As we discussed in the previous video, Netflix was willing to pay over $1 million for the best user rating algorithm, which shows how critical the recommendation system is to their business.

In this video, we'll discuss how recommendation systems work.

Let's start by thinking about the data.

When predicting user ratings, what data could be useful?

There are two main types of data that we could use.

The first is that for every movie in Netflix's database, we have a ranking from all users who have ranked that movie.

The second is that we know facts about the movie itself-- the actors in the movie, the director, the genre classifications of the movie, the year it was released, et cetera.

As an example, suppose we have the following user ratings for four users and four movies.

Our users are Amy, Bob, Carl, and Dan.

And our movies are Men in Black, Apollo 13, Top Gun, and Terminator.

The ratings are on a one to five scale, where one is the lowest rating and five is the highest rating.

The blank entries mean that the user has not rated the movie.

We could suggest to Carl that he watch Men in Black.

Since Amy rated it highly, she gave it a rating of five.

And Amy and Carl seem to have similar ratings for the other movies.

This technique of using other user's ratings to make predictions is called collaborative filtering.

Note that we're not using any information about the movie itself here, just the similarity between users.

Instead, we could use movie information to predict user ratings.

We saw on the table that Amy liked Men in Black.

She gave it a rating of five.

We know that this movie was directed by Barry Sonnenfeld, is classified in the genres of action, adventure, sci-fi, and comedy, and it stars actor Will Smith.

Based on this information, we could make recommendations to Amy.

We could recommend to Amy another movie by the same director, Berry Sonnenfeld's movie, Get Shorty.

We can instead recommend the movie Jurassic Park, which is also classified in the genres of action, adventure, and sci-fi.

Or we could recommend to Amy another movie starring Will Smith-- Hitch.

Note that we're not using the ratings of other users at all here, just information about the movie.

This technique is called content filtering.

There are strengths and weaknesses to both types of recommendation systems.

Collaborative filtering can accurately suggest complex items without understanding the nature of the items.

It didn't matter at all that our items were movies in the collaborative filtering example.

We were just comparing user ratings.

However, this requires a lot of data about the user to make accurate recommendations.

Also, when there are millions of items, it needs a lot of computing power to compute the user similarities.

On the other hand, content filtering requires very little data to get started.

But the major weakness of content filtering is that it can be limited in scope.

You're recommending similar things to what the user has already liked.

So the recommendations are often not surprising or particularly insightful.

Netflix actually uses what's called a hybrid recommendation system.

They use both collaborative and content filtering.

As an example, consider a collaborative filtering approach, where we determine that Amy and Carl have similar preferences.

We could then do content filtering as well, where we could find that the movie Terminator, which they both liked, is classified in almost the same set of genres as Starship Troopers.

So then we could recommend Starship Troopers to both Amy and Carl, even though neither of them have seen it before.

If we were only doing collaborative filtering, one of them would have had to have seen it before.

And if we were only doing content filtering, we would only be recommending to one user at a time.

So by combining the two methods, the algorithm can be much more efficient and accurate.

In the next video, we'll see how we can do content filtering by using a method called clustering.