




# Introduction to Machine Learning

## Recommendation Systems

Recommendation systems are everywhere

### LinkedIn People Recommendations

PEOPLE YOU MAY KNOW

-  **vidhya murali**, MTS at Oracle  
[Connect](#)
-  **Nishan Jain**, Research Assistant at University of  
[Connect](#)
-  **Dhwan Raj**, Machine Learning, Natural Language Processing,  
[Connect](#)

[See more »](#)

### Facebook People Recommendations

Are They Your Friends Too?

These people now have 1 or more friends in common with you.

-  1 mutual friend  
[Add Friend](#)
-  67 mutual friends  
[Add Friend](#)
-  19 mutual friends  
[Add Friend](#)
-  47 mutual friends  
[Add Friend](#)

[See All Suggestions](#)

### HotJobs Job Recommendations

**Recommended Jobs for Manish**

Job Description	Location	Date
Software Engineer Aruba Networks	Sunnyvale, CA	Feb 9
Software Engineer: Java Web Developer Remilon LLC	Mountain View, CA	Feb 2
Senior Software Engineer Optovue	Fremont, CA	Jan 26

[Show more jobs »](#)

### Bing Query Recommendations

bing

manish gupta **microsoft**

manish gupta **microsoft**

manish gupta **american express**

manish gupta **oncologist**

manish gupta **sylvania**

manish gupta **md**

manish gupta **power minister west bengal**

manish gupta **md las vegas nv**

### Amazon Product Recommendations

amazon

These recommendations are based on items you've viewed and more.

**Amazon Basics 3.0 Megapixel USB PC Camera**

by Amazon (May 21, 2014)  
Average Customer Review: **4.5 stars** (1,234)

**Link Power 3040**  
Price: \$1.99  
Amazon.com Price: \$2.99

### Netflix Movie Recommendations

Other Movies You Might Enjoy

-  **The Godfather**  
4.5 stars
-  **The Godfather Part II**  
4.5 stars
-  **The Godfather Part III**  
4.5 stars
-  **The Godfather: The Coda**  
4.5 stars

## Recommendation Systems

### Social overload

- Facebook – largest social network site
  - 600,000,000 users, half login every day
  - 35,000,000,000 online “friendships”
  - 900,000,000 objects people interact with
  - 30,000,000,000 shared content items / month
- YouTube – largest video sharing site
  - 2,000,000,000 views per day
  - 1,000,000 video hours uploaded per month
- Twitter – largest microblogging site
  - 200,000,000 users per month
  - 65,00,000 tweets per day (750 per second)
  - 8,000,000 followers of most popular user

## Recommendation Systems

### Social overload

- Information Overload
  - Blogs, microblogs, forums, wikis, news, bookmarked webpages, photos, videos, etc.
- Interaction Overload
  - Friends, followers, followees, commenters, co-members, voters, likers, taggers, review writers, etc.

## Recommendation Systems

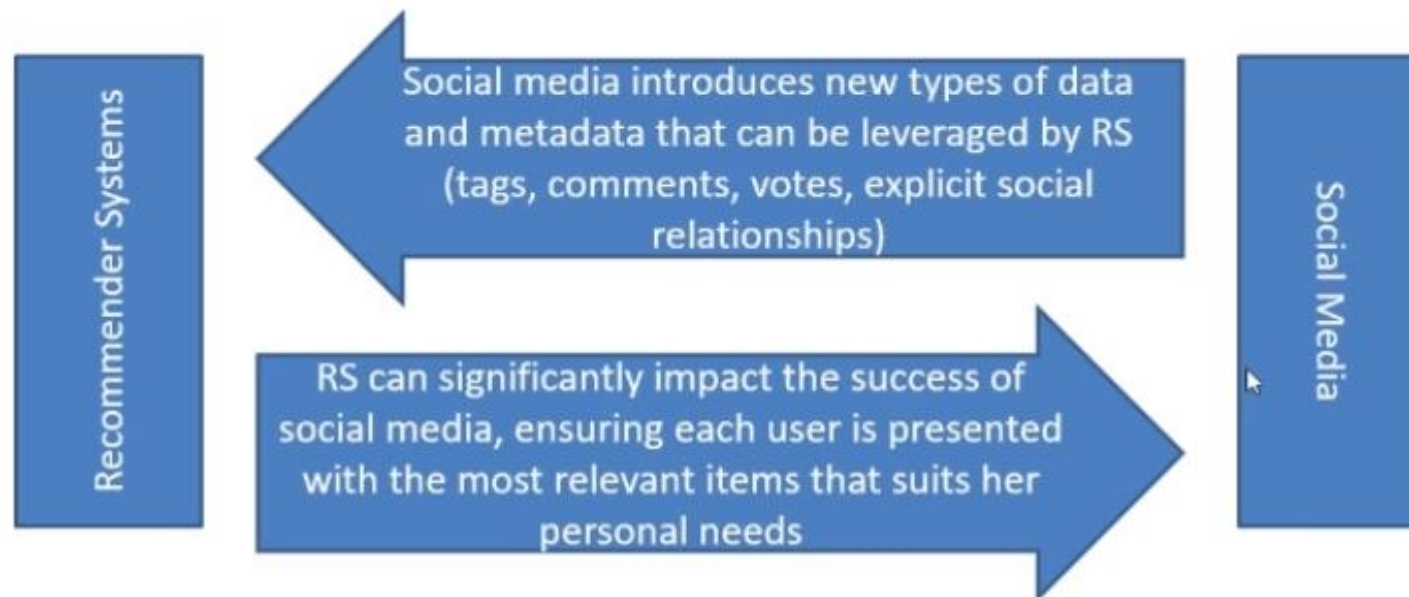
### Social Recommender Systems

- Recommender Systems that target the social media domain
- Aim at coping with the challenge of social overload by presenting the most attractive and relevant content
- Also aim at increasing adoption and engagement
- Often apply personalization techniques

