

# The Remarkable world of Recommender Systems

An overview of the Recommendation systems and how they provide an effective form of targeted marketing.



Parul Pandey

Follow

May 9 · 10 min read



Image by PatternPictures from Pixabay

*A lot of times, people don't know what they want  
until you show it to them : Steve Jobs*

Here is an excerpt from the book **The Long Tail** by **Chris Anderson**:  
“In 1988, a British mountain climber named Joe Simpson wrote a book called **“Touching the Void”**, a harrowing account of near death in the

*Peruvian Andes. It got good reviews but, only a modest success, it was soon forgotten. Then, a decade later, a strange thing happened. Jon Krakauer wrote ‘**Into Thin Air**’, another book about a mountain-climbing tragedy, which became a publishing sensation. Suddenly **Touching the Void** started to sell again”.*

The demand for “**Touching the Void**” was so high that it even outsold ‘**Into Thin Air**’ after some time. But what exactly happened here? Well, it turns out that since both the books were based on the same theme, Amazon suggested that readers who liked **Into Thin Air** would also like **Touching the Void**. When people took in the suggestions they actually liked the book and as a result, wrote positive reviews which resulted in more sales ultimately leading to more recommendations and thereby kicking in a positive feedback loop. This is the power of recommendation systems.

. . .

## Recommender Systems

Recommendation Engines try to make a product or service recommendation to people. In a way, Recommenders try to narrow down choices for people by presenting them with suggestions that they are most likely to buy or use. Recommendation systems are almost everywhere from Amazon to Netflix; from Facebook to LinkedIn. In fact, a large chunk of Amazon’s revenue is generated from recommendations alone. Companies like Youtube and Netflix depend on their recommendation engines to help users discover new content. Some examples of recommendations in our everyday lives are:

- **Amazon**

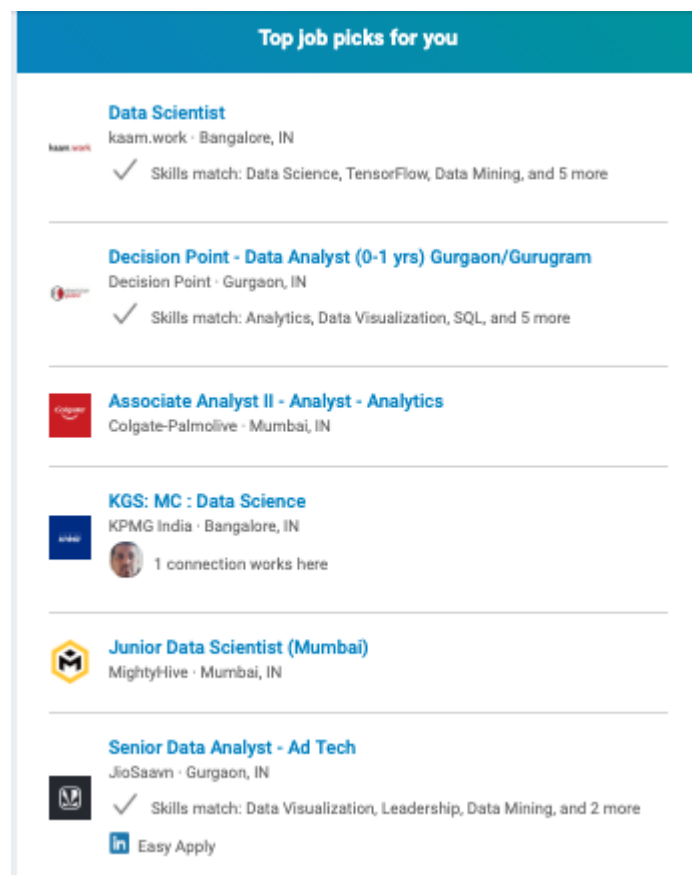
Amazon uses data from its millions of customers to identify which items are usually bought together and makes recommendations based on that. The recommendations in Amazon.com are provided on the basis of explicitly provided ratings, buying behaviour, and browsing history.



I intended to buy 'Show Dog' but ended up buying 'The Compound effect' too!

