations, dealing with missing values, deleting duplicated observations and normalizing text data.
2. Exploratory Data Analysis (EDA): Primarily, descriptive statistics and some measures of central tendency were used in an attempt to enhance data comprehension. This included:
3. Histograms for rating distributions-In this paper, histograms for rating distributions are done in order to illustrate the percentile, mean, and mode of the ratings for different products.
4. Non-scientific bar charts for depicting frequently used words in reviews.
5. Matrices of correspondence for the purpose of uncovering how different numeric variables are hinged.
6. Sentiment Analysis: In the analysis of textual reviews, both features and class label, the feature set and the aspect of sentiment analysis were decided by conducting an application of Natural Language Processing (NLP). Some of the library to be used include NLTK or TextBlob for categorizing sentiments into positive, negative or even neutral ones.
7. Feature Engineering: New features were derived from the existent data for example on the basis of sentiment score of the review, the text review was categorized on its positive, negative or neutral flavor.
8. Statistical Testing: Subsequently, to confirm hypotheses about the effects of the features on ratings, t-tests or when appropriate, an ANOVA was conducted.

**Conclusion**

So, in the context of this work, EDA of the Amazon product review dataset offered insights about customers that are useful and necessary. The analysis shows that although consumers express a positive attitude towards most products, there are potential problem areas that are revealed by a sentiment analysis and studies of interdependencies between features. The findings derived from this paper can help business entities to improve their products, as well as tailor their promotional strategies, in order to capture consumers’ attention and secure their loyalty by pinpointing what features are valued most by the clients.

Moreover, the knowledgeability of reviewer behavior allows in developing appropriate marketing approaches that encompass individual segments of buyers. In summary, this report endeavours to emphasize the necessity of capitalising on the data parameterising customer feedback for the enhancement of business performance and infographic improvement of product quality and customer satisfaction.

**References**

Google Colab Link: <https://colab.research.google.com/drive/16tLGInh8UOQT4Z66jJyEU8Gx2RBNOV9i#scrollTo=eKBMQ4eODRpn>