## Recommendation Systems

### Recommender Systems & Social Media

- Recommender Systems are an augmentation of the social process, in which we rely on advices or suggestions from other people
- Social Media and Recommender Systems can mutually benefit each other





## Recommendation Systems

### Fundamental Recommendation Approaches

- Collaborative filtering based Recommendation
  - Aggregate ratings of objects from users and generate recommendation based on inter-user similarity
- Demographic Recommendation
  - Categorize users based on personal attributes (age, gender, income..) and make recommendation based on demographic classes
- Content-based recommendation
  - A user profile is constructed based on the features of the items the user has rated/consumed. This profile is used to identify new interesting items for the user (that match his profile)
- Hybrid methods
  - Combine several approaches together

## Recommendation Systems

### Collaborative Filtering

#### Customers Who Bought This Item Also Bought

  <p>IPAD 2 Leather Case With Stand for Apple IPAD 2 (Black) Fits All Ipad2 Model</p> <p>★★★★☆ (886)</p> <p>\$6.50</p>	 <p>CANOPY 2 YEAR Accidental Protection Plan (\$400-\$450)</p> <p>★★★★☆ (29)</p> <p>\$74.99</p>	 <p>Ctech 360 Degrees Rotating Stand (black) Leather Case for iPad 2 2nd generation</p> <p>★★★★☆ (927)</p> <p>\$7.45</p>	 <p>3 Pack of Premium Crystal Clear Screen Protectors for Apple iPad</p> <p>★★★★☆ (2,153)</p> <p>\$4.44</p>
--	--	--	--

- In the real world we seek advices from our trusted people (friends, colleagues, experts)
- CF automates the process of “word-of-mouth”
  - Weight all users with respect to similarity with the active user.
  - Select a subset of the users (neighbors) to use as recommenders
  - Predict the rating of the active user for specific items based on its neighbors' ratings
  - Recommend items with maximum prediction

## Recommendation Systems

### User based Collaborative Filtering Algorithm

- The User x Item Matrix

	Shrek	Snow-white	Superman
Alice	Like	Like	Dislike
Bob	?	Dislike	Like
Chris	Like	Like	Dislike
John	Like	Like	?

- Shall we recommend Superman for John?
- John's taste is similar to both Chris and Alice tastes  $\Rightarrow$  Do not recommend Superman to John





## Recommendation Systems

### User based Collaborative Filtering Algorithm

- Let  $R$  be the rating matrix
  - $r_{uj}$  is then the vote of user  $u$  for item  $j$
- $I_u$  be the set of items for which user  $u$  has provided the rating
- Voting
  - Mean vote for user  $u$ :  $\bar{r}_u = \frac{1}{|I_u|} \sum_{i \in I_u} r_{ui}$
  - Prediction rating:  $p_{uj} = \bar{r}_u + \gamma \sum_{v=1}^n w(u, v)(r_{vj} - \bar{r}_v)$ 
    - $w(u, v)$  = similarity between users  $u$  and  $v$
    - $\gamma$  is a normalization constant  $\gamma = \frac{1}{\sum_{v=1}^n w(u, v)}$

## Recommendation Systems

### Similarity Functions

- Cosine based similarity between users

$$- w(u, v) = \frac{\sum_{i \in I} r_{ui} r_{vi}}{\sqrt{\sum_{i \in I} r_{ui}^2} \sqrt{\sum_{i \in I} r_{vi}^2}}$$

- Pearson based similarity between users

$$- w(u, v) = \frac{\sum_{i \in I} (r_{ui} - \bar{r}_u)(r_{vi} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{ui} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{vi} - \bar{r}_v)^2}}$$

## Recommendation Systems

### Collaborative Filtering: Practical Challenges

- Ratings data is often sparse, and pairs of users with few co-ratings are prone to skewed correlations
- Fails to incorporate agreement about an item in the population as a whole
  - Agreement about a universally loved item is much less important than agreement for a controversial item
    - Some algorithms account for global item agreement by including weights inversely proportional to an item's popularity
- Calculating a user's perfect neighborhood is expensive
  - requiring comparison against all other users
  - Sampling: a subset of users is selected prior to prediction computation
  - Clustering: can be used to quickly locate a user's neighbors

### Item-Based Nearest Neighbor Algorithm

- The transpose of the user-based algorithms
  - Generate predictions based on similarities between items
  - The prediction for an item is based on the user's ratings for similar items

	Shrek	Snow-white	Superman
Alice	Like	Like	Dislike
Bob	?	Dislike	Like
Chris	Like	Like	Dislike
John	Like	Like	?

- Bob dislikes Snow-white (which is similar to Shrek)  $\Rightarrow$  do not recommend Shrek to Bob
- Predicted rating:  $p_{uj} = \gamma \sum_{i=1}^m w(i,j)r_{ui}$
- Traverse over all m items rated by user u and measure their rating, averaged by their similarity to the predicted item
- $w(i,j)$  is a measure of item similarity - usually the cosine measure
- Average correction is not needed because the component ratings are all from the same target user