



METHOD FOR TACKLING COLD START PROBLEM IN RETAIL RECOMMENDER SYSTEMS

CIND 820
Big Data Analytics Project

Prepared By:
Abdur Rahman Mahbub

Submitted On:
28th September, 2023

Table Of Content

Introduction	1
Problem Statement	1
Research Questions	1
Dataset Selection	2
Techniques	2
Tools	2
References	3

Introduction:

The recommender systems help customers discover products they prefer in modern e-commerce and retail platforms, making them an essential part of the digital economy. It is a significant machine learning model that learns from historical data and provides personalized user recommendations, increasing user engagement and resulting in higher satisfaction. However, due to constraints or lack of available data, these systems often face difficulties in delivering reliable results. When it comes to a highly dynamic platform, many users typically join and adding new items makes it difficult for the system to predict appropriate outcomes. Overall, system fails to develop a relationship between users and objects due to a lack of historical data. This significant issue is known as the "Cold Start Problem."

Problem Statement:

Optimization of retail product recommendations is a vital issue we seek to address here. To solve problems like handling new product introductions and improving the accuracy of recommendations, this project will aim towards developing advanced recommender systems which can give retail customers more personalized and up-to-date product suggestions.

Research Questions:

The following research questions have been identified in order to guide this project:

1. To identify user preferences and recommend new products based on product resemblances, what advanced recommendation techniques work best in retail?

2. How can we use novel strategies when the user interaction is minimal to none or new products are entering the market to enhance the accuracy and personalized nature of product recommendations?
3. How can we assess the performance of the advanced recommender systems comprehensively and ensure that the provided recommendations are accurate for a diverse customer base?

Dataset Selection:

For this project, we are using the UC Irvine Machine Learning Repository's Online Retail Data Set.

Public access to this data set is available and can be viewed or accessed using the following link:

<https://archive.ics.uci.edu/dataset/352/online+retail>

This dataset contains all transactional data from an online retail store, including additional information to identify each transaction separately. It is the most suitable dataset, as it demonstrates real scenarios that give an idea of how possible mitigation techniques could be adopted to address the cold start problem.

Techniques:

The following techniques and instruments will be used to respond to the research issues, as well as develop more sophisticated recommendation systems for retailers' products.

- **Accommodating data and Preprocessing:** Collect and preprocess transactional data from the predefined dataset, cleaning and transforming it into a format suitable for recommendation modelling.

- **Collaborative Filtering:** Implementing collaborative filtering methods based on items and individuals to classify customer preferences and generate product recommendations.
- **Matrix Factorization:** Exploring matrix factorization techniques to capture itemized interactions, latent factors.
- **Hybrid Recommender Systems:** Combining the capacity of both content-based and collaborative filtering to provide more accurate recommendations, especially for users with limited interaction history.
- **Deep Learning:** Experimenting recommendation models to leverage sequential user behaviours and improve personalization.
- **Evaluation Metrics:** Using common evaluation metrics to measure the effectiveness of the recommender system.

Tools:

In the areas of data preparation, model development and evaluation, we will use industry-standard tools and libraries to efficiently deal with cold start problems. For data analysis, modelling and evaluation, Python will be the main programming language. Data manipulation and machine learning, as well as collaborative filtering will be facilitated through libraries and frameworks such as pandas, Scikitlearn or Surprise. The efficient handling of large quantities of data will benefit from cloud computing resources.

References:

Online Retail. (2015). UCI Machine Learning Repository. <https://doi.org/10.24432/C5BW33>.