- **Linkedin**

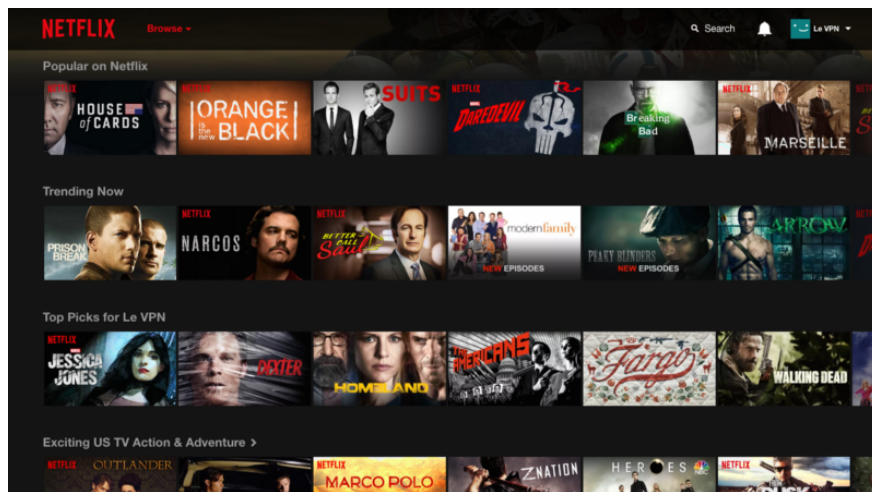
Linkedin utilises data from your past experience, current job titles and endorsements to suggest probable jobs to you.



- **Netflix**

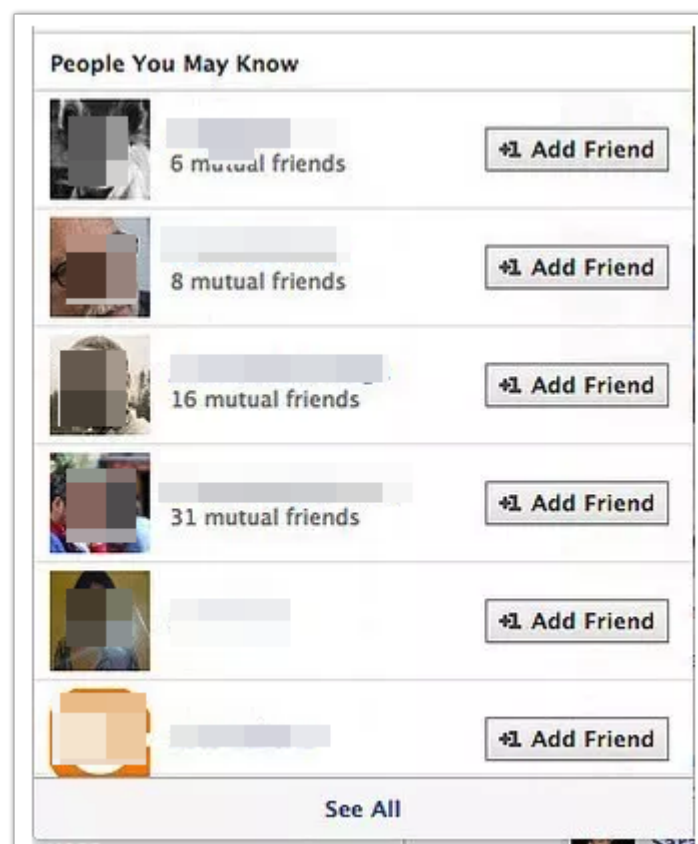
When we rate a movie or set up our preferences on Netflix, it uses this data and similar data from hundreds of other subscribers to recommend movies and shows. These ratings and actions are then used by Netflix to make recommendations.





- Facebook

Recommender systems, such as Facebook, do not directly recommend products but they recommend connections.



Apart from this, Spotify, Youtube, IMDB, Trip Advisor, Google News, and many other platforms continuously give out recommendations and suggestions to suit our needs.

. . .

## Why Recommenders

Today online stores are booming and we can get almost any item at the click of a mouse. However, in the brick-and-mortar era, there was limited physical space to store the goods, hence the owners only displayed the items which were most popular. This meant that a lot of products were not even displayed even though they had great quality like Books or CDs. In short, the shopkeepers had to pre-filter the contents.

