**CS584-Machine Learning: Project Final Report**

**Product Recommendation Model**

Illinois Institute of Technology

Computer Science Illinois Institute of Technology

Professor: Yan Yan

**Xiaoting Zhou**

**A20503290** [**xzhou70@hawk.iit.edu**](mailto:xzhou70@hawk.iit.edu)

**Bo**

**A20524770** [**hwang218@hawk.iit.edu**](mailto:hwang218@hawk.iit.edu)

**Nevil**

**A20474215** [**psubramanian@hawk.iit.edu**](mailto:psubramanian@hawk.iit.edu)

# Abstract

The Product Recommendation Model (PRM) [1] is a powerful tool for identifying associations between frequently co-occurring entities and objects, such as items in a shopper's basket. When used effectively, PRM can provide valuable insights into consumer behavior and influence it. Major retailers frequently use PRM to uncover relationships between items by analyzing combinations of items that occur together frequently in transactions [2]. This information can be used to improve sales by optimizing product placement, shelf arrangements, and up-sell, cross-sell, and bundling opportunities. Association Rules are commonly used to analyze retail basket or transaction data and identify strong rules using measures of interestingness based on the concept of strong rules [3]. PRM plays a critical role in online shopping by providing suggestions to customers during the order process. This analysis is made possible by electronic point-of-sale systems, which allow store owners to collect and analyze large amounts of transaction data to understand customer purchasing behavior and patterns. In this paper, we aim to help Instacart utilize their transaction data to understand customer purchase patterns, identify frequently purchased items, and maintain adequate product stock. Additionally, we seek to identify clusters and subgroups of customers with similar purchase behavior and provide actionable recommendations for improving revenue and customer experience through segmentation and prediction models. Our analysis will enable Instacart to enhance the user experience by suggesting the next likely product to purchase during the order process. Moreover, we will outline a marketing strategy for Instacart and similar retailers that includes personalized communications to customers highlighting predicted products. Overall, this paper provides valuable insights into PRM and its potential applications for improving business operations in the retail industry.

**Key Words**: item, frequency, association, recommend, transaction, data, predict

Contents

[Abstract 2](#_Toc133241394)

[1. Overview 4](#_Toc133241395)

[2. Problem Statement 4](#_Toc133241396)

[3. Related Work (TODO) 5](#_Toc133241397)

[4. Dataset and Features 5](#_Toc133241398)

[4.1. Data Preparation 5](#_Toc133241399)

[4.2. Data Processing 5](#_Toc133241400)

[5. Implementation 5](#_Toc133241401)

[5.1. Training and Testing Data Split 5](#_Toc133241402)

[5.2. Training and Evaluating Models 5](#_Toc133241403)

[5.3. Model Performance Metrics 5](#_Toc133241404)

[5.4. Challenges 5](#_Toc133241405)

[5.5. Refinement 5](#_Toc133241406)

[6. Results 5](#_Toc133241407)

[6.1. Model Evaluation and Validation 5](#_Toc133241408)

[6.2. Justification 5](#_Toc133241409)

[7. Conclusion and Future Work 5](#_Toc133241410)

[References 5](#_Toc133241411)

[Website Links: 6](#_Toc133241412)

[Appendix 6](#_Toc133241413)

# **Overview**

The aim of this project is to improve sales and gain insights into consumer behavior by analyzing the relationship between purchases and product preferences. Specifically, we develop a system that can reorder products based on user preferences and previous order records. To achieve this, we establish metrics for each product to reflect their quality and likelihood of being reordered based on their own merits. We also take into account unique user preferences, such as Mr. A's uncommon meal choices. Additionally, we analyze ordering patterns, preferred times for ordering, and other relevant metrics to create a comprehensive understanding of consumer behavior. To recommend products, we utilize support, confidence, and lift measures to establish association rules between products and items. Overall, this project seeks to optimize the purchasing experience for customers while increasing sales for the company.

# **Problem Statement**

Instacart is an American company that provides online grocery delivery and pick-up services through a website and mobile app. In order to improve the company's sales, we aim to develop a model that predicts which items will be the best seller in the future target period. By forecasting merchandise sales trends, adjusting stocks in time, and suggesting popular products to consumers, this model can help increase sales. Our approach involves analyzing data to identify entities and items with the highest frequency, identifying association rules of items that frequently appear together in the target period, building a model to predict which items are most popular based on past transaction history, and recommending those items to customers in the future to increase store revenue. The integration of personalized recommendations into our system offers the potential for increased user engagement and subsequent purchases. As users return to our platform to engage with the interactive features and recommended products, we are afforded the opportunity to gather more data on both the user and product offerings. This data informs opportunities for product improvement, which in turn enhances user acquisition and retention. Thus, the implementation of personalized recommendations serves as a perpetuating cycle, as illustrated in Figure 1.

