**AIE425 Intelligence Recommender System Fall semester 2024/2025**

**Assignment #2: Significance Weighting-based Neighborhood CF Filters**

**Student Name: Sondos Ahmed Ghoneim, Student ID: 221101426**

**WEEK 12**

## Table of Contents

1. Introduction  
2. Methodology  
3. Cosine Similarity Equation  
4. Data Preprocessing Steps  
5. Cosine Similarity Results Table  
6. Evaluation  
7. Conclusion

## Introduction

## The purpose of this assignment is to design, implement, and evaluate an intelligent recommender system. The project leverages a dataset containing product information, including prices, ratings, and reviews, to provide meaningful recommendations.

## Methodology

## 1.Data Cleaning and Preprocessing: -Removed non-numeric characters and handled missing data in 'Prices', 'Rating', and 'Number of Review' columns. - Converted cleaned data into numeric formats for further processing. - Selected the first 50 rows of the dataset for analysis. 2.Similarity Calculation: - Used cosine similarity as the metric to compute similarity between products based on numerical attributes. - Standardized numerical data for accurate similarity measurements. 3.Recommendation Algorithm: - Implemented a content-based recommendation system leveraging product similarities derived from the cosine similarity matrix.

## Results

## The cosine similarity matrix successfully calculates the pairwise similarity between products. The results can be used to identify the most similar products for each item in the dataset. For example, products with similar prices, ratings, and number of reviews are grouped effectively.

## Evaluation

## The recommendation system was evaluated qualitatively by analyzing the generated similarity matrix. Future iterations will include metrics such as precision, recall, and F1-score to assess recommendation accuracy.

## Cosine Similarity Equation

## The cosine similarity metric measures the cosine of the angle between two non-zero vectors in a multi-dimensional space. It is computed using the following formula: Cosine Similarity =

## Data Preprocessing Steps

## 1.Data Inspection:

## -Reviewed the dataset structure to identify numeric and non-numeric columns. - Identified issues such as non-numeric characters, missing values, and formatting inconsistencies. 2.Data Cleaning: - Removed the currency symbol '₹' and commas from the 'Prices' column. - Converted 'Rating' values like 'Not Available' into NaN and handled missing data. - Removed commas from the 'Number of Review' column for numeric conversion. 3.Data Transformation: - Converted cleaned columns ('Prices', 'Rating', and 'Number of Review') into numeric types. - Standardized the numerical data to normalize the scale of features before similarity computation.

## Steps Performed in the Assignment

## 1.Dataset Selection: - Used the provided dataset containing product details including prices, ratings, and reviews. - Selected the first 50 rows for analysis. 2.Data Preprocessing: - Performed cleaning and transformation as outlined in the preprocessing section. 3.Similarity Calculation: - Computed cosine similarity for the standardized data using numerical columns ('Prices', 'Rating', and 'Number of Review'). - Generated a similarity matrix to identify relationships between products. 4.Report Preparation: - Documented the methodology, results, and evaluation findings.

## Cosine Similarity Results Table

## Below is a preview of the cosine similarity matrix for the first five products in the dataset:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Product Name | Infinix Smart 7 HD (Green Apple, 64 GB) | Infinix SMART 7 (Azure Blue, 64 GB) | POCO C55 (Cool Blue, 64 GB) | POCO C50 (Country Green, 32 GB) | POCO C50 (Royal Blue, 32 GB) |
| Infinix Smart 7 HD (Green Apple, 64 GB) | 1.000 | 0.473 | 0.384 | 0.584 | 0.584 |
| Infinix SMART 7 (Azure Blue, 64 GB) | 0.473 | 1.000 | 0.549 | 0.720 | 0.720 |
| POCO C55 (Cool Blue, 64 GB) | 0.384 | 0.549 | 1.000 | 0.959 | 0.959 |
| POCO C50 (Country Green, 32 GB) | 0.584 | 0.720 | 0.959 | 1.000 | 1.000 |
| POCO C50 (Royal Blue, 32 GB) | 0.584 | 0.720 | 0.959 | 1.000 | 1.000 |

**Before Preprocessing**

|  |  |  |  |
| --- | --- | --- | --- |
| Product Name | Prices | Rating | Number of Review |
| OPPO Reno10 5G (Silvery Grey, 256 GB) | 32999 | Not Available | 0 |
| Infinix Smart 7 HD (Green Apple, 64 GB) | 6199 | 4.3 | 630 |
| Infinix SMART 7 (Azure Blue, 64 GB) | 7299 | 4.3 | 2400 |
| POCO C55 (Cool Blue, 64 GB) | 8499 | 4.2 | 2027 |
| POCO C50 (Country Green, 32 GB) | 5499 | 4.2 | 2039 |

## 

## Executive Summary

This report presents the design and implementation of an intelligent recommender system using cosine similarity to analyze and recommend products. It highlights the methodology used for preprocessing data, calculating similarities, and generating a functional recommendation system. The approach ensures a scalable and accurate recommendation framework tailored to the provided dataset.

## System Design

The system was developed to operate in three stages:

1. Data Preprocessing: Ensuring clean and structured input data.
2. Similarity Computation: Applying cosine similarity to identify relationships between products.
3. Recommendations Generation: Using similarity scores to suggest relevant products.