However, the online shopping industry changed this scenario. Since there was unlimited space the need to pre-filter was gone. On the contrary, this gave rise to a phenomenon which came to be known as the **Long Tail Effect**.



The Long Tail Phenomenon

The effect means that popular products are few can be found in both online and offline stores. On the contrary, less popular products are many and can only be found in online stores and they ultimately constitute the **long tail**. However, unpopular products can also be good and finding such products on a website is a herculean task and requires some form of a filter. Such a filter actually constitutes a Recommendation System.

. . .

## Formulation of the Recommendation problem

Recommendation systems are created primarily to answer either of the two problems:

- **Prediction version**

This version deals with predicting the rating value for a user-item combination. In such cases, we have training data consisting of ratings given by the user. The aim is to utilise this data and predict the ratings of the items with which the user hasn't interacted.

- **Ranking Version**

Honestly speaking, it isn't necessary to predict the rating of users for specific items in order to make recommendations. An online retailer or an e-commerce company isn't so much concerned about their user's predictions. Rather they would be more interested to produce a finite list of the best things to present to a given person. Also, customers don't want to see the system's ability to predict their rating for an item, they just want to see things they're likely to love.

*The success of a Recommendation engine depends on its ability to find the best top recommendations for people so it makes sense to focus on finding things people will love and not our ability to predict the items people will hate.*

. . .

## The Goal of Recommender Systems



Recommendations will only be valuable if they are relevant

The ultimate goal of recommender systems is to increase the sales of a company. To make that happen the recommendation systems should display or provide only meaningful items to the user. *Charu C Aggarwal* in his book *Recommender Systems* sums up the desired goals of recommendation engines in the following four points:

- **Relevance**

Recommended items will only make sense if they are relevant to the user. Users are more likely to buy or consume items they find interesting.

- **Novelty**

Along with relevance, novelty is another vital factor. Recommended items will make more sense if the items are something that the user has not seen or consumed before.

- **Serendipity**

Sometimes recommending items which are somewhat unexpected can also boost sales. Serendipity is however different from novelty. In the author's words:

*“if a new Indian restaurant opens in a neighbourhood, then the recommendation of that restaurant to a user who normally eats Indian food is novel but not necessarily serendipitous. On the other hand, when*

*the same user is recommended Ethiopian food, and it was unknown to the user that such food might appeal to her, then the recommendation is serendipitous”.*

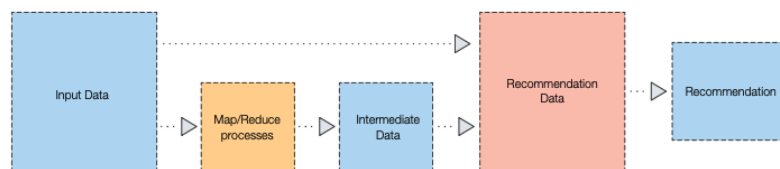
- **Diversity**

Also increasing diversity in recommendations is equally important. Simply recommending items which are similar to each other, isn't of much use.

. . .

## Working of Recommender Systems

So how do recommender systems work? Let's say Amazon wants to show you the top 10 recommendations in Books category. Here Amazon's recommender system will start with some kind of data about you so as to figure out your individual tastes and interests. It will then merge this data about you with the collective behaviour of everyone else like you to recommend stuff you might like. But where does this data about your likes and dislike comes from?



The flow of data into a recommendation engine

The user's preference data is collected in two ways:

- **Explicit Data**

Asking users to rate an item on a scale of one to five stars or rating content they see with a like or a thumbs down is an example of explicit data collection. In these cases, users are explicitly asked whether they like a particular item or not and this data is then used to build up a profile of that user's interest.

However, there is a drawback since not every user leaves feedback or rating and even if they leave ratings, it may mean different to different people. For instance, a 3 ★ rating may be great for one but average for another.



- **Implicit Data**

Implicit data is derived from a user's interaction with the site and interpreting them as indications of interest or disinterest. For example, buying a product from Amazon or watching a complete youtube clip is considered as a sign of positive interest. Implicit interactions can give you much more data to work with and in the case of purchase data, it might even be better data to start.

