Plagiarism Scan Report

8%
Plagiarism

2%
Exact Match

6%
Partial Match

92%
Unique

Words	861
Characters	5879
Sentences	57
Paragraphs	116
Read Time	5 minute(s)
Speak Time	6 minute(s)

Content Checked For Plagiarism

Methodology
The algorithms used Collaborative Filtering recommends companies based on either their preferences or similarities. Matrix Factorization helps predict the rating by factorizations of user-item interactions into latent space. Content-Based Recommender Systems and are employed to cluster companies and recommend those with similar features, such as ESG score or market cap. These methods cooperatively provide recommendations involving company suggestiveness and personalization.
1. Collaborative Filtering
Collaborative Filtering is the technique which helps in the recommendation of items (or companies) in my case based on the user preferences or similitude between the items.
User-Based Collaborative Filtering:
• Goal: Recommend companies to the user based on similar users' preferences.
Steps:
• Identify Users: If your dataset contains user interactions (e.g., user ratings or preferences for companies), then you can use that for input. Otherwise, consider simulating the user preferences based on dividend yield, market cap, or ESG scores.
• Find Similarity: Use similarity measures like cosine similarity or Pearson correlation to find similar users.
• Predict Ratings: Predict how a user would rate a company that he/she has not interacted with, on the basis of other similar users' ratings.
• Recommend: Recommend companies with the highest predicted ratings.

Item-Based Collaborative Filtering:
• Show companies similar to the ones that a user already has shown interest in.
Steps:
Page 12 of 48

• Identify Items (Companies): Represent each company by using some features like industry, market cap, dividend yield, or ESG scores.
• Calculate Similarity: Compute the similarity between companies using cosine similarity or the Pearson correlation.
• Recommend: For a particular company, recommend other companies that are most similar.
Example:
• If a user has interest in companies with high dividend yield and strong ESG

scores, you can recommend companies with similar traits.

2. Matrix Factorization
Matrix Factorization is a model-based approach in which the user-item interaction matrix is decomposed into latent factors.
Steps:
• Create User-Item Matrix: Construct a matrix where rows represent users, columns represent companies, and values represent user ratings or interactions.
• Decompose Matrix: Use Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), or similar techniques to decompose the matrix into latent factors.
• Predict Missing Values: Use these latent factors to predict ratings of companies for users they haven't previously interacted with.
• Recommend: Recommend companies with the highest predicted ratings.
Example:
• "With matrix factorization, it is possible to find latent factors (for example, high ESG scores, low betas) to recommend companies for a user who has interacted with companies of the healthcare sector."

Page 13 of 48

3.Content-Based Recommender Systems
Content-based recommender systems propose similar items based on the attributes of these items (in this case, the companies).
Steps:
• Feature Extraction: Construct a model using attributes such as industry, marketCap, dividendYield, ESG scores, and profitMargins for each company.
• Text Based Features: If it is text data (e.g., longBusinessSummary), the text can also be represented using TF-IDF or word embeddings.
• Find Similarity: Identify similar companies based on cosine similarity or other distance metrics.
• Recommend: Based on close associations to chosen companies, recommend other companies based on features similarity.
Example:
• You could recommend companies of a similar ESG profile where the user is looking for high environmentScore and low pesticide use.

4. Clustering and Neighborhood-Based Methods
In clustering methods, the idea is to first cluster similar companies or users and then recommend items within those clusters.
Steps:
• Cluster Companies: Using a clustering algorithm such as K-Means to cluster companies based on some features (marketCap, dividendYield, ESG scores, etc.).
• Cluster Users: If users' data are available, cluster users according to their preferences or actions.
• Using elbow method to find the optimal number of clusters
• Recommend: For a user lying within some clusters, the recommended companies would be those that are highly ranked within that cluster.
Example:
Page 14 of 48

• If the companies are clustered into groups such as "High ESG, Low Risk" or

"High Dividend, Low Growth," you can recommend companies within the same cluster.

5.Matrix Factorization with PCA

Matrix Factorization is a kind of dimensionality reduction that decomposes a user-item interaction matrix into latent factors. This latent factor represents the relationships within the data and enables prediction of unknown values (for, indeed, recommendations). Matrix factorization, in this implementation, involves PCA with two approaches:

• Mean Filling: Missing values are filled with the mean value of individual columns.

• Maximum Likelihood Estimation (MLE): Fullfills the unknown entries interactively with the help of statistical estimation.

Steps

Data Preparation:

• Load the dataset containing user-item interactions (e.g., user ratings for companies).

• Simulate a sparse user-item matrix with missing values that could be found in real life.

Handling Missing Data:

• Mean Filling: Replace missing values with the column mean.

• MLE (Iterative Imputer): An iterative method of estimating missing data based on the observed data.

PCA for Matrix Factorization:

• Apply PCA for the user-item matrix for dimensionality reduction.

• Using the principal components, reconstruct the matrix to predict the missing ratings.

Page 15 of 48

Recommendation:

• For a given user, predict ratings for unrated items using the reconstructed matrix.

• Recommend top-N items with the highest predicted ratings.

Matched Source

Similarity 4%

Title:Using elbow method to find the optimal number of clusters: In cluster ...

Using elbow method to find the optimal number of clusters: In cluster analysis, the elbow method is a heuristic used in determining the number of clusters in a data set. The method consists of plotting the explained variation as a function of the number of clusters, and picking the elbow of the curve as the number of cluster...

https://github.com/shova29/Regression_Seed_dataset_Clustering_Elbow

Similarity 4%

Title:Matrix Factorization: Enhance Data Analysis & ...

Nov 17, 2024 — Matrix Factorization is a technique used in data analysis that decomposes a matrix into a product of two or more smaller matrices.

https://www.lyzr.ai/glossaries/matrix-factorization

Similarity 3%

Title:Building a Recommendation System

User-Item Matrix Creation: Construct a user-item interaction matrix, where rows represent users, columns represent items, and the values ...

https://medium.com/muthoni-wanyoike/building-a-recommendation-system-11b13629a96

Similarity 2%

Title:Can someone please explain what to do next after getting ...

Think about it, you can take the first k (k < d) principal components (n x k matrix) and then reconstruct the image (multiply with k x d) to get ...

https://www.reddit.com/r/datascience/comments/yex3pq/can_someone_please_explain_what_to_do_next_afte

Similarity 2%

Title:ijcset.net · docs · VolumesRanking Based Techniques Using Optimised Recommendation Diversity

Recommender systems typically try to predict the ratings of strange items for every user, frequently using other users' ratings, and recommend top N items with the highest predicted ratings values. Accordingly, there have been numerous studies on developing novel algorithms that can advance the predictive accurateness...

https://ijcset.net/docs/Volumes/volume6issue11/ijcset2016061104.pdf