

Customer Purchase Journey Prediction

Aditya Sumbaraju

Bellevue University

DSC680- Applied Data Science

Dr. Brett Werner

December 2, 2021

<https://github.com/adityasumbaraju/DSC680>

Topic

The case study "**Customer Purchase Journey Prediction**" belongs to E-Commerce Retail Domain. Given my experience in the Retail-Ecomm domain- I decided to go with this dataset. The availability of significant transactions data and computing systems has provided an excellent opportunity to model and predict customer behavior. Deep learning techniques can assist marketing decision-makers in delivering more reliable and practical marketing strategic plans. In this project, I am going to propose a customer purchase journey prediction model using recurrent neural networks (RNNs) based on the client loyalty number (CLN), recency, frequency, and monetary (RFM) variables.

Business Problem

Customer purchasing optimization is a common problem for many companies, including large and small scale businesses. It is always an advantage to an organization that contains data wealth, and this will help them predict the possible purchasing behavior of a given customer. And it also allows clients to make informed decisions on maintaining inventory and providing customer recommendations on items, customer promotions, and predicting the customer's interests in buying a similar product from its competitor. In this case study, I would analyze and demonstrate some of these metrics on a publicly available dataset and develop an ML model to predict the future long-term purchasing in the form of customer lifetime values.

Project1: Milestone1

Research Questions:

1. Who are my customers?
2. What are customers typically buying?
3. Can we identify a customer's purchase window?
4. Is it possible to reduce the advertising cost by organizations' marketing by campaigning "more likely to buy" products specific to customer groups?

My project aims to analyze the purchase behavior and create a prototype to retain an existing customer, predict the next purchase, and roll out the relevant campaigns to the specific customer.

Datasets

Data Source Urls:

- <https://www.kaggle.com/chiranjivdas09/ta-feng-grocery-dataset>
- <https://stackoverflow.com/questions/25014904/download-link-for-ta-feng-grocery-dataset>

About Data: The Ta Feng Grocery Dataset

Column definition: Transaction date and time, Customer ID, Age Group, PIN Code (Region), Product subclass, Product ID, Amount, Asset, Sales price

Fields of the Dataset are:

- Transaction date and time
- Customer ID
- Age: 10 possible values

Project1: Milestone1

- Residence Area: 8 possible values
- Product subclass
- Product ID
- Amount
- Asset
- Sales price

The Ta Feng Grocery Dataset is a supermarket dataset containing

- 817741 transactions from November 2000 until the end of Feb 2001.
- The data set contains information about 119578 shopping baskets that belong to 32266 unique users.
- In total, 1129939 items were purchased from available 23812 products.

Methods

As a first step, I will be predicting RMF values. Pass all the features that could help along with split values of customer_id and Recency (R), Monetary Value (M), Frequency (F). Finally, I will split the data into train and test sets to pass them to the model. I would be using a simple recurrent neural network with 250 hidden units with Relu activation using L1 regularization and loss of mean square error.

Ethical Considerations

The project reminds me of a rational maxim, "good ethics is good business"; Does it mean that unethical conduct is penalized in the Retail-Ecomm domain and ethical conduct rewarded? It is proved that, in many circumstances, firms may be immune to marketplace sanctions and more inclined towards customer retention. It is worth examining the conditions under which the organizations are called to task "the indefinite campaigns" or reward the existing consumers, i.e., as reflected in consumer attitudes toward the brand, the organization, and purchase intentions. This includes investigating the relationship between shopping experience and consumer purchasing behavior. Different algorithms may give other output for a given dataset, provided their limitation is associated with the individual machine learning models. This seems to be a potential problem in choosing a model for my use-case, and I would pick the best-suited algorithm based on its characteristics.

Challenges/Issues

- Data contains Chinese characters; I might face charmap issues.
- Anticipating whitespaces in data and need to work on data alignment.
- Incorrect variable types.
- Python package-related issues.
- There is no detailed data description; it doesn't contain currency details for the sales price and the units used for the variables "amounts" and "Assets."

Project1: Milestone1

- I am relying on the Kaggle dataset as this is usually clean, and I anticipate no missing or insufficient data that needs to be substituted with dummy data.
- Since the data is not current, the modeling and forecasting may not apply to the present-day scenarios.

References

- Burke, S. J., & Milberg, S. J. (1993, January 1). *>the role of ethical concerns in consumer purchase behavior: Understanding Alternative Processes: ACR*. ACR North American Advances. Retrieved December 5, 2021, from <https://www.acrwebsite.org/volumes/7422/volumes/v20/NA-20>
- Agnihotri, R., Dingus, R., Hu, M. Y., & Krush, M. T. (2015, September 15). *Social Media: Influencing customer satisfaction in B2B sales*. Industrial Marketing Management. Retrieved December 5, 2021, from <https://www.sciencedirect.com/science/article/abs/pii/S0019850115002631?via%3Dihub>
- Bradlow, E. T., Gangwar, M., Kopalle, P., & Voleti, S. (2017, March 20). *The role of Big Data and predictive analytics in retailing*. Journal of Retailing. Retrieved December 5, 2021, from <https://www.sciencedirect.com/science/article/abs/pii/S0022435916300835?via%3Dihub>
- H. Salehinejad and S. Rahnamayan, "Customer shopping pattern prediction: A recurrent neural network approach," 2016 IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1-6, doi: 10.1109/SSCI.2016.7849921.