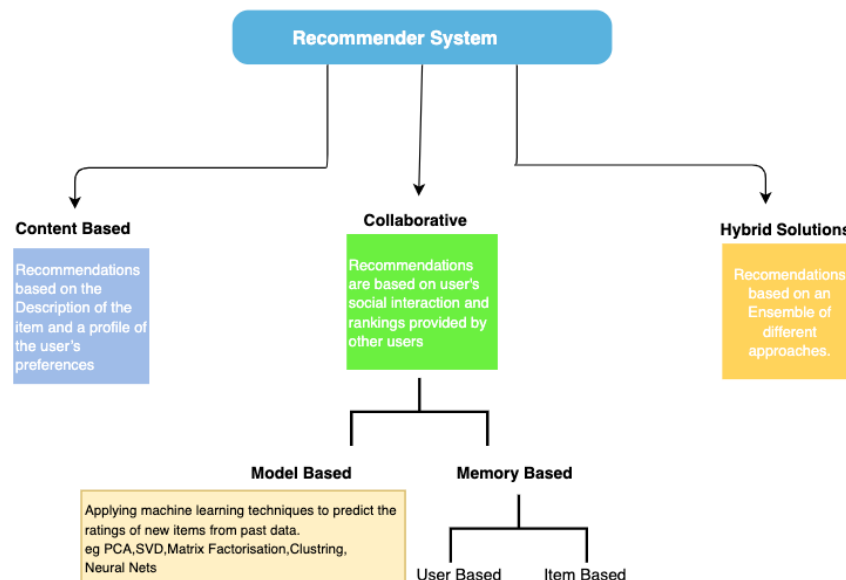
. . .

## Basic Models of Recommender Systems

There are many kinds of recommenders being employed in the industry today. The important decision, however, is to decide which type suits our needs and what kind of data is available with us. The selection primarily depends on :

- What we want to identify and,
- What type of relationship is specified in our data.

Some of the common approaches used for recommendations include:

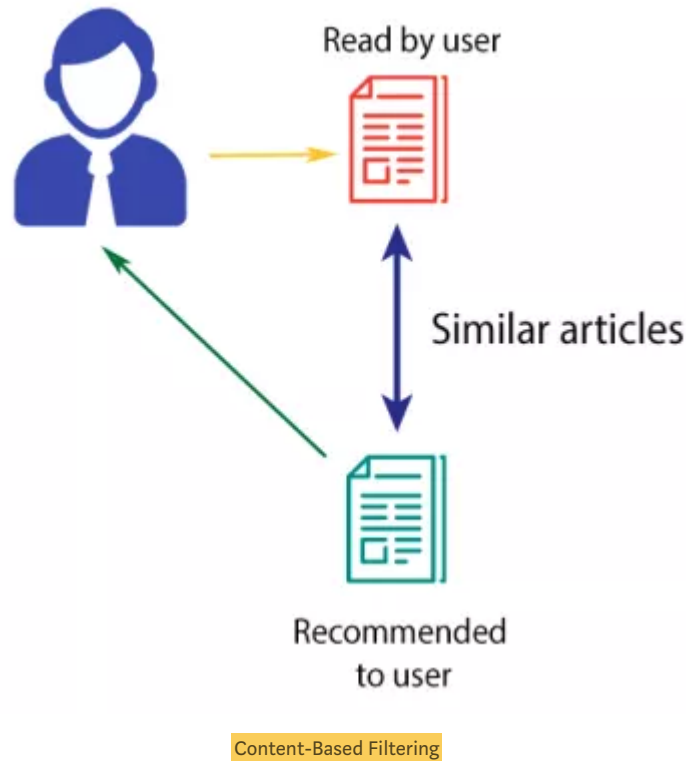


Approaches used for Recommendation systems

Let's have a brief overview of each one of them:

### Content-Based Filtering

Content-based filtering involves recommending items based on the attributes of the items themselves. Recommendations made by content-based filters use an individual's historical information to inform choices displayed. Such recommenders look for similarities between the items or products that a person had bought or liked in the past to recommend options in the future.



