

Women's E-Commerce Clothing Review Project – Version 1.

Introduction and Background of Project

It is the digital era, especially shopping online is the essential choice during pandemic at this moment. There are more than 100 years of history of shopping from home.[1] The research showed American woman is a vast growing group of online shoppers for clothing since the 1990s.[1] Clothing is one of the physiological needs, which are the most critical needs, such as air, food, and shelter based on Maslow's theory of needs.[2] Women can more easily impact each other's shopping preferences by sharing their attitudes toward clothing. [1] The women's e-commerce clothing review project will follow the data mining process to analyze the real commercial dataset for the predictive analysis with statistic modeling. [3]

The logistic regression techniques will apply to analyze the relationship between a categorical outcome binary variable and one or more categorical or continuous predictors.[4] The logistic regression is a sophisticated statistical tool that predicts the category of outcomes for individual cases with the most parsimonious models. [5] A logistic regression model will create with all predictor variables to predict the response variable for the project.

Data Resources

The dataset named Women's E-Commerce Clothing Reviews is extracted from Kaggle.com (<https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews>).[6] It is a real commercial data with data dictionary and excel dataset. It includes 23486 rows and 10 features variables.[6] Based on the data dictionary, recommended IND is the binary variable stating where the customer recommends the product where 1 is recommended, 0 is not recommended. [6] Thus, recommended IND is picked as an outcome dichotomous variable. The predictors also named independent variables are Age, Rating, Positive Feedback Count, Division Name, Department Name and Class Name.

Project Goals

The objectives of the project focus on the following areas:

- How the predictors impact the outcome variables Recommended IND
- What age most likely to offer a positive recommendation for women's online clothing shopping?
- Is a higher rating means definitely recommend the items to others online? If yes, what is the specific rating can be a threshold?
- How the Positive Feedback Count and affect the probability of Recommended IND?

References

- [1]. R. Goldsmith, L. Flynn, "Psychological and behavioral drivers of online clothing purchase," Emerald insight, March 1, 2004. Google Scholar. [Online] Available: <https://www.emerald.com/insight/content/doi/10.1108/13612020410518718/full/html>. [Accessed October 1, 2020]
- [2]. A. Duygun, E. Sen, "Evaluation of Consumer Purchasing Behaviors in the COVID-19 Pandemic Period in the Context of Maslow's Hierarchy of Needs,". ResearchGate, 2020. Google Scholar. [Online] Available: https://scholar.google.com/scholar?hl=en&as_sdt=0%2C50&q=Evaluation+of+Consumer+Purchasing+Behaviors+in+the+COVID-19+Pandemic+Period+in+the+Context+of+Maslow%E2%80%99s+Hierarchy+of+Needs&btnG=. [Accessed October 4, 2020]
- [3]. G. Shmueli, P. Bruce, I. Yahav, N. Patel, K. Lichtendahl, "Data mining for business analytics. concepts, techniques, and applications in R, ". John Wiley & Sons, Inc. 2018. Raynor Library, Marquette University. [Online] Available: <https://0-ebookcentral-proquest-com.libus.csd.mu.edu/lib/marquette/detail.action?docID=5014630>. [Accessed September 27, 2020]
- [4]. C. Peng, K. Lee & G. Ingersoll, "An Introduction to Logistic Regression Analysis and Reporting,". Indiana University-Bloomington. Google Scholar. [Online] Available: <http://sta559s11.pbworks.com/w/file/37766848/IntroLogisticRegressionPengEducResearch.pdf>. [Accessed October 5, 2020]
- [5]. S. Sarkar, "Importance of Assessing the Model Adequacy of Binary Logistic Regression,". Journal of Applied Sciences, June 2010. ResearchGate, Google Scholar. [Online] Available: https://www.researchgate.net/publication/44121586_Importance_of_Assessing_the_Model_Adequacy_of_Binary_Logistic_Regression. [Accessed October 2, 2020]
- [6]. Dataset, "Women's E-Commerce Clothing Reviews,". nicapotato, 2017 version 1, Kaggle.com. [Online] Available: <https://www.kaggle.com/nicapotato/womens-ecommerce-clothing-reviews/version/1>. [Accessed October 6, 2020]
- [7]. G. Condous, E. Okaro, A. Khalid, D. Timmerman, C. Lu, Y. Zhou, S. Van Huffel, T. Bourne, "The use of a new logistic regression model for predicting the outcome of pregnancies of unknown location,". Oxford Academic, August 2004, 2004. Google Scholar. [Online] Available: <https://academic.oup.com/humrep/article/19/8/1900/2356421>. [Accessed September 28, 2020]