Literature Review:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Paper Title | Year | Methods | Results and contribution | Weakness |
| RFM-Based Customer Analysis  and Product Recommendation System | 2021 | K-means for clustering.  Used Cognitive Similarity (Content + Collaborative) based approach to increase the efficiency and tackle the cold start. | 66.8% precision, 5% improvements | Didn’t make a comparison with existing hybrid recommendations. Used only explicit feedback. |
| A Filter is Better Than None: Improving Deep Learning-Based Product Recommendation Models by Using a User Preference Filter | 2021 | Deep learning approach (skipGram) for content based recommendation with RFM filtering | Both offline and online cases, a interactive hit rate (hr) and click through rate (ctr) is used. For offline the hitrate was greater without the filter but for online with filter the CTR performs 37% higher. | Couldn’t tackle the cold start problem,  Recussring purchase behavior have not been considered which may lead more accurate recommendation in the future. |
| Hybrid approaches to product recommendation based on customer lifetime value and purchase preferences | 2016 | K-means for clustering.  Proposed two hybrid model, Hybrid1 combines the customer lifetime value (RFM) with collaborative filtering and Hybrid2 uses the customer lifetime value with collaborative and association rule mining. | Compared the Hybrid1 and Hybrid2 with the WRFM, CF, KNN based methods. Hybrid1 and Hybrid2 both outperforms the other 3 and scored 0.528 and 0.533 f1-score for 20 clusters | Using of the collaborating filtering may cause sparsity problem for less transactional data,  Didn’t make comparison with exiting hybrid approach,  Only uses the binary choice of customer either “purchased” or not which is not so useful. |
| Recommender system based on customer segmentation (RSCS) | 2016 | AHP and EM for clustering,  Used Multiclass SVM and KNN for the recommendation | RFM segmentation with Collaborative filtering based recommendation improved the efficiency and accurteness. Improves the f1-score by 0.171 on average. | Didn’t consider comparing the existing hybrid method.  Couldn’t tackle the cold-start problem.  Used only explicit feedback. |
| RFM Ranking – An effective approach to Customer Segmentation | 2021 | Proposed a improved K -means algorithm |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Advantages of Collaborative Filtering:

* Implementation of RS using Memory-Based Collaborative filtering technique is easy.
* Additional of new data in an incremental manner is easy in Memory-Based Collaborative filtering.
* Improved prediction performance by a Model Based Collaborative filtering technique.

Disadvantages of Collaborative Filtering:

* Cold Start Problem: In the case of the new user the system either doesn’t know what to recommend or has very poor performance.
* Scalability: The CF technique generates recommendations over billions of users and products, which require a significantly huge amount of computational power.
* Sparsity: Only small subsets of the items are rated by the users from the available dataset of items. Hence very few ratings are available to generate a recommendation which leads in poor performance.

Advantages of Content-based filtering:-

* Content-based recommender system provides user independence through exclusive ratings which are used by the user to build their profile.
* Provides transparency to the active user by providing an explanation of working of CB filter.
* CB filter is good to recommend items that are not yet rated or viewed by any user. This will be advantageous for a new user.

Limitations of Content-based filtering :-

* In CB Filters, it is very difficult to generate characteristics of an item.
* CB Filters suffer from an over-specialization problem because it advocates the same types of items.
* It is more difficult to get feedback from users in CBF because users normally do not sort the items (as in CF) and therefore, it is not possible to determine if the recommendation is correct.

Research Methodology:

Transaction Dataset

RFM Feature Extraction

Build Cluster Algorithms

K-Means

Fuzzy C Means

Graph Based

Final Clusters

Clustering Model Evaluation and Comparison

Merged Customer, Product and Transaction Data

Merge Cluster Attribute

Final Cluster Labelling

Data Preprocessing and Selection

Feature Extraction

Interaction Matrix

User Feature Matrix

Item Feature Matrix

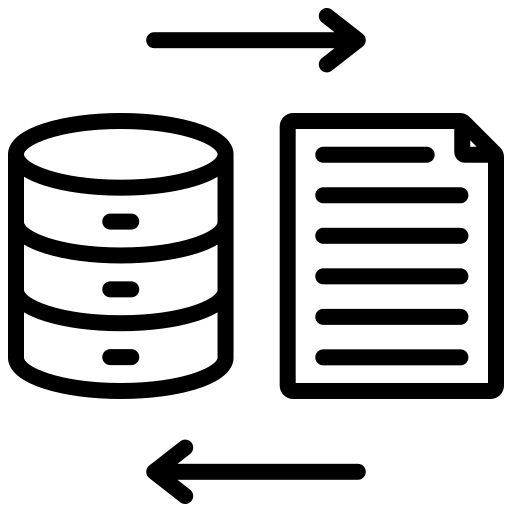
Sparse Matrix Transformation

Configure LightFM Model

Performance Evaluation

Existing Method Comparison

\



RFM Feature Extraction

Transaction Dataset

K-Means

Build Cluster Algorithms

Fuzzy C Means

Graph Based

Final Clusters

Clustering Model Evaluation and Comparison

Final Cluster Labelling

Merged Customer, Product and Transaction Data

Interaction Matrix

Merge Cluster Attribute

Data Preprocessing and Selection

Feature Extraction

User Feature Matrix

Item Feature Matrix

Sparse Matrix Transformation

Performance Evaluation

Existing Method Comparison

Configure LightFM Model

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Transaction ID | Customer ID | Tran\_date | Prod\_subcat\_code | Prod\_cat\_code | Qty | Rate | Tax |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Transaction ID | Customer ID | Tran\_date | Prod\_subcat\_code | Prod\_cat\_code | Qty | Rate | Tax | Total amt | Store Type |
| Unique ID | Unique ID | Date of trasaction | Product subcategory | Prod category | quantity | Unit price | Tax | Total amount | Store Type |
| N/A | N/A | N/A | 12 sub categories | 6 category | Min = -5,  Max = 5 | Min = -1499, max = 1500 | Min = 7.35, max =787.5 | Min = -8270, max = 8287 | Eshop = 9311,  MBR = 4661,  Flagship Store=4577,  Teleshop = 4504 |

|  |  |  |  |
| --- | --- | --- | --- |
| Customer ID | DOB | Gender | City Code |
| Unique ID | Date of Birth of Customer | Gender of Customer | Unique City Code of Customer |
| N/A | Mostly Unique | Male = 2892,  Woman = 2753 | 10 different cities dataset |

|  |  |  |  |
| --- | --- | --- | --- |
| Product Cat Code | Product Cat | Product Subcat Code | Product Subcat |
| Code for each product category | Product category | Code for each product subcategory | Product subcategory |
| 6 category | 6 category | 12 subcategory | 12 subcategory |