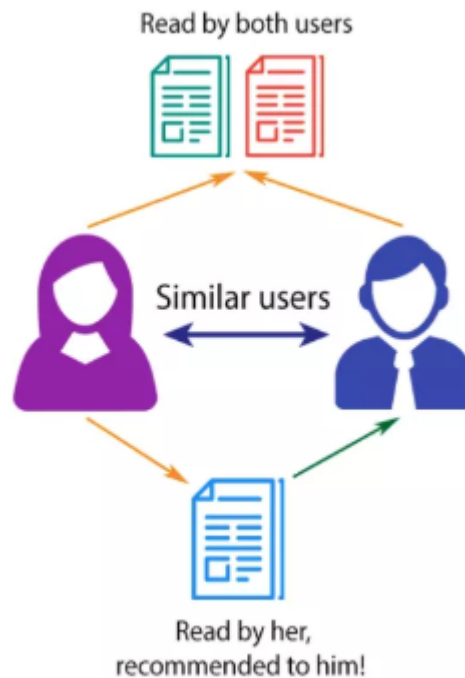
For instance, if a user likes a book in the '**Literature**' category, it makes sense to recommend books in the same category to the user. Also, recommending books released in the same **Year** and by the same **Author** would also be a great idea. This is how content-based filtering works.

The advantage of the content-based methods is that we don't really need a lot of transactions to build our models since we only need information on the products. The disadvantage, however, is that the model doesn't learn from the transactions, so there isn't much improvement in the performance of content-based systems over time.

. . .

## Collaborative Filtering

Collaborative filtering uses the combined power of ratings provided by many users/customers to present recommendations. It means recommending stuff based on other people's collaborative behaviour.



There are two approaches to collaborative filtering:

1. **Memory-based methods** which are also referred to as neighbourhood based collaborative filtering algorithms in which ratings of user-item combinations are predicted on the basis of their neighbourhoods. These neighbourhoods can further be defined in one of two ways:

- **User-based collaborative filtering:**

Finding other people like you and recommending items they liked.

- **Item-based collaborative filtering:**

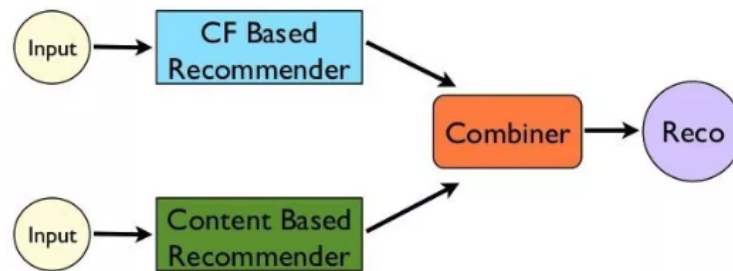
Recommending items that people bought who also bought items that you liked.

2. **Model-based methods** use machine learning methods to extract predictions for rating data by treating the problem as a normal machine learning problem. Techniques like PCA, SVD, Matrix Factorisation,

Clustering,  
Neural Nets etc can be used.

. . .