*Fig.1 Overview of Product Recommendation Model*

# **Related Work**

In the realm of recommender systems, the typical implementation involves a two-stage process: retrieval and ranking.

The retrieval stage is responsible for selecting a preliminary set of candidates from the vast pool of all possible options. The primary objective of this model is to efficiently eliminate candidates that are unlikely to interest the user. Given the potential enormity of the candidate pool, the retrieval model must be designed with computational efficiency in mind.

Subsequently, the ranking stage utilizes the results produced by the retrieval model and refines them to identify the best possible recommendations. The aim of this stage is to whittle down the list of potential items that the user might prefer to a small group of highly probable candidates.

Recent research has focused on improving the accuracy and efficiency of recommender systems. For example, Wang et al. (2021) proposed a graph-based recommender system that models user-item interactions as a graph and uses graph neural networks to make recommendations [4]. Zhang et al. (2020) developed a deep learning-based recommender system that uses attention mechanisms to capture the user's long-term and short-term preferences [5].

In comparison to traditional matrix factorization techniques, these approaches have shown promising results in terms of accuracy and efficiency. However, there is still a gap in the literature on how to handle cold-start and data sparsity problems in recommender systems.

# **Dataset and Features**

## Data Preparation

## Data Processing

# **Implementation**

## Training and Testing Data Split

## Training and Evaluating Models

## Model Performance Metrics

## Challenges

## Refinement

# **Results**

## Model Evaluation and Validation

## Justification

# **Conclusion and Future Work**

# **References**

[1] Chen, Yang, L., Rodríguez, R. M., Xiong, S., Chin, K., & Martínez, L. (2021). Power‐average‐operator‐based hybrid multiattribute online product recommendation model for consumer decision‐making. International Journal of Intelligent Systems, 36(6), 2572–2617. <https://doi.org/10.1002/int.22394>

[2] Lin, Chuang, W.-W., Chuang, C.-L., & Chang, W.-S. (2021). Applied Big Data Analysis to Build Customer Product Recommendation Model. Sustainability (Basel, Switzerland), 13(9), 4985–. https://doi.org/10.3390/su13094985

[3] Kim, Yang, G., Jung, H., Lee, S. H., & Ahn, J. J. (2019). An Intelligent Product Recommendation Model to Reflect the Recent Purchasing Patterns of Customers. Mobile Networks and Applications, 24(1), 163–170. https://doi.org/10.1007/s11036-017-0986-7

[4] Wang, C., Wang, Y., Zhou, K., Zheng, V. W., & Li, Y. (2021). Graph-based neural network for personalized item recommendation. IEEE Transactions on Neural Networks and Learning Systems, 32(4), 1465-1477.

[5] Zhang, W., Du, T., & Wang, J. (2020). Attentional sequential recommendation. In Proceedings of the 13th ACM Conference on Recommender Systems (pp. 328-336).

# **Website Links**

* [Market Basket Analysis - an overview | ScienceDirect Topics](https://www.sciencedirect.com/topics/computer-science/market-basket-analysis)
* [Association Rules in Data Mining | Learn the Algorithms, Types, and Uses (educba.com)](https://www.educba.com/association-rules-in-data-mining/)
* [What Is Data Analysis? Methods, Techniques, Types & How-To (datapine.com)](https://www.datapine.com/blog/data-analysis-methods-and-techniques/)
* [Beginner’s Guide to XGBoost for Classification Problems | Towards Data Science](https://towardsdatascience.com/beginners-guide-to-xgboost-for-classification-problems-50f75aac5390)

# **Appendix**

Tools

* **Software Packages:** VS Code, Python.
* **Development:** GitHub, Notion.
* **Project Planning:** Excel, Notion Kanban boards.
* **Libraries:** data.table, dplyr, ggplot2, knitr, stringr, DT, magrittr, grid, gridExtra, sqldf, Matrix, arules, tidyr, arulesViz, methods data.table, xgboost.
* **Documentation:** Microsoft Word, Notion, PPT, PDF.

Source Code

[GitHub | CS584\_Machine-Learning\_Project](https://github.com/May-Xiaoting-Zhou/CS584_Machine-Learning_Project)