Project Proposal

Kanya Kreprasertkul, Yelin Shin and Zhe Ren

Machine Learning on Sales Performance from E-commerce Wish

Background

Understanding the relationship between product information (e.g. price, ratings and merchant quality) and their sales performance has always been vital for e-Commerce platforms as well as the merchants there, which can be used to find direction and drive revenue. We want to use machine learning methods to find correlations and patterns between the 43 features provided in this dataset (e.g. Are people sensitive to price drops? and Does product rating influence people's decision to purchase?) and identify the important components that trigger product success.

Dataset Introduction

This dataset contains summer-related products and their very detailed information at around 2020 August. There are 1590 records and 43 columns included, such as product ratings, sales performance, price change details and merchant information.

Dataset can be accessed here: Sales of summer clothes in E-commerce Wish

Possible Methodology

We want to determine the important factors for sales performance, so we are going to build several models and predictions. Then, we are going to compare their adjusted R-squared and test MSE to find the best model.

1. A parametric method, such as linear regression model, lasso model and PCR model will be performed, and the nonlinear terms and the interaction between each variable will be considered.
2. A non-parametric method, such as a regression tree model, random forests and boosting will be introduced.