

In Class Activity

Recommender Systems – k-NN

A Recommender System is software that analyzes data about user preferences or behaviors and then makes personalized recommendations.

You have been exposed to recommender systems on streaming media platforms (Hulu, Netflix, YouTube, Disney+, Spotify), online shopping (Amazon, Walmart, Rakuten, Temu) and social media (Instagram, Facebook, LinkedIn).

A Recommender System is ... a *system*, not a Machine Learning model. A variety of Machine Learning models can be used to implement Recommender Systems. (And if you go way, way back to the beginning of your earliest understanding of technology you might recall that a system can be loosely described as the collective interaction between people, processes, and things.)

In today's Roadmap activity we are going to see the k-Nearest Neighbors (k-NN) model used to illustrate how an imaginary Recommender System.

Some important definitions:

Recommender System -- Software that analyzes data about user preferences or behaviors and then makes personalized recommendations using an algorithm.

Collaborative filtering – Collaboration filtering is a type of recommendation algorithm that makes predictions about user preferences based on the preferences of similar users or items. It relies on the idea that users who have expressed preferences in the past are likely to have similar preferences in the future.

There are two main types of **collaborative filtering**:

1) **User-based** collaborative filtering: This approach recommends items to a user based on the preferences of similar users. It identifies users who have rated items similarly to the target user and recommends items that similar users have liked. **(KEY IDEA: Find other users “like me” and show me what they liked best.)**

2) **Item-based** collaborative filtering: This approach recommends items to a user based on the similarity between items. It identifies items that are like those that the user has rated highly and recommends similar items. **(KEY IDEA: Find stuff that is similar to the other stuff I have liked before.)**

k-Nearest Neighbors (k-NN) – This is *one of many* possible M.L. models that can be used to generate recommendations.

We can get a quick overview of the k-NN algorithm here:

https://youtu.be/zeFt_JCA3b4?si=rvNuAiaLRLtu5Xj

The algorithm uses something called “cosine similarity” to determine the similarity between users (or items). (*Oh no, math again.*) The cosine similarity metric disregards the magnitude of the vectors and only considers their orientations, making it useful for comparing documents, user preferences, or any other high-dimensional data where the magnitude may not be relevant.

Here’s where we can get a quick overview of cosine similarity:

<https://youtu.be/e9U0QAFbLI?si=ZFUXQdVbSzHG1S3F>

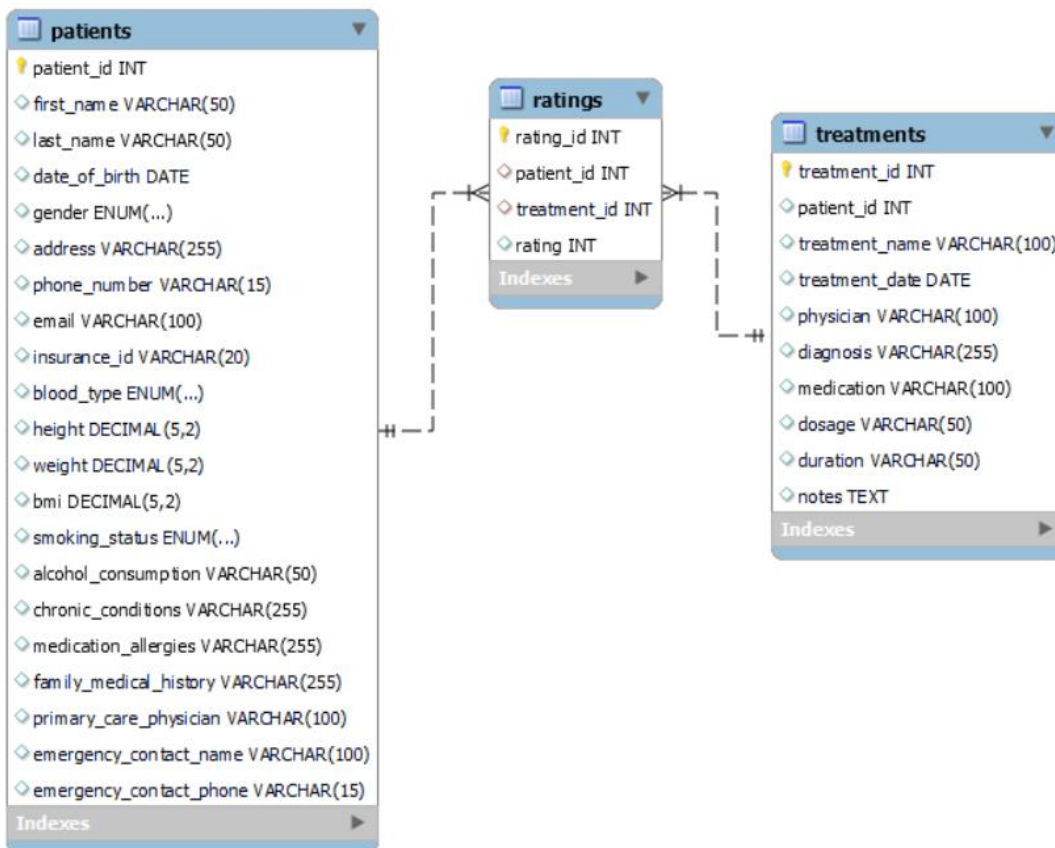
Recommender Systems are useful for more than streaming services, shopping, and social media. Here are some other interesting applications:

1. **News Aggregation:** News websites and apps use recommender systems to personalize news articles and headlines for users based on their interests, reading history, and behavior.
2. **Job Portals:** Job portals like LinkedIn and Indeed use recommender systems to suggest job listings to users based on their skills, experience, job history, and preferences.
3. **Travel and Hospitality:** Travel websites and booking platforms use recommender systems to suggest hotels, flights, vacation packages, and activities to users based on their travel history, preferences, and budget.
4. **Healthcare:** Recommender systems are increasingly being used in healthcare to suggest personalized treatment plans, medications, and healthcare providers to patients based on their medical history, symptoms, and preferences.
5. **Education:** E-learning platforms and educational websites use recommender systems to recommend courses, tutorials, and learning resources to students based on their learning objectives, progress, and preferences.
6. **Restaurant and Food Delivery:** Restaurant review websites and food delivery apps use recommender systems to suggest restaurants, dishes, and cuisines to users based on their taste preferences, dietary restrictions, and location.
7. **Financial Services:** Banks and financial institutions use recommender systems to suggest financial products, such as credit cards, loans, and investment options, to customers based on their financial history, goals, and risk appetite.

A side note for students of MIS2502 - Association Mining (a.k.a Market Basket Analysis) is *not* the same as **Collaborative Filtering**. While both are techniques used in recommendation systems and have similarities, they serve different purposes and operate on different types of data.

Instructions

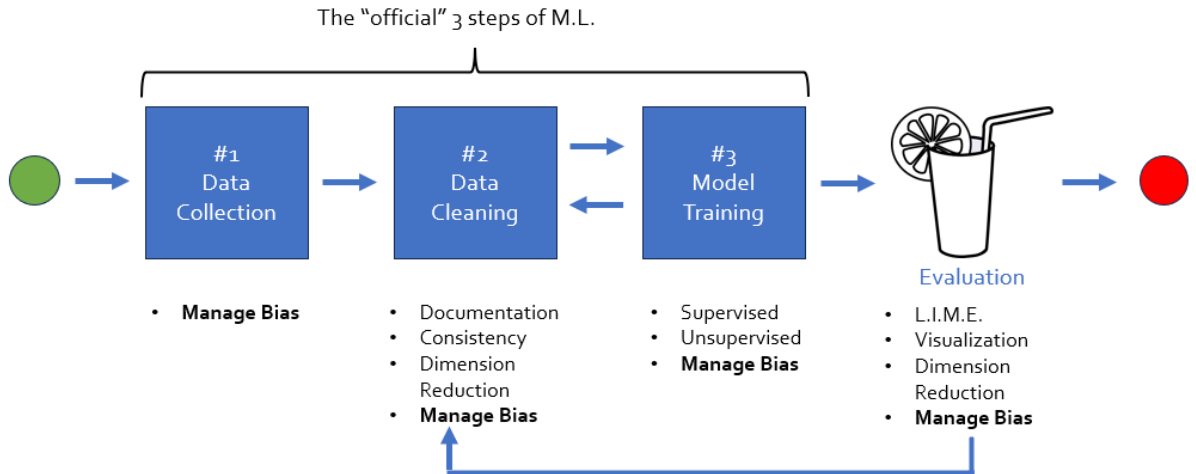
1. Our imaginary scenario is that of a healthcare provider. We want to recommend treatments based on similarities to other patients. That is, our process will recommend treatments based on the treatments received by similar users.,
2. Students should download ICA-KNN01.zip and unzip the ICA-KNN01 folder into their mis3536workspace. Use Anaconda Navigator to open up the Jupyter notebook file: kNN.ipynb
3. The “back end” database of our imaginary healthcare system looks like this:



4. Explore the data sources (URLs to datasets are in the script!)
5. Run the script. It is set up to generate recommendations for user 2.
6. Validate that the recommendations change per user.
7. Take a screenshot of your recommendation for a user (other than user 2) and upload it to Canvas.
8. DISCUSS: Look at the function “recommend_treatments” ... is this code implementing “user-based” or “item-based” collaborative filtering?
9. DISCUSS: Our imaginary healthcare provider is accused of over-prescribing ADHD medication to children of a particular ethnicity. How might we go about investigating this from a statistical point of view? What technical challenges would we face in such an investigation?

10. RECALL and DISCUSS: In our last lecture, we spoke about managing bias in machine learning models. We all agreed that bias (of different kinds) should be managed at different stages of machine learning development.

"Shafer's Roadmap" Semi-Supervised Machine Learning v1.1



DISCUSS: Where might we have a problem with bias in our medical recommendation system here? Describe where the bias might occur (step 1, 2, 3 or evaluation) and what kind of bias it might be (statistical, human, systemic).