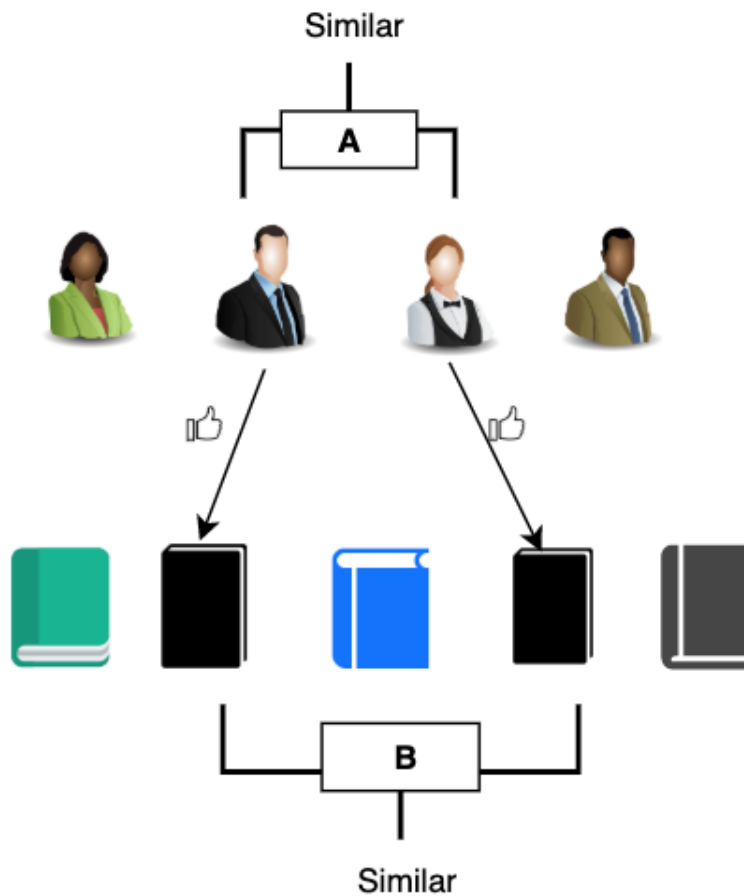
## Hybrid and Ensemble-Based



Hybrid filtering

Both Content-based and Collaborative approaches have their own strengths and weaknesses and one can end up with a better system by combining many algorithms together in what we call a hybrid approach. Hybrid systems leverage both item data and transaction data to give recommendations.

A great example of using a Hybrid approach is that of **Netflix**. At Netflix, recommendations are not only based on what people's watching and searching habits(collaborative systems)but also movies sharing similar characteristics(content-based) are also recommended.



How Hybrid Systems work

. . .

## Evaluating Recommender Systems: The Hype around accuracy

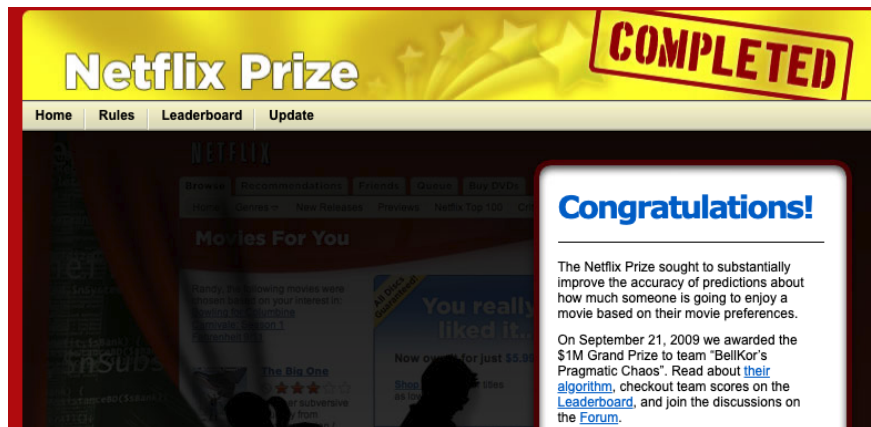
*Users don't really care about accuracy*

There isn't a straightforward way to measure how good the recommender systems are. Many of the research in this field tends to focus on the problem of predicting a user's ratings for everything they haven't rated already, good or bad. But that's very different from what recommender systems need to do in the real world. Measuring accuracy isn't really what we want our recommendation systems to do. So why is much importance is given to RMSE and accuracy in the recommendation system realm?

Well, a lot of it dates back to 2006 when Netflix announced the famous \$1 Million Prize Challenge. The race was on to beat their RMSE of 0.9525 with the finish line of reducing it to 0.8572 or less. Since the



focus of the Prize was RMSE, people focussed only on it and that effect has continued even till today.



<https://www.netflixprize.com/>

Interestingly, most of the algorithms that came out of the three-year competition were never integrated into Netflix. As discussed on the Netflix blog:

*You might be wondering what happened with the final Grand Prize ensemble that won the \$1M two years later... We evaluated some of the new methods offline but the additional accuracy gains that we measured did not seem to justify the engineering effort needed to bring them into a production environment.*

*Our business objective is to maximize member satisfaction and month-to-month subscription retention... Now it is clear that the Netflix Prize objective, accurate prediction of a movie's rating, is just one of the many components of an effective recommendation system that optimizes our members' enjoyment.*

. . .

## Conclusion

In this article, we had an overview of the Recommendation systems and how they provide an effective form of targeted marketing by creating a personalized shopping experience for each customer. However, we did not go deeper into the various methods of recommendations. This is because each of the methods is fairly extensive and deserve an article of its own. **So in the next article, I shall discuss in detail as to how**

## **Recommendations Methods work and their advantages and disadvantages.**

