SC1015 Project: An in-depth analysis of Amazon Review and Sales Data



Background

Companies tend to analyze their sales data to decide their sales direction. After all, this is the most accessible and commonly analyzed data. Meanwhile, online shopping websites contains a repository of reviews by shoppers. These reviews are used by other shoppers to decide whether a product is worth purchasing.

Problem Statement

We aim to put the underutilized data of user reviews to better use, by **building a recommendation system** based on **sentiment classification of reviews** to improve user experience. This can complement the existing sales data analysis, to improve customer satisfaction in online retail stores, as well as revenue.

Dataset

We have chosen a dataset from Kaggle which illustrates user reviews of Amazon products over a given period. The link to the data set is at <https://www.kaggle.com/datasets/karkavelrajaj/amazon-sales-dataset/data>. It is built from scraping product details from the official Amazon website. The data contains 1351 rows and is organized into 16 categories, which include key categories such as product\_name, category, discount percentage, rating and review content, which we analysed in the project.

A screenshot of a product list

Description automatically generated

Methods

We have followed the process of data analytics as taught in the SC1015 course, in the order of data extraction, data cleaning, exploratory data analysis, data visualization as well as sentiments analysis by means of text recognition. We conducted the analysis on the review scores to find out the popularity of products. We have included the use of Natural Language Toolkit (NLTK), a powerful and flexible library in performing sentiment analysis and natural language processing (NLP) tasks.

Rationale

As online