## Recommendations for Future Enhancements

To further enhance the recommender system, the following improvements can be made:

* Incorporating additional features such as product categories and user preferences\
* Implementing hybrid recommendation techniques combining collaborative and content-based filtering.
* Using advanced machine learning models like matrix factorization and neural networks.
* Adding evaluation metrics like precision, recall, and F1-score for quantitative assessment.
* Creating a user-friendly interface for real-time product recommendations.

## 

## Appendix

The appendix includes detailed code snippets and implementation notes for the cosine similarity calculation and preprocessing steps.

This section includes supplementary material used for the analysis:

* Full dataset preview (if applicable).
* Additional references and documentation.
* Sample Python code snippets used for the implementation.

**Results:**

## Before Preprocessing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Product Name | Product URL | Prices | Rating | Number of Review |
| OPPO Reno10 5G (Silvery Grey, 256 GB) | https://www.flipkart.com/.../reno10-5g | 32999.0 | nan | 0 |
| Infinix Smart 7 HD (Green Apple, 64 GB) | https://www.flipkart.com/.../smart-7-hd | 6199.0 | 4.3 | 630 |
| Infinix SMART 7 (Azure Blue, 64 GB) | https://www.flipkart.com/.../smart-7 | 7299.0 | 4.3 | 2400 |
| POCO C55 (Cool Blue, 64 GB) | https://www.flipkart.com/.../poco-c55 | 8499.0 | 4.2 | 2027 |
| POCO C50 (Country Green, 32 GB) | https://www.flipkart.com/.../poco-c50 | 5499.0 | 4.2 | 2039 |

### After Preprocessing

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Product Name** | **Product URL** | **Prices** | **Rating** | **Number of Review** |
| Infinix Smart 7 HD (Green Apple, 64 GB) | https://www.flipkart.com/.../smart-7-hd | 6199.0 | 4.3 | 630 |
| Infinix SMART 7 (Azure Blue, 64 GB) | https://www.flipkart.com/.../smart-7 | 7299.0 | 4.3 | 2400 |
| POCO C55 (Cool Blue, 64 GB) | https://www.flipkart.com/.../poco-c55 | 8499.0 | 4.2 | 2027 |
| POCO C50 (Country Green, 32 GB) | https://www.flipkart.com/.../poco-c50 | 5499.0 | 4.2 | 2039 |
| POCO C50 (Royal Blue, 32 GB) | https://www.flipkart.com/.../poco-c50-royal | 5499.0 | 4.2 | 2039 |

## Cosine Similarity Results

Below is the cosine similarity matrix for the first five products:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Product Name** | **Infinix Smart 7 HD** | **Infinix SMART 7** | **POCO C55** | **POCO C50 (Green)** | **POCO C50 (Blue)** |
| Infinix Smart 7 HD | 1.000 | 0.473 | 0.384 | 0.584 | 0.584 |
| Infinix SMART 7 | 0.473 | 1.000 | 0.549 | 0.720 | 0.720 |
| POCO C55 | 0.384 | 0.549 | 1.000 | 0.959 | 0.959 |
| POCO C50 (Green) | 0.584 | 0.720 | 0.959 | 1.000 | 1.000 |
| POCO C50 (Blue) | 0.584 | 0.720 | 0.959 | 1.000 | 1.000 |

## Top 5 Product Recommendations

Example product: Infinix Smart 7 HD (Green Apple, 64 GB)

* Infinix HOT 30i (Glacier Blue, 64 GB) (Similarity Score: 0.960)
* Infinix HOT 30i (Diamond White, 64 GB) (Similarity Score: 0.960)
* Infinix HOT 30i (Mirror Black, 64 GB) (Similarity Score: 0.960)
* Infinix HOT 30i (Marigold, 64 GB) (Similarity Score: 0.960)
* REDMI A1+ (Black, 32 GB) (Similarity Score: 0.887)

## Singular Value Decomposition (SVD)

The SVD components for the first five products are as follows:

|  |  |  |
| --- | --- | --- |
| **Product Name** | **SVD Component 1** | **SVD Component 2** |
| Infinix Smart 7 HD | 2.860 | 1.978 |
| Infinix SMART 7 | 0.898 | -1.275 |
| POCO C55 | 3.972 | -1.813 |
| POCO C50 (Green) | 3.798 | -1.287 |
| POCO C50 (Blue) | 3.798 | -1.287 |

## Singular Value Decomposition (SVD)

SVD reduces the dimensionality of the similarity matrix into two components, helping to visualize relationships between products.

The SVD results for the first 5 products are visualized below:

A screenshot of a computer

Description automatically generated

## SVD Results for All Products

A diagram of a product

Description automatically generated

## Conclusion

## This assignment demonstrated a structured approach to designing and implementing an intelligent recommender system. From data preprocessing to similarity computation, the project highlighted the importance of cleaning and preparing data to ensure meaningful outcomes. The cosine similarity-based system effectively identifies relationships between products and sets the stage for further enhancements, including integration with advanced recommendation algorithms and evaluation metrics. This assignment serves as a foundation for exploring more sophisticated recommendation systems.