DSE 2141– Data Analytics Lab

Lab 10 – Date: 20th October 2023

EXERCISE 1: Association Rule Mining

Use the "groceries.csv" dataset and answer the following:

1. How many transactions and items are there in the data set?
2. Prepare the data for finding association rules. Each transaction will contain a list of item in the transaction.
   [['citrus fruit', 'semi-finished bread', 'margarine', 'ready soups'],
   ['tropical fruit', 'yogurt', 'coffee'],…….
   ['whole milk']]
3. Use Python library *mlxtend* and convert the transactions into a format that can be used in the Apriori method for finding frequent itemsets.
   pip install mlxtend
   from mlxtend.preprocessing import TransactionEncoder
   from mlxtend.frequent_patterns import apriori, association_rules
4. Find top selling items with minimum support of 2%.
5. Find all frequent itemsets with minimum support of 5%.
6. Find all frequent itemsets of length 2 with minimum support of 2%.
7. Find the top 10 association rules with minimum support of 2%, sorted by confidence in descending order.
8. Find association rules with minimum support of 2% and lift of more than 1.0.

EXERCISE 1: Collaborative Filtering

1. Read about the movielens dataset and write down a summary of metadata.

User-Based Similarity

2. Read the "ratings.csv" file and create a pivot table with index='userId', columns='movieId', values = "rating.
3. sklearn.metrics.pairwise_distances can be used to compute distance between all pairs of users. pairwise_distances() takes a metric parameter for what distance measure to use. Use cosine similarity for finding similarity among users. Use the following packages.

```
4. from sklearn.metrics import pairwise_distances
5. from scipy.spatial.distance import cosine, correlation
```

6. Find the 5 most similar user for user with user Id 10.
7. Use the "movies" dataset to find out the names of movies, user 2 and user 338 have watched in common and how they have rated each one of them.
8. Use the movies dataset to find out the common movie names between user 2 and user 338 with least rating of 4.0

Item-Based Similarity

9. Create a pivot table for representing the similarity among movies using correlation.
10. Find the top 5 movies which are similar to the movie “Godfather”.