Determining Product Popularity from Online Ratings and Review Data

Background

Our company (a major electronics chain retailer) has had difficulty stocking the right amount of inventory, with it often being too much or too little. With the increase in Ecommerce purchases it has been hard to predict the demand of these purchases. Through our project we aim to provide a clearer picture on expected demand to allow the management to decide how much inventory to keep in stock.

Introduction

Our project seeks to investigate a large dataset of electronics product reviews collected from several online retail websites to determine the popularity of products and how much to keep in inventory.

We also want to predict the reasons for customers' non-recommendations on some electronic products in order to give feedback to our manufacturers, build better store warranties, educate our consumers on the products they are interested in purchasing and train our support team on helping disgruntled customers facing common product issues.

Dataset

The dataset is in CSV format and consists of 7000 reviews across 50 products from Amazon and Best Buy. It has fields for each product's brand and model name etc. and detailed review information including ratings, how many users found the review helpful, the review title and content, and the website where the review came from. In our pre-processing analysis we will use some Natural Language Processing techniques to sort the data and create extra fields to further analyse it. Thereafter we will apply some techniques learned in class such as regression to forecast expected demand of the products.