

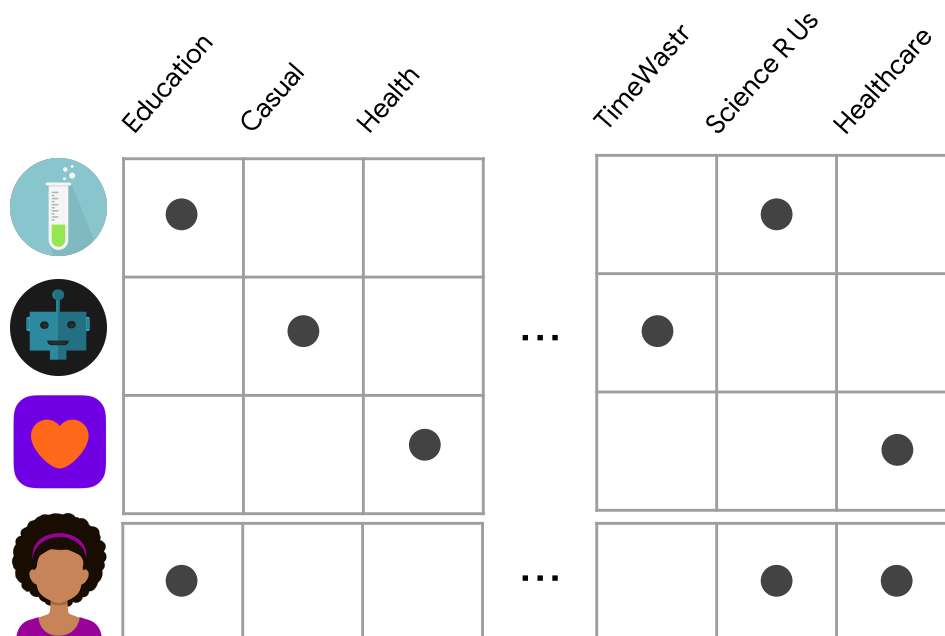
# Content-based Filtering

Content-based filtering uses item features to recommend other items similar to what the user likes, based on their previous actions or explicit feedback.

To demonstrate content-based filtering, let's hand-engineer some features for the Google Play store. The following figure shows a feature matrix where each row represents an app and each column represents a feature. Features could include categories (such as Education, Casual, Health), the publisher of the app, and many others. To simplify, assume this feature matrix is binary: a non-zero value means the app has that feature.

You also represent the user in the same feature space. Some of the user-related features could be explicitly provided by the user. For example, a user selects "Entertainment apps" in their profile. Other features can be implicit, based on the apps they have previously installed. For example, the user installed another app published by Science R Us.

The model should recommend items relevant to this user. To do so, you must first pick a similarity metric (for example, dot product). Then, you must set up the system to score each candidate item according to this similarity metric. Note that the recommendations are specific to this user, as the model did not use any information about other users.



## Using Dot Product as a Similarity Measure

Consider the case where the user embedding  $\mathbf{x}$  and the app embedding  $\mathbf{y}$  are both binary vectors. Since  $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{i=1}^d x_i y_i$ , a feature appearing in both  $\mathbf{x}$  and  $\mathbf{y}$  contributes a 1 to the sum. In other words,  $\langle \mathbf{x}, \mathbf{y} \rangle$  is the number of features that are active in both vectors

simultaneously. A high dot product then indicates more common features, thus a higher similarity.

## Try It Yourself!

Calculate the dot product for each app in the preceding app problem. Then use that information to answer the question below:

Which app should we recommend?	
The casual app created by TimeWastr.	▼
The educational app created by Science R Us.	▼
The health app created by Healthcare.	▼

[Previous](#)



[Candidate Generation Overview](#)

(/machine-learning/recommendation/overview/candidate-generation)

[Next](#)

[Advantages & Disadvantages](#)



(/machine-learning/recommendation/content-based/summary)

Except as otherwise noted, the content of this page is licensed under the [Creative Commons Attribution 4.0 License](https://creativecommons.org/licenses/by/4.0/) (<https://creativecommons.org/licenses/by/4.0/>), and code samples are licensed under the [Apache 2.0 License](https://www.apache.org/licenses/LICENSE-2.0) (<https://www.apache.org/licenses/LICENSE-2.0>). For details, see the [Google Developers Site Policies](https://developers.google.com/site-policies) (<https://developers.google.com/site-policies>). Java is a registered trademark of Oracle and/or its affiliates.

Last updated 2021-02-05 UTC.